UC Riverside

UC Riverside Electronic Theses and Dissertations

Title

Proteome-Wide Discovery and Characterizations of Nucleotide-Binding Proteins With Affinity-Labeled Chemical Probes

Permalink

https://escholarship.org/uc/item/7xw1s20p

Author

XIAO, Yongsheng

Publication Date

2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Proteome-Wide Discovery and Characterizations of Nucleotide-Binding Proteins With
Affinity-Labeled Chemical Probes

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

Doctor of Philosophy

in

Chemistry

by

Yongsheng Xiao

March 2014

Dissertation Committee:

      Dr. Yinsheng Wang, Chairperson
      Dr. Ryan Julian
      Dr. Wenwan Zhong

The Dissertation of Yongsheng Xiao is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGMENTS

DEDICATION

To my wife Lei Guo, my parents Mingxing XIAO and Yong Chen. I appreciate all of your love, encourage, patience, and support.

COPYRIGHT ACKNOWLEDGEMENTS

The text and figures in Chapter 2, in part or full, are a reprint of the material as it appears in Analytical Chemistry 2013, 85:3198-3206. The coauthor (Dr. Yinsheng Wang) listed in that publication directed and supervised the research that forms the basis of this chapter.

The text and figures in Chapter 3, in part or full, are a reprint of the material as it appears in Analytical Chemistry 2013, 85:7478–7486. The coauthor (Dr. Yinsheng Wang) listed in that publication directed and supervised the research that forms the basis of this chapter.

The text and figures in Chapter 4, in part or full, are a reprint of the material as it appears in Analytical Chemistry 2014, submitted. The coauthor (Dr. Yinsheng Wang) listed in that publication directed and supervised the research that forms the basis of this chapter.

The text and figures in Chapter 5, in part or full, are a reprint of the material as it appears in Mol Cell Proteomics 2014, in press. The coauthor (Dr. Yinsheng Wang) listed in that publication directed and supervised the research that forms the basis of this chapter.

ABSTRACT OF THE DISSERTATION

Proteome-Wide Discovery and Characterizations of Nucleotide-Binding Proteins With
Affinity-Labeled Chemical Probes

by

Yongsheng Xiao

Doctor of Philosophy, Graduate Program in Chemistry
University of California, Riverside, March 2014
Dr. Yinsheng Wang, Chairperson

Nucleotide-binding proteins such as ATP- and GTP- binding proteins play pivotal
roles in many cellular processes. However, targeted study of nucleotide-binding proteins,
especially protein kinases and GTPases, remained challenging, which is mainly because
of the lack of efficient enrichment approaches to selectively capture nucleotide-binding
proteins. In this dissertation, we reported a general strategy in using affinity-labeled
chemical probes to enrich, identify, and quantify ATP- and GTP- as well as other
nucleotide binding proteins in the entire human proteome.

In Chapter two, we described the synthesis and application of biotin-conjugated
ATP and GTP affinity probes. The ATP/GTP affinity probes facilitated the identification
of 100 GTP-binding proteins and 206 kinases with the use of low mg quantities of cell
lysate. Additionally, our strategy led to the identification of three and one unique
nucleotide-binding motifs for kinases and GTP-binding proteins, respectively. We also

assessed the ATP/GTP binding selectivities of nucleotide-binding proteins at the global proteome scale.

In Chapter three, an affinity profiling strategy to comprehensively characterize ATP-protein interactions at the entire proteome scale was developed by comparing the labeling behaviors of lysine residues with the use of low and high concentrations of the ATP affinity reagents. This novel quantitative ATP-affinity profiling strategy is particularly useful for unveiling previously unrecognized nucleotide-binding property and binding sites in whole proteome.

In Chapter four, we extended the use of ATP/GTP affinity probes and introduced an orthogonal strategy encompassing the nucleotide-affinity profiling assay and nucleotide-binding competition assay to comprehensively characterize ^{S}GTP-binding proteins along with the specific binding sites from the entire human proteome. More than 160 proteins involved in different biological processes were determined to be ^{S}GTP-binding targets.

In Chapter five, we introduced a novel global kinome profiling method, based on a newly developed isotope-coded ATP-affinity probe and a targeted proteomic method using multiple-reaction monitoring (MRM), for assessing simultaneously the expression of more than 300 kinases in human cells and tissues. This facile and accurate kinome profiling assaycould provide invaluable knowledge to predict the effectiveness of kinase inhibitor drugs and offer the opportunity for individualized cancer chemotherapy.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

General Overview

1.1 Introduction

In post-genomic era, proteomics has been emerging as an indispensable technology to interpret the information encoded in genomes (1, 2). Proteomics is the study of the identities, quantities, structures, and cellular functions of the entire complement of proteins expressed in an organism, organelle, or cell population (3). In traditional protein biochemistry, due to technical limitations, biochemists often can only deal with individual proteins or a handful of related protein components in a specific biochemical pathway. Therefore, over the past decades, molecular biology paradigm follows the "one gene-one protein-one function" rule which assumes a linear pathway between individual proteins and their associated functions (4). However, this uni-directional pathway-based paradigm focusing on the properties of individual proteins has been proved to be inadequate (5). More and more evidence has indicated that proteins do not act alone and the dynamically modulated network comprised of interacting proteins determines the genotype-phenotype relationship (6-8). In recent years, there is a clear trend in life sciences towards the study of global proteome and proteomics has run on the forefront in systems biology research (9, 10).

Owing to the high degrees of complexity of cellular proteomes and the relative low abundance of many of the proteins with significant functions, comprehensive study of whole proteome faces big challenges and requires the technological breakthroughs for protein measurement (11). Mass spectrometry (MS) has increasingly become the method of choice for analysis of complex protein samples due to its high specificity, accuracy and high throughput (12). With the recent development of mass spectrometry instrumentation (13) in combination with new sample preparation (14) and quantification methods (15), as well as refined bio-informatics tools (16), proteomics has evolved as a powerful tool for solving cellular biological and real medical problems systematically.

Currently, MS-based proteomics mainly focus on three types of experimental results. First, expression proteomics determines the relative or absolute amounts of expressed proteome from two or more physiological states of a biological system (17). Quantitative comparison of proteomes under different conditions can be invaluable for unraveling the underlying mechanism for specific biological processes and has been widely used for biomarker discovery. In addition, MS-based proteomics is well suited for post-translational modification (PTM) study of recombinant protein or endogenous proteome (18, 19). MS measures mass-to-charge ratio (m/z), yielding the molecular weight and unique fragmentation pattern of peptides to unambiguously determine of the primary structure of proteins. Recently, MS-based proteomics has been widely used to support protein interaction studies and biological network mapping (20). The high-throughput of MS-based proteomics provides the new opportunities for biologist to

integratively explore the biological process, focusing not only the protein of interest but also the interacting networks between individual proteins. One thing needs to be noted is that this MS-based interaction proteomics is not restricted to protein-protein interaction but encompasses other interactions of proteins with small molecules (21), or RNA and DNA macromolecules (22).

In this chapter, I will first review commonly used MS-based strategies to study quantitative proteomics. I will then discuss the application of quantitative proteomics technologies in global nucleotide-binding protein studies, which is the main area for my Ph.D. work.

1.2 MS-based quantitative proteomics

Quantification is of central importance in MS-based proteomics and above-mentioned three major applications of proteomics all heavily rely on the quantitative capability of modern LC-MS/MS methods (15). For example, it is the change in protein expression or the change in the level of a particular PTM upon a certain perturbation that is of biological significance. Additionally, in interaction proteomics or affinity-purification mass spectrometry, high-sensitivity MS often identify hundreds of background proteins due to non-specific interaction as well as the real targets in pull-down samples, which necessitates quantification of proteins in bait pull-downs against a control (23).

In order to achieve accurate quantification, isotope labeling techniques are often incorporated in various proteomic workflows to derive quantitative data from mass

spectrometry data. Given that a mass spectrometer can differentiate the mass shift between the labeled and unlabeled forms of a peptide, quantification can be achieved by comparing their respective MS intensity. These isotopic mass tags can be introduced into proteins or peptides by either metabolic or chemical labeling (15). Moreover, the quantitative measurement of these isotopically labeled samples derived from different physiological conditions can be achieved using data-dependant analysis (DDA) in discovery MS mode (24), or using multiple-reaction monitoring analysis (MRM) in targeted MS mode (25). In this section, I will summarize several widely used isotope-labeling approaches for MS-based quantitative proteomics as well as the principle and merits for DDA and MRM MS analysis.

1.2.1 Quantitative strategies for MS-based proteomics

1.2.1.1 Stable isotope labeling by amino acids in cell culture (SILAC)

In metabolic labeling, a stable isotope signature was introduced into proteins during cell growth and division. SILAC is the most widely used metabolic labeling method developed in Mann's lab in 2002 (26). Figure 1.1 shows an outline for a standard SILAC experiment allowing for relative quantification of protein abundance. In SILAC, two cell populations under different physiological conditions are cultured in pre-defined isotope media separately. Arginine and lysine in light or heavy forms are often supplemented to the SILAC medium, which ensures that all tryptic peptides, except the C-terminal peptides of some proteins, carry at least one labeled amino acid. As cells proliferate, endogenous proteins begin to incorporate the isotope-labeled analogs of

4

Figure 1.1 General procedures of SILAC experiment for quantitative proteomics.
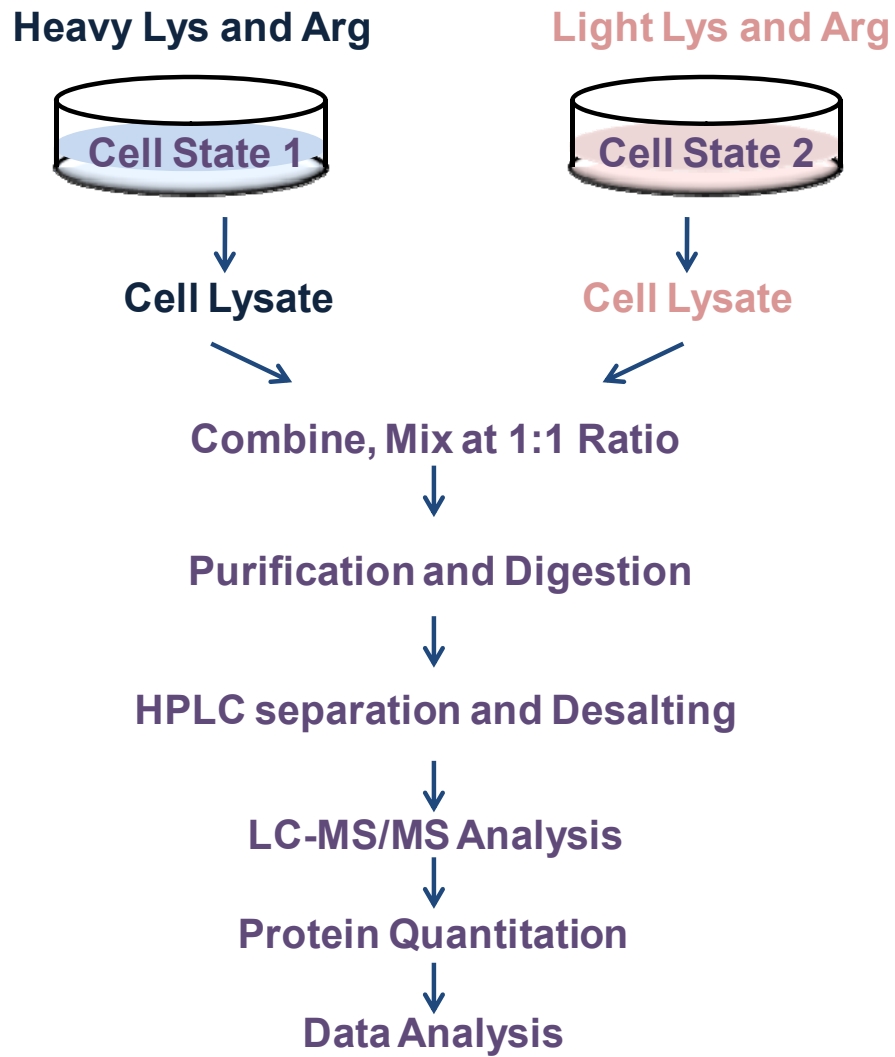
lysine or arginine. After 6-8 cell doublings, more than 99% of endogenous proteins will be labeled by isotope-labeled arginine or lysine. Both the heavy and light cells are then harvested, and the extracts are mixed directly at certain ratios, usually in equal amounts. Protein mixtures can be further purified or fractionated prior to proteolytic digestion. The resulting peptide mixture is finally subjected to LC-MS/MS analysis and the quantification is based on the MS signal intensity of the peptide pair with clearly defined mass difference.

The main advantage of SILAC labeling method is that the differently treated cell samples can be combined immediately after cell lysis. This early combination excludes all the variability in sample processing and mass spectrometry detection, rendering SILAC the most accurate isotopic labeling strategy for MS-based quantitative proteomics. Therefore, SILAC labeling is particularly suitable for assessing subtle changes in protein expression of cells under internal or external stimulus (27). However, some cell lines that are sensitive to media composition and cannot grow or be maintained in SILAC medium. Moreover, clinical samples (e.g. biological fluids, tissue samples) are not amenable to SILAC labeling. Another major drawback for SILAC labeling is the limited number of available labels. Generally, a maximum of three independent labeling experiments can be simultaneously conducted in a single SILAC experiment (15).

1.2.1.2 Isotope-coded affinity tag (ICAT)

In addition to metabolic labeling, chemical tagging strategies serve as efficient approaches for quantitative proteomic analysis, in which side chains of lysine and

cysteine are the primary targets used for labeling (28). In 1999, Aebersold and co-workers developed the first chemical tagging strategy for quantitative proteomics termed ICAT (29). As shown in Figure 1.2A, the ICAT reagent consists of three elements: an affinity tag (biotin); a light or heavy isotope labeled (diethylene glycol) linker containing either hydrogens or deuterons; and a thiol reactive group which will specifically derivatize cysteine residues. The typical workflow for ICAT experiment is shown in Figure 1.2B. First, the side chains of cysteine in a protein sample representing two cell states are derivatized with either isotopically labeled light or heavy form of the ICAT reagent. The two samples are then combined and enzymatically cleaved to generate peptide fragments. Only the tagged cysteine-containing peptides are isolated by avidin affinity chromatography and analyzed by LC-MS/MS. The quantification is based on the MS signal intensities of the peptide pairs.

Since its introduction, ICAT has been widely used for MS-based quantitative analysis, especially for those samples that are not amenable to SILAC labeling. However, since only cysteine-tagged peptides will be trapped and analyze, ICAT is not suitable for quantifying those proteins without cysteine residues. Additionally, ICAT may only be useful to compare global protein expression and is of limited use for PTM proteomics study (15). This limitation of ICAT can be circumvented by other chemical tagging strategy targeting other functional groups on specific amino acids. For example, isotopic dimethyl labeling approach was recently developed to introduce the stable isotopic tags by methylation of all primary amino groups on lysine residues or peptide N-

7

Figure 1.2 (A) The structure of ICAT reagent. (B) General procedures of ICAT experiment for quantitative proteomics.

(A)

Biotin
(Enrichment moiety)

Linker
Heavy: X=deuterium
Light: X=hydrogen

**Thio-specific
reactive group**
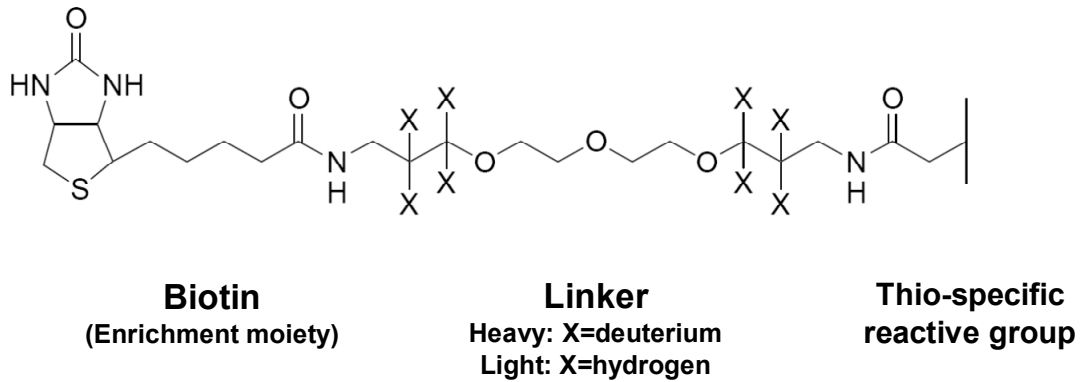
(B)
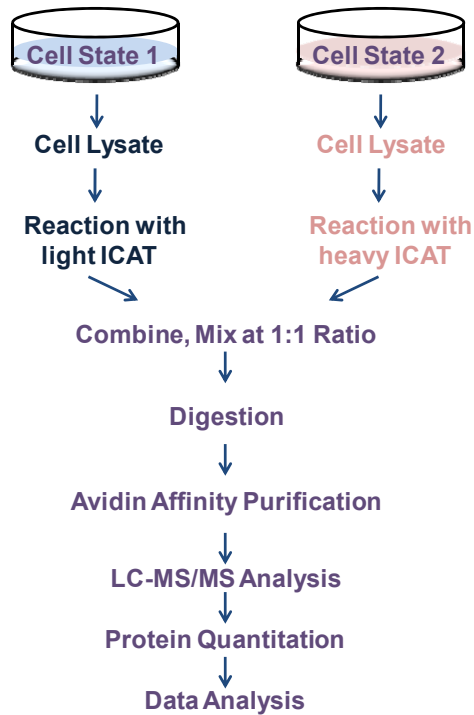
terminals (30, 31). Dimethyl labeling approach is very fast, reliable and cheap, which has been widely used in high-throughput proteomics experiments (32).

1.2.1.3 Isobaric tag for relative and absolute quantitation (iTRAQ)

Besides above mentioned ICAT or Dimethyl labeling, isobaric mass tagging is another popular implementations of chemical derivatization for quantitative proteomics. The isobaric tag for relative and absolute quantitation (iTRAQ) strategy was developed by Applied Biosystems in 2004 (33). Figure 1.3A shows the structure of the iTRAQ reagents. The complete molecule consists of a reporter group (based on N-methylpiperazine), a mass balance group (carbonyl) and a peptide-reactive group (NHS ester). When treated with a peptide, the tag forms an amide bond to any peptide amino group located on N-terminal or lysine side chain. In CID fragmentation during MS analysis, the balance group is lost from peptide backbone and the charge is retained on the reporter group fragments. These resulting reporter ions include at least four isotope-coded variants ranging in mass from m/z 114.1 to 117.1, which allows for simultaneous quantification of relative protein expression of multiple experimental conditions (Figure 1.3B). Recently, Thermo Fisher Scientific has also released a similar product, tandem mass tags (TMT) isobaric mass tagging kit, which can provide simultaneous quantification of up to 6 different samples (34).

There are several advantages of using these isobaric labeling approaches for quantitative proteomics. First, iTRAQ strategy generates isobaric peptides with the same molecular mass and elution behavior in liquid chromatography separation, which cannot

9

Figure 1.3 (A) The structure of iTRAQ reagent. (B) General procedures of iTRAQ experiment for quantitative proteomics.

(A)

Isobaric tag (total mass=145)
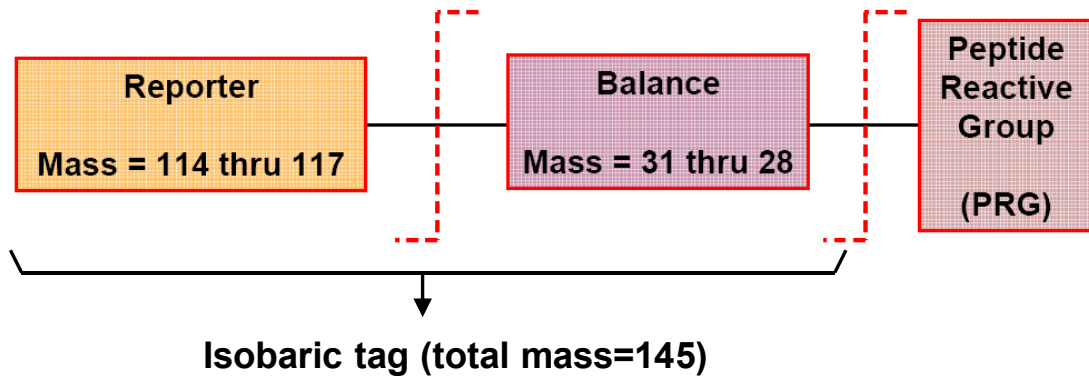
(B)
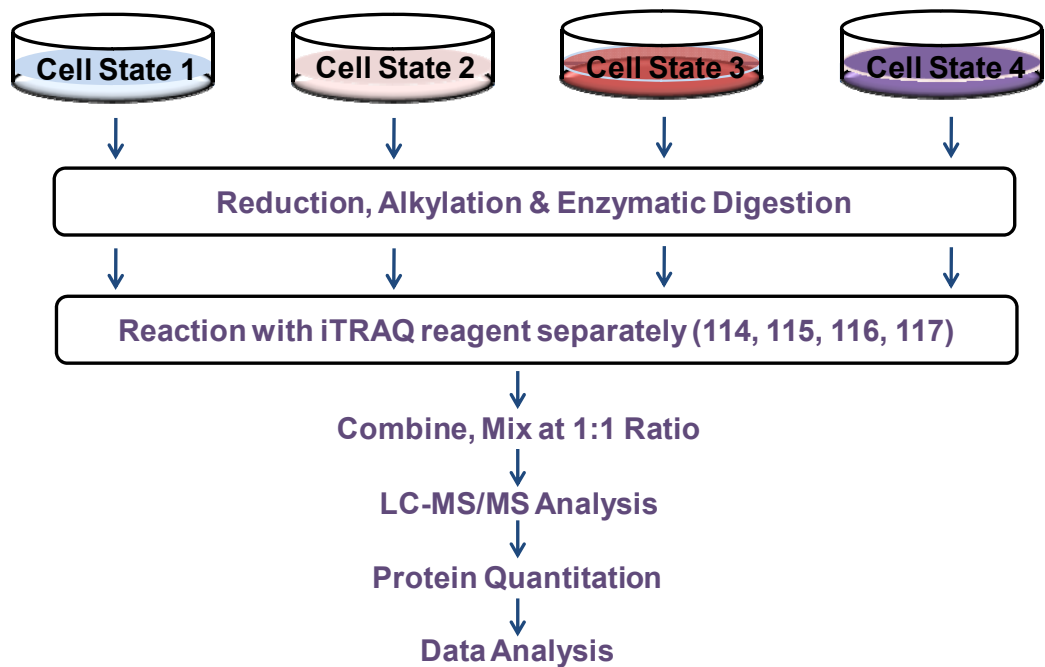
be differentiated in MS survey spectra (MS1). This nature of iTRAQ technique results in concomitant increase in precursor ion intensity and reduction in sample complexity relative to multiplexed, precursor-based quantitative methods such as ICAT or dimethyl labeling. Therefore, more MS/MS spectra can be acquired for protein identification and quantification, which is especially helpful for current mass spectrometer with relatively slow scanning rate. It has been reported that larger numbers of proteins can be identified and quantified by iTRAQ labeling relative to other labeling techniques including SILAC or ICAT (15). 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.   However, iTRAQ labeling is usually performed at the peptide level, which is in the late stage of the sample preparation.   Therefore, the late combination of samples is prone to additional experimental errors and renders iTRAQ quantification less accurate than SILAC(15).

1.2.1.4 Absolute quantification using isotope-labeled internal standards

The use of isotope-labeled synthetic standards based on stable isotope dilution theory has a long history in mass spectrometry analysis. For quantitative proteomics, this strategy is generally referred as AQUA (absolute quantification of proteins) (35). In AQUA strategy, a known quantity of AQUA peptide, which is simply a synthetic heavy-isotope labeled tryptic peptide corresponding to a normal peptide of interest, is added to a biological protein sample.   The mixture is then digested and analyzed by LC-MS/MS (Figure 1.4). The AQUA peptides are chemically identical to their native counterparts

Figure 1.4 General procedures of AQUA-based quantitative proteomics.

Cell State 1

↓

Cell Lysate

↓

Add AQUA peptides

↓

Digestion

↓

LC-MS/MS Analysis

↓

Protein Quantitation

↓

Data Analysis
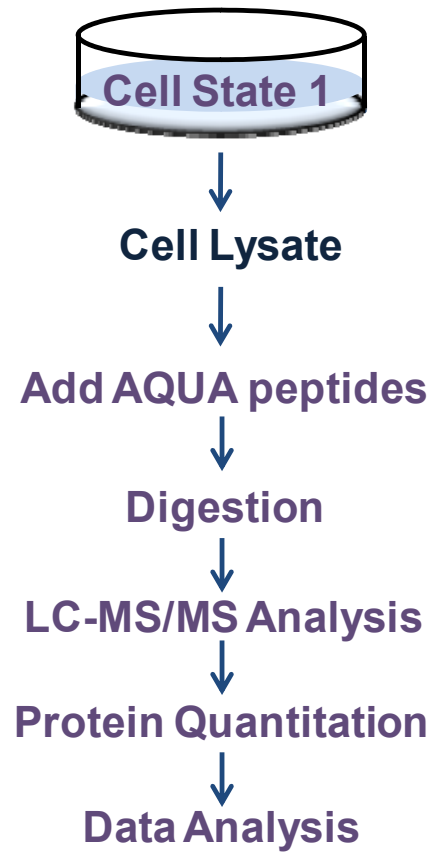
derived from protein mixture, but the labeled and native forms of these peptide pairs can be easily differentiated by molecular weight in MS analysis. The absolute quantity of native peptide as well as corresponding proteins can be subsequently determined by integrating the MS peak areas of the native peptide and the synthetic AQUA peptide.

Unlike metabolic or chemical labeling methods, AQUA cannot provide global quantification for the whole proteome. AQUA is generally used to quantify a very small number of proteins of interest; therefore, AQUA is specifically attractive for those targeted proteomics experiment which repeated measurements are required. In recent years, AQUA in combination with multiple reaction monitoring mass spectrometry (MRM-MS) has been widely used for analysis and validation of potential biomarkers in clinical samples (36).

1.2.2 Protein quantification in discovery MS mode and targeted MS mode

1.2.2.1 Shotgun or discovery proteomics

In quantitative proteomics, "shotgun" or discovery proteomics is the most widely used approach. In shotgun proteomics, proteins are first digested to peptides and the resulting peptide mixture is separated with different HPLC, followed by LC-MS/MS analysis (37). Because there are many peptides that coelute from the column and reach the mass spectrometer at the same time, the MS analysis is often operated in data-dependant acquisition (DDA) mode(13). In this mode, the mass spectrometer continuously repeats a data acquisition cycle consisting of a full-scan mass spectrum, followed by selection and fragmentation of the nth (typically 10-20th) most abundant m/z

ratios for peptide identification and quantification (Figure 1.5). The first full-scan mass spectrum is generally obtained at high resolution, with a typical value of 60,000 for Orbitrap mass analyzers and 20,000 for time-of-flight analyzers. The ultra high resolution and accuracy in full-scan MS allows for unambiguously distinguishing peptides with similar nominal molecular mass, which is crucial for both peptide quantification and identification in complex proteome samples (16, 38). In the following MS/MS analysis, the fragment ions from auto-selected peptide were detected in either high-resolution or low-resolution mode (39).

Coupling with several stable isotope labeling techniques, shotgun or discovery proteomics achieved great success by enabling large-scale quantitative comparison of protein expression or PTM level between biological samples. For example, Mann and coworkers compared protein levels of essentially all endogenous proteins in haploid yeast cells to their diploid counterparts by shotgun proteomics, which spans more than four orders of magnitude in protein abundance with no discrimination against membrane or low level regulatory proteins (40). The Gygi lab can routinely detect more than 30,000 phosphorylation sites from human or mouse tissue by shotgun proteomics due to the advances in instrument performance, computational software, and sample preparation (41). Additionally, shotgun proteomics does not require any prior knowledge of the composition of proteome samples, and thus it provides the potential to uncover novel sites and proteins.

Figure 1.5 Typical CID-MS/MS scan events in Orbitrap-Velos mass spectrometer for shotgun proteomics (adopted form Ref (39)).
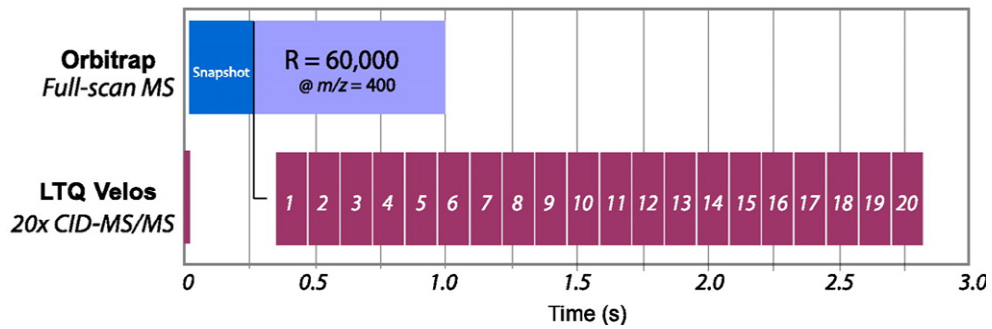
1.2.2.2 Targeted proteomics by multiple reaction monitoring mass spectrometry (MRM-MS)

Sample complexity, together with inherent variation in automated peak selection, results in compromised sensitivity and reproducibility for protein quantification using shotgun proteomics strategy. As a result, only partially overlapping sets of proteins can be identified even from substantially similar samples in data-dependent analysis (42). The inadequate sensitivity and reproducibility of shotgun proteomics strategies severely hamper their utility in biomarker discovery and clinical studies. As an alternative to a data-dependent operation of the mass spectrometer, targeted proteomics by multiple reaction monitoring mass spectrometry (MRM-MS) has gained increasing popularity in recent years (43, 44). MRM-MS analysis is often performed on triple quadrupole (QQQ) MS instrument. In MRM mode, the first and the third quadrupoles act as mass filters to specifically select predefined m/z values corresponding to the peptide ion and a specific fragment ion of the peptide, whereas the second quadrupole serves as the collision cell (Figure 1.6A). The specific pair of m/z values associated with the precursor and fragment ions selected is referred to as a "transition" and generally at least three transitions are required to accurately quantify a single peptide. It is important to note that MRM requires previously acquired tandem mass spectra of target peptides to define transitions for MRM analysis. These transitions can be determined either from synthetic peptides (45), or from high-quality MS/MS spectra generated by previous large-scale discovery proteomics experiment(46).

Figure 1.6 (A) SRM/MRM analysis on triple quadrupole (QQQ) mass spectrometer. (B)

Different MS operation modes for shotgun proteomics and targeted proteomics.

(A)
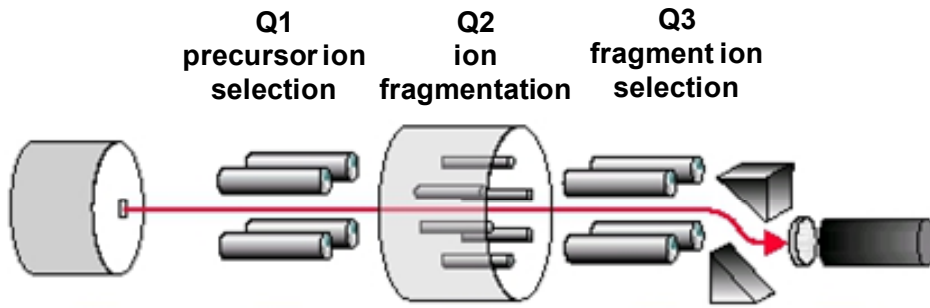
(B)
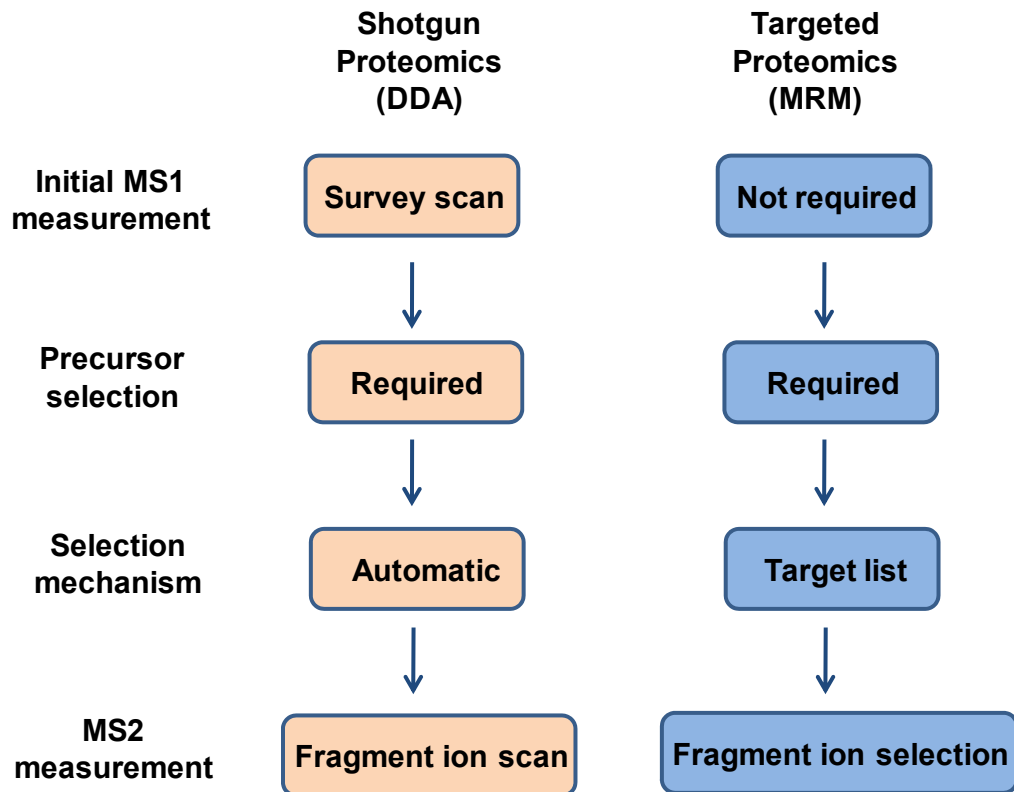
The MRM-based targeted MS analysis permits rapid and continuous monitoring of specific ions of interest, which enhances the sensitivity for peptide detection by up to 100-fold as compared with MS analysis in DDA-based discovery mode(25). More importantly, unlike shotgun proteomic technique, no full mass spectra are recorded in QQQ-based SRM analysis, which results in a linear response over a wide dynamic range up to five orders of magnitude (Figure 1.6B) (25). Thus, the MRM-based targeted proteomic approach often generates much more reproducible quantification results than shotgun proteomics, which renders MRM-MS particularly attractive in biological applications requiring repeated measurement of predefined set of proteins (25). Currently, the key challenge for MRM-based targeted proteomics lies in increasing throughput for large scale peptide quantification (47-49), as well as the development of robust statistics to filter false-positive background signals (50).

1.3 Affinity-based proteomics for study of ATP-, GTP- and other nucleotide-binding proteins

Nucleotide-binding protein, such as ATP-, GTP-binding proteins are one of the most important families of proteins that are involved in a variety of pivotal cellular processes, like cell signaling, proliferation, differentiation, and apoptosis. For example, numerous ATP-binding cassette transporters (ABC transporters) catalyze the hydrolysis of ATP to provide required energy for translocation of various substrates across cell membranes (51). Additionally, kinases, as one of the largest enzyme superfamilies in higher eukaryotes, mediate cellular protein and lipid phosphorylation to regulate

18

downstream signaling cascade.(52) Moreover, GTP-binding proteins are also important signal transducing molecules in cells (53). Heterotrimeric G proteins and small GTPases, along with G protein-coupled receptors and protein kinases, mediate important signal-transduction cascades in almost every organism. A variety of elongation factors, which are also GTP-binding proteins, play an important role in initiation, elongation and termination of protein biosynthesis (53).

Despite the importance of nucleotide-binding proteins in cellular functions, the current picture of nucleotide-protein interactions is far from complete. Individual nucleotide-binding proteins can be characterized using immunoblot (54); however, global study of the structure, function, and expression level of nucleotide-binding proteins *in vivo* can hardly be achieved with the use of conventional methods owing to enormous diversity of the nucleotide-binding protein family. Therefore, comprehensive identification and quantification of ATP/GTP-binding proteins as well as dynamic analysis of nucleotide-protein interactions at the proteomic scale are important for understanding better the regulatory mechanisms of nucleotide-binding proteins.

The development of mass spectrometry (MS) instrumentation and bioinformatic tools provides the opportunity to identify and quantify a large number of proteins in complex samples. Up to several thousand proteins can be routinely identified and quantified using MS (19), which could greatly facilitate the extensive study of the perturbations of protein expression in an organism upon external stimuli. However, proteomic studies of specific family of proteins, including nucleotide-binding proteins, by

MS are still a big challenge owing to the extreme complexity of the proteome and the relatively low abundance of some proteins. For example, although the group of human kinase is the largest enzyme complement in mammals, it only constitutes 1.7% of the human genome (52). This limitation can be partially overcome by combining MS with powerful separation techniques, such as polyacrylamide gel electrophoresis (PAGE) (55) or multi-dimensional liquid chromatography (56). However, none of these approaches permit selective enrichment of nucleotide-binding proteins from cellular extracts. Therefore, selective enrichment of ATP/GTP-binding proteins from cellular extracts is essential for the comprehensive identification and quantification of ATP/GTP-binding proteins.

In this section, I summarize general and widely used affinity approaches to capture and detect nucleotide-binding proteins, as well as the applications of these techniques in global profiling of nucleotide-binding protein.

1.3.1 Global kinome enrichment platforms

Protein phosphorylation, one of the most important types of post-translational modifications (PTMs), is catalyzed by protein kinases (collectively referred to as the kinome), which are encoded by over 500 genes in higher eukaryotes (Figure 1.7) (52). Aberrant expression and/or activation/deactivation of kinases have been implicated as among the major mechanisms through which cancer cells escape normal physiological constraints of cell growth and survival (57). Additionally, dynamic kinome reprogramming has been found to be closely associated with resistance toward cancer

Figure 1.7 The dendrogram of the human kinome (adopted form Ref (52)).
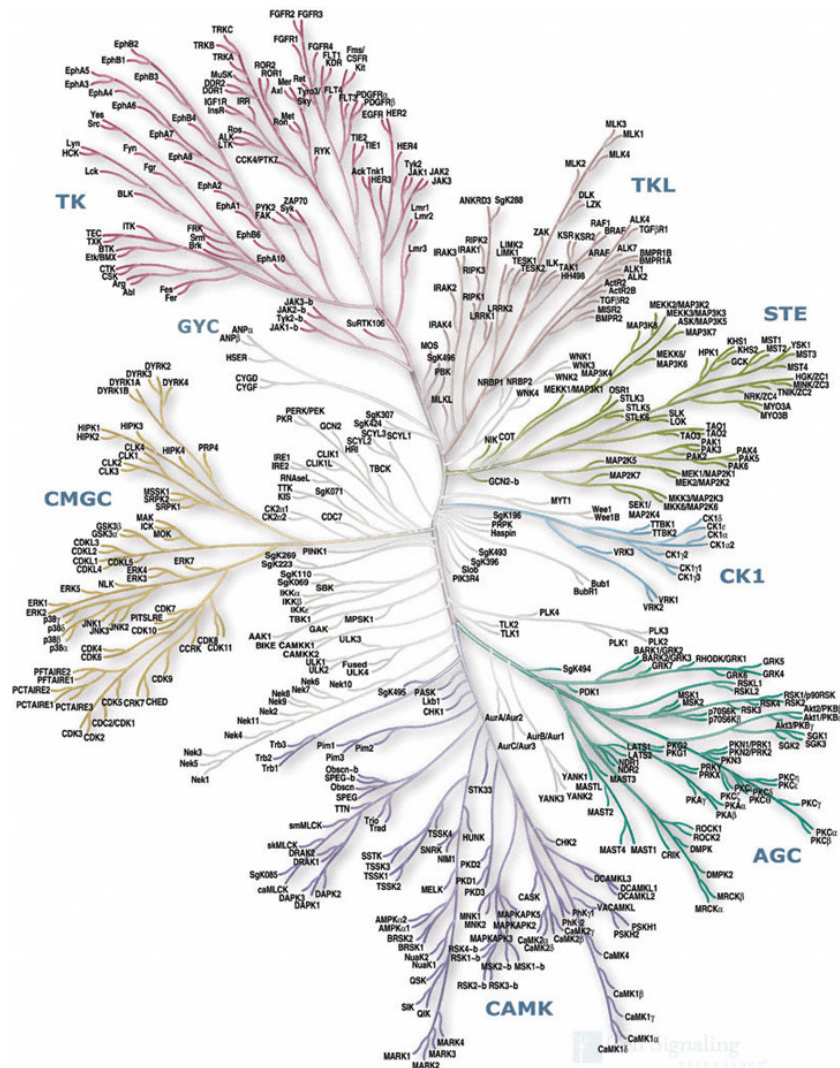
chemotherapy (58). Owing to their crucial roles in cancer development, kinases have become one of the most intensively pursued enzyme superfamilies as drug targets for cancer chemotherapy and more than 130 distinct kinase inhibitors have been developed for phase 1-3 clinical trials (59). Recently, inhibitor potency and selectivity for more than 400 kinases have been reported, which provided a comprehensive target-inhibition profile for the majority of the human kinome (60-62). Therefore, the kinome-inhibitor interaction networks coupled with comprehensive profiling of global kinome expression and activity associated with certain types of cancer could be invaluable for understanding the mechanisms of carcinogenesis and for designing rationally novel kinase-directed anti-cancer chemotherapies.

Unfortunately, currently there is no optimal strategy for profiling the expression levels of the entire kinome at the protein level. Traditional methods for measuring kinase expression rely primarily on antibody-based immunoassays due to their high specificity and sensitivity (63). The immunoassays, however, are limited by the availability of high-quality antibodies; therefore, these methods are only useful for assessing a small number of kinases in low-throughput. The development of mass spectrometry (MS) instrumentation and bioinformatic tools enables the identification and quantification of a significant portion of the human proteome from complex samples (3). However, proteomic studies of global kinome by MS are still very challenging, which is largely attributed to the fact that, similar as other regulatory enzymes, protein kinases are generally expressed at low levels in cells (64). This analytical challenge is further

22

aggravated in shotgun proteomics approach where even more complex mixtures of peptides instead of proteins from whole cell or tissue extracts are analyzed (44). Therefore, selective enrichment of protein kinases from cellular extracts is essential for the comprehensive identification and quantification of the global kinome.

1.3.1.1 Kinome enrichment by affinity capture chromatography

The use of affinity chromatography is increasingly seen as a choice for purification of a specific protein or a group of sub-proteome from complex mixtures. Although there are more than 500 protein kinases spreading over nine categories, kinases share several structural and functional features in common. A very important attribute of protein kianses is that they all employ ATP as the phosphate donor. Therefore, all protein kinases exhibit strong ATP binding affinity. Additionally, the availability of the crystal structures of many kinase-ATP complexes further reveals that most kinases share very similar ATP binding domains. In ATP binding domains of many protein kinases, the adenine ring of ATP is buried deeply in a hydrophobic cleft, while the α, β, and γ phosphates interact with residues near the mouth of the cleft (65). The interaction of the adenine ring and α and β phosphates with conserved amino acid residues is important for ATP binding, hence γ phosphate modified ATP is considered to best mimic the native nucleotide triphosphate to study interactions between the ATP and kinases. Based on these observations, purification of protein kinases by ATP affinity resins has been developed by several vendors (Figure 1.8).

Figure 1.8 Structure of some ATP-affinity Resin with (A) or without (B) spacer

(A)
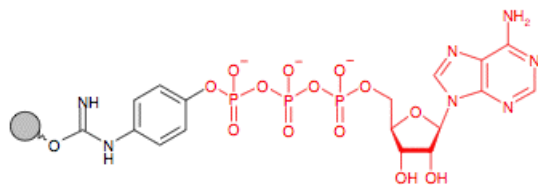
(B)
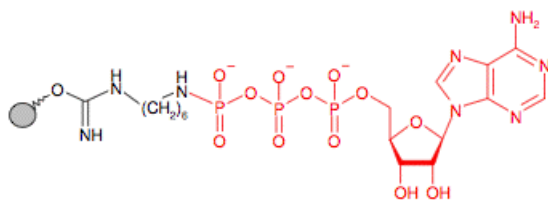
In earlier studies, researchers used affinity-phosphate-linked ATP-sepharose beads to bind and isolate the entire purine binding proteome from an mouse or human cell lysate (66). When the bounded proteins were displaced by 20 mM NADH, AMP, or ADP, respectively, only a low amount and a small number of protein bands were observed from the supernatants. In contrast, 20 mM ATP elution from sample yielded an abundant complex sample. Importantly, if ATP was linked to Sepharose through adenosine at N6 (N-6 linked resin), very few proteins were recovered from protein extracts. From mouse and human whole red blood cells, total of 72 mouse and 15 human ATP binding proteins were identified using ATP-sepharose beads. Bound proteins identified include protein and nonprotein kinases, dehydrogenases, DNA ligases, mononucleotide ATPases, and nonconventional purine-binding proteins. In a similar study, the same approach was applied to study the ATP-binding proteome of plant mitochondria (67).

Due to the competitive binding by other ATP-binding proteins, enrichment methods using immobilized ATP as the ligand resulted in the capture of fewer than ten kinases and was instead dominated by the binding of highly abundant ATPases, such as heat shock proteins (68). To overcome these limitations, Bantscheff et al. (68) developed an affinity resin immobilized with a combination of seven kinase inhibitor, which was named as kinobeads, to efficiently discriminate between protein kinases and other classes of ATP-binding proteins. These kinase inhibitors immobilized on kinobeads were chosen to display little selectivity and interacted with kinases from different kinome group;

25

therefore, a large portion of the expressed kinome can be efficiently captured. Using mass spectrometric analysis, Bantscheff et al. (68) were able to identify a total of 174 and 183 protein kinases, from HeLa and K562 cells, respectively, in single pull-down experiments. This group also found slightly higher coverage within the tyrosine kinase family compared to serine/threonine kinase family, since most of the immobilized kinase inhibitors were designed to target tyrosine kinases. Although kinobeads also bind some ATP- and purine-binding proteins such as chaperones, helicases, ATPases, it was estimated that kinases account for almost 80% of the total captured protein amount based on the total mass spectrometric signal.

Recently, Henrik et al. (69) developed a similar kinome enrichment strategy by loading the cell lysates onto a series of three affinity columns displaying a total of five different immobilized non-selective kinase inhibitors (Figure 1.9). The relative expression of more than 200 protein kinases from S and M phase-arrested human cancer cells was successfully quantified (69). In addition, Zhang et al. (70) introduced another sepharose-supported kinase capture reagent immobilized with a novel kinase inhibitor bisanilino pyrimidine, CTx-0294885 to enables the identification of large proportions of the expressed kinome by mass spectrometry.

1.3.1.2 Kinome enrichment by chemical reactive affinity probes

Activity-based proteomics (ABP), or activity-based protein profiling (ABPP) is the use of chemical reactive affinity probes to study a specific family of subproteome based on their unique structural or functional similarities (71). Therefore, at the heart of

Figure 1.9 Protein kinase enrichment by multicolumn affinity chromatography immobilized with five different non-selective kinase inhibitors (adopted form Ref (69)).
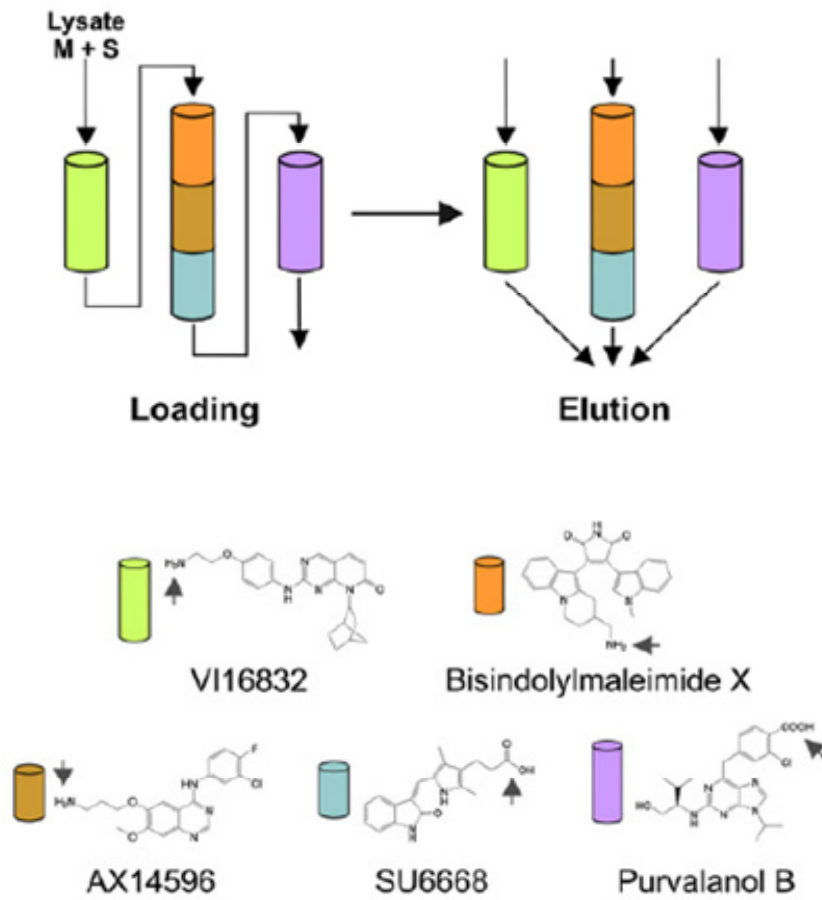
the ABP technology are the small molecules known as chemical reactive affinity probes that are designed to interact with active-site residues of enzymes resulting in the formation of a stable covalent bond. A typical chemical reactive probe usually consists of at least three basic elements(72) (Figure 1.10): (1) a binding group for binding the active sites of a given enzyme class (2) a reactive functional group that conjugates with the protein active site to form a stable covalent bond (3) a chemical tag or reporter group, like biotin and/or a fluorophore, used to facilitate the rapid isolation and detection of probe-labeled enzymes.

Recently, activity-based proteomics involving specific labeling of proteins with functional similarities has emerged as an important technique in targeted detection of protein kinases. In this context, the entire proteome is treated with a chemical affinity probe, which incorporates an enrichment tag to a specific functional group on protein kinase family to facilitate downstream purification. For example, in pioneering work, 5'-*p*-fluorosulfonylbenzoyladenosine (FSBA), a reactive ATP analog, was employed as an activity-based probe to target nucleotide-binding proteins from whole cell lysates (73). As other chemical reactive probes, FSBA contain three components: a reactive group (fluorosulfonyl), a binding moiety (adenosine), and a tag (sulfonylbenzoic acid). FSBA modifications mainly occurred on tyrosine (66.98%) and lysine (32.53%), rarely on serine (0.48%), but not on histidine or cysteine. When applied to a whole proteome, 185 different FSBA-labeled sites in a human Jurkat cell lysate were identified, including six kinases. Although FSBA is a poor mimic for nucleotide with weak binding affinity for

Figure 1.10 The general structure of a chemical reactive affinity probe
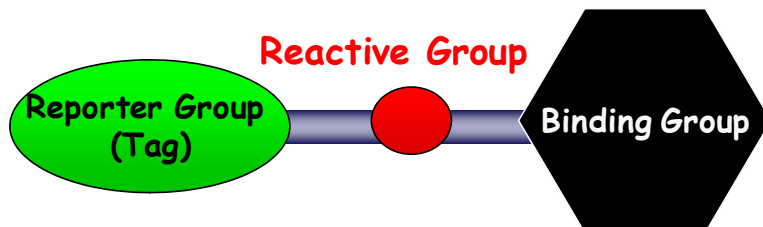
ATP binding sites, this approach highlighted the potential of activity-based protein profiling in kinome studies.

Around 2007, we and others reported the application of biotin-conjugated acyl nucleotide probe for the enrichment and identification of ATP-binding proteins, especially protein kinases, from complex protein mixtures (74, 75). Bioinformatics studies showed that most ATP/GTP-binding proteins carry a consensus amino acid sequence motif termed phosphate-binding loop (P-loop), which constitutes the nucleotide-binding site and is responsible for ATPase/GTPase activity (76). A conserved motif of GxxxxGK, in which 'x' represents any amino acid, is often found in the P-loop region of ATP/GTP-binding proteins (77). In addition, sequence comparisons have shown that virtually all protein kinases have at least one conserved lysine residue within their active sites (78). An invariant lysine residue involving in ATP binding can be localized in a region of subdomain II for all protein kinases except WNK kinase. In addition, most kinases including WNK kinase show another conserved motif of HRDxKxxN located in subdomain VIB, which participates in ATP binding (79, 80). Cocrystal structures of protein kinase catalytic domains bound to ATP reveal that these lysine residues are positioned in the proximity of the β- and γ-phosphates of bound ATP (Figure 1.11). In light of these previous findings, an ATP analog bearing an acyl phosphate moiety is developed to target the lysine residue in the nucleotide-binding site (Figure 1.12A). This ATP-affinity probe contains three components, namely, a binding moiety consisted of nucleotide of interest (i.e. ATP), an enrichment moiety of biotin or its

analog desthiobiotin which facilitates downstream purification, and an acyl phosphate group which targets side-chain amino group of the P-loop lysine residue to form a stable amide bond. When the ATP affinity probe interacts with ATP-binding proteins, the ATP moiety binds to the P-loop region, which facilitates the ε-amino group of the P-loop lysine residue to react with the acyl phosphate component to yield a stable amide bond (Figure 1.12B).

After labeling reaction, the labeled proteome can be processed in two ways. First, similar as pull down procedure using kinobeads, the labeled proteins are isolated and enriched by avidin affinity purification. The purified proteins are subsequently digested with trypsin, and the resulting peptide mixtures are analyzed by LC-MS/MS for target identification. With enrichment by ATP-affinity probe, around 100 kinases can be readily detected using 1 mg of cellular protein from K562 cell lysates. The kinobead pulldown sample using the same amount of K562 cell lysates resulted in the identification of similar number of protein kinases (81). However, the kinobeads exhibit much higher selectivity for kinases than the ATP probe. More than 40% of the assigned spectra from kinobeads originate from kinases, whereas only 15% of the assigned spectra from the ATP probe stem from proteins associated with kinase activity. These observations can be easily explained by the way the two affinity reagents are designed. The ATP probe can potentially label any ATP binding protein; however, kinobeads are designed to target protein kinases with relatively high selectivity. This notion is further elaborated by the fact that 44% of the spectra from the ATP probe experiment belong to

Figure 1.11 The proximity of the two conserved lysine residues to the phosphate groups

of ATP (adapted form Ref. (75)).
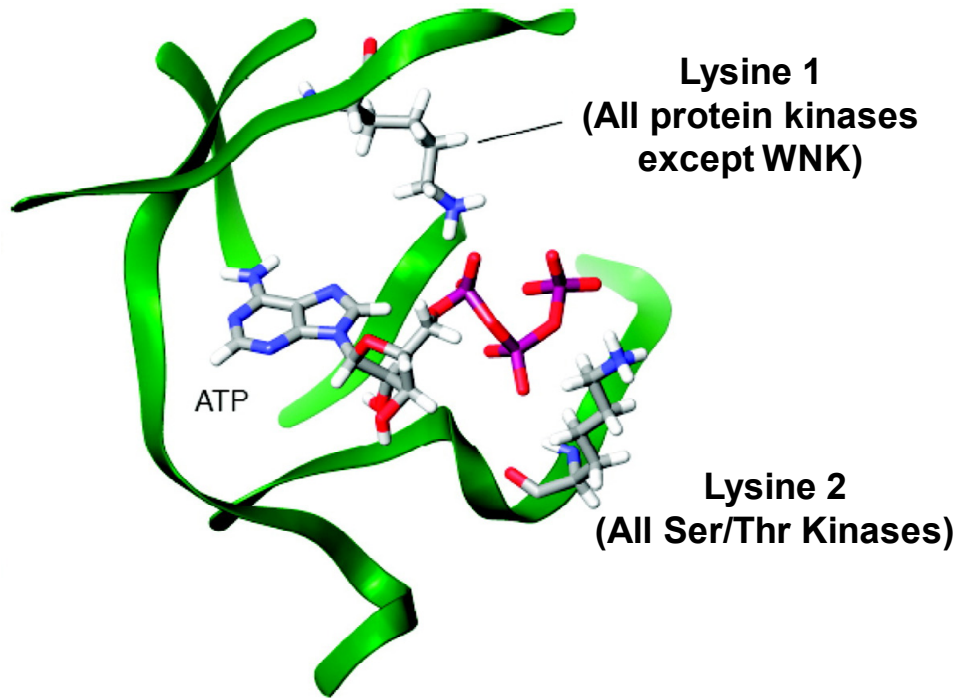
Figure 1.12 (A) Structures of ATP-affinity probes with biotin or desthiobiotin tag (B) A schematic diagram showing the reaction between ATP-affinity probe with an ATP-binding protein.

(A)

Biotin-LC-ATP   **Desthiobiotin-ATP**
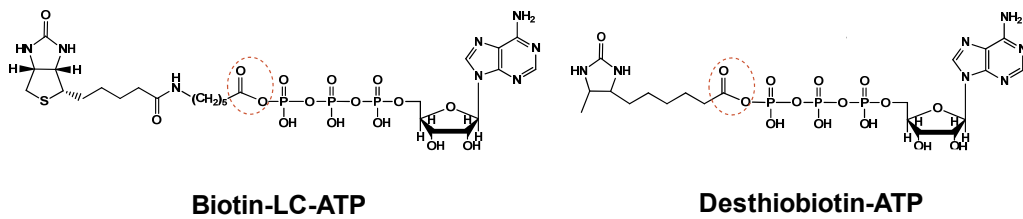
(B)
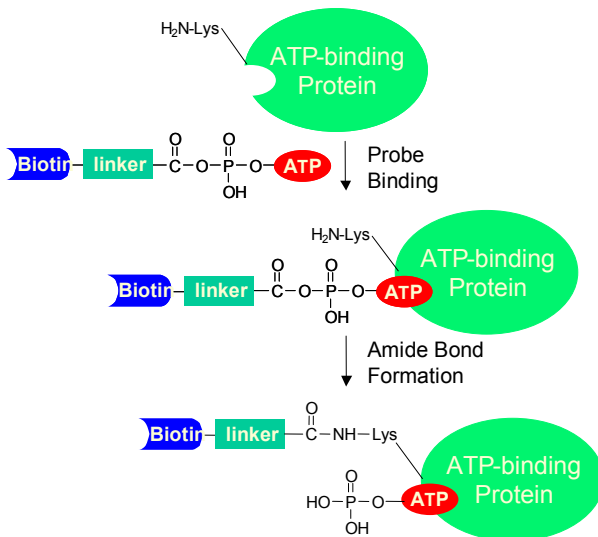
proteins with described nucleotide binding activity, whereas this figure is only 12% in the experiment using kinobeads. In line with this finding, the author also observed that increasing the amount of input material facilitated the detection of more kinases using kinobeads; however, increasing the sample amount beyond 1 mg quantity in experiments using the ATP/ADP probes led to detrimental effects because the background of contaminating proteins increased significantly (81). Moreover, a significantly higher number of tyrosine kinases (TKs) are found to be enriched using kinobeads, whereas members of the STE kinase group (mainly covering the MAP kinases) are better represented by the ATP probes. This observation may not be surprising as several of the small molecule inhibitors immobilized on kinobeads were originally developed to target tyrosine kinases. Based on this observation, an integrated enrichment procedure combining these two methods using the same sample was developed, where cell lysate were first enriched by kinobeads and unbound protein fraction was then subjected to a second pulldown using the ATP probe and avidin agorose. This sequential enrichment leads to an increase of 27% in protein kinases discovery (81).

Instead of enrichment in the protein level, the biotin-labeled ATP-binding proteins can be first digested with trypsin to peptides, some of which are probe-modified. Those biotin-labeled peptides can then be purified by avidin agarose and analyzed by LC-MS/MS to determine the identities of the labeled proteins as well as the labeling sites. In this vein, it is worth noting that biotin conjugation with the side chain of cysteine or lysine was found not to result in significant alteration in peptide backbone fragmentation

upon low-energy collision activation and peptide identification based on the resulting MS/MS (82, 83). Distinguished from kinase capture strategy using kinobeads, this enrichment and analysis workflow for ATP affinity probe-labeled peptides should allow for the determination of lysine residue at the nucleotide binding site of enriched proteins, which could potentially provide invaluable information about site-specific interactions between ATP and its binding proteins. Using this workflow, around 90 protein kinases can be routinely identified in single proteome (74, 75). In addition, in a large-scale analysis of more than 100 human, mouse, rat, and dog proteomes using this workflow with ATP affinity probe, 322 different protein kinases were identified with 77% (247 kinases) being labeled on one of the two conserved active site lysines described above(75). However, the identification and quantification accuracy for this workflow with enrichment in peptide level may be compromised, because only single lysine on each kinase is often labeled generating a single labeled peptide for protein determination.

1.3.1.3 Application of Kinome profiling techniques

These kinome enrichment approaches exhibit considerable potential in future signal transduction and kinase drug target analysis. First, combination of kinase-selective enrichment with quantitative MS techniques, such as SILAC or iTRAQ reagents could enable comprehensive profiling of the expression or activity for global kinome. For example, using kinobeads, James et al. (58) assessed kinome activity in response to MEK inhibition in triple-negative breast cancer (TNBC) cells and genetically engineered mice (GEMMs). They found that MEK inhibition caused acute ERK activity loss, resulting in

rapid c-Myc degradation that induced over-expression of several receptor tyrosine kinases (RTKs), which suggested that a kinase inhibitor combination therapy could effectively reduced single agent resistance. Moreover, coupled with the further phosphopeptide purification technique, these kinome enrichment strategies can also allow for global survey for the activity of kinome. For example, kinobeads and TiO_2 phosphopeptide enrichment enabled the quantification of activities for 219 protein kinases with more than 1000 phosphorylation sites from S and M phase-arrested human cancer cells (69). Similarly, McAllister et al. (84) successfully compared the activities for more than 200 kinases across the six cell lines, representative of different breast cancer clinical subtypes by performing phosphopeptide enrichment for these ATP-probe labeled proteins.

On the other hand, these kinome enrichment strategies allow for the parallel quantitative determination of protein-affinity and -selectivity profiles for kinase inhibitors in any cell type or primary tissue. Recently, inhibitor potency and selectivity for more than 400 kinases have been reported (60, 61); however, these data relied on recombinant expression and purification of full-length or catalytically active fragment of kinases, which may not be capable of reflecting the complexity of kinase function and regulation in native biological systems. Therefore, it is highly desirable to perform these protein-kinase inhibitors interaction studies using native protein extracts. To this end, different concentrations of drugs are titrated into a protein extract, and subsequently kinase enrichment is performed by either kinobeads or ATP probes. The free inhibitor in the

lysate competes with the affinity matrix binding or affinity labeling for the kinase target. Kinases with high affinity to the free inhibitor will therefore show a reduced binding to the affinity matrix or labeling efficiency, which can be identified and quantified by mass spectrometry. For example, Matthew et al. (85) developed an ATP probe-based platform called KiNativ to profile several well studied kinase inhibitors against >200 kinases in native cell proteomes and revealed that biological targets for some of these inhibitors were strikingly different between native and recombinant kinase inhibitory profiles. Similar kinase inhibitory profiling experiment was also conducted using kinobeads (68).

1.3.2 GTP-binding protein enrichment platforms

GTP-binding proteins are part of a superfamily that comprises of more than 100 proteins, which includes heterotrimeric G proteins, small GTPases and a variety of other GTP-binding proteins such as elongation factors and tubulins. For example, the dynamic interplay of GDP-bound inactive and GTP-bound active form of heterotrimeric G proteins is closely related to signal amplitude by G-protein-related pathway (86). The regulation of the activity of small GTPases plays vital roles in cell signaling and their dysregulation is believed to be closely associated with the development of various types of cancer (87). However, the properties and functions of a number of GTP-binding proteins are still unknown because of the low abundance of GTP-binding target. To achieve functional proteomic analysis of GTP-binding proteins, efficient separation of a family of GTP-binding proteins would be highly desirable.

The obvious choice of a scaffold for a GTP affinity beads or probes is GTP nucleotide, since GTP is a common ligand for all GTPases with reasonable affinity at a conservative motif sequence. To this end, GTP-immobilized affinity beads are commercially available. In addition, a variety of chemical reactive GTP analogues as affinity probes have also been developed. For example, similar as acyl phosphate based ATP probe, desthiobiotin-GTP probe was also synthesized to enable selective labeling and enrichment of small GTPases and large G-protein subunits (88). Kaneda et al. (89) developed a simple and efficient photoaffinity method with the use of diazirine-carrying GTP analogue to facilitate proteomic analysis of GTP-binding proteins. Recently, Elizabeth et al. (90) synthesized and characterized a photoreactive GTP-BP-yne probe affinity probe that covalently photocross-links to protein targets and has an alkyne handle for click chemistry conjugation to reporter tags, which facilitated the identification of more than thirty GTP-binding proteins by mass spectrometry, including small GTPases and members of the GTP1/OBG family.

Small GTPase or Ras superfamily, with the molecular weight of 21 kDa, is one of most important GTP binding proteins. It has been revealed that high conserved amino acid sequence (30-55% homologous to each other) exist in at least five families, namely, the Ras, Rho, Rab, Sar1/Arf, and Ran families (53). In addition to the known p-loop motif present in most GTP-binding proteins, there is a unique amino acid sequence CAAX (A represents aliphatic acid, X represents any amino acid) near the C-termini of small GTPases, which has been post-translationally modified with lipid, such as farnesyl

or geranylgeranyl group. This modification facilitated the translocation of Ras protein from the cytosol to cytoplasmic face of the plasma membrane (91). Based on this unique post-translational modification (PTM), a detection and enrichment of small GTPases involving the metabolic incorporation of a synthetic azido-farnesyl analog was developed (92). The farnesyl group for post-translational modification of Ras protein is from endogenous synthesis of farnesyl pyrophosphate (FPP). To replace endogenous FPP, the FPP analogues, FPP-azide and F-azide-OH, have been synthesized and incubated with tumor cells for metabolic labeling (Figure 1.13). Since the similar structure of FPP and FPP-azide, no difference of enzymatic farnesylation activity of FPP-azide from FPP has been observed. Next, the F-azide-modified proteins are selectively ligated to a phosphine biotin probe via the Staudinger reaction. Finally, the farnesylated protein-phosphine biotin conjugates are affinity purified by avidin beads and detected by liquid chromatography mass spectrometry (LC-MS). Preliminary studies have showed that 18 Ras superfamily proteins, including H-ras, K-ras, N-ras, Rheb, and Rap2, were identified by using this approach.

1.4 Scope of the dissertation

Nucleotide-binding proteins such as ATP- and GTP- binding proteins play pivotal roles in many cellular processes. However, targeted study of nucleotide-binding proteins, especially protein kinases and GTPases, remained challenging, which is mainly because of the lack of efficient enrichment approaches to selectively capture nucleotide-binding proteins. In this dissertation, we reported a general strategy in using affinity-labeled

39

Figure 1.13 Scheme of farnesylated proteins detection by F-azide metabolic labeling, Biotin-phosphine probe conjugation and avidin beads purification. (A) Farnesylation of protein by FPP and FPP-azide (B) structure of FPP and FPP-azide (C,D) Biotin phosphine probe conjugation with azide-F modified protein (E) Purification of F-azide modified protein by avidin beads. (adapted form Ref(92)).

(A)
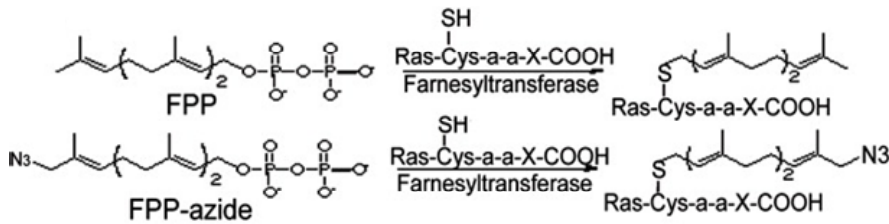
(B)
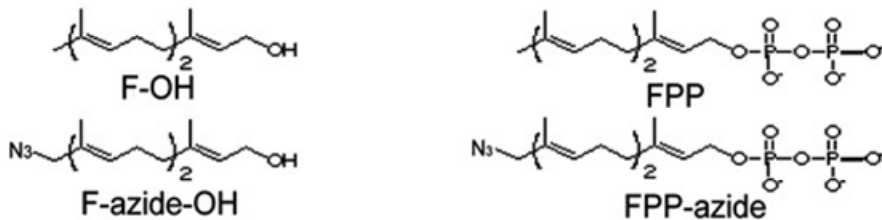
(C)
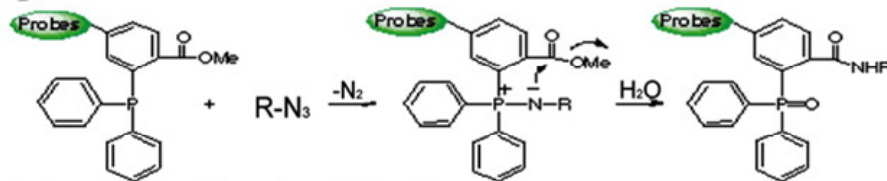
Probes = Biotin or Photo-cleavable beads.
R-N3 = a F-azide-modified protein.

(D)
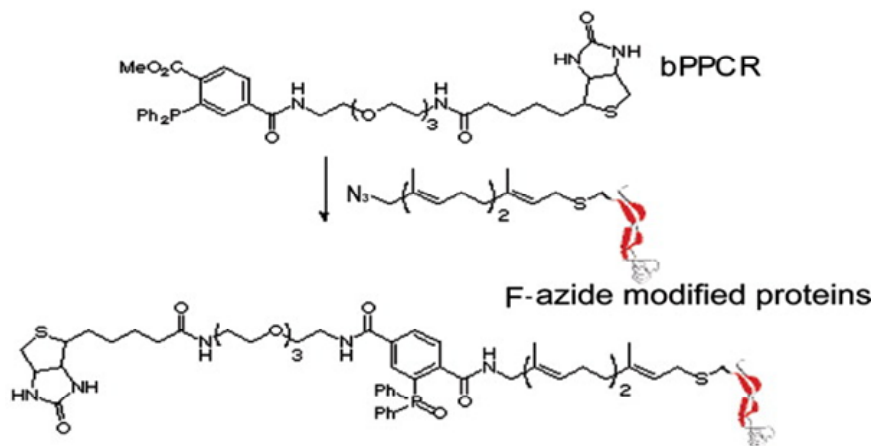
(E)
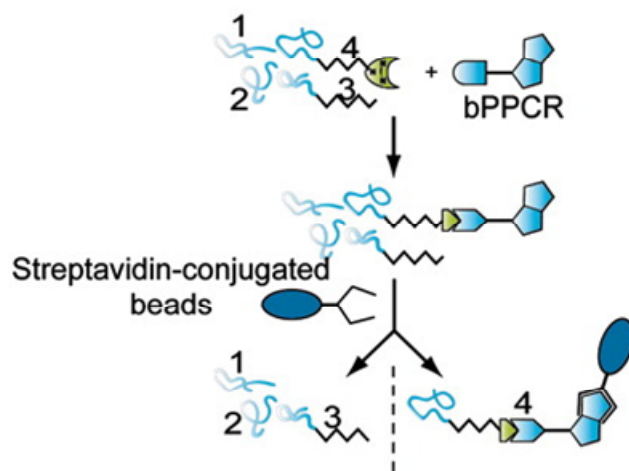
chemical probes to enrich, identify, and quantify ATP- and GTP- as well as other nucleotide binding proteins in the entire human proteome.

In Chapter two, we mainly focused on the method development. We described the synthesis and application of biotin-conjugated ATP and GTP affinity probes. The ATP/GTP affinity probes facilitated the identification of 100 GTP-binding proteins and 206 kinases with the use of low mg quantities of cell lysate. Additionally, our strategy led to the identification of three and one unique nucleotide-binding motifs for kinases and GTP-binding proteins, respectively. In combination with the use of SILAC-based quantitative proteomics method, we also assessed the ATP/GTP binding selectivities of nucleotide-binding proteins at the global proteome scale. Our results confirmed known and, more importantly, unveiled new ATP/GTP-binding preferences of hundreds of nucleotide-binding proteins.

In Chapter three, we applied this nucleotide-affinity probe to facilitate the protein-nucleotide interaction study. An affinity profiling strategy to comprehensively characterize ATP-protein interactions at the entire proteome scale was developed by comparing the labeling behaviors of lysine residues with the use of low and high concentrations of the ATP affinity reagents, which could effectively minimize false-positive identification of ATP-binding sites derived from non-specific labeling. This novel quantitative ATP-affinity profiling strategy is particularly useful for unveiling previously unrecognized nucleotide-binding property and binding sites in whole proteome. In addition, we introduced a novel isotope-coded ATP-affinity probe (ICAP)

as acylating agent to simultaneously enrich and introduce isotope label to ATP-binding proteins. Chapter three built a solid foundation for further studies in Chapter four and five.

In Chapter four, we extended the use of ATP/GTP affinity probes and introduced an orthogonal strategy encompassing the nucleotide-affinity profiling assay and nucleotide-binding competition assay to comprehensively characterize ^{S}GTP-binding proteins along with the specific binding sites from the entire human proteome. With the simultaneous use of ^{S}GTP and GTP affinity probes, we identified 165 ^{S}GTP-binding proteins that are involved in several different biological processes. We also examined the binding selectivities of these proteins toward ^{S}GTP and GTP, which allowed for the revelation of the relative binding affinities of the two nucleotides toward the nucleotide-binding motif sequence of proteins. Our results suggest that ^{S}GTP mainly targets GTPases, with strong binding affinities observed for multiple heterotrimeric G proteins. We also demonstrated that ^{S}GTP binds to several cyclin-dependent kinases (CDKs), which may perturb the CDK-mediated phosphorylation and cell cycle progression. Together, this represents the first comprehensive characterization of ^{S}GTP-binding property for the entire human proteome. We reason that a similar strategy can be generally employed for the future characterization of the interaction of other modified nucleotides with the global proteome.

In Chapter five, we introduced a novel global kinome profiling method, based on a newly developed isotope-coded ATP-affinity probe and a targeted proteomic method

43

using multiple-reaction monitoring (MRM), for assessing simultaneously the expression of more than 300 kinases in human cells and tissues. This MRM-based assay displayed much better sensitivity, reproducibility and accuracy than the discovery-based shotgun proteomic method. Approximately 250 kinases could be routinely monitored detected in the lysate of a single cell line. Additionally, the incorporation of iRT into MRM kinome library rendered our MRM kinome assay easily transferrable across different instrument platforms and laboratories. We further employed this approach for profiling kinase expression in two melanoma cell lines, which revealed substantial kinome reprogramming during cancer progression and demonstrated an excellent correlation between the anti-proliferative effects of kinase inhibitors and the expression levels of their target kinases. Therefore, this facile and accurate kinome profiling assay, together with the kinome-inhibitor interaction map, could provide invaluable knowledge to predict the effectiveness of kinase inhibitor drugs and offer the opportunity for individualized cancer chemotherapy.

References

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

2.   Pandey A, Mann M. Proteomics to study genes and genomes. Nature. 2000;405:837-46.

3.   Cox J, Mann M. Quantitative, High-Resolution Proteomics for Data-Driven Systems Biology. Annu Rev Biochem. 2011;80:273-99.

4.   Beadle GW, Tatum EL. Genetic Control of Biochemical Reactions in Neurospora. Proceedings of the National Academy of Sciences. 1941;27:499-506.

5.   Cazier J-B, Tomlinson I. General lessons from large-scale studies to identify human cancer predisposition genes. The Journal of Pathology. 2010;220:255-62.

6.   Bensimon A, Heck AJR, Aebersold R. Mass Spectrometry–Based Proteomics and Network Biology. Annu Rev Biochem. 2012;81:379-405.

7.   Kreeger PK, Lauffenburger DA. Cancer systems biology: a network modeling perspective. Carcinogenesis. 2010;31:2-8.

8.   Vidal M, Cusick Michael E, Barabási A-L. Interactome Networks and Human Disease. Cell. 2011;144:986-98.

9.   Weston AD, Hood L. Systems Biology, Proteomics, and the Future of Health Care: Toward Predictive, Preventative, and Personalized Medicine. J Proteome Res. 2004;3:179-96.

10.   Kolch W, Pitt A. Functional proteomics to dissect tyrosine kinase signalling pathways in cancer. Nat Rev Cancer. 2010;10:618-29.

11.   Aebersold R, Goodlett DR. Mass Spectrometry in Proteomics. Chem Rev. 2001;101:269-96.

12.   Yates JR, Ruse CI, Nakorchevsky A. Proteomics by Mass Spectrometry: Approaches, Advances, and Applications. Annu Rev Biomed Eng. 2009;11:49-79.

13.   Olsen JV, Schwartz JC, Griep-Raming J, Nielsen ML, Damoc E, Denisov E, et al. A dual pressure linear ion trap orbitrap instrument with very high sequencing speed. Mol Cell Proteomics. 2009;8:2759-69.

14.   Wisniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. Nat Meth. 2009;6:359-62.

15.   Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B. Quantitative mass spectrometry in proteomics: a critical review. Anal Bioanal Chem. 2007;389:1017-31.

16.   Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotech. 2008;26:1367-72.

17.   Blackstock WP, Weir MP. Proteomics: quantitative and physical mapping of cellular proteins. Trends Biotechnol. 1999;17:121-7.

18.   Villen J, Gygi SP. The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. Nat Protocols. 2008;3:1630-8.

19.   Choudhary C, Kumar C, Gnad F, Nielsen ML, Rehman M, Walther TC, et al. Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. Science. 2009;325:834-40.

20.   Havugimana Pierre C, Hart GT, Nepusz T, Yang H, Turinsky Andrei L, Li Z, et al. A Census of Human Soluble Protein Complexes. Cell. 2012;150:1068-81.

21.   Ong S-E, Schenone M, Margolin AA, Li X, Do K, Doud MK, et al. Identifying the proteins to which small-molecule probes and drugs bind in cells. Proceedings of the National Academy of Sciences. 2009.

22.   Mittler G, Butter F, Mann M. A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements. Genome Res. 2009;19:284-93.

23.   Butter F, Scheibe M, Mörl M, Mann M. Unbiased RNA–protein interaction screen by quantitative proteomics. Proceedings of the National Academy of Sciences. 2009;106:10626-31.

24.   Bateman NW, Goulding SP, Shulman N, Gadok AK, Szumlinski KK, Maccoss MJ, et al. Maximizing peptide identification events in proteomic workflows utilizing data-dependent acquisition. Mol Cell Proteomics. 2013.

25.   Lange V, Picotti P, Domon B, Aebersold R. Selected reaction monitoring for quantitative proteomics: a tutorial. Mol Syst Biol. 2008;4.

26.   Ong S-E, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics. 2002;1:376-86.

27.   Mann M. Functional and quantitative proteomics using SILAC. Nat Rev Mol Cell Biol. 2006;7:952-8.

28.   Leitner A, Lindner W. Current chemical tagging strategies for proteome analysis by mass spectrometry. Journal of Chromatography B. 2004;813:1-26.

29.   Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotech. 1999;17:994-9.

30.   Hsu J-L, Huang S-Y, Chow N-H, Chen S-H. Stable-Isotope Dimethyl Labeling for Quantitative Proteomics. Anal Chem. 2003;75:6843-52.

31.   Boersema PJ, Raijmakers R, Lemeer S, Mohammed S, Heck AJR. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. Nat Protocols. 2009;4:484-94.

32.   Kovanich D, Cappadona S, Raijmakers R, Mohammed S, Scholten A, Heck AR. Applications of stable isotope dimethyl labeling in quantitative proteomics. Anal Bioanal Chem. 2012;404:991-1009.

33.   Wiese S, Reidegeld KA, Meyer HE, Warscheid B. Protein labeling by iTRAQ: A new tool for quantitative mass spectrometry in proteome research. Proteomics. 2007;7:340-50.

34.   Dayon L, Hainard A, Licker V, Turck N, Kuhn K, Hochstrasser DF, et al. Relative Quantification of Proteins in Human Cerebrospinal Fluids by MS/MS Using 6-Plex Isobaric Tags. Anal Chem. 2008;80:2921-31.

35.   Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proceedings of the National Academy of Sciences. 2003;100:6940-5.

36.   Mayya V, Rezual K, Wu L, Fong MB, Han DK. Absolute Quantification of Multisite Phosphorylation by Selective Reaction Monitoring Mass Spectrometry: Determination of Inhibitory Phosphorylation Status of Cyclin-Dependent Kinases. Mol Cell Proteomics. 2006;5:1146-57.

37.   Wu CC, MacCoss MJ. Shotgun proteomics: tools for the analysis of complex biological systems. Curr Opin Mol Ther. 2002;4:242-50.

38.   Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res. 2011;10:1794-805.

39.   Nagaraj N, D'Souza RCJ, Cox J, Olsen JV, Mann M. Feasibility of Large-Scale Phosphoproteomics with Higher Energy Collisional Dissociation Fragmentation. J Proteome Res. 2010;9:6786-94.

40.   de Godoy LMF, Olsen JV, Cox J, Nielsen ML, Hubner NC, Frohlich F, et al. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. Nature. 2008;455:1251-4.

41.   Huttlin EL, Jedrychowski MP, Elias JE, Goswami T, Rad R, Beausoleil SA, et al. A Tissue-Specific Atlas of Mouse Protein Phosphorylation and Expression. Cell. 2010;143:1174-89.

42.   Wolf-Yadlin A, Hautaniemi S, Lauffenburger DA, White FM. Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. Proceedings of the National Academy of Sciences. 2007;104:5860-5.

43.   Keshishian H, Addona T, Burgess M, Mani DR, Shi X, Kuhn E, et al. Quantification of Cardiovascular Biomarkers in Patient Plasma by Targeted Mass Spectrometry and Stable Isotope Dilution. Mol Cell Proteomics. 2009;8:2339-49.

44.   Marx V. Targeted proteomics. Nat Meth. 2013;10:19-22.

45.   Hüttenhain R, Surinova S, Ossola R, Sun Z, Campbell D, Cerciello F, et al. N-Glycoprotein SRMAtlas: A RESOURCE OF MASS SPECTROMETRIC ASSAYS FOR N-GLYCOSITES ENABLING CONSISTENT AND MULTIPLEXED PROTEIN QUANTIFICATION FOR CLINICAL APPLICATIONS. Mol Cell Proteomics. 2013;12:1005-16.

46.   Prakash A, Tomazela DM, Frewen B, MacLean B, Merrihew G, Peterman S, et al. Expediting the Development of Targeted SRM Assays: Using Data from Shotgun Proteomics to Automate Method Development. J Proteome Res. 2009;8:2733-9.

47.   Kiyonami R, Schoen A, Prakash A, Peterman S, Zabrouskov V, Picotti P, et al. Increased Selectivity, Analytical Precision, and Throughput in Targeted Proteomics. Mol Cell Proteomics. 2011;10.
48.   Unwin RD, Griffiths JR, Whetton AD. A sensitive mass spectrometric method for hypothesis-driven detection of peptide post-translational modifications: multiple reaction monitoring-initiated detection and sequencing (MIDAS). Nat Protocols. 2009;4:870-7.

49.   Escher C, Reiter L, MacLean B, Ossola R, Herzog F, Chilton J, et al. Using iRT, a normalized retention time for more targeted measurement of peptides. Proteomics. 2012;12:1111-21.

50.   Reiter L, Rinner O, Picotti P, Huttenhain R, Beck M, Brusniak M-Y, et al. mProphet: automated data processing and statistical validation for large-scale SRM experiments. Nat Meth. 2011;8:430-5.

51.   Chene P. ATPases as drug targets: learning from their structure. Nat Rev Drug Discov. 2002;1:665-73.

52.   Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. Science. 2002;298:1912-34.

53.   Takai Y, Sasaki T, Matozaki T. Small GTP-binding proteins. Physiol Rev. 2001;81:153-208.
54.   Horscroft NJ, Roy P. NTP binding and phosphohydrolase activity associated with purified bluetongue virus non-structural protein NS2. J Gen Virol. 2000;81:1961-5.

55.   Dong X, Xiao Y, Jiang X, Wang Y. Quantitative proteomic analysis revealed lovastatin-induced perturbation of cellular pathways in HL-60 cells. J Proteome Res. 2011;10:5463-71.

56.   Taylor P, Nielsen PA, Trelle MB, Andersen MB, Vorm O, Moran MF, et al. Automated 2D peptide separation on a 1D Nano-LC-MS system. J Proteome Res. 2009;8:1610-6.

57.   Blume-Jensen P, Hunter T. Oncogenic kinase signalling. Nature. 2001;411:355-65.

58.   Duncan James S, Whittle Martin C, Nakamura K, Abell Amy N, Midland Alicia A, Zawistowski Jon S, et al. Dynamic reprogramming of the kinome in response to targeted MEK inhibition in triple-negative breast cancer. Cell. 2012;149:307-21.

59.   Zhang J, Yang PL, Gray NS. Targeting cancer with small molecule kinase inhibitors. Nat Rev Cancer. 2009;9:28-39.

60.   Karaman MW, Herrgard S, Treiber DK, Gallant P, Atteridge CE, Campbell BT, et al. A quantitative analysis of kinase inhibitor selectivity. Nat Biotech. 2008;26:127-32.

61.   Fabian MA, Biggs WH, Treiber DK, Atteridge CE, Azimioara MD, Benedetti MG, et al. A small molecule-kinase interaction map for clinical kinase inhibitors. Nat Biotech. 2005;23:329-36.

62.   Davis MI, Hunt JP, Herrgard S, Ciceri P, Wodicka LM, Pallares G, et al. Comprehensive analysis of kinase inhibitor selectivity. Nat Biotech. 2011;29:1046-51.

63.   Pulford K, Lamant L, Morris SW, Butler LH, Wood KM, Stroud D, et al. Detection of anaplastic lymphoma kinase (ALK) and nucleolar protein nucleophosmin (NPM)-ALK proteins in normal and neoplastic cells with the monoclonal antibody ALK1. Blood. 1997;89:1394-404.

64.   Oppermann FS, Gnad F, Olsen JV, Hornberger R, Mann M, Daub H. Large-scale proteomics analysis of the human kinome. Mol Cell Proteomics. 2009;8:1751-64.
65.   Bartlett S, Beddard GS, Jackson RM, Kayser V, Kilner C, Leach A, et al. Comparison of the ATP Binding Sites of Protein Kinases Using Conformationally Diverse Bisindolylmaleimides. J Am Chem Soc. 2005;127:11699-708.

66.   Graves PR, Kwiek JJ, Fadden P, Ray R, Hardeman K, Coley AM, et al. Discovery of Novel Targets of Quinoline Drugs in the Human Purine Binding Proteome. Mol Pharmacol. 2002;62:1364-72.

67.   Ito J, Heazlewood JL, Millar AH. Analysis of the soluble ATP-binding proteome of plant mitochondria identifies new proteins and nucleotide triphosphate interactions within the matrix. J Proteome Res. 2006;5:3459-69.

68.   Bantscheff M, Eberhard D, Abraham Y, Bastuck S, Boesche M, Hobson S, et al. Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. Nat Biotech. 2007;25:1035-44.

69.   Daub H, Olsen JV, Bairlein M, Gnad F, Oppermann FS, Stemmann O, et al. Kinase-selective enrichment enables quantitative phosphoproteomics of the kinome across the cell cycle. Mol Cell. 2008;31:438-48.

70.   Zhang L, Holmes IP, Hochgräfe F, Walker SR, Ali NA, Humphrey ES, et al. Characterization of the Novel Broad-Spectrum Kinase Inhibitor CTx-0294885 As an

Affinity Reagent for Mass Spectrometry-Based Kinome Profiling. J Proteome Res. 2013;12:3104-16.

71.   Nomura DK, Dix MM, Cravatt BF. Activity-based protein profiling for biochemical pathway discovery in cancer. Nat Rev Cancer. 2010;10:630-8.

72.   Cravatt BF, Wright AT, Kozarich JW. Activity-Based Protein Profiling: From Enzyme Chemistry to Proteomic Chemistry. Annu Rev Biochem. 2008;77:383-414.

73.   Ohta T, Nagano K, Yoshida M. The active site structure of Na+/K+-transporting ATPase: location of the 5'-(p-fluorosulfonyl)benzoyladenosine binding site and soluble peptides released by trypsin. Proc Natl Acad Sci U S A. 1986;83:2071-5.

74.   Qiu H, Wang Y. Probing adenosine nucleotide-binding proteins with an affinity-labeled nucleotide probe and mass spectrometry. Anal Chem. 2007;79:5547-56.

75.   Patricelli MP, Szardenings AK, Liyanage M, Nomanbhoy TK, Wu M, Weissig H, et al. Functional interrogation of the kinome using nucleotide acyl phosphates. Biochemistry (Mosc). 2006;46:350-8.

76.   Saraste M, Sibbald PR, Wittinghofer A. The P-loop--a common motif in ATP- and GTP-binding proteins. Trends Biochem Sci. 1990;15:430-4.

77.   Deyrup AT, Krishnan S, Cockburn BN, Schwartz NB. Deletion and site-directed mutagenesis of the ATP-binding motif (P-loop) in the bifunctional murine Atp-sulfurylase/adenosine 5-phosphosulfate kinase enzyme. J Biol Chem. 1998;273:9450-6.

78.   Hanks SK, Hunter T. Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. . The FASEB Journal 1995 9 576-96

79.   Johnson LN, Noble MEM, Owen DJ. Active and inactive protein kinases: structural basis for regulation. Cell. 1996;85:149-58.
80.   Nolen B, Taylor S, Ghosh G. Regulation of protein kinases: controlling activity through activation segment conformation. Mol Cell. 2004;15:661-75.

81.   Lemeer S, Zörgiebel C, Ruprecht B, Kohl K, Kuster B. Comparing Immobilized Kinase Inhibitors and Covalent ATP Probes for Proteomic Profiling of Kinase Expression and Drug Selectivity. J Proteome Res. 2013;12:1723-31.

82.   Borisov OV, Goshe MB, Conrads TP, Rakov VS, Veenstra TD, Smith RD. Low-Energy Collision-Induced Dissociation Fragmentation Analysis of Cysteinyl-Modified Peptides. Anal Chem. 2002;74:2284-92.

83.   Sioud S, Genestie B, Jahouh F, Martin P, Banoub J. Gas-phase fragmentation study of biotin reagents using electrospray ionization tandem mass spectrometry on a quadrupole orthogonal time-of-flight hybrid instrument. Rapid Commun Mass Spectrom. 2009;23:1941-56.

84.   McAllister FE, Niepel M, Haas W, Huttlin E, Sorger PK, Gygi SP. Mass Spectrometry Based Method to Increase Throughput for Kinome Analyses Using ATP Probes. Anal Chem. 2013;85:4666-74.

85.   Patricelli Matthew P, Nomanbhoy Tyzoon K, Wu J, Brown H, Zhou D, Zhang J, et al. In situ kinase profiling reveals functionally relevant properties of native kinases. Chem Biol. 2011;18:699-710.

86.   Preininger AM, Hamm HE. G Protein Signaling: Insights from New Structures. Sci STKE. 2004;2004:re3-.

87.   Cox AD, Der CJ. Ras Family Signaling: Therapeutic Targeting. Cancer Biol Ther. 2002;1:599-606.

88.   Rosenblum JS, Nomanbhoy TK, Kozarich JW. Functional interrogation of kinases and other nucleotide-binding proteins. FEBS Lett. 2013;587:1870-7.

89.   Kaneda M, Masuda S, Tomohiro T, Hatanaka Y. A Simple and Efficient Photoaffinity Method for Proteomics of GTP-Binding Proteins. Chembiochem. 2007;8:595-8.

90.   George Cisar EA, Nguyen N, Rosen H. A GTP Affinity Probe for Proteomics Highlights Flexibility in Purine Nucleotide Selectivity. J Am Chem Soc. 2013;135:4676-9.

91.   Casey PJ, Seabra MC. Protein prenyltransferases. J Biol Chem. 1996;271:5289-92.

92.   Kho Y, Kim SC, Jiang C, Barma D, Kwon SW, Cheng J, et al. A tagging-via-substrate technology for detection and proteomics of farnesylated proteins. Proc Natl Acad Sci U S A. 2004;101:12479-84.

Chapter 2

Proteome-wide Discovery and Characterizations of Nucleotide-binding Proteins with Affinity-labeled Chemical Probes

Introduction

Adenine and guanine nucleotides are abundant and they bind to numerous proteins involved in pivotal cellular processes, including cell signaling, proliferation, differentiation, and apoptosis (1, 2). Despite the importance of nucleotide-binding proteins in cellular functions, the current picture of nucleotide-protein interactions is far from complete. Therefore, comprehensive identification of ATP/GTP-binding proteins and dynamic analysis of nucleotide-protein interactions at the proteomic scale are important for understanding better the regulatory mechanisms of nucleotide-binding proteins.

The development of mass spectrometry (MS) instrumentation and bioinformatic tools provides the opportunity to identify and quantify up to several thousand proteins in complex samples (3). However, proteomic studies of specific family of proteins, including nucleotide-binding proteins, by MS are still a big challenge owing to the extreme complexity of the proteome and the relatively low abundance of some proteins. This limitation can be partially overcome by combining MS with various separation

53

techniques, such as polyacrylamide gel electrophoresis (PAGE)(4) or multi-dimensional liquid chromatography (5). However, none of these approaches permit selective enrichment of nucleotide-binding proteins from cell lysates.

Affinity chromatography is commonly used for fractionating complex protein mixture to yield functional sub-groups of proteins. Ito et al.(6) used γ phosphate-linked ATP media to enrich ATP-binding proteins from the soluble fraction of *Arabidopsis* mitochondria. Additionally, Mann et al.(7) employed kinase-selective affinity column with immobilized kinase inhibitors as capture ligands to facilitate the identification and quantification of approximately 200 protein kinases. On the other hand, chemical tagging methods involving specific labeling of proteins with functional similarities have emerged as an important technique in targeted proteomics (8). For example, 5'-*p*-fluorosulfonylbenzoyladenosine, a reactive ATP analog, was employed as an activity-based probe to target nucleotide-binding proteins from whole cell lysates (9). Additionally, a photo-reactive GTP analog possessing a diazirine moiety was developed for the detection of GTP-binding proteins (10). Others and we also reported the application of biotin-conjugated acyl nucleotide probe for the enrichment and identification of ATP-binding proteins from complex protein mixtures (11, 12). In principle, this reactive affinity probe-based enrichment strategy, which involved labeling reaction, enzymatic digestion, affinity purification and LC-MS/MS analysis should be generally applicable for the identification and characterization of other nucleotide-binding proteins.

Here, we extended the use of the biotin-based nucleotide affinity probes as acylating agents to selectively label and enrich ATP- and GTP-binding proteins from the entire human proteome. Together with the use of extensive separation techniques, the method allowed for the identification of a significant number of nucleotide-binding proteins. In addition, the nucleotide-binding protein enrichment approach, along with quantitative proteomics using stable isotope labeling by amino acids in cell culture (SILAC) (13), facilitated the characterizations of nucleotide-protein interactions at the entire proteome scale.

Experimental Details

Preparation of Biotinylated Nucleotide Affinity Probe

The biotinylated nucleotide affinity probes were prepared according to previously published procedures with minor modifications (11). Briefly, probes were synthesized by directly conjugating biotin-LC or desthiobiotin with ATP or GTP. To render nucleotides soluble in organic solvent, the commercially available sodium salt form of nucleotides were first converted to the tributylammonium form by passing the nucleotides through a cation-exchange column packed with Spectra/Gel IE 50×8 resin (40-75 μm) at 4°C once. The tributylammonium form of nucleotide fractions were collected and lyophilized. Biotin-LC (10 mg) or desthiobiotin (6 mg), dissolved in a 1-mL solvent mixture of ice-cold dry CH_2Cl_2 and DMF (4:1, v/v), was mixed with tri-*n*-butylamine (11 μL) and ethyl chloroformate (5 μL). After stirring at 0°C for 5 min, the mixture was stirred at room temperature under argon atmosphere for another 60 min. Tributylammonium form of

55

nucleotides (50 mg), dissolved in a 1.25-mL solution of CH_2Cl_2 and DMF (4:1, v/v), was then added to the above reaction mixture. The reaction was continued at room temperature and under argon atmosphere for 18 hr. The CH_2Cl_2 was then removed by argon purging for 10 min and the remaining 200 μL solution was directly subjected to HPLC purification with a YMC ODS-AQ column (4.8×250 mm, 120 Å in pore size, 5 μm in particle size, Waters). The flow rate was 0.8 mL/min, and a 45-min linear gradient of 0-30% acetonitrile in 50 mM triethylammonium acetate (pH 6.8) was used for the purification. A UV detector was set at 265 nm to monitor the effluents (Figure 2.2 A). Appropriate HPLC fractions were pooled, lyophilized, and stored at -80°C. The structures of the products were confirmed by ESI-MS and MS/MS (Figure 2.2 B).

Cell Lysate Preparation and Labeling with Nucleotide Affinity Probe

HL-60 cells (ATCC, Manassas, VA) were cultured in Iscove's modified minimal essential medium (IMEM) supplemented with 10% fetal bovine serum (FBS, Invitrogen, Carlsbad, CA) and penicillin (100 IU/mL). For SILAC experiments, the IMEM medium without L-lysine or L-arginine was custom-prepared according to ATCC formulation. The complete light and heavy IMEM media were prepared by the addition of light or heavy lysine and arginine, along with dialyzed FBS (Invitrogen), to the above lysine, arginine-depleted medium. The HL-60 cells were cultured in heavy IMEM medium for at least 5 cell doublings to achieve complete isotope incorporation. Approximately 2×10^7 cells were harvested and washed three times with cold PBS. The cells were then lysed in 1 mL lysis buffer, which contained 0.7% CHAPS, 50 mM HEPES (pH 7.4), 0.5 mM

EDTA, 100 mM NaCl, and 10 µL (1:100) protease inhibitor cocktail on ice for 30 min. The cell lysates were centrifuged at 16000g at 4°C for 30 min, and the resulting supernatants were collected and subjected to gel filtration separation using NAP-25 columns (Amersham Biosciences) to remove free endogenous nucleotides. Cell lysates were eluted into a 2-mL buffer containing 50 mM HEPES (pH 7.4), 75 mM NaCl, and 5% glycerol. The resulting proteins in cell lysates were quantified using Quick Start Bradford Protein Assay (Bio-Rad, Hercules, CA) and stored at -80°C.

Immediately prior to the labeling reaction, $MgCl_2$, $MnCl_2$, and $CaCl_2$ were added to the concentrated cell lysate until their final concentrations reached 50, 5, and 5 mM, respectively. Approximately 1 mg cell lysate was treated with biotin-nucleotide affinity probe at concentrations ranging from 15 to 100 µM. Labeling reactions were carried out at room temperature with gentle shaking for 25 min. After the reaction, the remaining probes in the cell lysates were removed by buffer exchange with 25 mM NH_4HCO_3 solution (pH 8.5) using Amicon Ultra-4 filter (10,000 NMWL, Millipore).

In-solution Enzymatic Digestion and Affinity Purification

After addition of 8 M urea for protein denaturation and dithiothreitol (DTT) and iodoacetamide (IAM) to reduce and block cysteines, the labeled proteins were digested with modified sequencing-grade trypsin (Roche Applied Science) at an enzyme/substrate ratio of 1:100 in 25 mM NH_4HCO_3 (pH 8.5) and at 37°C for overnight. The peptide mixture was subsequently dried in a Speed-vac and redissolved in 1 mL of 100 mM potassium phosphate and 0.15 M NaCl (pH 7.5, PBS buffer). Avidin-agarose resin

(Sigma-Aldrich) was used to capture the biotin-labeled peptides. Prior to the binding, 200 µL resin was washed with 20 mM potassium phosphate and 0.15 M NaCl (pH 7.5) for 3 times. After adding to the digested peptide solution, the mixture was then incubated at 25°C for 1 hr with gentle shaking. To remove the unbound peptides, agarose resin was washed sequentially with 3 mL PBS buffer and 3 mL pure H_2O. Following washing, the labeled peptides were eluted with 1% TFA in CH_3CN/H_2O (7/3, v/v) at 65°C.   The eluates were dried in a Speed-vac and stored at -20°C prior to LC-MS/MS analysis.

SDS-PAGE Separation and In-Gel Digestion

The labeled proteins were denatured by boiling in Laemmli loading buffer for 5 min, and separated by a 12% SDS-PAGE with 4% stacking gel. The gel was stained with Coomassie blue; after destaining, the gel was cut into 8 slices, and the proteins were reduced in-gel with DTT and alkylated with IAM. The proteins were digested in-gel with modified sequencing-grade trypsin overnight, after which peptides were extracted from the gels with 5% acetic acid in H_2O and subsequently with 5% acetic acid in CH_3CN/H_2O (1:1, v/v). The resultant fractions of peptide mixtures were dried and stored at −20°C for further avidin enrichment.

Fractionation of Biotinylated Peptides with SCX Chromatography

Biotinylated peptides were desalted using C18 OMIX tips (Agilent) and subsequently loaded onto SCX TopTips (PolyLC). The bound peptides were sequentially eluted into 9 fractions with increasing KCl concentrations from 0.01 to 0.5 M.   The

peptides in each fraction were subjected to desalting again using C18 OMIX tips prior to LC-MS/MS analysis.

LC-MS/MS Analysis

LC-MS/MS analysis was performed on an LTQ-Orbitrap Velos mass spectrometer equipped with a nanoelectrospray ionization source (Thermo Fisher Scientific, San Jose, CA). Samples with or without SDS-PAGE and SCX prefractionation (procedures shown in the Supporting Information) were automatically loaded from a 48-well microplate autosampler using an EASY-nLC system (Proxeon Biosystems, Odense, Denmark) at 3 µL/min onto a home-made 4 cm trapping column (150 µm i.d.) packed with 5 µm C18 120 Å reversed-phase material (ReproSil-Pur 120 C18-AQ, Dr. Maisch). The trapping column was connected to a 20 cm fused silica analytical column (PicoTip Emitter, New Objective, 75 µm i.d.) with 3 µm C18 beads (ReproSil-Pur 120 C18-AQ, Dr. Maisch). The peptides were then separated with a 120-min linear gradient of 2-35% acetonitrile in 0.1% formic acid and at a flow rate of 250 nL/min. The LTQ-Orbitrap Velos was operated in data-dependent scan mode. Full-scan mass spectra were acquired in the Orbitrap analyzer with a resolution of 60000 with lock mass option enabled for the ion of *m/z* 445.120025 (14). Up to 20 most abundant ions found in MS with charge state ≥ 2 were sequentially isolated and sequenced in the linear ion trap with a normalized collision energy of 35, an activation q of 0.25 and an activation time of 10 ms.

Online 2D LC Separation and LC-MS/MS analysis

The fully automated 7-cycle on-line two-dimensional LC-MS/MS was set up as described (5). Briefly, the C18 trapping column was replaced with a biphasic precolumn (150 μm i.d.) comprised of a 3.5-cm column packed with 5 μm C18 120 Å reversed-phase material (ReproSil-Pur 120 C18-AQ, Dr. Maisch) and 3.5-cm column packed with Luna 5 μm SCX 100 Å strong cation exchange resin (Phenomenex, Torrance, CA) while all other setups remained the same. Enriched biotinylated peptides were first loaded onto the biphasic precolumn. Ammonium acetate at concentrations of 0, 25, 50, 75, 125, 200 and 500 mM were then sequentially injected using a 48-well autosampler from the sample vial to elute bound peptides from precolumn to analytical column with reversed-phase separation. LC-MS/MS experiments were also performed with a 120-min linear gradient of 2-35% acetonitrile in 0.1% formic acid.

Data Processing and Analysis.

For peptide identification, the raw data acquired were processed with the Maxquant search engine (version 1.2.0.18) (15) against human IPI protein database version 3.68 which contained 87,062 entries. Initial precursor mass tolerance of 10 ppm and fragment mass deviation of 0.8 Th were set as the search criteria. The maximum number of miss-cleavages for trypsin was set as two per peptide. Cysteine carbamidomethylation was considered as a fixed modification, whereas methionine oxidation and lysine biotinylation (+339.161662 Da) or desthiobiotinylation (+196.121178 Da) were included as variable modifications. The reverse database search

60

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

For SILAC-based comparison experiment, the raw data were converted to mzXML files and DTA files using ReAdW (http://sourceforge.net/projects/sashimi/files/) and MzXML2Search (http://tools.proteomecenter.org/wiki/index.php?title=Software:MzXML2Search) programs, respectively. Bioworks 3.2 was used for protein identification by searching the DTA files against the human IPI protein database version 3.68 and its reversed complement. Aside from the search parameters described above, lysine (+8 Da) and arginine (+6 Da) mass shifts introduced by heavy isotope labeling were considered as variable modifications. The search results were then filtered with DTASelect (16) to achieve a peptide false discovery rate of 1%. Census was employed for peptide and protein quantification (17). Extracted-ion chromatograms were first generated for peptide ions based on their m/z values and peptide intensity ratios were subsequently calculated in Census from peak areas found in each pair of extracted-ion chromatograms.

Results and Discussion

1. Design of nucleotide-affinity probe

Bioinformatics studies showed that most ATP/GTP-binding proteins carry a consensus amino acid sequence motif termed phosphate-binding loop (P-loop), which constitutes the nucleotide-binding site and is responsible for ATPase/GTPase activity

61

(18). It has been frequently observed that there is at least one lysine residue at nucleotide binding sites. For example, a conserved motif of GxxxxGK, in which 'x' represents any amino acid, is often found in the P-loop region of ATP/GTP-binding proteins (19). In light of these previous findings, a nucleotide analog bearing an acyl phosphate moiety is developed to target the lysine residue in the nucleotide-binding site (Figure 2.1). These nucleotide-binding probes contain three components, namely, a binding moiety consisted of nucleotide of interest (e.g., GTP or ATP), an enrichment moiety of biotin or its analog desthiobiotin which facilitates downstream purification, and an acyl phosphate group which targets side-chain amino group of the P-loop lysine residue to form a stable amide bond. For instance, when the ATP affinity probe interacts with ATP-binding proteins, the ATP moiety binds to the P-loop region, which facilitates the ε-amino group of the P-loop lysine residue to react with the acyl phosphate component to yield a stable amide bond (Figure 2.3). After the reaction, the protein mixture is digested with trypsin and the resulting biotin-labeled peptides can be enriched using an avidin agarose column. Finally, the affinity-purified peptides with the biotin tag are analyzed by LC-MS/MS (Figure 2.4). In this vein, it is worth noting that biotin conjugation with the side chain of cysteine or lysine was found not to result in significant alteration in peptide backbone fragmentation upon low-energy collision activation and peptide identification based on the resulting MS/MS(20, 21). We also did not observe any apparent difference in the percentage of spectra with successful peptide identification for the biotin-conjugated peptides vs. unmodified peptides.

Figure 2.1 (A) The design of nucleotide-affinity probe. (B) The structures of biotin-based ATP/GTP probes and desthiobiotin-based ATP/GTP probes. (C) The representative HPLC trace for the purification of the crude product of nucleotide-affinity probe. (D) Representative of ESI-MS of purified biotin-based ATP/GTP probes and desthiobiotin-based ATP/GTP probes

(A)
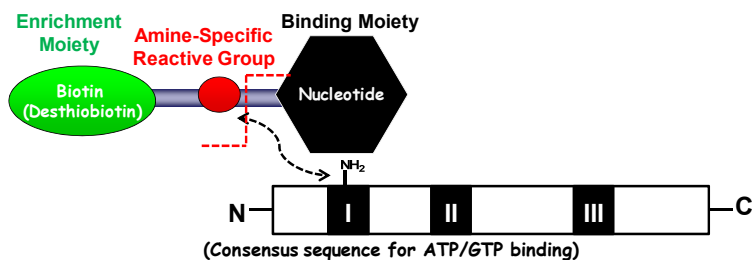
(B)
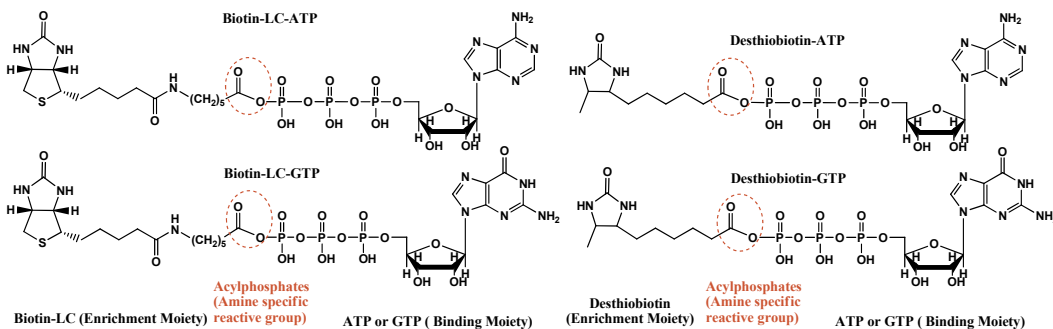
Figure 2.2 (A) The representative HPLC trace for the purification of the crude product of nucleotide-affinity probe. (B) Representative of ESI-MS of purified biotin-based ATP/GTP probes and desthiobiotin-based ATP/GTP probes

(A)
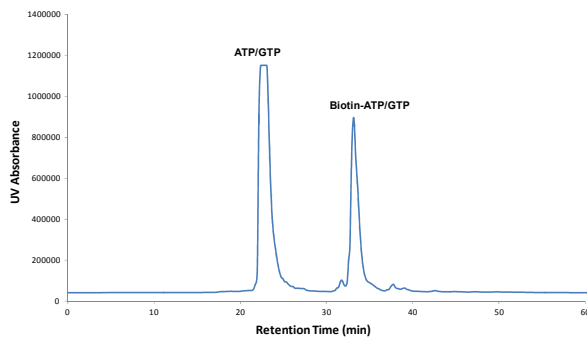
(B)
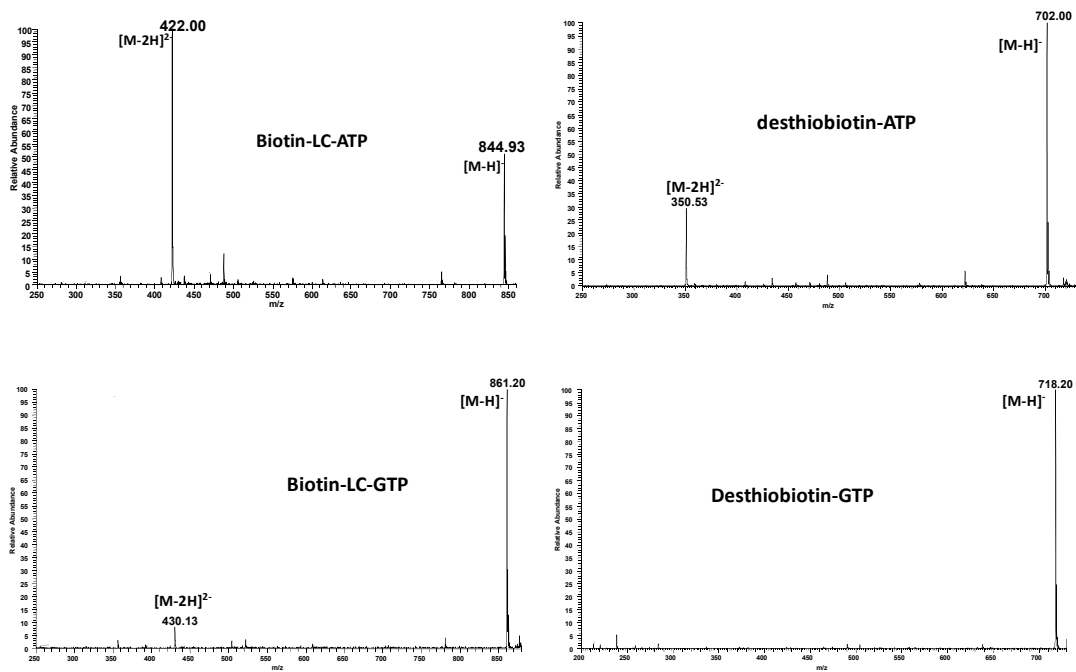
Figure 2.3 A schematic diagram showing the reaction between biotin-LC-ATP affinity probe with an ATP-binding protein.
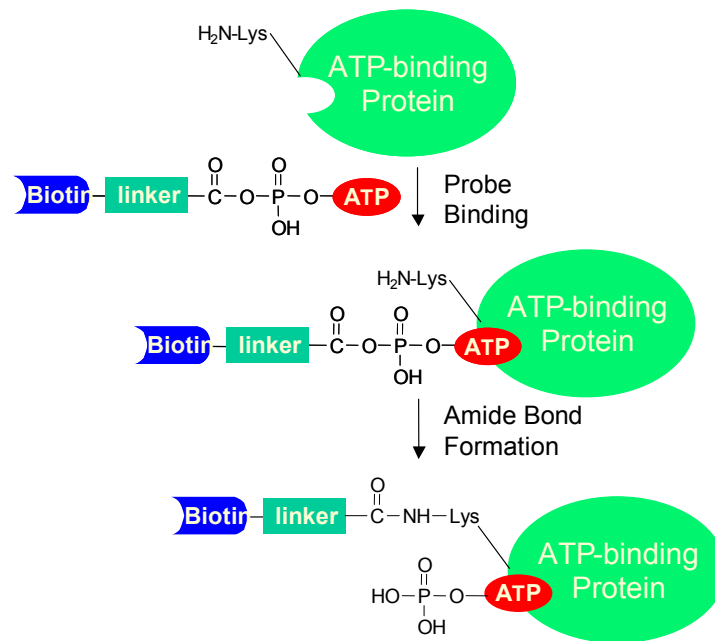
Figure 2.4 A scheme showing the enrichment and identification of nucleotide-binding proteins from human whole proteome with biotin-based ATP/GTP affinity probe.
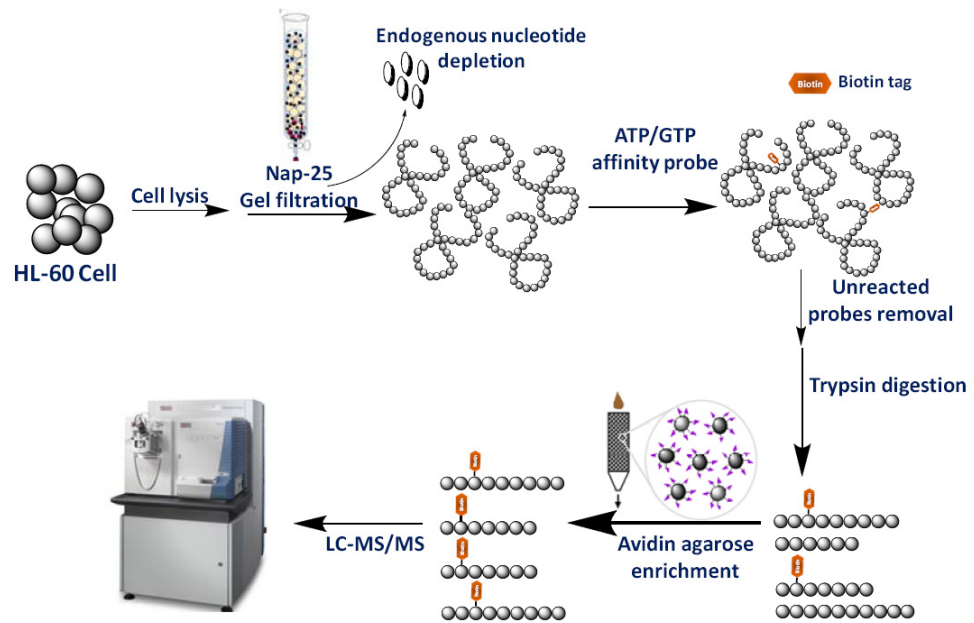
Initial experiment with 50 and 100 μM biotin-LC-ATP/GTP affinity probe and 1-mg cell lysate led to the identification of similar numbers of ATP- or GTP-binding proteins, though more nucleotide-binding proteins apart from ATP/GTP-binding proteins were identified in 100 μM probe experiment (Figure 2.5). However, if the GTP probe concentration was decreased to 15 μM, the total number of identified GTP-binding proteins decreased by more than 4 fold (Data not shown). Thus, we employed a probe concentration of 100 μM for subsequent experiments.

Due to the very high affinity and specificity of biotin-avidin interaction, biotinylation is the most widely used chemical modification for the affinity purification and detection of proteins and peptides (22). However, the extremely high stability of biotin-avidin complex renders it difficult to elute the biotinylated molecules from the avidin resin. In this vein, a similar ATP affinity probe carrying a desthiobiotin in lieu of biotin has been introduced by Pierce (Active X). The desthiobiotin moiety with lower affinity to avidin agarose may afford a better recovery of labeled peptides and higher detection sensitivity, thereby minimizing sample loss during the avidin enrichment step (23). We compared the performance of biotin-LC- and desthiobiotin-based nucleotide affinity probes, and it turned out that there is no substantial difference for these two types of probes in the identification of nucleotide-binding proteins (Figure 2.5).

Figure 2.5 Optimization of the performance of nucleotide-affinity probes for protein detection by varying concentrations of ATP/GTP-affinity probe (A, B) and enrichment moiety (C, D).
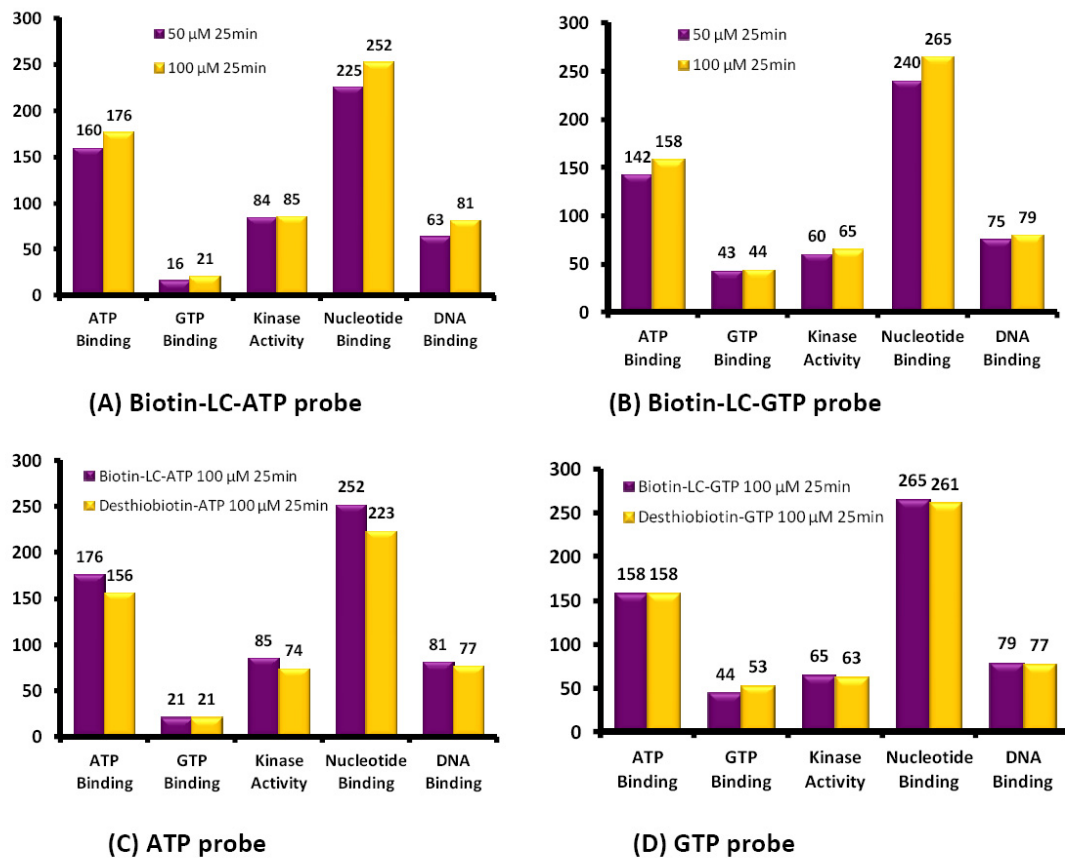
2. Extensive prefractionation for Large-scale profiling of nucleotide-binding proteins from the whole human proteome

Owing to the extreme complexity of human proteome, MS-based identification of specific subfamilies of proteins, including nucleotide-binding proteins, from complex samples necessitates the use of powerful separation techniques, such as SDS-PAGE or multi-dimensional LC for sample prefractionation (4, 5). Thus, we employed three different pre-fractionation methods, i.e. SDS-PAGE, offline SCX chromatography and online 2D-LC with SCX-C18 separation, for experiments with desthiobiotin-based ATP affinity probe. For SDS-PAGE fractionation, we separated ATP affinity probe-treated whole cell lysate using 12% SDS-PAGE, cut the gel into 8 slices, digested the proteins in-gel with trypsin, enriched the biotin-labeled peptides from each fraction using avidin agarose, and analyzed the enriched peptides with LC-MS/MS. This method allowed for the identification of 1072 proteins with biotin-labeled peptides (Figure 2.6). Offline SCX separation, during which the digested peptide mixture from labeled whole cell lysate was separated into 6 fractions and then subjected individually to avidin enrichment, led to the identification of 953 proteins with biotin-labeled peptides from 1-mg cell lysates (Figure 2.6). On the other hand, online 2D-LC system with SCX and C18 separation results in the identification of 946 proteins with the probe-labeled peptides (Figure 2.6). Relative to the results obtained from the unfractionated lysates, the total number of probe-labeled proteins increased by approximately two-fold with each single prefractionation method

Figure 2.6 The Venn diagrams showing all proteins (i.e., identified with probe-labeled peptides, A), known ATP-binding proteins (B), and kinases (C) identified from the reaction of ATP affinity probe with HL-60 cell lysate followed by SDS-PAGE, offline SCX, or online 2D-LC separation.

(A) Total proteins (B) ATP-binding proteins (C) Kinases
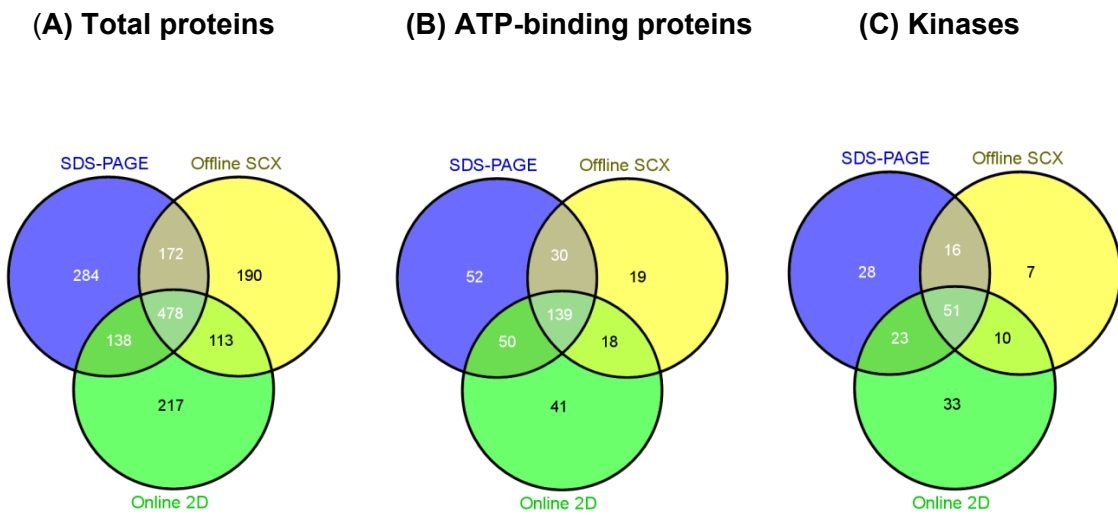
and almost a 3-fold increase was achieved by combining all three prefractionation strategies.

Altogether we identified a total of 1538 unique proteins with desthiobiotin modification from these three separation strategies (Figure 2.7). Among these proteins, 349 (23%), 168 (11%) and 53 (3%) are known ATP-binding proteins, kinases, and GTP-binding proteins, respectively. Additionally, more than 510 known nucleotide-binding proteins were unambiguously detected, highlighting the great potential of this method in nucleotide-binding protein studies at the global proteome scale. We also subjected biotin-based GTP-probe-labeled peptides to online 2D-LC-MS/MS analysis alone and we were able to identify a total of 1355 proteins, among which 66 (5%) are known GTP-binding proteins (Figure 2.7) and 441 possess nucleotide-binding capability. The relatively low percentage of known GTP-binding proteins among all the identified proteins could be attributed, in part, to the presence of many as-yet characterized or annotated GTP-binding proteins.

3. A strategy to compare the relative ATP/GTP-binding affinities of proteins from whole cell lysates

Nucleotide-binding affinity and phosphohydrolase activity are key features of various enzymes such as kinases, helicases and G proteins. However, the selectivity in binding toward different nucleotides, is an often overlooked but non-trivial property of nucleotide-binding proteins. Both ATP and GTP are purine nucleoside triphosphates and share similar structures as well as biological functions. Some kinases, e.g. casein kinase

Figure 2.7 A summary of proteins identified with biotin-ATP probe (A) and biotin-GTP probe (B) from HL-60 whole cell lysate with extensive separations.
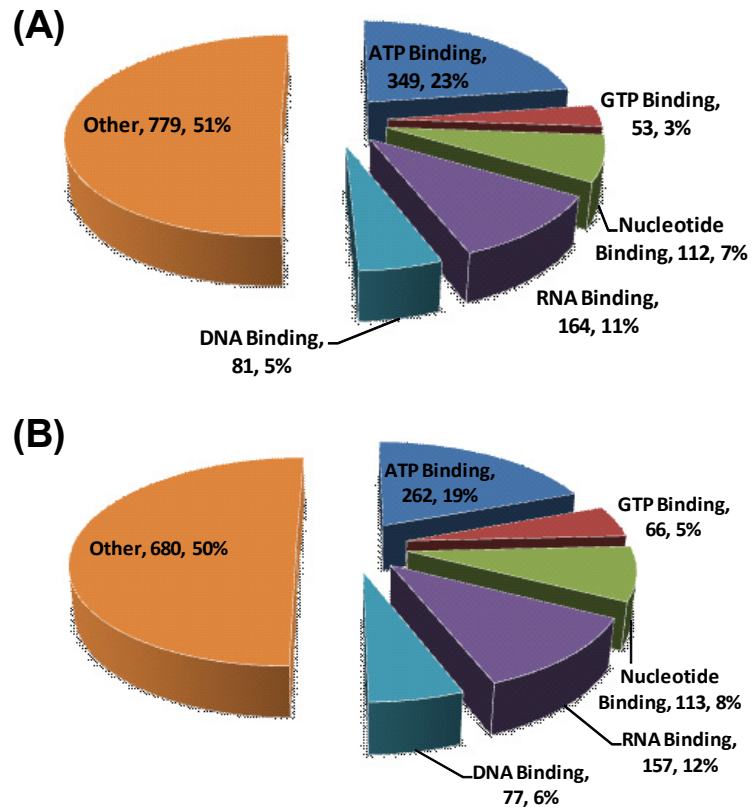
2 (24), may employ GTP as phosphate donor to phosphorylate substrates, whereas ATP may also bind to GTP-binding proteins to regulate their activities. Therefore, direct comparison of the binding affinity of signal transduction proteins such as kinases and G-proteins toward ATP/GTP is essential for elucidating better the underlying mechanism of signal activation and transduction in cells. Although various assays can be used to assess the binding selectivities of nucleotides for individual proteins, systematic comparison of ATP- and GTP-binding affinities of proteins, especially at the entire proteome scale, has not yet been performed.

Apart from nucleotide selectivity, the nucleotide-binding site, or 'nucleotide-interacting residues', is another aspect of significant interest in nucleotide-protein interaction studies. The ATP/GTP binding sites are particularly important because many anti-cancer drugs also target nucleotide-binding sites in ATPases and GTPases (25). The experimental determination of residues that interact with ATP/GTP often relies on X-ray crystal structure or site-directed mutagenesis, which is costly and time-consuming. We demonstrated previously that lysine residues labeled by the biotin-ATP probe may represent the specific nucleotide-binding site of proteins, which directly interacts with γ-phosphate group of the bound ATP (11). We reason that this site-directed labeling strategy using nucleotide-affinity probe, in combination with quantitative proteomics, may allow for the assessment of the selectivities of ATP and GTP in binding toward specific sequence segments of nucleotide-binding proteins.

To explore the nucleotide-binding affinity of ATP/GTP-binding proteins at specific binding sites and to uncover novel nucleotide-binding targets, we employed SILAC together with our nucleotide affinity probe to investigate the nucleotide-binding property of proteins at the whole proteome scale. In a pilot experiment, light- and heavy-labeled cell lysates were treated with the same concentrations of biotin-based ATP and GTP probes (100 μM each), respectively. After the reaction, light- and heavy-labeled cell lysates were mixed prior to any further steps of sample manipulation as described in regular affinity experiments. To minimize the bias introduced by SILAC, we also performed reverse SILAC experiment, where heavy and light cell lysates were treated with ATP and GTP probes, respectively (Figure 2.8). All the biotin-labeled peptides were directly analyzed by 1D-LC/MS/MS. Peak intensity ratios of light and heavy biotin-labeled tryptic peptides were subsequently used to derive ATP/GTP binding affinity ratio, $R_{ATP/GTP}$, which reflects the relative binding affinities of ATP and GTP towards specific lysine residues in proteins of interest.

The median $R_{ATP/GTP}$ ratio for all the quantified peptides from a single protein was employed to represent its relative affinity towards ATP and GTP. The medium was selected to minimize the effect of outlier and non-specific binders (15). For example, ADP-ribosylation factor-like protein 2 (ARL2), a member of the Ras small GTPase superfamily, exhibits strong binding towards GTP in both forward and reverse SILAC labeling experiments as the median of $R_{ATP/GTP}$ for all the detected biotin-labeled peptides from ARL2 is only 0.11. In this regard, one biotin-labeled lysine with K29 being

Figure 2.8 A general strategy for comparing ATP/GTP-binding properties of proteins from SILAC cell lysates.
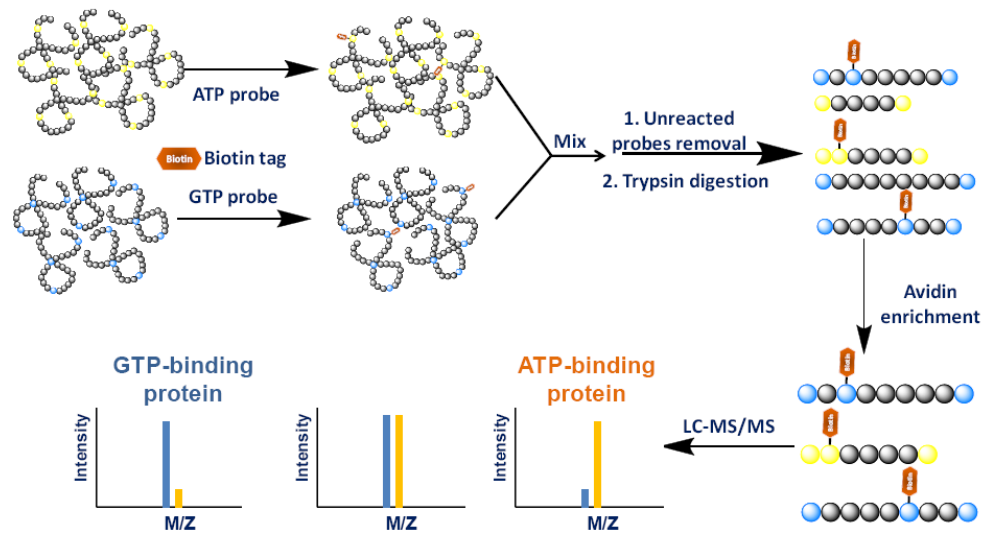
conjugated, i.e., R.LLMLGLDNAGK#TTILK.K, was found for ARL2 (Figure 2.9, MS/MS shown in Figure 2.10). The crystal structure of the ARL2-GTP complex (PDB entry: 3DOE) revealed that K29 contacts directly the γ-phosphate group of GTP (26), demonstrating that biotin-labeled lysine with $R_{ATP/GTP}$ ratio deviating substantially from 1 corresponds to real nucleotide-binding site, or more precisely, the lysine residue that is in close proximity to the γ-phosphate group of the nucleotide.

It is worth noting that the nucleotide affinity probes with an acyl phosphate linkage have relatively high reactivity; therefore, apart from the lysine residue(s) located in the nucleotide binding site, other lysine residues may also be labeled with biotin because of their non-specific electrostatic interaction with phosphate group of the probe (11). Nevertheless, because of the structural similarity of ATP and GTP probes, peptides housing these non-specifically labeled lysine residues are expected to exhibit a $R_{ATP/GTP}$ ratio being close to unity. For instance, we found two biotin-labeled lysine sites in adenylate kinase 2 (AK2, MS/MS shown in Figure 2.10), where K28 displays a clear ATP binding preference with an $R_{ATP/GTP}$ of 2.3. This is consistent with the observation that K28 is responsible for ATP binding and phosphohydrolase activity (PDB entry: 3TLX, Figure 2.9). The other labeling site (K85), however, exhibits similar ATP/GTP binding affinity ($R_{ATP/GTP}$=1.09). This finding, together with the absence of prior report supporting K85's involvement in ATP binding, suggests that modification of K85 is likely derived from non-specific interaction. Meanwhile, the median $R_{ATP/GTP}$ ratio for all peptides identified for AK2 is 2.26 (Table 2.1), supporting that the median $R_{ATP/GTP}$ ratio

Figure 2.9 Forward- and reverse-SILAC combined with LC-MS/MS for the quantitative comparison of ATP/GTP binding affinity towards ADP-ribosylation factor-like protein 2 (A, B) and adenylate kinase 2 (C, D). # indicates the biotin-labeling site. Crystal structures of ARL2 bound with GTP (E) and AK2 bound with ATP (F) demonstrate the direct contact of nucleotide with identified lysine residues.

(A) MS of ARL2, Forward SILAC

(D) MS of AK2, Forward SILAC

(B) MS of ARL2, Reverse SILAC

(E) MS of AK2, Reverse SILAC

(C) Crystal structure of ARL2

(F) Crystal structure of AK2
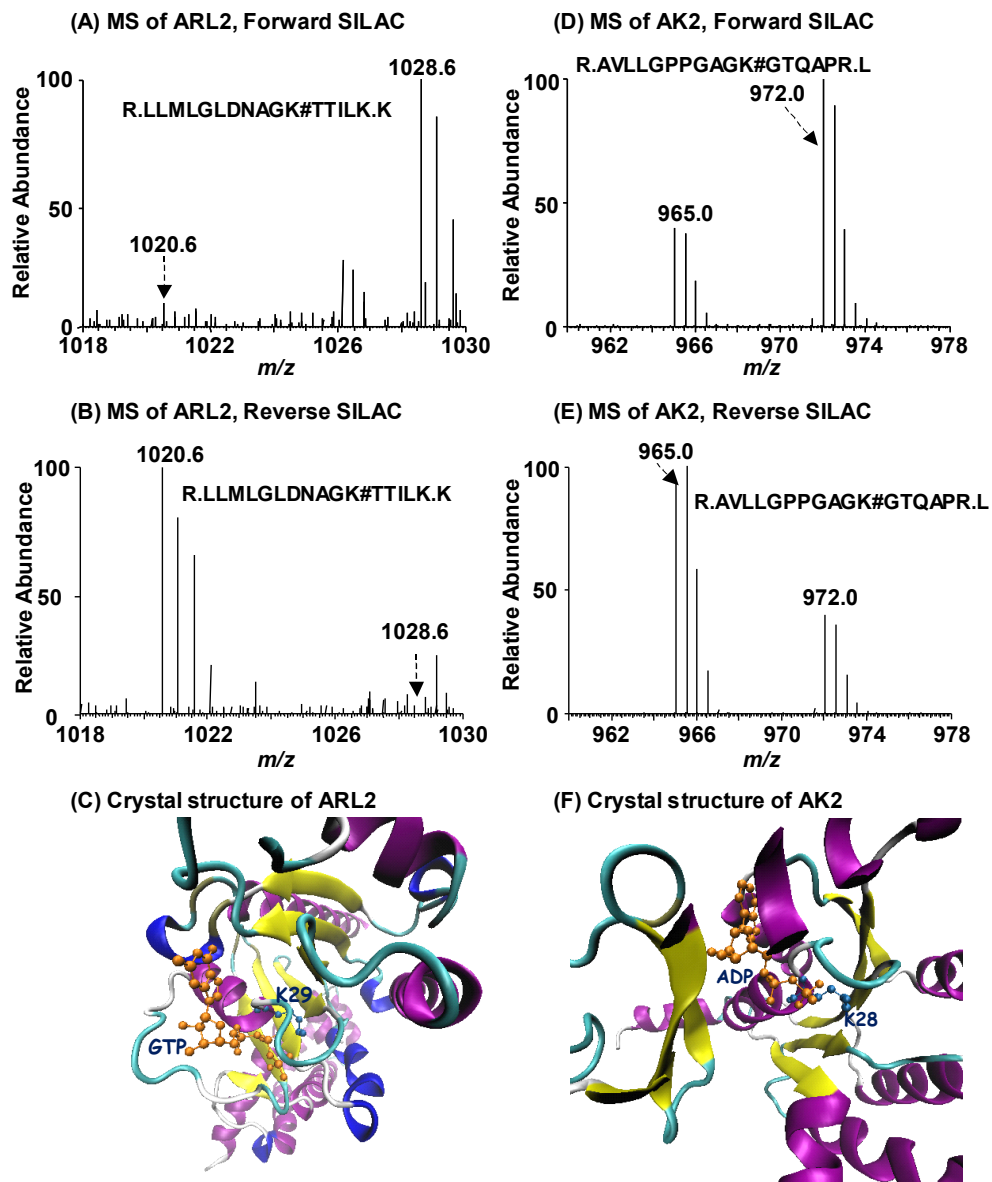
Figure 2.10 MS/MS of the light- and heavy-labeled LLMLGLDNAGK#TTILK (A, B) from ARL2 and AVLLGPPGAGK#GTQAPR (C, D) from AK2.

(A)
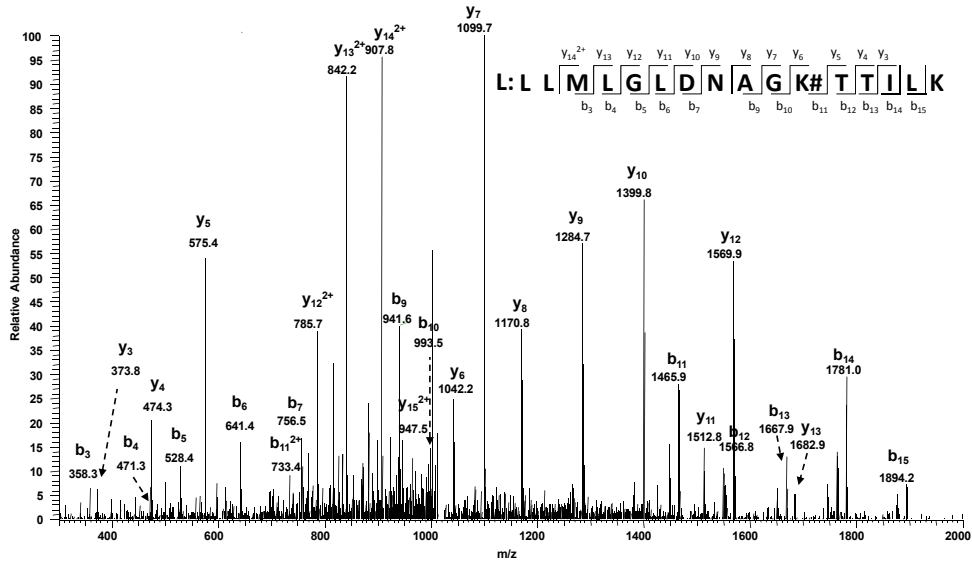
(B)
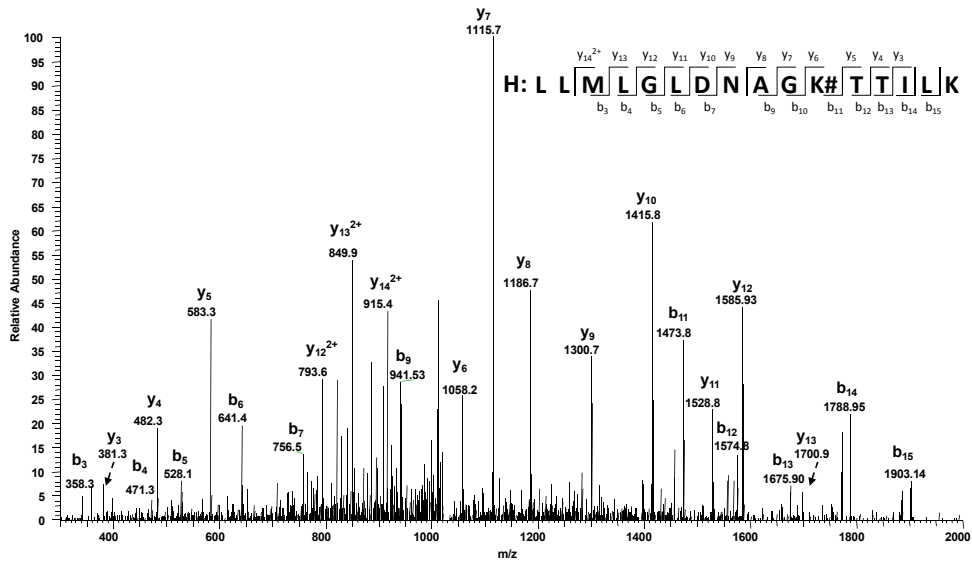
(C)
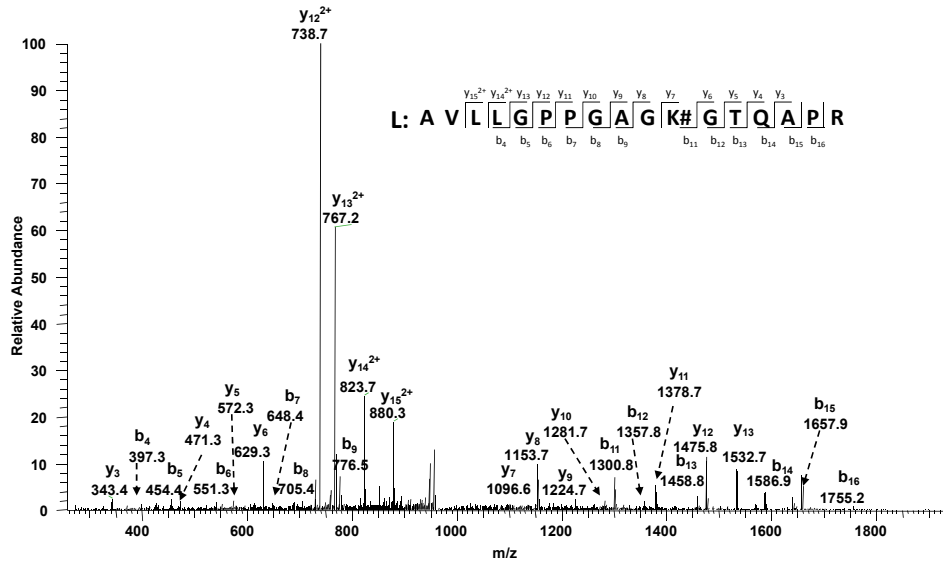
(D)
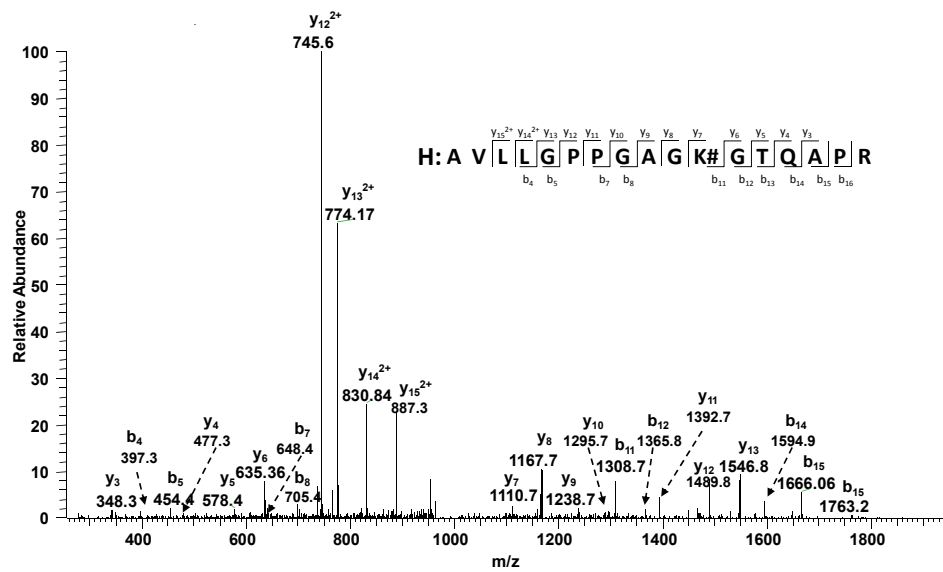
Table 2.1 Quantification results for AK2 in six runs of SILAC-based ATP/GTP (1:1 ratio)

competition experiment; Ratio indicates the RATP/GTP ratio

Protein Name	SEQUENCE	RATIO	DETERMINANT FACTOR	XCorr	FILE NAME	SCAN
Adenylate kinase 2	K.LVSDEMVVELIEK#NLETPLCK.N	0.99	0.66	3.25	ATPGTP-1	37797
Adenylate kinase 2	K.LVSDEMVVELIEK#NLETPLCK.N	1.13	0.57	3.32	ATPGTP-2	32476
Adenylate kinase 2	R.AVLLGPPGAGK#GTQAPR.L	2.65	0.99	4.94	ATPGTP-1	23939
Adenylate kinase 2	R.AVLLGPPGAGK#GTQAPR.L	2.53	0.99	3.41	ATPGTP-1	23808
Adenylate kinase 2	R.AVLLGPPGAGK#GTQAPR.L	3.04	0.89	2.78	ATPGTP-1	23632
Adenylate kinase 2	R.AVLLGPPGAGK#GTQAPR.L	2.76	1.00	5.03	ATPGTP-2	21322
Adenylate kinase 2	R.AVLLGPPGAGK#GTQAPR.L	2.30	0.99	3.03	ATPGTP-3	27337
Adenylate kinase 2	R.AVLLGPPGAGK#GTQAPR.L	2.29	0.99	5.26	ATPGTP-3	27210
Adenylate kinase 2	R.AVLLGPPGAGK#GTQAPR.L	2.29	0.99	3.97	ATPGTP-3	27023
Adenylate kinase 2	R.AVLLGPPGAGK#GTQAPR.L	2.13	0.99	4.41	GTPATP-1	23364
Adenylate kinase 2	R.AVLLGPPGAGK#GTQAPR.L	2.04	0.98	5.18	GTPATP-2	18985
Adenylate kinase 2	R.AVLLGPPGAGK#GTQAPR.L	2.04	0.98	3.73	GTPATP-2	19105
Adenylate kinase 2	R.AVLLGPPGAGK#GTQAPR.L	2.26	0.99	2.70	GTPATP-3	27822
Adenylate kinase 2	R.AVLLGPPGAGK#GTQAPR.L	2.24	0.99	5.25	GTPATP-3	27803
Adenylate kinase 2	R.AVLLGPPGAGK#GTQAPR.L	2.24	0.99	4.49	GTPATP-3	27597
Adenylate kinase 2	R.AVLLGPPGAGK#GTQAPR@.L	2.53	0.99	2.98	ATPGTP-1	23809
Adenylate kinase 2	R.AVLLGPPGAGK#GTQAPR@.L	2.65	0.99	5.14	ATPGTP-1	23574
Adenylate kinase 2	R.AVLLGPPGAGK#GTQAPR@.L	2.76	0.99	5.10	ATPGTP-2	20711
Adenylate kinase 2	R.AVLLGPPGAGK#GTQAPR@.L	2.61	0.99	3.38	ATPGTP-2	20755
Adenylate kinase 2	R.AVLLGPPGAGK#GTQAPR@.L	2.30	0.99	3.14	ATPGTP-3	27042
Adenylate kinase 2	R.AVLLGPPGAGK#GTQAPR@.L	2.29	0.99	5.11	ATPGTP-3	26836
Adenylate kinase 2	R.AVLLGPPGAGK#GTQAPR@.L	2.04	0.99	3.06	GTPATP-1	23376
Adenylate kinase 2	R.AVLLGPPGAGK#GTQAPR@.L	2.13	0.99	5.25	GTPATP-1	23732
Adenylate kinase 2	R.AVLLGPPGAGK#GTQAPR@.L	1.11	0.86	2.76	GTPATP-2	19059
Adenylate kinase 2	R.AVLLGPPGAGK#GTQAPR@.L	2.08	0.99	4.92	GTPATP-2	19527
Adenylate kinase 2	R.AVLLGPPGAGK#GTQAPR@.L	1.92	0.99	2.99	GTPATP-2	18995
Adenylate kinase 2	R.AVLLGPPGAGK#GTQAPR@.L	2.26	0.99	3.10	GTPATP-3	27675
Adenylate kinase 2	R.AVLLGPPGAGK#GTQAPR@.L	2.24	0.99	5.14	GTPATP-3	27425

of all the quantified peptides from a single protein allows for the prediction of the protein's overall nucleotide-binding property.

4. Global profiling of ATP/GTP binding selectivity of nucleotide-binding proteins from the whole human proteome

Previous studies revealed that the cellular concentration of ATP is approximately 6-7 times higher than GTP (27). Thus, those proteins with similar or even slightly better affinity towards GTP at the same concentrations of ATP and GTP may be associated with more ATP than GTP in cells. To better mimic physiological conditions and to identify true novel GTP-binding proteins, we carried out similar ATP/GTP competition experiment using ~ 7:1 (100:15 μM) desthiobiotin-based ATP/GTP probe concentration ratio. We employed online 2D-LC-MS/MS with SCX-C18 biphasic column in 7:1 ATP/GTP competition experiment to obtain better coverage of nucleotide-binding proteins. As a result, a larger number (a total of 1695) of proteins were successfully quantified. For example, we were able to determine the $R_{ATP/GTP}$ ratios of 206 kinases from four trials of 7:1 ATP/GTP competition experiments (Figure 2.11). Notably, these 206 kinases, including 153 protein kinases, were identified from a single cell line (HL-60 cells) and is estimated to cover approximately half of the expressed kinome in a given mammalian cell, which consists of up to 300 distinct protein kinases (28). Moreover, only approximately 1 mg light and heavy cell lysates were used in our approach, whereas 20-40 mg cell lysates were used for kinase enrichment previously (28-30). This

Figure 2.11 A heatmap showing the relative selectivity of kinases towards ATP/GTP binding based on median R$_{ATP/GTP}$ ratio. Dark blue and white boxes indicate significant ATP-binding preference with high R$_{ATP/GTP}$ ratio and significant GTP-binding preference with small R$_{ATP/GTP}$ ratio, respectively, as indicated by the scale bar above the heatmap.
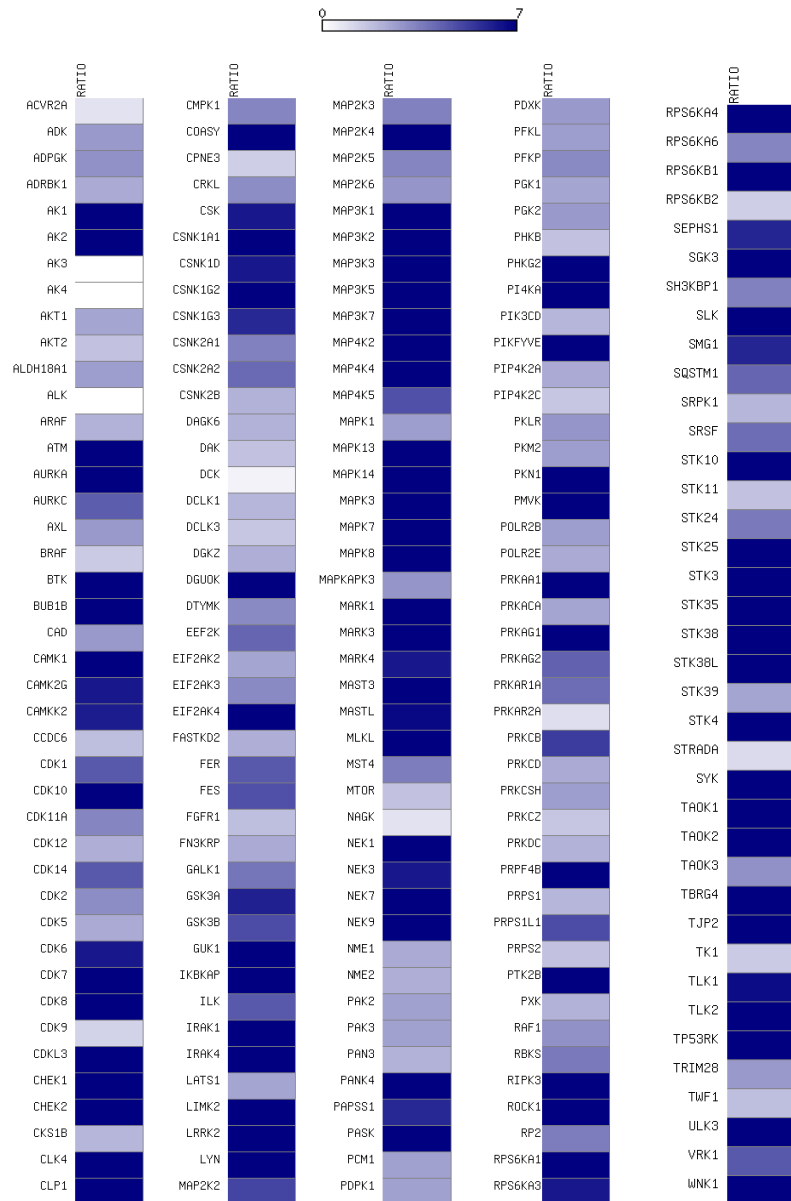
demonstrates the sensitivity of our approach in detecting kinases from complex whole cell lysates.

As expected, a vast majority of $R_{ATP/GTP}$ ratios for the quantified nucleotide-binding proteins are well above 1. For example, the average $R_{ATP/GTP}$ ratio for all the proteins with ATP-binding GO annotation (GO: 0005524) reaches 6.08 in 7:1 ratio experiment, which is accompanied with an increase in average $R_{ATP/GTP}$ ratio for all the known kinases to 12.2. These results suggest that most purine nucleotide-binding proteins bind preferentially to ATP *in vivo*. Because of the pivotal role of nucleotide binding in regulating kinases activity in signal transduction cascades, we next sought to systematically assess the ATP/GTP binding affinity of known kinases from LC-MS/MS results based on 7:1 ATP/GTP competition experiment. A heatmap was generated to better visualize the relative ATP/GTP binding affinity of the 206 identified kinases in 7:1 ATP/GTP competition experiment according to the median $R_{ATP/GTP}$ ratios for all the quantified peptides for each kinase (Figure 2.11). White box was set for a minimum $R_{ATP/GTP}$ ratio of 0. Because the instrument and software limitations in detecting extremely large ratios in SILAC experiment, a $R_{ATP/GTP}$ ratio of 7 or greater was represented with dark blue box. Some insights into ATP/GTP selectivity of kinases can be gleaned from the heatmap. As shown in Figure 2.11, even though most kinases display blue or dark blue color (i.e., with preferential binding to ATP over GTP), kinases from different subgroups with various structure and function assume distinct ATP/GTP binding selectivity. For example, all four identified NimA-related kinases (NEK1,

83

NEK3, NEK7, NEK9) show dark blue color in our heatmap, indicating their significant preference toward ATP-binding. This is consistent with the fact that these four kinases are known to adopt ATP as phosphate donor (31). However, *N*-acetyl-D-glucosamine (GlcNAc) kinase (NAGK) and nucleoside diphosphate kinases (NME1, NME2) display light blue color in the heatmap, translating to moderate ATP/GTP binding selectivity. Indeed, previous studies indicate that NAGK and NME kinases may have the capacity to bind GTP, and GTP may serve as the phosphate donor for NAGK to convert endogenous GlcNAc to GlcNAc-6-phosphate (32).

We also found a significant number of kinases in the same subfamily exhibiting differential selectivity toward binding to ATP and GTP. For instance, among 7 casein kinases detected (shown as CSNK in heatmap), four of them belong to casein kinases 1 and all show dark blue colors with strong preference for ATP binding. However, the other three identified casein kinases display significantly lighter blue color and all of them belong to casein kinase 2. This is consistent with the previous findings that casein kinase 2 can effectively utilize both ATP and GTP in the phosphotransferase reaction, whereas casein kinase 1 is only known to bind to ATP (24, 33). Among the cyclin-dependent kinases, CDK5, CDK9 and CDK12 display strong preference for GTP binding; in contrast, CDK6, CDK7, CDK8 and CDK10 exhibit more favorable binding toward ATP than GTP. Taken together, our SILAC-based competition experiment enables global profiling of binding selectivity of kinases toward ATP and GTP.

On the other hand, even in 7:1 ATP/GTP competition experiment, some proteins still display preferential binding toward GTP with $R_{ATP/GTP}$ ratios smaller than one. Therefore, we considered those proteins with medium $R_{ATP/GTP}$ ratio smaller than 1 in 7:1 ATP/GTP competition experiment as true GTP-binding proteins, which include 119 unique proteins. GO analysis using DAVID (34) showed that 42 out of those 119 proteins were with GTP-binding GO (GO: 0005525), suggesting a 14-fold increase in enrichment factor relative to entire human proteome with a p-value of 6.2E-34. Other than the known GTP-binding proteins, 14 known ATP-binding proteins including 6 kinases still show median ratio < 1. Six of these 14 proteins with known ATP-binding GO were also detected in 1:1 ATP/GTP competition experiment and all of them show clear preference for GTP binding, demonstrating the reproducibility of our ATP/GTP competition approach. Furthermore, aside from AK3, AK4 and SUCLG2, the remaining 11 proteins with ATP-binding GO annotation were not annotated as GTP-binding proteins. However, there are still a few reports supporting the GTP binding affinity of some of these proteins. For example, it was found that GTP serves as a slightly better phosphate donor for deoxycytidine kinase (DCK) than ATP, though most studies only considered ATP as its phosphate donor (35). In our 7:1 ATP/GTP competition experiments, we found that DCK exhibits preferential binding toward GTP over ATP with the biotin-labeled peptide (K.ISIEGNIAAGK#STFVNILK.Q). In another example, methylmalonic aciduria type A (MMAA) is annotated as an ATP-binding protein in Uniprot (Accession #: Q495G5). However, 7:1 ATP/GTP competition experiments led

to the identification of only one peptide from MMAA (R.VGLSGPPGAGK#STFIEYFGK.M) with a very small $R_{ATP/GTP}$ ratio of 0.19. This is in keeping with a recent study revealing that MMAA shares similar sequence motif as GTPase chaperons and possesses GTPase activity (36). Therefore, ATP/GTP competition experiment not only confirms the nucleotide-binding affinity of known ATP/GTP-binding proteins, but also serves as a discovery tool to explore novel nucleotide-binding proteins without prior knowledge of GO annotation.

5. Discovery of nucleotide binding motif in proteins

As mentioned above, our nucleotide competition strategy distinguishes from other nucleotide-binding assay by virtue of site-specific differentiation of ATP and GTP binding. In this regard, we sought to analyze local sequence context surrounding the biotin-modified lysine, which is considered as the nucleotide-binding site in proteins (Figure 2.12). First, we examined all the biotin-labeled peptides from known ATP-binding proteins. The well-known P-loop sequence motif of GxxxxGK was successfully identified by motif-X (37) search with a 9.9-fold enrichment with respect to the occurrence frequency in our input sequence compared to the entire proteome. In addition, another two unique motifs of HRDxKxxN and VAxK, which were identified with enrichment factors (352 and 12.3, respectively) greater than that for the well-known GxxxxGK motif, are derived exclusively from kinases. For peptides from GTP-binding protein group, only GxxxxGK motif was identified, which agrees with the previous

Figure 2.12 Unique binding motifs found for known kinases (A, B, C) and known GTP-binding proteins (D) in 7:1 ATP/GTP competition experiment.
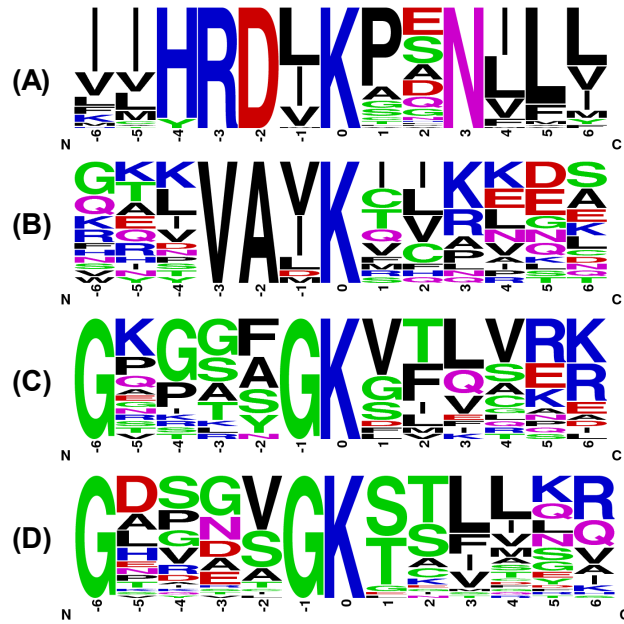
observation that GxxxxGK motif is the conserved nucleotide binding site in GTP-binding proteins (18).

The proximity of nucleotide γ-phosphate to the catalytic domain is a prerequisite for the enzymatic activity of kinases. Previous studies revealed that there are three major sequence components in kinases that are essential for nucleotide binding (38). First, the glycine-rich loop of subdomain I at the N-terminus of the kinase domain contains the consensus motif of GxxxxGKV. The second part of the kinase domain that has been implicated in ATP binding localizes in a region of subdomain II with an invariant lysine residue. In addition, most kinases show another conserved motif of HRDxKxxN located in subdomain VIB, which participates in ATP binding (39, 40). It is of note that the nucleotide-binding site of kinases or other nucleotide-binding proteins are conformationally dynamic, and ATP may bind to different sites in nucleotide-binding proteins under different physiological states (25). Therefore, it is very difficult to predict the most relevant amino acid residues involved in ATP binding solely based on motif analysis, especially a significant number of kinases possess at least two of the three possible binding sequences. Interestingly, our motif analysis of peptides from ATP-binding protein groups in 7:1 ATP/GTP competition experiments uncovers the aforementioned three possible ATP binding sites in kinases successfully (K in VAxK motif corresponds to the invariant lysine in subdomain II). The motif with the most significant ATP-binding preference may represent the most relevant region involved in

kinase's binding toward γ-phosphate of ATP, thereby providing important knowledge basis for designing mutants to modulate kinase activity.

Among all of 206 kinases identified, 123 contained the biotin-labeled peptides covering at least one of these three motifs in 7:1 ATP/GTP competition experiment. A heatmap was created to better visualize the ATP/GTP binding selectivity of these three motifs for each kinase (Figure 2.13). Meanwhile, we map the kinases to the human kinome dendrogram with color annotation to specify the motif with the most profound ATP/GTP selectivity (i.e., with the largest $R_{ATP/GTP}$ ratio). As shown in Figure 2.13, the reactive lysine most frequently resides in the consensus motif of HRDxKxxN in kinases, except for tyrosine kinases. For example, both GxxxxGK and HRDxKxxN motifs are present in ribosomal protein S6 kinase alpha-1 (RPS6KA1), whereas K94 in RPS6KA1 is the equivalent lysine in the VAxK motif except that V is replaced with a Y. However, two biotin-labeled peptides derived from the HRDxKxxN motif, namely, R.DLK#PENILLDEEGHIK.L including K189 ($R_{ATP/GTP}$ ratio= 61.3) and R.DLK#PSNILYVDESGNPECLR.I including K537 ($R_{ATP/GTP}$ ratio= 51.2), were detected in our ATP/GTP competition experiment with significant selectivity of ATP over GTP. This finding is consistent with the X-ray crystal structure of RPS6KA1 (PDB entry: 2Z7Q, residue 33-353), which reveals the close proximity of K189 to the reactive γ-phosphate group of ATP; however, K94 and K75 in the VAxK and GxxxxGK motifs are more proximal to non-transferable β- and α-phosphates, respectively (Figure 2.14).

Figure 2.13 (A) A heatmap displaying the relative selectivity of ATP/GTP towards each possible binding motif of kinases. Dark blue and white indicate significant ATP-binding preference with high $R_{ATP/GTP}$ ratio and significant GTP-binding preference with small $R_{ATP/GTP}$ ratio, respectively (See scale bar above the heatmap). (B) Protein kinases were mapped to the human kinome dendrogram with color annotation to specify the motif with the highest $R_{ATP/GTP}$ ratio. Red, green and blue represent kinases with HRDxKxxN, VAxK, and GxxxxGK motifs, respectively.
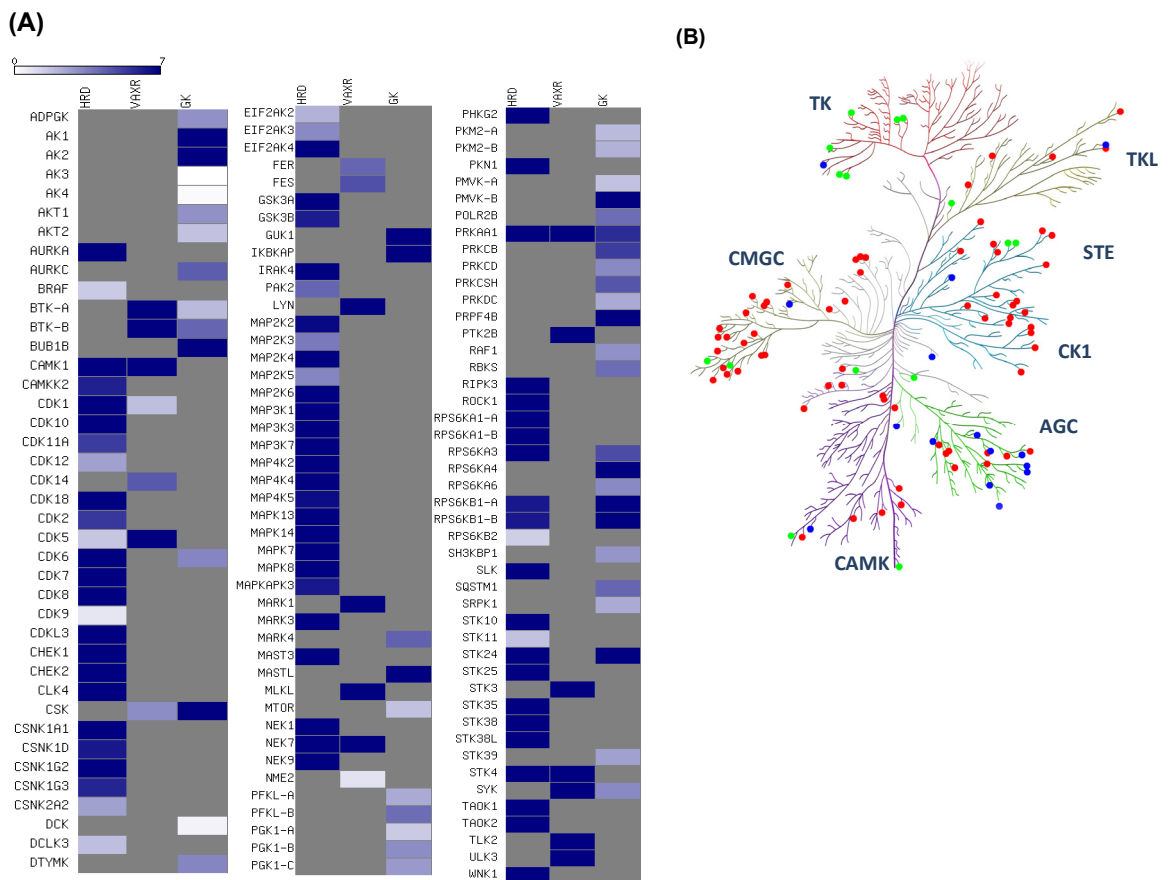
Figure 2.14 The crystal structures of RPS6KA1 (left) and Lyn (right) bound with ATP
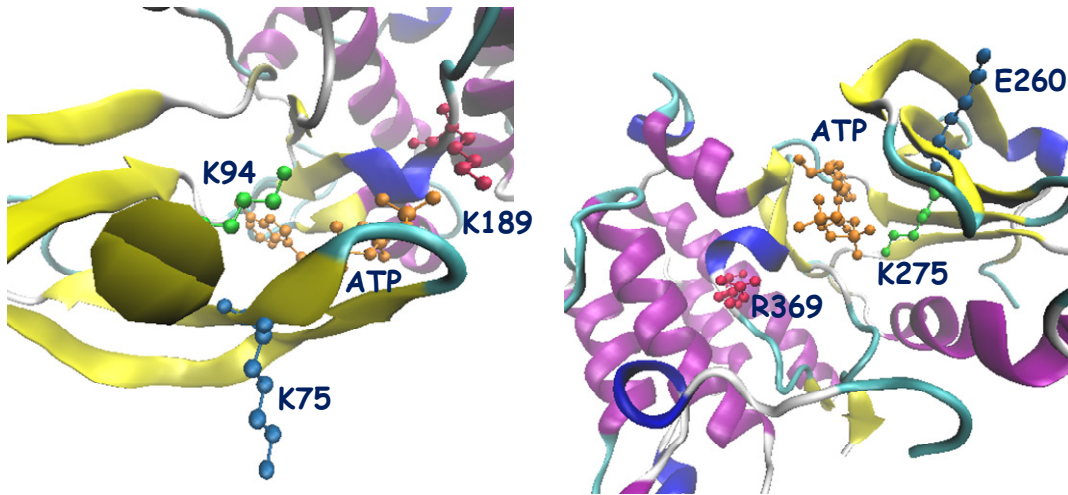
The identification of HRDxKxxN motif with a moderate $R_{ATP/GTP}$ ratio could raise the possibility that other sequence may contribute to ATP binding, even though HRDxKxxN is the only motif identified. For instance, one peptide with K416 residing in the HRDxKxxN motif of double-stranded RNA-activated protein kinase (PKR or EIF2AK2) was quantified with a moderate $R_{ATP/GTP}$ ratio of 2.08, whereas another peptide containing K296 (K.TYVIK#R.V) without any clear motif feature shows a much larger $R_{ATP/GTP}$ ratio of 61.4. Site-directed mutagenesis studies showed that substitution of K296, but not K416, with an arginine gave rise to a catalytically dead mutant that completely abolishes the kinase activity of PKR (41). These examples demonstrate the capability of our approach in differentiating the selectivity of nucleotide binding affinity to both known binding motifs and previously unrecognized nucleotide-binding site of kinases. The latter facilitates the identification of the lysine residue that is most relevant for ATP binding and hydrolysis even in the absence of structural information of the protein.

Most identified tyrosine kinases carry the VAxK, instead of HRDxKxxN, as the motif exhibiting the greatest selectivity toward ATP binding. For instance, protein tyrosine kinase Lyn has one biotin-modified peptide K.VAVK#TLKPGTMSVQAFLEEANLMK.T with the VAxK motif, in which the biotin-labeled K275 is proximal to the transferable phosphate (PDB entry: 2ZV8). Careful examination of all the tyrosine kinases identified from our competition experiment revealed that the lysine residue in the HRDxKxxN motif is substituted with an arginine in

tyrosine kinases. This arginine is in closer proximity to γ-phosphate of ATP than the lysine residue in the VAxK motif (Figure 2.14). However, arginine does not react with our amine-reactive probe and K275 is the only available reactive lysine in Lyn sequence. Interestingly, the VAxK motif peptide is rarely detected for serine-threonine kinases, despite the fact that they are generally present in the primary sequence of these kinases, which distinguishes serine-threonine protein kinases from tyrosine kinases.

GxxxxGK motif generally serves as the most selective binding motif for metabolic kinases, which lack the HRDxKxxN and VAxK motifs. For example, we detected four kinases in the adenylate kinase (AK) subfamily and biotin-labeled peptides from all of them contain a lysine in the GxxxxGK motif, which was found to be the closest lysine to γ-phosphate of ATP as revealed by crystal structure analysis(18). This finding allowed us to precisely characterize the ATP/GTP selectivity of proteins in the same subfamily with similar structures by targeting unique nucleotide binding motif. AK1 and AK2, which use ATP as the phosphate donor, are colored in dark blue in heatmap. By contrast, AK3 and AK4 with very similar binding sequences of R.AVIMGAPGSGK#GTVSSR.I ($R_{ATP/GTP}$ =0.05) and R.AVILGPPGSGK#GTVCQR.I ($R_{ATP/GTP}$ =0.11) exhibit robust GTP binding preference. Similar to AK2, AK3 is a phosphotransferase enzyme located in mitochondria and it induces AMP phosphorylation; however, AK3 can only use GTP in lieu of ATP as the phosphate donor (42). Although AK4 is annotated as an ATP-AMP transphosphorylase in Uniprot (Accession #: P27144), recombinant human AK4 exhibited greater efficiency in

phosphorylating AMP in the presence of GTP than in the presence of ATP (43, 44). Moreover, Blast search revealed that AK4 has approximately 60% sequence similarity to AK3 but only around 40% similarity to AK2. Besides, we also note that most protein kinases exhibiting the greatest probe-labeling selectivity toward GxxxxGK motif belong to the AGC kinase group. For example, protein kinase C beta (PKCb) and protein kinase C delta (PKCd), belonging to AGC kinase group, only have one peptide within GxxxxGK motifs quantified with $R_{ATP/GTP}$ ratio of 5.26 and 3.27, respectively, despite the fact that they both possess the HRDxKxxN motif. This finding suggests that the lysine in GxxxxGK motif may play an important role in deciding the specificity of nucleotide binding of protein kinase C in aqueous solution. However, further experiments are needed to validate this notion.

The capabilities of our approach in assessing quantitatively the nucleotide binding selectivities towards unique binding motifs can also be extended to GTP-binding proteins. In this vein, GxxxxGK is the only conserved nucleotide-binding motif for GTP-binding proteins, and a heatmap was generated to better visualize the ATP/GTP binding selectivity of this motif for each identified known GTP-binding protein (Figure 2.15). As depicted on the map, the GxxxxGK motif in most known GTP-binding proteins, including four heterotrimeric G proteins as well as four translation initiation and elongation factors, shows significant binding preference toward GTP. In addition, 21 small GTPases, including 14 Rab proteins, have been quantified and the GxxxxGK motif in all of them displays significant binding preference to GTP. In contrast, GxxxxGK

94

Figure 2.15 A heatmap showing the relative selectivity of ATP/GTP towards the

GxxxxGK motif of GTP-binding proteins.
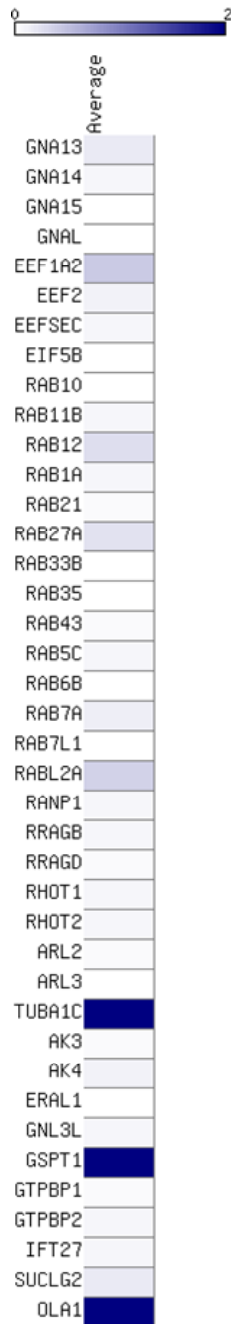
motif in GTP-binding protein 9 (GTPBP9) exhibits significant preference toward ATP-binding. GTPBP9 is the human homologue of bacterial YchF, which is considered a GTPase and involved in protein translation. However, Roland et al. (45) found that GTPBP9 binds and hydrolyzes ATP more efficiently than GTP and hence renamed human GTPBP9 to Obg-like ATPase 1. Our result further demonstrated the ATP binding preference for GTPBP9 protein, which furnishes another line of evidence to support that the GxxxxGK motif is the actual ATP-binding site in GTPBP9. The identification of a GxxxxGK motif with preferential binding toward ATP in a previously documented GTP-binding protein further illustrates the ability of our approach in exploiting novel nucleotide-protein interactions and in uncovering the actual nucleotide binding sites.

Conclusions

Here, we report a general strategy using affinity chemical probe to enrich and identify ATP/GTP-binding proteins in human cell proteome. Our ATP/GTP affinity probe mainly has two applications. First, it can be used as an enrichment method to facilitate the subsequent LC-MS/MS identification and quantification of ATP/GTP-binding proteins at the global proteome scale; more than 100 GTP-binding proteins and 206 kinases from cell lysate of a single cell line could be identified and quantified with low mg quantity of lysate. Therefore, the ATP and GTP affinity probes constitute powerful reagents for the enrichment and subsequent detection of nucleotide-binding

proteins from human proteome, which will enable sensitive and accurate profiling of cell signaling pathway perturbed by extracellular stimuli.

Second, this approach, in conjunction with quantitative proteomics, can be employed to characterize, at the entire proteome level, nucleotide-protein interactions and to identify specific nucleotide-binding sites in proteins. ATP/GTP binding selectivity of nucleotide-binding proteins in the entire proteome was systematically investigated by quantifying peak intensity ratios of light and heavy biotin-labeled tryptic peptides from ATP or GTP probe reaction, demonstrating the great potential of our ATP/GTP affinity probes in quantitative proteomic analysis. A key advantage in charactering nucleotide-protein interaction with our affinity probe over traditional binding assay lies in that our strategy allows for site-specific determination of relative binding affinities of proteins towards different nucleotides (e.g. ATP and GTP). Thus, nucleotide-protein interaction studies can be extended to quantitative surveys of specific binding regions in proteins of interest. Furthermore, it can be envisaged that similar nucleotide-affinity probes can be generally applied for the capture and characterization of proteins that can bind to other nucleotides, and such experiments are currently being pursued in our laboratory.

References

1.   Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. Science. 2002;298:1912-34.

2.   Takai Y, Sasaki T, Matozaki T. Small GTP-binding proteins. Physiol Rev. 2001;81:153-208.

3.   Choudhary C, Kumar C, Gnad F, Nielsen ML, Rehman M, Walther TC, et al. Lysine acetylation targets protein complexes and co-regulates major cellular functions. Science. 2009;325:834-40.

4.   Dong X, Xiao Y, Jiang X, Wang Y. Quantitative proteomic analysis revealed lovastatin-induced perturbation of cellular pathways in HL-60 cells. J Proteome Res. 2011;10:5463-71.

5.   Taylor P, Nielsen PA, Trelle MB, Andersen MB, Vorm O, Moran MF, et al. Automated 2D peptide separation on a 1D Nano-LC-MS system. J Proteome Res. 2009;8:1610-6.

6.   Ito J, Heazlewood JL, Millar AH. Analysis of the soluble ATP-binding proteome of plant mitochondria identifies new proteins and nucleotide triphosphate interactions within the matrix. J Proteome Res. 2006;5:3459-69.

7.   Daub H, Olsen JV, Bairlein M, Gnad F, Oppermann FS, Stemmann O, et al. Kinase-selective enrichment enables quantitative phosphoproteomics of the kinome across the cell cycle. Mol Cell. 2008;31:438-48.

8.   Adam GC, Sorensen EJ, Cravatt BF. Trifunctional chemical probes for the consolidated detection and identification of enzyme activities from complex proteomes. Mol Cell Proteomics. 2002;1:828-35.

9.   Ohta T, Nagano K, Yoshida M. The active site structure of Na+/K+-transporting ATPase: location of the 5'-(p-fluorosulfonyl)benzoyladenosine binding site and soluble peptides released by trypsin. Proc Natl Acad Sci U S A. 1986;83:2071-5.

10.   Kaneda M, Masuda S, Tomohiro T, Hatanaka Y. A simple and efficient photoaffinity method for proteomics of GTP-binding proteins. Chembiochem. 2007;8:595-8.

11.   Qiu H, Wang Y. Probing adenosine nucleotide-binding proteins with an affinity-labeled nucleotide probe and mass spectrometry. Anal Chem. 2007;79:5547-56.

12.   Patricelli MP, Szardenings AK, Liyanage M, Nomanbhoy TK, Wu M, Weissig H, et al. Functional interrogation of the kinome using nucleotide acyl phosphates. Biochemistry (Mosc). 2006;46:350-8.

13.   Ong S-E, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, et al. Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. Mol Cell Proteomics. 2002;1:376-86.

14.   Olsen JV, de Godoy LMF, Li G, Macek B, Mortensen P, Pesch R, et al. Parts per Million Mass Accuracy on an Orbitrap Mass Spectrometer via Lock Mass Injection into a C-trap. Mol Cell Proteomics. 2005;4:2010-21.

15.   Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotech. 2008;26:1367-72.

16.   Tabb DL, McDonald WH, Yates JR. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. J Proteome Res. 2002;1:21-6.

17.   Park SK, Venable JD, Xu T, Yates JR. A quantitative analysis software tool for mass spectrometry-based proteomics. Nat Methods. 2008;5:319-22.

18.   Saraste M, Sibbald PR, Wittinghofer A. The P-loop--a common motif in ATP- and GTP-binding proteins. Trends Biochem Sci. 1990;15:430-4.

19.   Deyrup AT, Krishnan S, Cockburn BN, Schwartz NB. Deletion and site-directed mutagenesis of the ATP-binding motif (P-loop) in the bifunctional murine Atp-sulfurylase/adenosine 5-phosphosulfate kinase enzyme. J Biol Chem. 1998;273:9450-6.

20.   Borisov OV, Goshe MB, Conrads TP, Rakov VS, Veenstra TD, Smith RD. Low-Energy Collision-Induced Dissociation Fragmentation Analysis of Cysteinyl-Modified Peptides. Anal Chem. 2002;74:2284-92.

21.   Sioud S, Genestie B, Jahouh F, Martin P, Banoub J. Gas-phase fragmentation study of biotin reagents using electrospray ionization tandem mass spectrometry on a quadrupole orthogonal time-of-flight hybrid instrument. Rapid Commun Mass Spectrom. 2009;23:1941-56.

22.   Rybak JN, Scheurer SB, Neri D, Elia G. Purification of biotinylated proteins on streptavidin resin: a protocol for quantitative elution. Proteomics. 2004;4:2296-9.

23.   Hirsch JD, Eslamizar L, Filanoski BJ, Malekzadeh N, Haugland RP, Beechem JM, et al. Easily reversible desthiobiotin binding to streptavidin, avidin, and other biotin-binding proteins: uses for protein labeling, detection, and isolation. Anal Biochem. 2002;308:343-57.

24.   Jakobi R, Traugh JA. Analysis of the ATP/GTP binding site of casein kinase II by site-directed mutagenesis. Physiol Chem Phys Med NMR. 1995;27:293-301.

25.   Soti C, Vermes A, Haystead TAJ, Csermely P. Comparative analysis of the ATP-binding sites of Hsp90 by nucleotide affinity cleavage: a distinct nucleotide specificity of the C-terminal ATP-binding site. Eur J Biochem. 2003;270:2421-8.

26.   Zhang T, Li S, Zhang Y, Zhong C, Lai Z, Ding J. Crystal structure of the ARL2-GTP-BART complex reveals a novel recognition and binding mode of small GTPase with effector. Structure (London, England : 1993). 2009;17:602-10.

27.   Pillwein K, Chiba P, Knoflach A, Czermak B, Schuchter K, Gersdorf E, et al. Purine metabolism of human glioblastoma in vivo. Cancer Res. 1990;50:1576-9.

28.   Oppermann FS, Gnad F, Olsen JV, Hornberger R, Mann M, Daub H. Large-scale proteomics analysis of the human kinome. Mol Cell Proteomics. 2009;8:1751-64.

29.   Bantscheff M, Eberhard D, Abraham Y, Bastuck S, Boesche M, Hobson S, et al. Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. Nat Biotech. 2007;25:1035-44.

30.   Duncan James S, Whittle Martin C, Nakamura K, Abell Amy N, Midland Alicia A, Zawistowski Jon S, et al. Dynamic reprogramming of the kinome in response to targeted MEK inhibition in triple-negative breast cancer. Cell. 2012;149:307-21.

31.   O'Regan L, Blot J, Fry A. Mitotic regulation by NIMA-related kinases. Cell Div. 2007;2:25.
32.   Datta A. Studies on hog spleen N-acetylglucosamine kinase. I. Purification and properties of N-acetylglucosamine kinase. Biochim Biophys Acta. 1970;220:51-60.

33.   Srinivasan N, Antonelli M, Jacob G, Korn I, Romero F, Jedlicki A, et al. Structural interpretation of site-directed mutagenesis and specificity of the catalytic subunit of protein kinase CK2 using comparative modelling. Protein Eng. 1999;12:119-27.

34.   Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protocols. 2008;4:44-57.

35.   Coleman C, Stoller R, Drake J, Chabner B. Deoxycytidine kinase: properties of the enzyme from human leukemic granulocytes. Blood. 1975;46:791-803.

36.   Takahashi-Iniguez T, Garcia-Arellano H, Trujillo-Roldan MA, Flores M. Protection and reactivation of human methylmalonyl-CoA mutase by MMAA protein. Biochem Biophys Res Commun. 2011;404:443-7.

37.   Schwartz D, Gygi SP. An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. Nat Biotech. 2005;23:1391-8.

38.   Hanks SK, Hunter T. Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. . The FASEB Journal 1995 9 576-96

39.   Johnson LN, Noble MEM, Owen DJ. Active and inactive protein kinases: structural basis for regulation. Cell. 1996;85:149-58.

40.   Nolen B, Taylor S, Ghosh G. Regulation of protein kinases: controlling activity through activation segment conformation. Mol Cell. 2004;15:661-75.

41.   Dey M, Cao C, Dar AC, Tamura T, Ozato K, Sicheri F, et al. Mechanistic link between PKR dimerization, autophosphorylation, and eIF2± substrate recognition. Cell. 2005;122:901-13.

42.   Schricker R, Magdolen V, Strobel G, Bogengruber E, Breitenbach M, Bandlow W. Strain-dependent occurrence of functional GTP:AMP phosphotransferase (AK3) in saccharomyces cerevisiae. J Biol Chem. 1995;270:31103-10.

43.   Panayiotou C, Solaroli N, Johansson M, Karlsson A. Evidence of an intact N-terminal translocation sequence of human mitochondrial adenylate kinase 4. The International Journal of Biochemistry & Cell Biology. 2010;42:62-9.

44.   Liu R, Strom A-L, Zhai J, Gal J, Bao S, Gong W, et al. Enzymatically inactive adenylate kinase 4 interacts with mitochondrial ADP/ATP translocase. The International Journal of Biochemistry & Cell Biology. 2009;41:1371-80.

45.   Koller-Eichhorn R, Marquardt T, Gail R, Wittinghofer A, Kostrewa D, Kutay U, et al. Human OLA1 defines an ATPase subfamily in the Obg family of GTP-binding proteins. J Biol Chem. 2007;282:19928-37.

Chapter 3

Isotope-coded ATP Probe for Quantitative Affinity Profiling of ATP-binding Proteins

Introduction

Adenosine triphosphate (ATP)-binding proteins, including ATP-binding cassette transporters, chaperones, and kinases, are one of the most important superfamily of proteins that are involved in a variety of pivotal cellular processes, like cell signaling, proliferation, differentiation, and apoptosis. For example, numerous ATP-binding cassette transporters (ABC transporters) catalyze the hydrolysis of ATP to provide required energy for translocation of various substrates across cell membranes.[1] Additionally, kinases, as one of the largest enzyme superfamilies in higher eukaryotes, mediate cellular protein and lipid phosphorylation to regulate downstream signaling cascade.[2]

Despite the importance of ATP-binding proteins in cellular functions, the proteome-wide picture of ATP-binding proteins is far from complete. The abundance of individual ATP-binding proteins may be assessed by using immunoblot;[3] however, global study of the expression levels of ATP-binding proteins can hardly be achieved with the use of conventional methods owing to the enormous diversity of the ATP-binding protein family. The development of mass spectrometry (MS) instrumentation

and bioinformatic tools provides the opportunity to identify and quantify up to several thousand proteins in complex samples. However, proteomic studies of ATP-binding proteins, including different ATPases and kinases, by MS are still a challenge, which is mainly attributed to the under-representation of some ATP-binding proteins in the proteome. For example, although the group of human kinase is the largest enzyme complement in mammals, it only constitutes 1.7% of the human genome(2). Therefore, selective enrichment of ATP-binding proteins from cellular extracts is essential for the comprehensive identification and quantification of ATP-binding proteins.

Much remains to be learnt about protein-ATP interactions in human proteome. Although a large number of known ATP-binding proteins are documented in gene ontology (GO) annotation database, numerous proteins with available sequence information remain unannotated with respect to their nucleotide-binding affinity.(4) Furthermore, experimental characterization of ATP-protein interaction often relies on radioactivity-based ultrafiltration assay(5) or fluorescence-based binding assay.(6) These traditional methods are usually costly and time-consuming because they require the use of purified proteins, which prevents high-throughput studies at the whole proteome level. Most importantly, none of these approaches permit the robust discovery of ATP-binding site, or 'ATP-interacting residues', which is another aspect of significant interest in nucleotide-protein interaction studies. The ATP binding sites, where ATP is captured and hydrolyzed to ADP, are particularly important because they are responsible for changing the conformation and/or modulating the catalytic activity of proteins.(1)

Additionally, many anti-cancer drugs target ATP-binding sites in ATPases and kinases.(1) Therefore, new methods for robust and systematic characterizations of protein-ATP interactions as well as for the identification of ATP-interacting residues at a global proteome scale are invaluable for understanding better the regulatory mechanisms of ATP-related protein functions.

Recently, we and others reported the application of biotin-conjugated acyl nucleotide probes for the enrichment and identification of ATP-binding proteins from complex protein mixtures.(7, 8) In this context, the entire proteome is treated with an acyl phosphate-linked ATP affinity probe, which allows for the incorporation of a biotin tag to lysine residue on a subset of proteins with ATP-binding affinity. It was demonstrated that this ATP-affinity probe can be harnessed as an enrichment tool to enable the subsequent LC-MS/MS identification and quantification of ATP-binding proteins at the global proteome scale. Moreover, this approach could be employed to characterize, at the entire proteome level, site-specific interaction between ATP and its binding proteins. We demonstrated previously that the lysine residue labeled by the biotin-ATP probe may represent the specific lysine residue that directly contacts the γ-phosphate group of the bound ATP. However, both applications rely on stable isotope labeling by amino acids in cell culture (SILAC) because proteins or peptides in the different experimental groups need to be differentiated to achieve quantitative comparison. However, the SILAC labeling approach may not be compatible with some applications of the ATP affinity probe. For instance, clinical samples (e.g. biological

fluids, tissue samples) are not amenable to SILAC.(9) In addition, chemical isotopic labeling methods like dimethyl labeling(10) are prone to additional experimental errors because the isotopic tags are usually incorporated in the late stage of the sample preparation.

To overcome this limitation, we introduced the desthiobiotin-based Isotope-Coded ATP-affinity Probe (ICAP) as an acylating agent to selectively label and enrich ATP-binding proteins from the human proteome. With the use of this novel probe, ATP-binding proteins from whole cell lysate can be simultaneously enriched and isotopically labeled. In addition, we developed a novel quantitative profiling strategy with the use of ICAP to characterize ATP-protein interactions and determine specific ATP-binding sites in ATP-binding proteins at the entire proteome scale.

Experimental Procedures

Preparation of the Biotinylated Nucleotide Affinity Probe

The biotinylated nucleotide affinity probes were prepared according to previously published procedures with minor modifications(11). In this regard, we first prepared light or heavy form of desthiobiotinyl-aminobutyric acid (desthiobiotin-C4) following previously published procedures. Isobutyl chloroformate (0.19 mL) was added to a solution containing 150 mg desthiobiotin, 15 mL DMF, and 0.38 mL tri-*n*-butylamine. After being incubated at room temperature for 10 min, the mixture was slowly added to a suspension of 230 mg light or heavy γ-aminobutyric acid (sigma) in DMF (7.5 mL) at 5 °C. After stirring at 5 °C for 2 h, the solvent was removed under reduced pressure and the

crude product was dissolved in 5 mL water at 40 °C. The solution pH was adjusted to 2 with 2.0 N HCl, and the mixture was kept at 0 °C for 12 h. The desired desthiobiotin-C4 was precipitated out of solution. The precipitate was filtered, washed with water, and dried under vacuum.

To render nucleotide soluble in organic solvent, the commercially available sodium salt form of ATP was first converted to the tributylammonium form by passing the nucleotides through a cation-exchange column packed with Spectra/Gel IE 50×8 resin (40-75 μm) at 4°C once. Fractions containing tributylammonium form of ATP were collected and lyophilized. Desthiobiotin-C4 (8 mg), dissolved in a 1-mL solvent mixture of ice-cold dry CH_2Cl_2 and DMF (4:1, v/v), was mixed with tri-n-butylamine (11 μL) and ethyl chloroformate (5 μL). After stirring at 0 °C for 5 min, the mixture was stirred at room temperature under argon atmosphere for another 60 min. Tributylammonium form of ATP (50 mg), dissolved in a 1.25 mL solution of CH_2Cl_2 and DMF (4:1, v/v), was then added to the above reaction mixture. The reaction was continued at room temperature and under argon atmosphere for 18 h. The CH_2Cl_2 was then removed by argon purging for 10 min and the remaining 200 μL solution was directly subjected to HPLC purification with a YMC ODS-AQ column (4.8×250 mm, 120 Å in pore size, 5 μm in particle size, Waters). The flow rate was 0.8 mL/min, and a 45 min linear gradient of 0-30% acetonitrile in 50 mM triethylammonium acetate (pH 6.8) was used for the purification. A UV detector was set at 265 nm to monitor the effluents. Appropriate HPLC fractions

were pooled, lyophilized, and stored at -80 °C. The structures of the products were confirmed by ESI-MS (Figure 3.1)(11).

Cell lysate preparation and labeling with the nucleotide affinity probe

HeLa-S3 cells were purchased from the National Cell Culture Center (Minneapolis, MN). Jurkat-T cells (ATCC; Manassas, VA) were cultured in RPMI-1640 medium supplemented with 10% fetal bovine serum (FBS, Invitrogen, Carlsbad, CA) and penicillin (100 IU/mL). Cells were maintained in a humidified atmosphere with 5% CO_2 at 37°C. Approximately 2×10^7 cells were harvested, washed with cold PBS for three times, and lysed in 1-mL lysis buffer, which contained 0.7% CHAPS, 50 mM HEPES (pH 7.4), 0.5 mM EDTA, 100 mM NaCl, and 10 μL (1:100) protease inhibitor cocktail on ice for 30 min. The cell lysates were centrifuged at 16000g at 4°C for 30 min, and the resulting supernatants were collected and subjected to gel filtration separation using NAP-25 columns (Amersham Biosciences) to remove free endogenous nucleotides. Cell lysates were eluted into a 2 mL buffer, containing 50 mM HEPES (pH 7.4), 75 mM NaCl, and 5% glycerol. The resulting proteins in cell lysates were quantified using Quick Start Bradford Protein Assay (Bio-Rad, Hercules, CA) and stored at -80°C. Immediately prior to the labeling reaction, $MgCl_2$, $MnCl_2$, and $CaCl_2$ were added to the concentrated cell lysate until their final concentrations reached 50, 5, and 5 mM, respectively. Approximately 1 mg cell lysate was treated separately with light and heavy labeled desthiobiotin-ATP affinity probe at concentrations of 10 and 100 μM, respectively. Labeling reactions were carried out with gentle shaking at room temperature for 1.5 h.

Figure 3.1 The structure of isotope-coded ATP affinity probe (ICAP). (B) A scheme showing the synthesis of isotope-coded ATP affinity probes (C) Representative HPLC trace for the purification of the crude product of ATP-affinity probe. (D) Representative of ESI-MS of purified desthiobiotin-based light and heavy ATP affinity probes.

(A)

(B)

(C)

(D)

After the reaction, the remaining probes in the cell lysates were removed by buffer exchange with 25 mM NH$_4$HCO$_3$ (pH 8.5) using Amicon Ultra-4 filter (10,000 NMWL, Millipore).

In-solution enzymatic digestion and affinity purification

After addition of 8 M urea for protein denaturation, and dithiothreitol and iodoacetamide for cysteine reduction and alkylation, the labeled proteins were digested with modified sequencing-grade trypsin (Roche Applied Science) at an enzyme/substrate ratio of 1:100 in 25 mM NH$_4$HCO$_3$ (pH 8.5) at 37°C for overnight. The peptide mixture was subsequently dried in a Speed-vac and redissolved in 1 mL of 100 mM potassium phosphate and 0.15 M NaCl (pH 7.5, PBS buffer), to which solution was subsequently added 200 μL avidin-agarose resin (Sigma-Aldrich). The mixture was then incubated at 25°C for 1 hr with gentle shaking. The agarose resin was then washed with 3 mL PBS buffer and 3 mL H$_2$O to remove unbound peptides, and the labeled peptides were subsequently eluted with 1% TFA in CH$_3$CN/H$_2$O (7:3, v/v) at 65°C.

Fractionation of biotinylated peptides with high-pH reversed-phase

chromatography (High-pH-RP-LC)

The above enriched desthiobiotinylated peptides were redissolved in 20 μL H$_2$O and subjected to high-pH-RP-LC separation on an Agilent 1100 system using a 0.5 mm × 150 mm Zorbax SB-C18 column (5 μm in particle size, Agilent Technologies). Mobile phase A was 15 mM aqueous ammonium formate (pH 7.5), and mobile phase B contained 15 mM ammonium formate in water/acetonitrile (1:9). Peptides were eluted

with a 60-min linear gradient of 5-40% B and six fractions were collected. Each fraction was dried and reconstituted in 0.1% formic acid prior to LC-MS/MS analysis.

LC-MS/MS Analysis

LC-MS/MS analysis was performed on an LTQ-Orbitrap Velos mass spectrometer equipped with a nanoelectrospray ionization source (Thermo Fisher Scientific, San Jose, CA). Samples were automatically loaded from a 48-well microplate autosampler using an EASY-nLC system (Proxeon Biosystems, Odense, Denmark) at 3 µL/min onto a home-made 4 cm trapping column (150 µm i.d.) packed with 5 µm C18 120 Å reversed-phase material (ReproSil-Pur 120 C18-AQ, Dr. Maisch). The trapping column was connected to a 20 cm fused silica analytical column (PicoTip Emitter, New Objective, 75 µm i.d.) with 3 µm C18 beads (ReproSil-Pur 120 C18-AQ, Dr. Maisch). The peptides were then separated with a 120-min linear gradient of 2-35% acetonitrile in 0.1% formic acid and at a flow rate of 250 nL/min. The LTQ-Orbitrap Velos was operated in a data-dependent scan mode. Full-scan mass spectra were acquired in the Orbitrap analyzer with a resolution of 60000 with lock mass option enabled for the ion of *m/z* 445.120025 (12). Up to 20 most abundant ions found in MS with charge state ≥ 2 were sequentially isolated and sequenced in ion trap with a normalized collision energy of 35, an activation q of 0.25 and an activation time of 10 ms.

Data Processing and Analysis.

The raw data were first converted to mzXML files and DTA files using ReAdW (http://sourceforge.net/projects/sashimi/files/) and MzXML2Search

programs, respectively. Bioworks 3.2 was used for protein identification by searching the DTA files against the human IPI protein database version 3.68 (87,062 entries) and its reversed complement. Initial precursor mass tolerance of 10 ppm and fragment mass deviation of 0.8 Th were set as the search criteria. The maximum number of miss-cleavages for trypsin was set as two per peptide. Cysteine carbamidomethylation was considered as a fixed modification, whereas methionine oxidation and light or heavy desthiobiotinylation of lysine (+281.17394 Da and +287.21160 Da, respectively) were included as variable modifications. The search results were then filtered with DTASelect (13) to achieve a peptide false discovery rate of 1%. Census was employed for peptide and protein quantification (14). Extracted-ion chromatograms were first generated for peptide ions based on their m/z values and peptide intensity ratios were subsequently calculated in Census from peak areas found in each pair of extracted-ion chromatograms.

Nucleotide-binding assay

Human proliferating cell nuclear antigen (PCNA) was kindly provided by Prof. Zhihao Zhuang (University of Delaware).(15) Aliquots of TNP-ATP (Fisher Scientific) at a final concentration of 25 µM were mixed with 6 µM protein samples in 50 mM PBS (pH 7.5) containing 20 mM $MgCl_2$ at 4°C and the mixture was incubated for 20 min prior to fluorescence measurements. Proteins at a final concentration of 3 µM were used in all titration experiments. All fluorescence spectroscopic measurements were conducted using a monochromator-based Synergy Mx Microplate Reader (BioTek, Winooski, VT).

112

To monitor TNP-ATP fluorescence, the excitation wavelength ($\lambda_{excitation}$) was set at 408 nm and emission spectra were recorded in the wavelength range of 480-650 nm. Apparent K_d values for TNP-ATP interaction with proteins were obtained from the best fit of the data to a hyperbolic curve of the equation: $F = F_{max}X/(K_d + X)$, where F_{max} is the maximum fluorescence intensity, X is the concentration of TNP-ATP, and K_d is the dissociation constant for the TNP-ATP-protein complex.(16-18)

Results and Discussion

1. Working principle of the isotope-coded ATP-affinity probe (ICAP)

Bioinformatic studies showed that most ATP-binding proteins carry at least one lysine residue in a consensus amino acid sequence motif known as the phosphate-binding loop (P-loop), which constitutes a nucleotide-binding site and is important for the ATPase activity.(19) Therefore, nucleotide affinity probes bearing an acyl phosphate moiety were developed to target the lysine residue in the nucleotide binding site.(7) Here, we devised a novel ICAP-based strategy to simultaneously enrich and isotopically label ATP-binding proteins from cell lysates. The ICAP reagent contains three components (Figure 3.1), namely, a binding moiety (ATP), an enrichment moiety (i.e. desthiobiotin) facilitating downstream purification, and an isotope-coded linker present in light (contains six hydrogens) or heavy form (contains six deuterons). The isotope-coded linker and ATP are conjugated through an acyl phosphate. Upon binding to ATP-binding proteins, the acyl-phosphate component of the ICAP reacts with the ε-amino group of the

113

P-loop lysine residue to yield a stable amide bond. As a result, desthiobiotin along with light or heavy isotope-coded linker is covalently attached to the lysine residue on ATP-binding proteins (Figure 3.2).

A general experimental procedure with the use of ICAP includes the following steps (Figure 3.3): (1) ATP-binding proteins from a protein sample representing one experimental state are derivatized with the isotopically light form of the ATP affinity probe, whereas the ATP-binding proteins from protein sample representing another experiment state are modified with the heavy isotope-coded ATP probe. (2) The two protein samples are combined and digested with trypsin and the resulting light/heavy desthiobiotin-labeled peptide pairs can be enriched with avidin agarose. (3) The affinity-purified peptides with the desthiobiotin tag are analyzed by LC-MS/MS. The MS peak intensity ratios of the light and heavy forms of desthiobiotin-modified peptide pairs can be utilized to obtain accurate quantification results of the corresponding ATP-binding proteins originating from the two experiment states.

2. Strategy for proteome-wide characterization of ATP-protein interactions

Building upon the quantitative capability of the ICAP-based method, we first applied these probes for the characterization of ATP-protein interactions at the global proteome scale. As reported previously,[7] owing to its relatively high reactivity, apart from the lysine residue(s) located at the nucleotide binding site, other lysine residues may also be modified by the nucleotide affinity probe through non-specific electrostatic interactions. To distinguish specific from non-specific labelings, we developed a novel

Figure 3.2 A schematic diagram showing the reaction between isotope-coded ATP affinity probe with an ATP-binding protein.
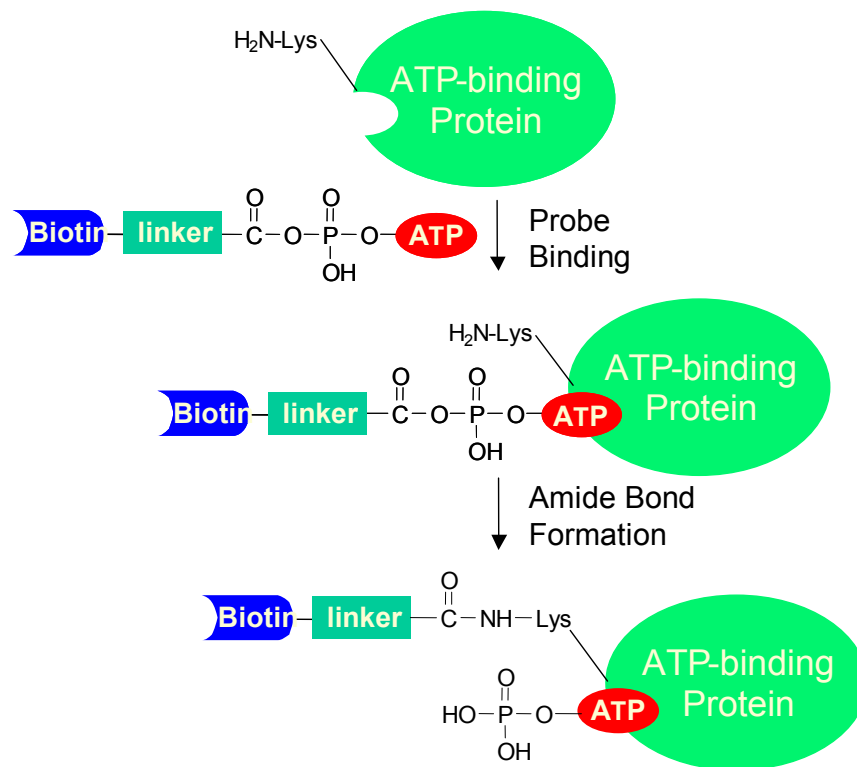
Figure 3.3 A schematic diagram showing a general experimental procedure with the use of ICAP in quantitative proteomic analysis.
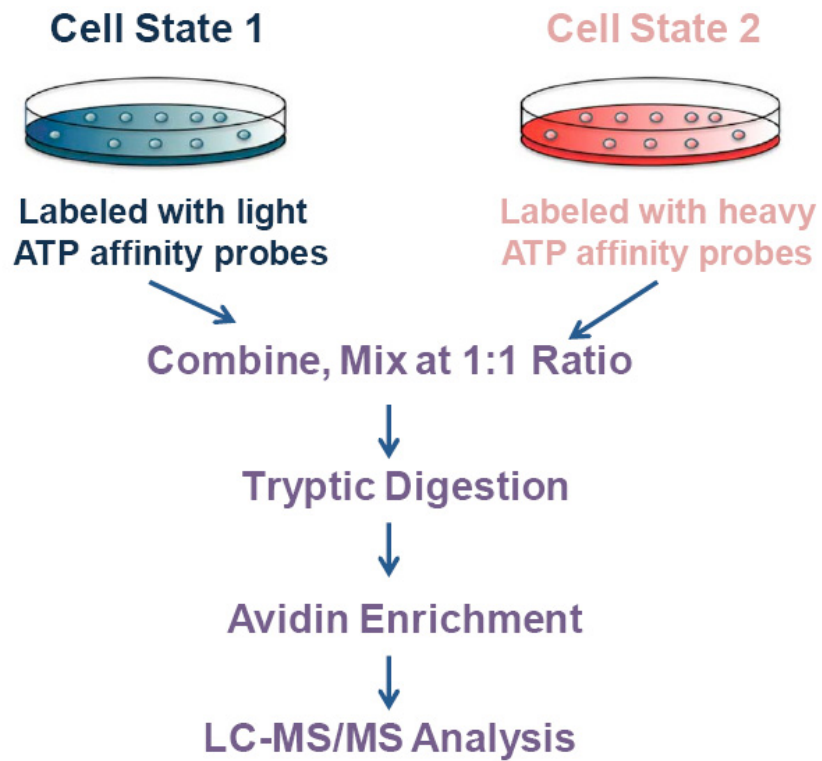
strategy to unambiguously characterize ATP-binding affinity of proteins at the entire proteome scale.

Intrigued by a previous study in reactivity profiling of functional cysteines by comparing the extent of cysteine alkylation under different probe concentrations,(20) we reasoned that the binding of the ATP component of the probe to a protein renders the acyl phosphate moiety conducive for coupling with the lysine residue at the ATP binding site of the protein, thereby greatly promoting the rate for amide bond formation.(21) Therefore, lysine residues involved in ATP binding and those that are not would exhibit distinct labeling behaviors with low and high concentrations of ATP affinity probe. At low probe concentration, the lysine residue at the ATP-binding site possesses hyper-reactivity and will be completely labeled. However, lysines not involved with ATP binding will only be partially labeled because limited amount of labeling reagent reacts preferentially with the lysine at the ATP-binding site. At high probe concentration, the ATP-binding lysine will still be labeled to completion, whereas non-ATP binding lysine will be labeled to a much greater extent than that at low probe concentration.

Based on the above hypothesis, we devised an affinity profiling strategy by allowing a low (10 μM) concentration of the light ATP probe and a high (100 μM) concentration of the heavy ATP probe to react separately with the same amount of cell lysates (forward experiment). To minimize the bias introduced by the labeling, reverse labeling experiment was also performed, where the same amount of cell lysates were treated individually with a high (100 μM) concentration of the light ATP probe and a low

(10 μM) concentration of the heavy ATP probe. As a result, peak intensity ratios of light and heavy desthiobiotin-labeled peptides can be used to derive ATP-binding affinity ratio, $R_{ATP10/1}$, which reflects the relative binding affinities of ATP towards specific lysine residues in individual proteins (Figure 3.4). With this assumption, specific ATP-binding lysine will generate an $R_{ATP10/1}$ close to 1 since similar amount of ATP-binding lysine will be labeled regardless of the probe concentration. By contrast, non-specifically labeled lysine will show concentration-dependent increase in ATP probe labeling, which results in a $R_{ATP10/1} \gg 1$.

3. Proteome-wide profiling of ATP-binding proteins

To test this proposed strategy, we performed quantitative ATP-affinity profiling of whole cell lysate from Hela-S3 cells using paired light and heavy ICAP reagents at 10 μM and 100 μM concentrations. We were able to quantify a total of 1575 proteins, including more than 4600 light or heavy desthiobiotin tag-modified lysine residues. As depicted in Figure 3.5A, a large number of peptides with desthiobiotin modification exhibited significantly different probe labeling efficiency under low and high probe concentrations, generating $R_{ATP10/1} \gg 1$. However, a small portion of the desthiobiotin-modified peptides bear $R_{ATP10/1}$ close to 1, indicating that they share similar labeling efficiency at low and high probe concentrations.

We next sought to assess if these peptides with low $R_{ATP10/1}$ ratios are indeed derived from ATP-binding proteins. Similar as the previous quantitative reactivity profiling of functional cysteines in proteins,(20) we arbitrarily consider lysine residue in

Figure 3.4 A general strategy for quantitative ATP-affinity profiling using ICAP.
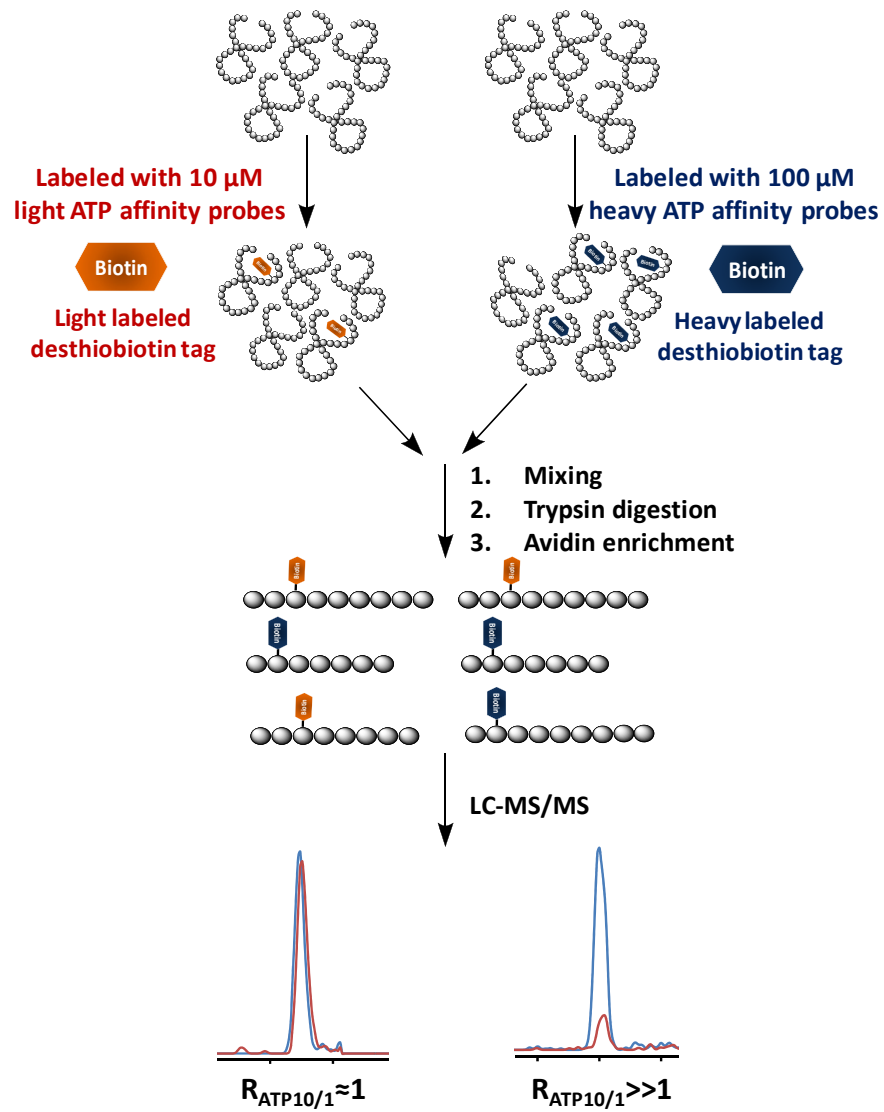
peptides with $R_{ATP10/1}$<2 as ATP-binding lysine. In addition, only those modified lysine

residues which are successfully quantified at least twice are retained on the list. With the

use of these criteria, we obtained 496 unique desthiobiotin-modified peptides,

representing approximately 11% of total identified peptides, from 333 unique proteins.

We considered these 333 proteins as possible ATP-binding proteins. GO analysis using

DAVID(22) indicated that ATP binding and kinase activity are significantly over-

represented in this protein group. As indicated in Figure 3.5B and Table 3.1, more than

50% of the 333 proteins are with known ATP-binding GO, suggesting a 5-fold increase

in enrichment relative to the entire human proteome with a p-value of 1.9E-77. In

contrast, the percentage of known ATP-binding proteins in the protein group containing

peptides with $R_{ATP10/1}$>2 is only 11%. More significantly, among the 496 peptides from

the possible ATP-binding protein group, peptides of known ATP-binding proteins were

on average identified and quantified three times more frequently than those of proteins

without ATP-binding GO, suggesting even more pronounced ATP-binding enrichment

efficiency (77% of all quantification events). Accordingly, the percentages of proteins

with kinase activity GO and protein kinase activity GO in the possible ATP-binding

protein group are 5- and 6-fold higher than those in the rest protein group. By contrast,

54 out of 65 identified known GTP-binding proteins were excluded from possible ATP-

binding protein groups, indicating that a large proportion of non-specific labeling arising

from electrostatic interaction can be discerned based on the large $R_{ATP10/1}$ ratio. For

example, K134 from GTP-binding nuclear protein Ran (Uniprot: P62826) was found to

Figure 3.5 (A) Measured $R_{ATP10/1}$ ratio from Hela-S3 cell lysates with low (10 μM) and

high (100 μM) concentrations of ICAP; (B) GO analysis of proteins with different

$R_{ATP10/1}$ ratios using both Hela-S3 (top) and Jurkat-T (bottom) cell lysates.

(A)
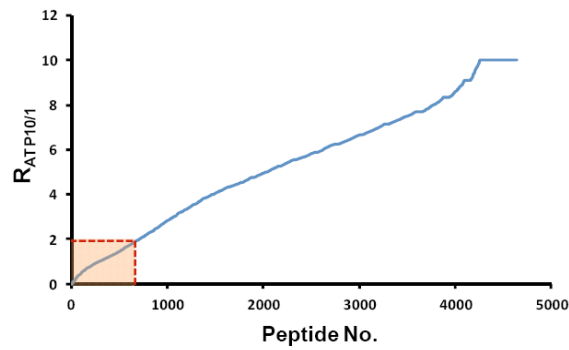
(B)
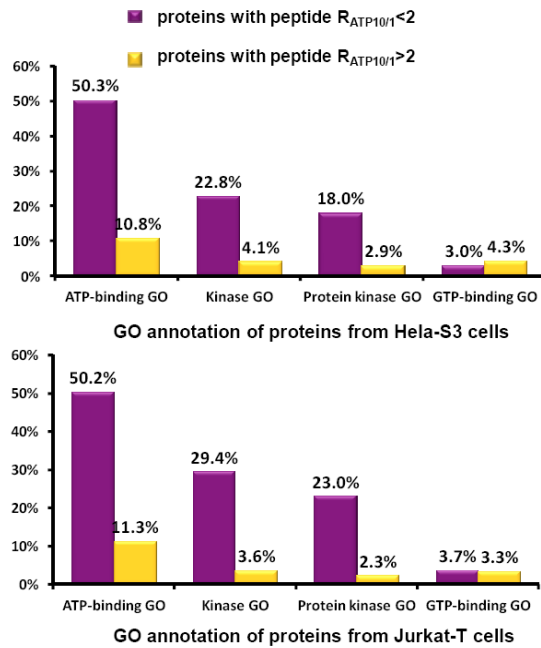
be modified with desthiobiotin with a large $R_{ATP10/1}$ ratio of 6.64, indicating that labeling of this lysine arises likely from non-specific binding. These findings demonstrated that the above-described quantitative affinity profiling could serve as an effective strategy to validate and discover ATP-binding proteins by effectively eliminating false-positive targets, which often compromise the application of this type of affinity reactive probes in protein-ligand binding studies.

Similar results were obtained when the same procedure was applied for quantitative profiling of ATP-protein interaction in the whole cell lysate of Jurkat-T cells. We were able to obtain a total of 504 unique desthiobiotin-labeled peptides from 323 unique proteins satisfying the aforementioned criteria . According to GO annotation (Figure 3.5B and Table 3.1), 50% and 29% of the 323 proteins are known ATP-binding proteins and kinases, respectively. This result reveals that our quantitative ATP-affinity profiling strategy is generally applicable for protein mixtures derived from any biological samples and not restricted to a particular cell line.

By combining the quantitative ATP-affinity profiling results from Hela-S3 and Jurkat-T cells (Figure 3.6), a total of 258 proteins with known ATP-binding affinity were validated with this method, among which 71 proteins were commonly found for both cell lines. In addition, 138 proteins in our predicted ATP-binding protein list are known to possess kinase activity. Furthermore, 166 proteins from Hela-S3 cells and 161 proteins from Jurkat-T cells are predicted to be ATP-binding proteins by our quantitative affinity profiling strategy, although they do not possess documented ATP-binding GO. These

122

Table 3.1 Details of GO analysis of proteins with different RATP10/1 ratio using both Hela-S3 (A) and Jurkat-T (B) cell lysates.

(A) GO annotation of proteins from Hela-S3 cell

		Total quantified protein (1575)	Possible ATP binding protein (333)	Other quantified protein (1242)
	Number	301	167	134
	percentage	19.1%	50.3%	10.8%
	p value	8.00E-42	1.9E-77	>0.1
ATP binding protein	enrichment factor	2.1	5.5	1.18
	Number	127	76	51
	percentage	8.1%	22.8%	4.1%
	p value	3.80E-09	7.7E-26	>0.1
kinases	enrichment factor	1.7	4.8	0.8
	Number	96	60	36
	percentage	6.1%	18.0%	2.9%
	p value	6.10E-07	9.6E-24	>0.1
Protein Kinases	enrichment factor	1.7	5.0	0.8
	Number	65	11	54
	percentage	4.1%	3.3%	4.3%
	p value	2.00E-06	9.1E-02	1.12E-06
GTP binding protein	enrichment factor	1.8	1.4	1.9

(B) GO annotation of proteins from Jurkat-T cell

		Total quantified protein (872)	Possible ATP binding protein (323)	Other quantified protein (549)
	Number	224	162	62
	percentage	25.7%	50.2%	11.3%
	p value	8.80E-45	9.6E-71	>0.1
ATP binding protein	enrichment factor	2.7	5.3	1.18
	Number	115	95	20
	percentage	13.2%	29.4%	3.6%
	p value	2.20E-20	1.9E-42	>0.1
kinases	enrichment factor	2.6	5.8	0.7
	Number	87	74	13
	percentage	9.9%	23.0%	2.3%
	p value	3.70E-17	1.8E-36	>0.1
Protein Kinases	enrichment factor	2.7	6.3	0.6
	Number	30	12	18
	percentage	3.4%	3.7%	3.3%
	p value	5.80E-02	>0.1	7E-02
GTP binding protein	enrichment factor	1.4	1.5	1.4

Figure 3.6 The Venn diagrams showing known ATP-binding proteins (A), kinases (B), protein kinases (C) and proteins without ATP-binding GO annotation identified in ATP affinity profiling experiment of Hela-S3 and Jurkat-T cell lysate

(A) ATP-binding proteins
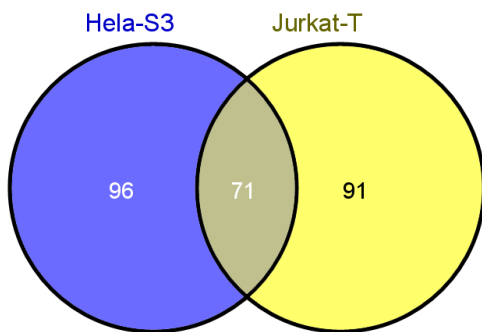
(B) Kinases
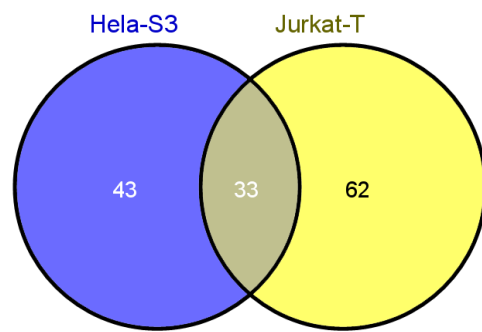
(C) Protein kinases
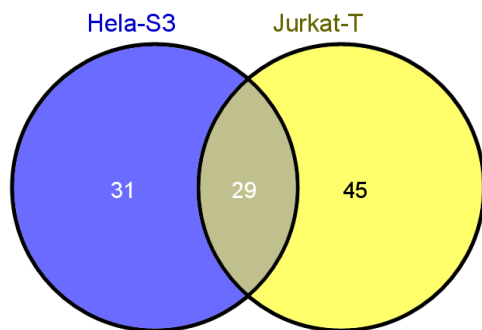
(D) Proteins without ATP-binding GO annotation
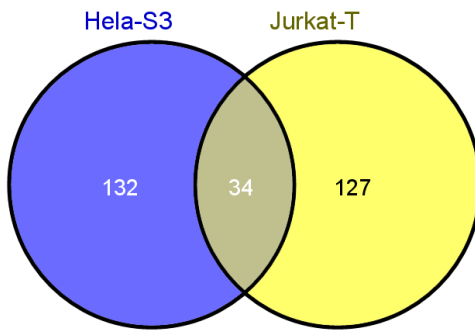
results provide a wealth of information to further validate ATP-protein interaction and discover novel ATP-binding sites.

4. Validation and discovery of ATP-binding sites in known ATP-binding proteins

As mentioned above, aside from the study of ATP affinity at the protein level, the ATP-binding site, or 'ATP-interacting residues', is also important in nucleotide-protein interaction studies because these specific residues generally play crucial roles in maintaining the ATP-binding affinity. Therefore, mutations of these identified ATP-interacting residues serve as an effective means to diminish the ATPase and/or kinase activities of the ATP-binding proteins.(23) However, identification of residues that interact with ATP often relies on X-ray crystal structure(24) or site-directed mutagenesis,(25) which is often costly and time-consuming.

Due to the specificity and precise quantification capability, our proposed ATP-affinity profiling strategy could serve as a quantitative and direct method to accurately assign ATP-binding lysine residues, which directly interact with γ-phosphate group of the bound ATP,(7) without the availability of 3-dimensional structure information of the protein. For example, we found one desthiobiotin-labeled lysine site from two peptides (AVLLGPPGAGK#GTQAPR and AVLLGPPGAGK#) in ATP-binding protein adenylate kinase 2 (AK2), where K28 displays a low $R_{ATP10/1}$ of 1.2 (Figure 3.7). The small $R_{ATP10/1}$ suggests that K28 is a possible ATP-binding lysine in AK2, which is consistent with the observation that K28 is responsible for ATP binding and phosphohydrolase activity (PDB entry: 3TLX). Moreover, although only a single

Figure 3.7 Light- and heavy-labeled peptides from forward- and reverse-affinity profiling experiments using ICAP. (A, B) Peptide AVLLGPPGAGK$_{28}$#GTQAPR with a low $R_{ATP10/1}$ ratio from adenylate kinase 2; (C, D) Peptide DLK$_{130}$#PQNLLIDDK with a low $R_{ATP10/1}$ ratio from cyclin-dependent kinase 1; (E, F) Peptide K$_{201}$#PLFHGDSEIDQLFR with a high $R_{ATP10/1}$ ratio from cyclin-dependent kinase 1. "#" indicates the desthiobiotin-labeling site.
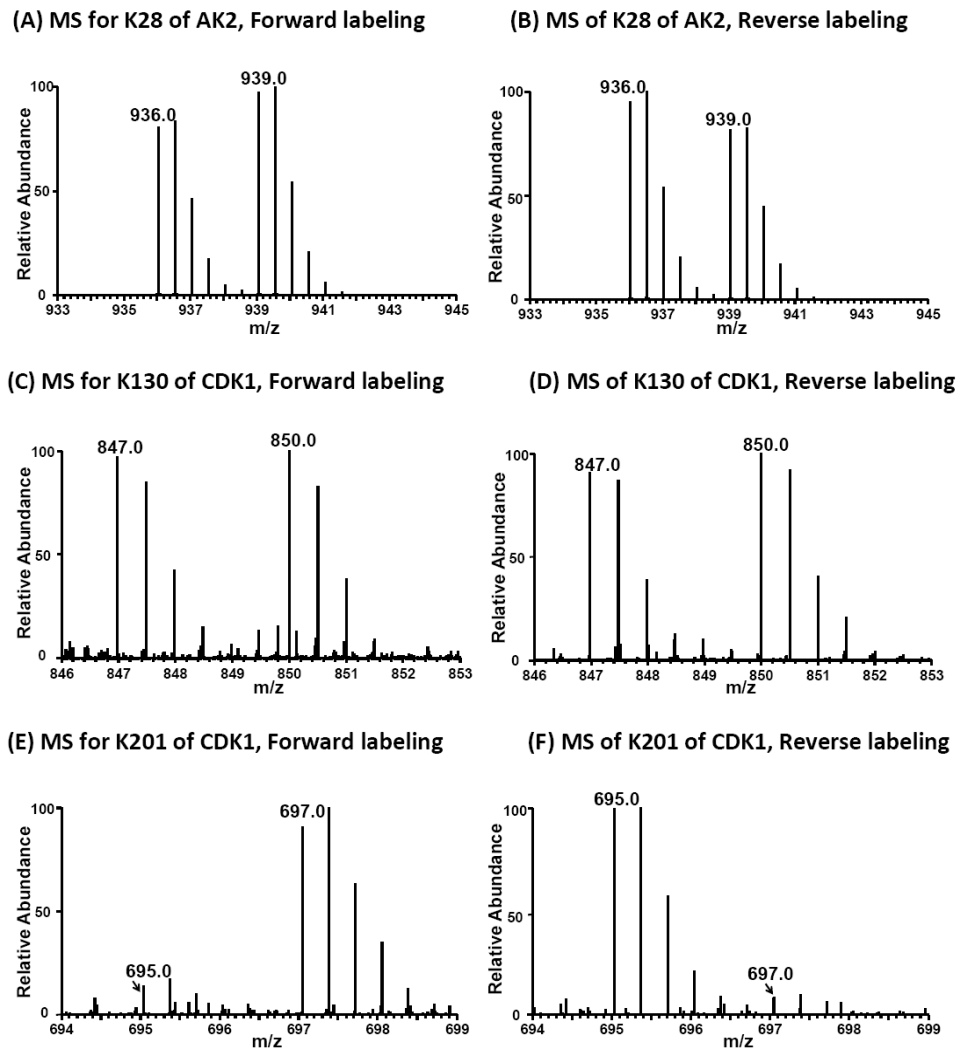
modified lysine was detected in the majority of identified possible ATP-binding proteins, multiple lysines were labeled for some proteins. However, these lysine residues often exhibited markedly different $R_{ATP10/1}$ ratios, which can be utilized to differentiate ATP-binding lysine from those that are non-specifically labeled. For instance, two lysine residues, K130 and K201, from cyclin-dependent kinase 1 were labeled in our quantitative profiling experiment; however, K201 shows a much larger $R_{ATP10/1}$ than K130, indicating that K130 is the true ATP-binding lysine and the labeling of K201 emanates from non-specific binding (Figure 3.7). Crystal structure of Cdk1 homolog, Cdk2 (PDB entry: 1HCK), reveals that K130 is in close proximity to the γ-phosphate group of ATP. Therefore, our strategy allows for site-specific determination of relative ATP-binding affinities of different binding sites in proteins by monitoring $R_{ATP10/1}$, thereby minimizing false-positive identification of ATP-binding sites.

Benefited from this high-throughput proteome-wide quantitative analysis, we sought to analyze local sequence context surrounding the desthiobiotin-modified lysine, which is considered as the ATP-binding site in proteins. First, we examined all the desthiobiotin-labeled peptides with $R_{ATP10/1}$ ratios being smaller than 2 from profiling experiments of Hela-S3 and Jurkat-T cells. Similar as our previous results,(7) three unique motifs were successfully identified using motif-X (Figure 3.8)(26). The well-known P-loop sequence motif of GxxxxGK(19) displays a 39-fold enrichment with respect to the occurrence frequency in our input sequence compared to the entire proteome. In addition, another two unique motifs of HRDxKxxN and xAVK, which

Figure 3.8 Unique binding motifs found from all peptides with $R_{ATP10/1}$ ratio < 2.
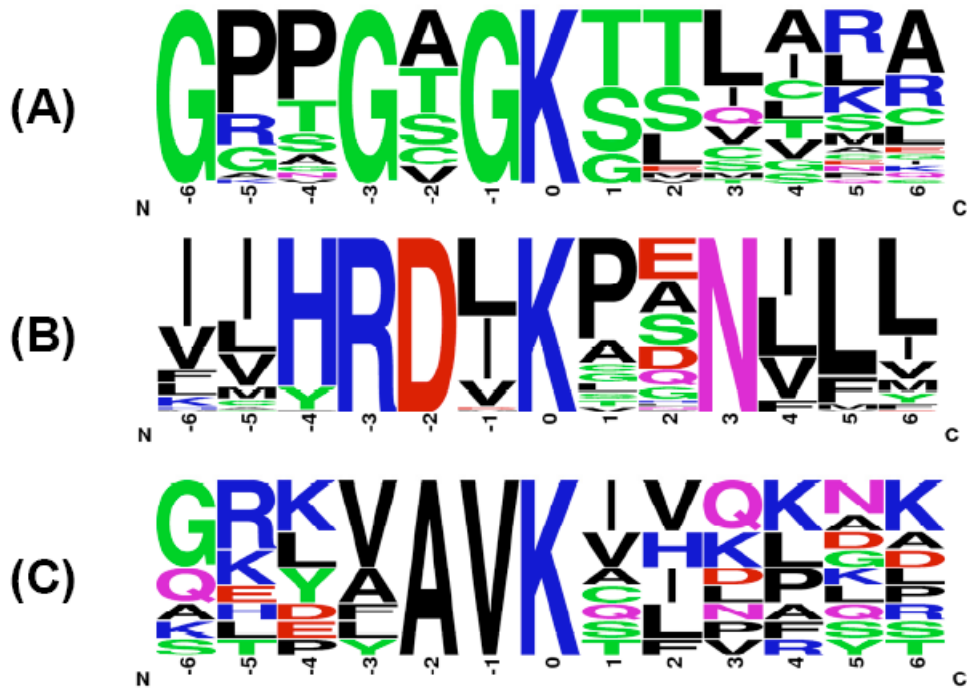
were identified with significant enrichment factors (185 and 1.8, respectively), are derived exclusively from kinases. These identified motifs provide invaluable information to study the functions of ATPase and kinases by generating catalytically inactive mutants, especially in large scale.(23)

More importantly, this novel quantitative ATP-affinity profiling strategy is particularly useful for unveiling the lysine residues that are involved in ATP-binding, but not present in any of the known ATP-binding motif. For instance, K176 in ribose-phosphate pyrophosphokinase 3 (PRPS1L1), which is not situated in any consensus motif known to be important in ATP-binding, was labeled by the desthiobiotin-tag with a low $R_{ATP10/1}$ of 1.3. This result strongly suggests that K176 is involved with ATP binding, which is supported by the crystal structure of the PRPS1L1-AMP complex (PDB entry: 2HCR).

We also noted that some ATP-binding proteins possess multiple lysine residues with low $R_{ATP10/1}$ ratios, raising the possibility that more than one lysine may contribute to ATP-protein interaction. For instance, for double-stranded RNA-activated protein kinase (PKR), apart from K416 residing in the HRDxKxxN motif ($R_{ATP10/1}$ ratio of 1.19), another peptide containing K296 (K.TYVIK#R.V) without any clear motif feature also displays ATP-affinity with an $R_{ATP10/1}$ ratio of 1.32 (Figure 3.9A). Indeed site-directed mutagenesis studies showed that substitution of K296 with an arginine gave rise to a catalytically dead mutant that completely abolishes the kinase activity of PKR.(27) In another example, we found two lysine residues (K112 and K356) from heat shock protein

Figure 3.9 Chromatograms represent multiple lysine sites with low $R_{ATP10/1}$ identified from PKR (A) and HSP90 (B) in quantitative ATP-affinity profiling assay. Blue and red profiles depict chromatograms of peptides from cell lysates labeled with high and low concentrations of the probe, respectively. Two sequence fragments containing intact C terminal domain generated from ATP binding-induced, iron-catalyzed chemical cleavage of HSP90 are also shown in (B).

(A)

PKR

$R_{ATP10/1}$=1.32 $R_{ATP10/1}$=1.19

N — C

K296   K415

(B)

HSP90

$R_{ATP10/1}$=0.92 $R_{ATP10/1}$=1.08

N — C

K112   K356

C C39

C C70
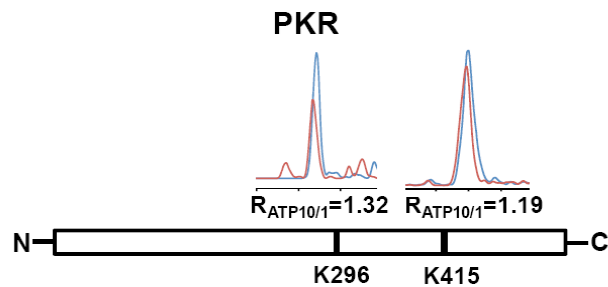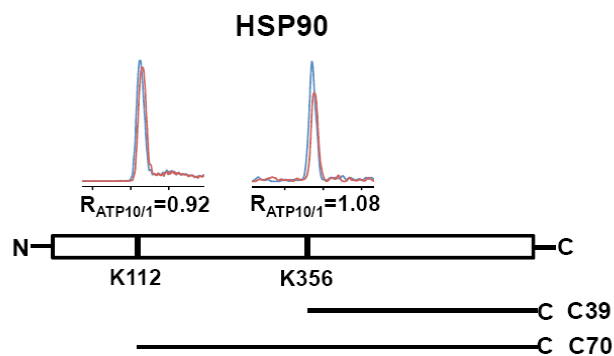
90 (HSP90), an ATP-binding chaperone, exhibit low $R_{ATP10/1}$ ratios of 0.92 and 1.08, respectively (Figure 3.9B). These two lysine residues are remote from each other in the crystal structure of HSP90 (Figure 3.10, PDB entry: 2CG9), suggesting that HSP90 may contain two distinct ATP-binding sites. Interestingly, a previous study demonstrated that, in addition to the N-terminal ATP-binding site, a second C-terminal ATP-binding site in HSP90 becomes accessible after the first N-terminal site is occupied with ATP.[28] Consistent with the low $R_{ATP10/1}$ ratio observed for K112, crystal structure (Figure 3.10) revealed the direct contact between ATP and K112. However, due to the lack of conclusive experimental evidence, the exact position of C-terminal ATP-binding site on HSP90 remains elusive. On the basis of our quantitative proteomic results, we deduce that K356 may represent the other ATP-binding site located in C-terminal domain of HSP90. This notion is further supported by a previous study, where ATP binding induces the generation of two C-terminal fragments of 70 kDa (C70) and 39 kDa (C39) of HSP90 upon iron-catalyzed chemical cleavage.[29] Mapping these two fragments to the sequence of HSP90 revealed that C70 fragment matches the cleavage site around N-terminal ATP-binding site of K112, whereas the C39 fragment suggests that C-terminal ATP-binding site may localize in the region harboring amino acid residues 350-450. This is consistent with our finding that K356 may constitute the second ATP-binding lysine (Figure 3.9). Therefore, our results provided the first line of direct evidence to support that the second ATP-binding site resides on the C-terminal domain of HSP90 and K356 is involved with this ATP binding.

Figure 3.10 Crystal structure of full-length HSP90 bound with ATP displays the structural relationship between K112, K356 and ATP-binding sites on N-terminal of HSP90.
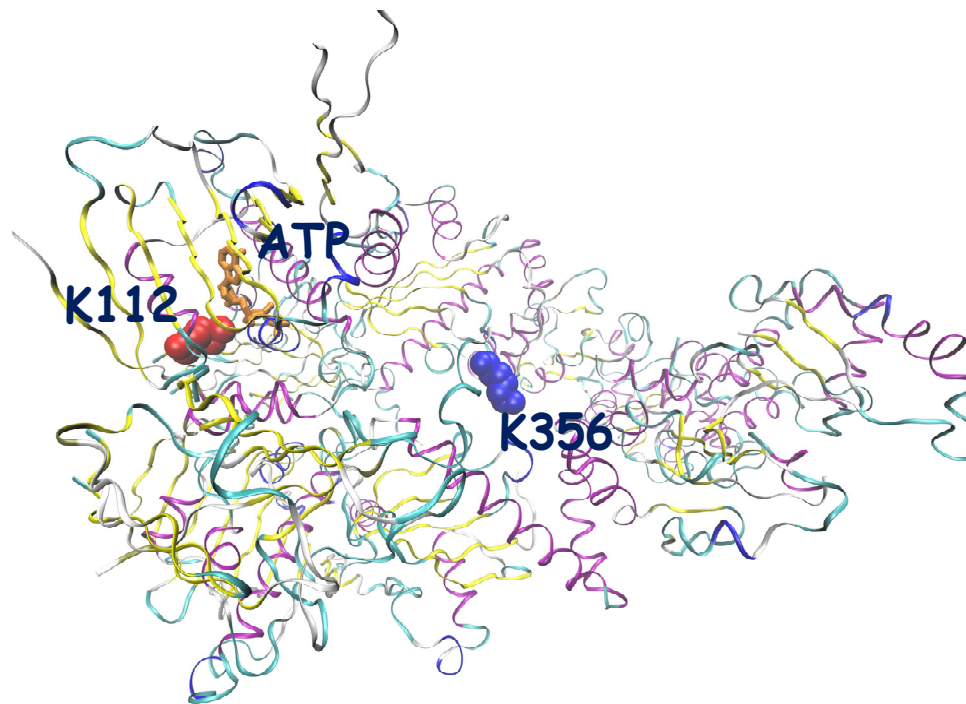
These examples clearly illustrated the capability of our approach in characterizing the nucleotide-binding affinity of both known binding motifs and, more importantly, previously unrecognized nucleotide-binding site of ATP-binding proteins. This facilitates the identification of the lysine residues that are the most relevant for ATP-related function of proteins, which may provide invaluable information to generate activity-deficient mutants for functional studies of ATPase and kinases.

5. Exploration of novel ATP-binding proteins and ATP-binding sites

As mentioned above, more than 160 proteins without documented ATP-binding GO from Hela-S3 cells and Jurkat-T cells, respectively, are predicted to be ATP-binding proteins based on our quantitative affinity profiling results. To explore novel ATP-binding proteins and ATP-binding sites, we first examined the 22 proteins that were identified with the same peptides from the two cell lines.

GO analysis reveals that most of the 22 proteins have reported nucleotide-binding affinity. Four of them belong to NAD-binding proteins, such as NAD-dependent malic enzyme and glyceraldehyde-3-phosphate dehydrogenase, which is consistent with our previous findings that this ATP-affinity probe can also capture NAD-binding proteins.(11) Only 8 proteins do not possess any nucleotide-related GO annotations. However, literature search reveals that some of them are also closely associated with ATP-binding. For example, K212 in phosphopantothenate-cysteine ligase (PPCS) was identified with low $R_{ATP10/1}$ ratios in affinity profiling experiments with both Hela-S3 and Jurkat-T lysates (1.24 and 0.87, respectively). This is in accordance with the previous

finding that PPCS catalyzes the conversion of pantothenate to coenzyme A by utilizing ATP as a substrate.(30) In another example, K375 in protein disulfide-isomerase (PDI) exhibits low $R_{ATP10/1}$ ratios in ATP profiling experiments for both cell lines, strongly suggesting that this protein possesses ATP-binding property. PDI is a multifunction protein and it is involved in protein folding.(31) Although the sequence of PDI does not overlap bear any motifs or homologies of kinases, ATPases or chaperones, Guthapfel et al.(32) demonstrated that PDI was directly associated with ATP-dependent autophosphorylation with clear ATP binding and hydrolysis properties. This notion is supported by our results. Moreover, our study also provides the information about the ATP-binding site in PDI.

We also made an interesting observation with proliferating cell nuclear antigen (PCNA). In eukaryotes, including humans, PCNA forms a homotrimer, which complexes with the five-subunit replication factor C (RFC) as clamp loader to encircle duplex DNA, forming mobile tethers for polymerases and thereby enabling processive DNA synthesis.(33) Previous studies suggest that ATP binding and hydrolysis in ATPase cycle is closely associated with every step of the DNA-clamp loading reaction. RFC, with conserved ATP binding/hydrolysis motifs, is therefore believed to serve as the primary ATP-binding sites to enable utilization of ATP for driving the complex clamp loading reaction. Intriguingly, our ATP affinity profiling with both Hela-S3 and Jurkat-T cell lysates strongly suggest that PCNA itself may also be involved in ATP-binding. We note that K80 in PCNA was quantified multiple times with an average $R_{ATP10/1}$ ratio of

1.16 (Figure 3.11). Careful examination of crystal structure of RFC-ATP-PCNA complex reveals five ATP-binding sites on RFC (Figure 3.11C, PDB entry: 1SXJ). In our ATP-affinity profiling assay, we successfully quantified lysine residues located in GxxxxGK motifs from RFC2 ($R_{ATP10/1}$ = 1.4) and RFC4 ($R_{ATP10/1}$ = 1.1) with low $R_{ATP10/1}$ ratios, which are consistent with the ATP-binding sites of RFC as revealed by the crystal structure. Furthermore, as indicated in the crystal structure, these five binding sites on RFC are far from K80 in PCNA, indicating that desthiobiotin labeling on PCNA is unlikely derived from crossover labeling from the bound ATP probe on RFC protein.

To examine PCNA's ATP-binding ability, we performed a binding assay using a widely used fluorescent ATP analog, 2',3'-O-(2,4,6-trinitrophenyl)-ATP (TNP-ATP) (Supporting Information).(34, 35) TNP-ATP is only moderately fluorescent in aqueous solution, while a substantial fluorescence enhancement and blue shift are observed upon binding to ATP-binding proteins. As revealed in Figure 3.11D, 25 μM of TNP-ATP displays minimal fluorescence in solution with a λ_{max} of 540 nm. However, when 6 μM PCNA was added to TNP-ATP, the relative fluorescence increases by approximately 4.5-fold, indicating the direct binding of PCNA with the ATP analog. These effects were diminished when 2 mM free ATP was added. To estimate the binding affinity of TNP-ATP to PCNA, we titrated a constant concentration of PCNA solution with increasing amounts of TNP-ATP. Fluorescence intensity versus TNP-ATP concentration was plotted and fit into single binding model of a hyperbolic function(17, 18) to obtain a dissociation constant (K_d) of 13 μM for interaction between PCNA and TNP-ATP. For

135

Figure 3.11 (A, B) Light- and heavy-labeled peptide ILK_{80}#CAGNEDIITLR with low $R_{ATP10/1}$ ratio from PCNA in quantitative ATP-affinity profiling assay; "#" indicates the desthiobiotin labeling site. (C) Crystal structure of RFC-PCNA complex bound with ATP. Displayed is the relationship between K80 in PCNA and ATP-binding sites on RFC complex. (D) Fluorescence emission spectra of 50 mM PBS (red), PCNA alone (orange), TNP-ATP alone (yellow), or TNP-ATP together with PCNA (green) and a mixture of TNP-ATP, PCNA and ATP (blue).
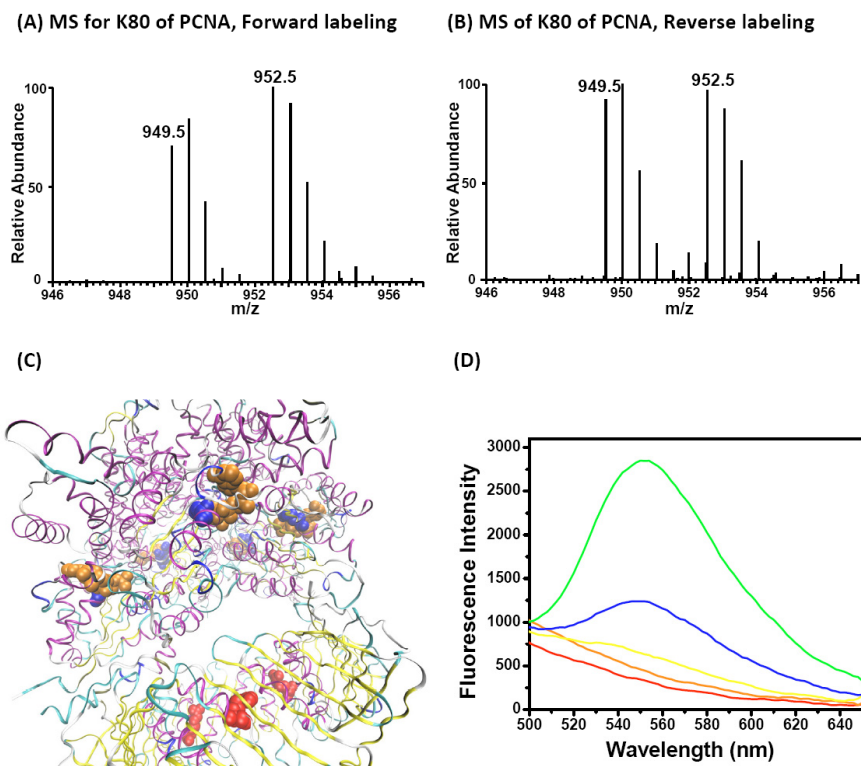
comparison, similar experiments were performed with known ATP-binding proteins, YADH and creatine phosphokinase, and similar fluorescence enhancement response and K_d values of 13 and 27 μM were determined for the two proteins (Figure 3.12). In this context, Fukuda et al.(36) observed that the mutation of K80 in human PCNA to an alanine results in a greater than 2-fold reduction of its capability in stimulating polymerase δ-mediated DNA synthesis. Thus, our results suggest that the direct binding of ATP to PCNA may play a significant role in PCNA's ability to promote processive DNA replication. Moreover, our results also unveiled the lysine residue in PCNA that is involved in ATP-binding.

Conclusions

Here, we introduced ICAP to simultaneously enrich and isotopically label ATP-binding proteins in the whole proteome. It has several advantages for studying nucleotide-binding proteins. First, unlike other chemical labeling methods including dimethyl labeling, the label is introduced onto the protein in the early stage of experiment, which minimizes variations in subsequent steps of sample preparation such as protein reduction, alkylation and proteolytic digestion. Furthermore, the ICAP-based strategy should be applicable for any biological samples, including clinical samples that are not amenable to metabolic labeling. Therefore, ICAP exhibits great potential in quantitative studies of nucleotide-binding proteins in the whole proteome scale.

This multifunctional affinity probe may be applied for two aspects of nucleotide-binding protein studies. In this vein, we have devised a quantitative affinity profiling

Figure 3.12 (A) Fluorescence emission spectra of 50 mM PBS (black), TNP-ATP along (red), in the presence of YADH (green) and YADH plus free ATP (blue). (B) Fluorescence emission spectra of 50 mM PBS (black), TNP-ATP along (red), in the presence of creatine phosphokinase (green) and creatine phosphokinase plus free ATP (blue). (C) Dependence of fluorescence intensity on TNP-ATP concentration. Protein (3 µM) was incubated with increasing concentration of TNP-ATP in 50 mM PBS solution, pH 7.4. Red, blue, orange and pink symbols represent YADH, PCNA, creatine phosphokinase and TNP-ATP alone, respectively. Apparent K_d values for TNP-ATP with proteins were obtained from the best fit of the data to a hyperbolic function.

(A)
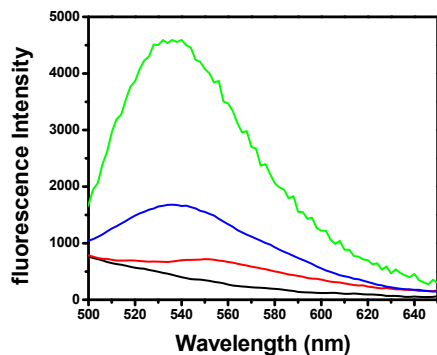
(B)
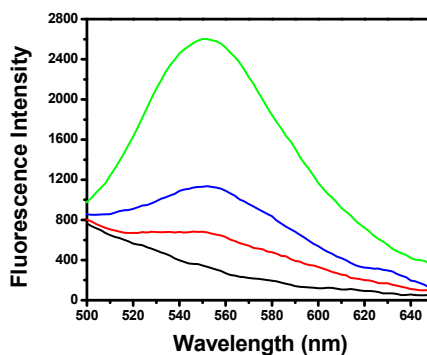
(C)
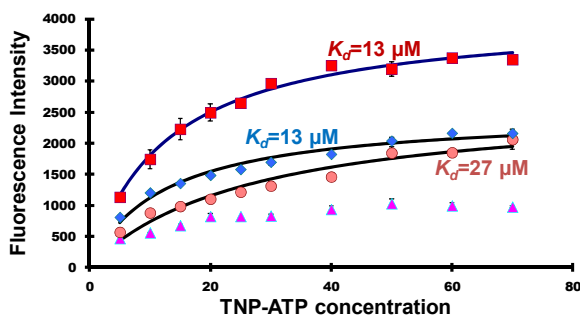
strategy, encompassing the use of ICAP, to comprehensively characterize ATP-protein interactions at the entire proteome scale. A key advantage in charactering ATP-protein interaction with our affinity probe over previous binding assay lies in that our strategy allows for unambiguous determination of the interaction between ATP ligand and lysine residues in proteins of interest. False-positive identification of ATP-binding sites and the corresponding proteins due to non-specific labeling can be effectively minimized due to the specificity and quantitative nature of our strategy. Thus, ATP-protein interaction studies can be, for the first time, extended to quantitative surveys of specific binding regions of proteins at the entire proteome scale. It can be envisaged that similar nucleotide-affinity probes can be generally applied for the capture and characterization of proteins that can bind to other nucleotides. Moreover, the isotope-coded ATP probe with both enrichment and labeling capability facilitates future quantitative comparison of expression levels of ATPase and kinases originating from any biological samples.

References

1.   Chene P. ATPases as drug targets: learning from their structure. Nat Rev Drug Discov. 2002;1:665-73.

2.   Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. Science. 2002;298:1912-34.

3.   Daub H, Olsen JV, Bairlein M, Gnad F, Oppermann FS, Stemmann O, et al. Kinase-selective enrichment enables quantitative phosphoproteomics of the kinome across the cell cycle. Mol Cell. 2008;31:438-48.

4.   Zhang YN, Yu DJ, Li SS, Fan YX, Huang Y, Shen HB. Predicting protein-ATP binding sites from primary sequence through fusing bi-profile sampling of multi-view features. BMC Bioinformatics. 2012;13:118.

5.   Ormö M, Sjöberg B-M. An ultrafiltration assay for nucleotide binding to ribonucleotide reductase. Anal Biochem. 1990;189:138-41.

6.   Guarnieri MT, Blagg BS, Zhao R. A high-throughput TNP-ATP displacement assay for screening inhibitors of ATP-binding in bacterial histidine kinases. Assay Drug Dev Technol. 2011;9:174-83.

7.   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;85:3198-206.

8.   Patricelli MP, Szardenings AK, Liyanage M, Nomanbhoy TK, Wu M, Weissig H, et al. Functional interrogation of the kinome using nucleotide acyl phosphates. Biochemistry (Mosc). 2006;46:350-8.

9.   Ong S-E, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics. 2002;1:376-86.

10.   Boersema PJ, Raijmakers R, Lemeer S, Mohammed S, Heck AJR. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. Nat Protocols. 2009;4:484-94.

11.   Qiu H, Wang Y. Probing adenosine nucleotide-binding proteins with an affinity-labeled nucleotide probe and mass spectrometry. Anal Chem. 2007;79:5547-56.

12.   Olsen JV, de Godoy LMF, Li G, Macek B, Mortensen P, Pesch R, et al. Parts per Million Mass Accuracy on an Orbitrap Mass Spectrometer via Lock Mass Injection into a C-trap. Mol Cell Proteomics. 2005;4:2010-21.

13.   Tabb DL, McDonald WH, Yates JR. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. J Proteome Res. 2002;1:21-6.

14.   Park SK, Venable JD, Xu T, Yates JR. A quantitative analysis software tool for mass spectrometry-based proteomics. Nat Methods. 2008;5:319-22.

15.   Chen J, Ai Y, Wang J, Haracska L, Zhuang Z. Chemically ubiquitylated PCNA as a probe for eukaryotic translesion DNA synthesis. Nat Chem Biol. 2010;6:270-2.
16.   de Wet H, McIntosh DB, Conseil G, Baubichon-Cortay H, Krell T, Jault J-M, et al. Sequence Requirements of the ATP-Binding Site within the C-Terminal Nucleotide-Binding Domain of Mouse P-Glycoprotein: Structure−Activity Relationships for Flavonoid Binding†. Biochemistry (Mosc). 2001;40:10382-91.

17.   Ni Q, Shaffer J, Adams JA. Insights into nucleotide binding in protein kinase A using fluorescent adenosine derivatives. Protein Sci. 2000;9:1818-27.

18.   Marek M, Milles S, Schreiber G, Daleke DL, Dittmar G, Herrmann A, et al. The Yeast Plasma Membrane ATP Binding Cassette (ABC) Transporter Aus1: PURIFICATION, CHARACTERIZATION, AND THE EFFECT OF LIPIDS ON ITS ACTIVITY. J Biol Chem. 2011;286:21835-43.

19.   Saraste M, Sibbald PR, Wittinghofer A. The P-loop--a common motif in ATP- and GTP-binding proteins. Trends Biochem Sci. 1990;15:430-4.

20.   Weerapana E, Wang C, Simon GM, Richter F, Khare S, Dillon MBD, et al. Quantitative reactivity profiling predicts functional cysteines in proteomes. Nature. 2010;468:790-5.

21.   Li X, Liu DR. DNA-Templated Organic Synthesis: Nature's Strategy for Controlling Chemical Reactivity Applied to Synthetic Molecules. Angewandte Chemie International Edition. 2004;43:4848-70.

22.   Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protocols. 2008;4:44-57.

23.   Varjosalo M, Björklund M, Cheng F, Syvänen H, Kivioja T, Kilpinen S, et al. Application of Active and Kinase-Deficient Kinome Collection for Identification of Kinases Regulating Hedgehog Signaling. Cell. 2008;133:537-48.

24.   Ali MMU, Roe SM, Vaughan CK, Meyer P, Panaretou B, Piper PW, et al. Crystal structure of an Hsp90–nucleotide–p23/Sba1 closed chaperone complex. Nature. 2006;440:1013-7.

25.   Obermann WM, Sondermann H, Russo AA, Pavletich NP, Hartl FU. In vivo function of Hsp90 is dependent on ATP binding and ATP hydrolysis. J Cell Biol. 1998;143:901-10.

26.   Schwartz D, Gygi SP. An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. Nat Biotech. 2005;23:1391-8.

27.   Dey M, Cao C, Dar AC, Tamura T, Ozato K, Sicheri F, et al. Mechanistic link between PKR dimerization, autophosphorylation, and eIF2± substrate recognition. Cell. 2005;122:901-13.

28.   Soti C, Vermes A, Haystead TAJ, Csermely P. Comparative analysis of the ATP-binding sites of Hsp90 by nucleotide affinity cleavage: a distinct nucleotide specificity of the C-terminal ATP-binding site. Eur J Biochem. 2003;270:2421-8.

29.   Soti C, Racz A, Csermely P. A Nucleotide-dependent molecular switch controls ATP binding at the C-terminal domain of Hsp90. N-terminal nucleotide binding unmasks a C-terminal binding pocket. J Biol Chem. 2002;277:7066-75.

30.   Manoj N, Strauss E, Begley TP, Ealick SE. Structure of human phosphopantothenoylcysteine synthetase at 2.3 A resolution. Structure. 2003;11:927-36.
31.   Noiva R, Lennarz WJ. Protein disulfide isomerase. A multifunctional protein resident in the lumen of the endoplasmic reticulum. J Biol Chem. 1992;267:3553-6.

32.   Guthapfel R, Guéguen P, Quéméneur E. ATP Binding and Hydrolysis by the Multifunctional Protein Disulfide Isomerase. J Biol Chem. 1996;271:2663-6.

33.   Chen S, Levin MK, Sakato M, Zhou Y, Hingorani MM. Mechanism of ATP-driven PCNA clamp loading by S. cerevisiae RFC. J Mol Biol. 2009;388:431-42.

34.   Kubala M, Plasek J, Amler E. Fluorescence competition assay for the assessment of ATP binding to an isolated domain of Na+, K(+)-ATPase. Physiol Res. 2004;53:109-13.

35.   Stewart RC, VanBruggen R, Ellefson DD, Wolfe AJ. TNP-ATP and TNP-ADP as probes of the nucleotide binding site of CheA, the histidine protein kinase in the chemotaxis signal transduction pathway of Escherichia coli. Biochemistry (Mosc). 1998;37:12269-79.

36.   Fukuda K, Morioka H, Imajou S, Ikeda S, Ohtsuka E, Tsurimoto T. Structure-function relationship of the eukaryotic DNA replication factor, proliferating cell nuclear antigen. J Biol Chem. 1995;270:22527-34.

Chapter 4

Comprehensive Characterization of ^{S}GTP-binding Proteins by Orthogonal Quantitative ^{S}GTP-affinity Profiling and ^{S}GTP/GTP Competition Assays

Introduction

      Thiopurine drugs, including 6-mercaptopurine, 6-thioguanine, and azathioprine, are widely used as cancer therapeutic and immunosuppressive agents(1). Although the exact mechanisms underlying the cytotoxic effects of these thiopurines remain elusive, it is generally accepted that thiopurines are pro-drugs and require metabolic activation to exert their toxicity. After cellular uptake, the thiopurine drugs can be metabolically activated to yield 6-thioguanosine triphosphate (^{S}GTP) and 6-thio-2'-deoxyguanosine triphosphate, which can be incorporated into RNA or DNA(2). In this vein, it was proposed that DNA 6-thioguanine may be spontaneously methylated by *S*-adenosyl-L-methionine to give S^6-methylthioguanine (S^6mG), which directs the misincorporation of dTMP during DNA replication(3). The resulting S^6mG:T mispair can trigger the postreplicative mismatch repair (MMR) pathway, thereby inducing cell death(4).

The triggering of MMR pathway may not be the sole mechanism contributing to the cytotoxic effects of the thiopurine drugs considering the fact that the MMR-deficient leukemia cells were also sensitive toward thiopurines(5). In this context, 6-thioguanine was found to reactivate epigenetically silenced genes in leukemia cells by inducing DNMT1 degradation(6), and 6-thioguanine could also induce mitochondrial dysfunction and reactive oxygen species generation in cultured human cells(7, 8). In addition, ^{S}GTP, a thiopurine metabolite, was shown to block the activation of Rac1, a GTP-binding protein, in human T lymphocytes, which leads to the inactivation of its target genes such as MEK, NF-κB, and Bcl-xL and the induction of mitochondrial pathway of apoptosis(9, 10). Considering the involvement of a variety of nucleotide-binding proteins such as protein kinases(11) and small GTPases(12) in cell signaling, investigation of the interaction of ^{S}GTP with cellular nucleotide-binding proteins may unveil novel mechanisms of action of the thiopurine drugs.

Currently there is no proteome-wide characterization of cellular proteins that can recognize ^{S}GTP. Traditional studies on nucleotide-protein interaction often rely on radioactivity-based ultrafiltration assay(13) or fluorescence-based binding assay(14). These methods are usually costly and time-consuming because they require the use of purified proteins. Moreover, none of these approaches permit the robust discovery of nucleotide-binding site, or "nucleotide-interacting residues", in proteins. Recently, we developed a quantitative affinity profiling strategy, encompassing the use of low and high concentrations of desthiobiotin-conjugated acyl ATP probes, to comprehensively

145

characterize ATP-protein interactions at the entire proteome scale(15). The method allows for the minimization of false-positive identification of ATP-binding targets arising from non-specific labeling and facilitates the identification of previously unrecognized ATP-binding sites in ATP-binding proteins.

Here, we devised a quantitative profiling strategy with the use of [S]GTP-affinity probe to unambiguously discover novel [S]GTP-binding proteins along with the specific binding sites from the entire human proteome. Additionally, we characterized the binding selectivities of these proteins toward [S]GTP and GTP. Many known GTP-binding proteins, including multiple heterotrimeric G proteins, exhibit strong binding preference toward [S]GTP. We also observed that [S]GTP displays robust binding toward multiple cyclin-dependent kinases (CDKs), which may perturb the CDK-mediated phosphorylation and cell cycle progression.

Experimental Procedures

Synthesis of 6-thioguanosine triphosphate

All reagents were obtained from Sigma-Aldrich unless otherwise indicated. 6-Thioguanosine was phosphorylated to give the corresponding 5'-monophosphate following published procedures(16). A reaction mixture containing 6-thioguanosine-5'-monophosphate (225 mg, 0.6 mmol), triphenylphosphine (0.396 g, 1.5 mmol), diphenyl thioether (0.33 g, 1.5 mmol) and L-methylimidazole (0.5 mL, 6.2 mmol) in DMF/DMSO (1:2, 10 mL) was incubated at room temperature for 15 min. To the resulting solution

was subsequently added 0.5 M bis(tri-*n*-butylammonium) pyrophosphate in DMF (1.5 mmol, 3.0 mL)(17). The solution mixture was incubated at room temperature for 50 min, and the product was precipitated by the addition of acetone (100 mL). The precipitate was washed twice with acetone (20 mL each). The product was subsequently purified by using an anion exchange column packed with DEAE Sephadex G-25, and the ^{S}GTP was eluted with 200 mM NH_4HCO_3. Fractions were pooled, lyophilized, dissolved in water, and lyophilized again to give ^{S}GTP as a white solid (54 mg, yield 17%). ^{1}H NMR (D_2O, 300 MHz): δ 8.13 (s, 1H), 5.82 (s, 1H), 4.65 (s, 1H), 4.51–4.46 (m, 1H), 4.23 (brs, 1H), 4.12–4.08 (m, 2H). ^{31}P NMR (D_2O, 80 MHz): δ −8.5, −10.2, −21.7.

Preparation of the Biotinylated Nucleotide Affinity Probe

The biotinylated nucleotide affinity probes were prepared according to previously published procedures with minor modifications(18). In this regard, we first prepared desthiobiotinyl-aminobutyric acid (desthiobiotin-C4) following previously published procedures. Isobutyl chloroformate (0.19 mL) was added to a solution containing 150 mg desthiobiotin, 15 mL DMF, and 0.38 mL tri-*n*-butylamine. After being incubated at room temperature for 10 min, the mixture was slowly added to a suspension of 230 mg γ-aminobutyric acid (Sigma) in DMF (7.5 mL) at 5°C. After stirring at 5°C for 2 h, the solvent was removed under reduced pressure and the crude product was dissolved in 5 mL water at 40°C. The solution pH was adjusted to 2 with 2.0 N HCl, and the mixture was kept at 0°C for 12 h. The desired desthiobiotin-C4 was precipitated out of solution. The precipitate was filtered, washed with water, and dried under vacuum.

To render nucleotide soluble in organic solvent, the commercially available sodium salt form of GTP or synthesized [S]GTP was first converted to the tributylammonium form by passing the nucleotides through a cation-exchange column packed with Spectra/Gel IE 50×8 resin (40-75 μm) at 4°C once. Fractions containing the tributylammonium form of GTP or [S]GTP were collected and lyophilized. Desthiobiotin-C4 (8 mg), dissolved in a 1-mL solvent mixture of ice-cold dry CH_2Cl_2 and DMF (4:1, v/v), was mixed with tri-n-butylamine (11 μL) and ethyl chloroformate (5 μL). After stirring at 0°C for 5 min, the mixture was stirred at room temperature under argon atmosphere for another 60 min. Tributylammonium form of GTP or [S]GTP (50 mg), dissolved in a 1.25 mL solution of CH_2Cl_2 and DMF (4:1, v/v), was then added to the above reaction mixture. The reaction was continued at room temperature and under argon atmosphere for 18 h. The CH_2Cl_2 was then removed by purging with argon for 10 min and the remaining 200 μL solution was directly subjected to HPLC purification with a YMC ODS-AQ column (4.8×250 mm, 120 Å in pore size, 5 μm in particle size, Waters). The flow rate was 0.8 mL/min, and a 45 min linear gradient of 0-30% acetonitrile in 50 mM triethylammonium acetate (pH 6.8) was used for the purification. A UV detector was set at 253 and 340 nm to monitor the effluents for GTP and [S]GTP probes, respectively. Appropriate HPLC fractions were pooled, lyophilized, and stored at -80°C. The structures of the products were confirmed by ESI-MS (Figure 4.2)(18).

Cell lysate preparation and labeling with the nucleotide affinity probe

The desthiobiotinylated ^{S}GTP and GTP affinity probes (Figure 4.1) were prepared following previously published procedures(15, 19) with minor modifications (see Supporting Information). For stable isotope labeling by amino acids in cell culture (SILAC) experiments, lysine, arginine-depleted RPMI-1640 medium (Pierce) was supplemented with light or heavy ($[^{13}C_6, ^{15}N_2]$-L-lysine and $[^{13}C_6]$-L-arginine) lysine and arginine, along with dialyzed FBS (Invitrogen) and penicillin (100 IU/mL) to give the complete SILAC medium. The Jurkat-T acute lymphoblastic leukemia cells (ATCC; Manassas, VA) were cultured in heavy RPMI-1640 SILAC medium for at least 5 cell doublings to achieve complete isotope incorporation. Cells were maintained in a humidified atmosphere with 5% CO_2 at 37°C. Approximately 2×10^7 cells were harvested, washed with cold PBS for three times, and lysed in a 1-mL lysis buffer, which contained 0.7% CHAPS, 50 mM HEPES (pH 7.4), 0.5 mM EDTA, 100 mM NaCl, and 10 μL (1:100) protease inhibitor cocktail on ice for 30 min. The cell lysates were centrifuged at 16000g at 4°C for 30 min, and the resulting supernatants were collected and subjected to gel filtration separation using NAP-25 columns (Amersham Biosciences) to remove free endogenous nucleotides. Cell lysates were eluted into a 2 mL buffer, containing 50 mM HEPES (pH 7.4), 75 mM NaCl, and 5% glycerol. The resulting proteins in cell lysates were quantified using Quick Start Bradford Protein Assay (Bio-Rad, Hercules, CA) and stored at -80°C. Immediately prior to the labeling reaction, $MgCl_2$, $MnCl_2$, and $CaCl_2$ were added to the concentrated cell lysate until their final concentrations reached 50, 5,

and 5 mM, respectively. Approximately 1 mg cell lysate was treated with 10 or 100 μM desthiobiotin-^{S}GTP or -GTP affinity probe. Labeling reactions were carried out at room temperature with gentle shaking for 1.5 h. After the reaction, the remaining probes in the cell lysates were removed by buffer exchange with 25 mM NH$_4$HCO$_3$ (pH 8.5) using Amicon Ultra-4 filter (10,000 NMWL, Millipore).

In-solution enzymatic digestion and affinity purification

After addition of 8 M urea for protein denaturation, and dithiothreitol and iodoacetamide for cysteine reduction and alkylation, the labeled proteins were digested with modified sequencing-grade trypsin (Roche Applied Science) at an enzyme/substrate ratio of 1:100 in 25 mM NH$_4$HCO$_3$ (pH 8.5) at 37°C for overnight. The peptide mixture was subsequently dried in a Speed-vac and redissolved in 1 mL of 100 mM potassium phosphate and 0.15 M NaCl (pH 7.5, PBS buffer), to which solution was subsequently added 200 μL avidin-agarose resin (Sigma-Aldrich). The mixture was incubated at 25°C for 1 hr with gentle shaking. The agarose resin was then washed with 3 mL PBS and 3 mL H$_2$O to remove unbound peptides, and the labeled peptides were subsequently eluted with 1% TFA in CH$_3$CN/H$_2$O (7:3, v/v) at 65°C. The eluates were dried in a Speed-vac and stored at -20°C prior to LC-MS/MS analysis.

LC-MS/MS Analysis

LC-MS/MS analysis was performed on an LTQ-Orbitrap Velos mass spectrometer equipped with a nanoelectrospray ionization source (Thermo Fisher Scientific, San Jose, CA). Samples were automatically loaded from a 48-well microplate

autosampler using an EASY-nLC system (Proxeon Biosystems, Odense, Denmark) at 3 µL/min onto a home-made 4 cm trapping column (150 µm i.d.) packed with 5 µm C18 120 Å reversed-phase material (ReproSil-Pur 120 C18-AQ, Dr. Maisch). The trapping column was connected to a 20 cm fused silica analytical column (PicoTip Emitter, New Objective, 75 µm i.d.) with 3 µm C18 beads (ReproSil-Pur 120 C18-AQ, Dr. Maisch). The peptides were then separated with a 120-min linear gradient of 2-35% acetonitrile in 0.1% formic acid and at a flow rate of 250 nL/min. The LTQ-Orbitrap Velos was operated in a data-dependent scan mode. Full-scan mass spectra were acquired in the Orbitrap analyzer with a resolution of 60000 with lock mass option enabled for the ion of *m/z* 445.120025 (20). Up to 20 most abundant ions found in MS with charge state ≥ 2 were sequentially isolated and sequenced in ion trap with a normalized collision energy of 35, an activation q of 0.25 and an activation time of 10 ms.

Online 2D LC Separation and LC-MS/MS analysis

The fully automated 7-cycle on-line two-dimensional LC-MS/MS was set up as described (21). Briefly, the C18 trapping column was replaced with a biphasic precolumn (150 µm i.d.) comprised of a 3.5-cm column packed with 5 µm C18 120 Å reversed-phase material (ReproSil-Pur 120 C18-AQ, Dr. Maisch) and 3.5-cm column packed with Luna 5 µm SCX 100 Å strong cation exchange resin (Phenomenex, Torrance, CA) while all other setups remained the same. Enriched peptides were first loaded onto the biphasic precolumn. Ammonium acetate at concentrations of 0, 25, 50, 75, 125, 200 and 500 mM were then sequentially injected using a 48-well autosampler from the

sample vial to elute bound peptides from precolumn to analytical column with reversed-phase separation. LC-MS/MS experiments were also performed with a 120-min linear gradient of 2-35% acetonitrile in 0.1% formic acid.

Data Processing and Analysis.

The raw data were first converted to mzXML files and DTA files using ReAdW (http://sourceforge.net/projects/sashimi/files/) and MzXML2Search (http://tools.proteomecenter.org/wiki/index.php?title=Software:MzXML2Search) programs, respectively. Bioworks 3.2 was used for protein identification by searching the DTA files against the human IPI protein database version 3.68 (87,062 entries) and its reversed complement. Initial precursor mass tolerance of 10 ppm and fragment mass deviation of 0.8 Th were set as the search criteria. The maximum number of miss-cleavages for trypsin was set as two per peptide. Cysteine carbamidomethylation was considered as a fixed modification, whereas methionine oxidation and desthiobiotinylation of lysine (+281.17394 Da) were included as variable modifications. Aside from the search parameters described above, lysine (+8 Da) and arginine (+6 Da) mass shifts introduced by heavy isotope labeling were also considered as variable modifications. The search results were then filtered with DTASelect (22) to achieve a peptide false discovery rate of 1%. Census was employed for peptide and protein quantification (23). Extracted-ion chromatograms were first generated for peptide ions based on their m/z values and peptide intensity ratios were subsequently calculated in Census from peak areas found in each pair of extracted-ion chromatograms.

***In vitro* kinase activity assay for CDK6**

CDK6 substrate peptide EGLPTPT$_{821}$KMTPPFR derived from retinoblastoma-associated protein (Rb) was purchased from Genemed Synthesis, Inc. *In vitro* kinase assays for CDK6 are performed using mixture of 0.5 μg of CDK6/Cyclin D3 complex (SignalChem, Catalog #: C35-10H), 5 μM of CDK 6 peptide substrate EGLPT$_{821}$PTKMTPPFR, 1×Kinase Dilution Buffer III (SignalChem, Catalog #: K23-09) and 1×Kinase Assay Buffer I (SignalChem, Catalog #: K01-09). To each reaction mixture, 500 μM of ATP, GTP, or ^{S}GTP was added separately and the final volume was adjusted to 40 μL by adding H$_2$O. To test the inhibitory effect of ^{S}GTP on CDK6 phosphorylation, a series of ^{S}GTP solutions at concentrations of 0, 125 μM, 250 μM, 500 μM, 1 mM and 2.5 mM were added to individual CDK6 reaction mixture in the presence of a constant concentration of ATP (500 μM). *In vitro* CDK6 phosphorylation reactions were continued at 25°C for 45 min. The reactions were quenched with 10% TFA (4 μL). The resulting peptides were further desalted by ZipTip and then analyzed by LC-MS/MS on an LTQ-Orbitrap Velos mass spectrometer.

Results and Discussion

1. Strategy for proteome-wide characterization of ^{S}GTP-protein interactions

Here, we extended the use of biotin-labeled nucleotide affinity probes as acylating agents to selectively label and enrich ^{S}GTP-binding proteins from the entire human proteome. Similar as the previously reported ATP or GTP affinity probes(19), ^{S}GTP

affinity probe harbors a binding moiety (^{S}GTP) and an enrichment moiety (i.e. desthiobiotin) that are conjugated through an acyl phosphate linkage (Figure 4.1). Upon binding to proteins, the acyl phosphate component of the affinity probe reacts with the ε-amino group of the specific lysine residue at the nucleotide-binding site to yield a stable amide bond, which results in the covalent attachment of desthiobiotin to the lysine residue on ^{S}GTP-binding proteins (Figure 4.2).

It is worth noting that, owing to its relatively high reactivity, lysine residues not involved in nucleotide binding may also be modified by the nucleotide affinity probe via non-specific electrostatic interactions. To distinguish specific from non-specific labeling, we applied a nucleotide affinity profiling strategy developed previously in our lab(15) utilizing a high and a low concentrations of ^{S}GTP affinity probe, along with SILAC-based quantitative proteomics platform(24), to unambiguously characterize ^{S}GTP-binding affinities of proteins at the entire proteome scale. Along this line, the binding of the ^{S}GTP component of the probe to a protein promotes the acyl phosphate moiety to couple with the lysine residue at the nucleotide binding site.(25) Therefore, lysine residues involved with ^{S}GTP binding and those that are not would exhibit distinct labeling behaviors with low and high concentrations of the ^{S}GTP affinity probe: At low probe concentration (10 μM), the former lysine residues possess high reactivity and are completely labeled and the latter are partially labeled because of the limited amount of labeling reagent present; however, at high probe concentration (100 μM), both types of lysines are labeled to almost completion.

Figure 4.1 The structures of the ^SGTP and GTP affinity probes.

Desthiobiotin
(Enrichment moiety)

Nucleotide
(Binding moiety)

^SGTP Affinity Probe
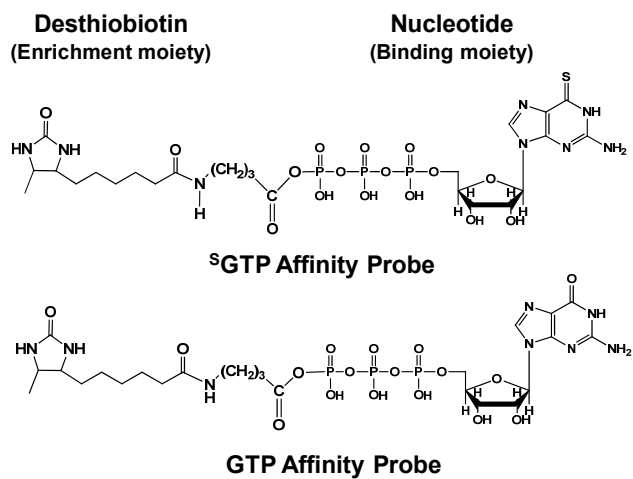
GTP Affinity Probe

Figure 4.2 (A) A schematic diagram showing the reaction between ^{S}GTP affinity probe with an ^{S}GTP-binding protein. (B) Representative of ESI-MS of purified desthiobiotin-based ^{S}GTP affinity probes.

(A)
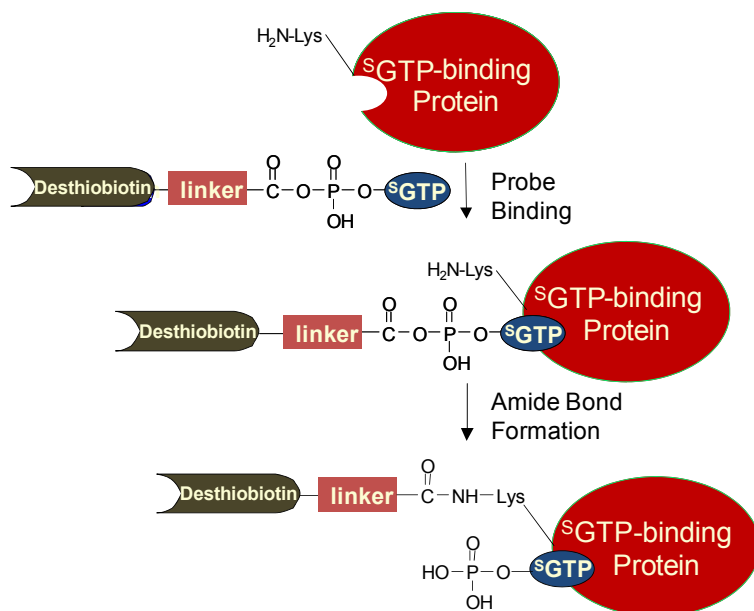
(B)
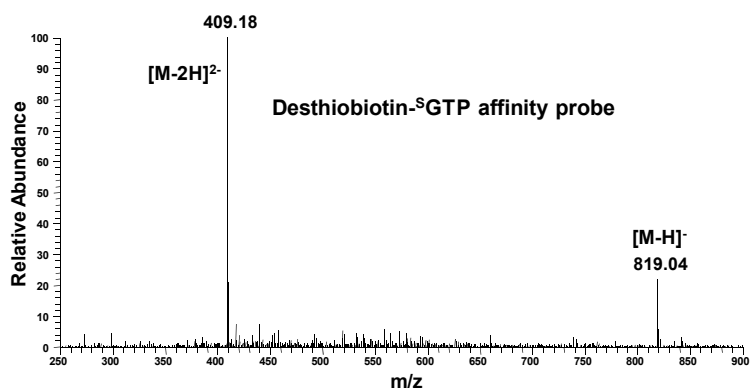
After labeling the light and heavy SILAC cell lysates with 10 and 100 μM of the probe, respectively (forward SILAC, and the labeling experiment was also carried out in an opposite way in reverse SILAC, Figure 4.3), we combined the two protein samples, digested the protein mixture with trypsin, enriched the resulting desthiobiotin-labeled peptides with avidin agarose, and subjected the enriched peptides to LC-MS/MS analyses. The peak intensity ratios of desthiobiotin-labeled light and heavy peptides were then employed to derive ^{S}GTP-binding affinity ratio, $R_{SGTP10/1}$, for specific lysine residues in individual proteins (Figure 4.3), where specific ^{S}GTP-binding lysine will display an $R_{SGTP10/1}$ close to 1 since similar amount of ^{S}GTP-binding lysine will be labeled regardless of the probe concentration. By contrast, non-specifically labeled lysine will show concentration-dependent increase in ^{S}GTP probe labeling, which will yield a $R_{SGTP10/1} \gg 1$.

Although the above-proposed ^{S}GTP affinity profiling assay is effective in identifying specific ^{S}GTP-binding proteins, it provides little information about the binding selectivities of these proteins towards ^{S}GTP versus other endogenous nucleotides (i.e., ATP and GTP). To address this, we also employed SILAC together with our ^{S}GTP and GTP affinity probes to compare directly the ^{S}GTP- and GTP-binding affinities of proteins at the entire proteome scale. In this experiment, light- and heavy-labeled cell lysates were treated with 100 μM each of desthiobiotin-based ^{S}GTP and GTP probes, respectively, and, to minimize the bias introduced by SILAC, we also conducted the labeling experiments in the opposite way (i.e., reverse SILAC experiment). After the

Figure 4.3 Comprehensive characterization of ^{S}GTP-binding proteins by parallel quantitative ^{S}GTP-affinity profiling assay and ^{S}GTP/GTP competition assay.
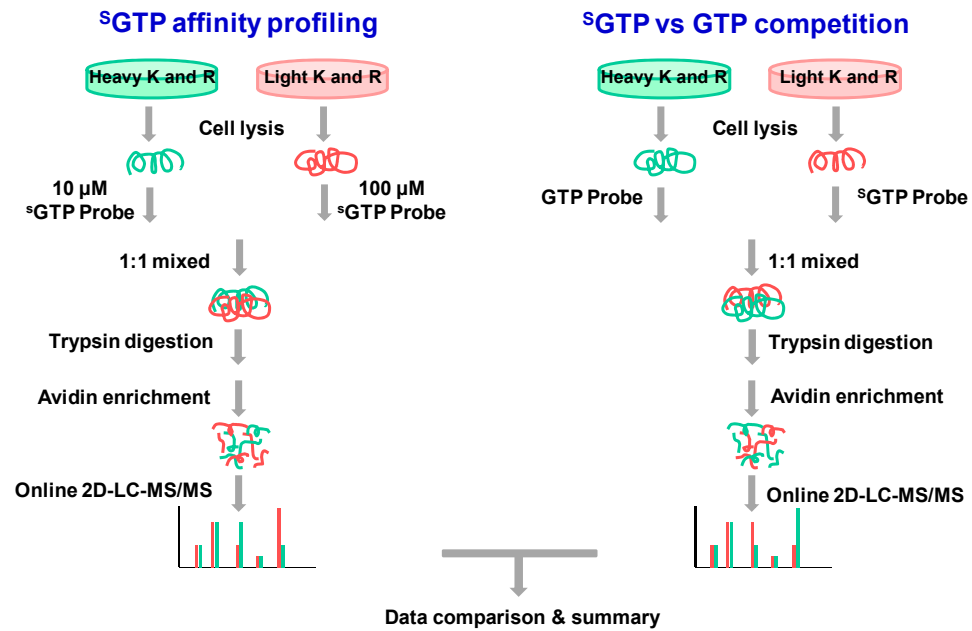
reaction, light- and heavy-labeled cell lysates were mixed prior to any further steps of sample manipulation as described above. Peak intensity ratios of light and heavy desthiobiotin-labeled tryptic peptides were subsequently used to derive ^{S}GTP/GTP binding affinity ratio, $R_{SGTP/GTP}$, which reflects the relative binding affinities of ^{S}GTP and GTP towards specific lysine residues in proteins of interest.

2. Proteome-wide profiling of ^{S}GTP-binding proteins

Our quantitative ^{S}GTP-affinity profiling of whole cell lysate from Jurkat-T cells led to the quantification of a total of 1925 proteins, which include more than 5400 light or heavy desthiobiotin-modified lysine residues. As depicted in Figure 4.4A, a large number of peptides with desthiobiotin modification exhibited significantly different probe labeling efficiency when 10 and 100 µM of probe were employed for the labeling reactions, with $R_{SGTP10/1} \gg 1$. However, a small portion of the desthiobiotin-modified peptides bear $R_{SGTP10/1}$ close to 1, indicating that they possess comparable labeling efficiency at low and high probe concentrations.   Similar as the previous quantitative affinity profiling assay for ATP-binding proteins(15), we arbitrarily consider lysine residue in peptides with $R_{SGTP10/1} < 2$ as ^{S}GTP-binding lysine. In addition, only those modified lysine residues that were successfully quantified at least twice are retained on the list. With these criteria, we obtained 199 unique desthiobiotin-modified peptides, representing approximately 4% of total identified peptides, from 165 unique proteins. We considered these 165 proteins as ^{S}GTP-binding proteins.

Figure 4.4 (A) Measured $R_{SGTP10/1}$ ratio from Jurkat-T cell lysates with low (10 μM) and high (100 μM) concentrations of ^{S}GTP probe in ^{S}GTP affinity profiling assay; (B) Molecular function GO analysis for proteins with different $R_{SGTP10/1}$ ratios using Jurkat-T cell lysates. (C) Biological process GO analysis for targeted ^{S}GTP-binding proteins.

(A)

(B)

(C)
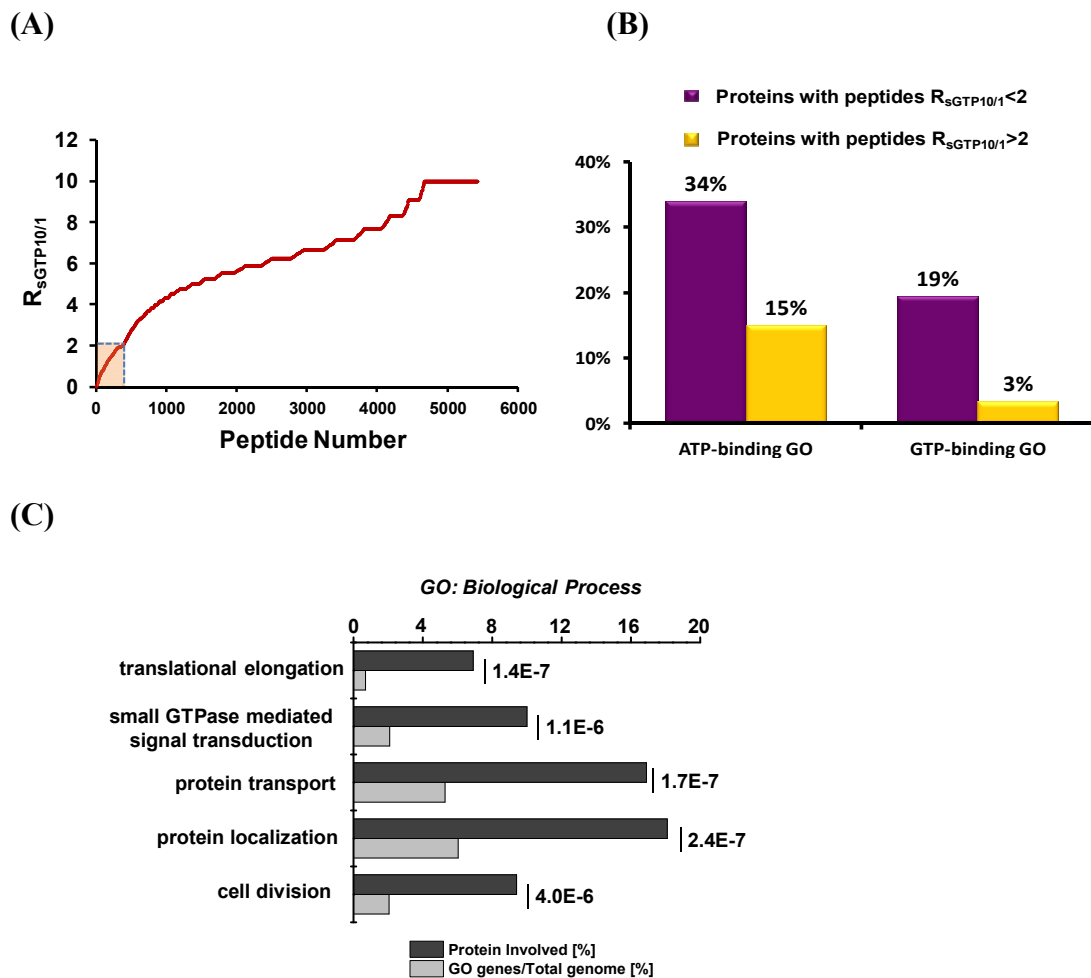
Owing to the structural similarity of ^{S}GTP and GTP, the ^{S}GTP affinity probe may also bind to, and conjugate with GTP-binding proteins. We surveyed the 165 identified ^{S}GTP-binding targets for their functional characteristics using DAVID(26), which revealed that 19% of the 165 proteins are with known GTP-binding gene ontology (GO), suggesting a 8.4-fold enrichment relative to the entire human proteome with a p-value of 8.1E-20 (Figure 4.4B and Table 4.1). In contrast, the percentage of known GTP-binding proteins in the protein group containing peptides with $R_{SGTP10/1}>2$ is only 3%. More significantly, among the 199 peptides from the candidate ^{S}GTP-binding protein group, peptides from known GTP-binding proteins were on average identified and quantified 2-3 times more frequently than those from proteins lacking GTP-binding GO, suggesting an even more pronounced GTP-binding enrichment efficiency (44% of all quantification events). Additionally, we found that 56 out of 165 (34%) identified ^{S}GTP-binding proteins are known ATP-binding proteins, indicating a 3.4-fold enrichment relative to the entire human proteome with a p-value of 1.5E-15. However, in contrast to the GTP-binding protein group, a large percentage of known ATP-binding proteins with desthiobiotin labeling identified in our profiling experiment were excluded from ^{S}GTP-binding protein groups (Figure 4.4B and Table 4.1). Only 18% (56 out of a total of 318) of identified known ATP-binding proteins are ^{S}GTP-binding proteins, suggesting the necessity in revealing non-specific probe labeling arising from electrostatic interaction based on the $R_{SGTP10/1}$ ratio. Thus, ^{S}GTP targets a wide array of known GTP- and ATP-binding proteins.

Table 4.1 Details of GO analysis of proteins with different RSGTP10/1 ratio using Jurkat-T cell lysates.

		Total quantified protein (1925)	Possible ^{S}GTP binding protein (165)	Other quantified protein (1760)
	Number	318	56	262
	percentage	16.5%	33.9%	14.9%
	p value	4.4E-33	1.5E-15	1.2E-21
ATP binding protein	enrichment factor	1.6	3.4	1.5
	Number	91	32	59
	percentage	4.7%	19.4%	3.3%
	p value	1.0E-12	8.1E-20	3.1E-4
GTP binding protein	enrichment factor	2.0	8.4	1.4

GO analysis also unveiled a strong enrichment of multiple GTP-related pathways including translational elongation, protein transport and localization, and small GTPase-mediated signal transduction (Figure 4.4C). This is in keeping with the over-representation of GTP-binding GO of [S]GTP-binding targets, suggesting that [S]GTP may interfere significantly with the GTP-related cell function. For example, GTP serves as an important co-factor for elongation factors 1 and 2 to regulate the GTP hydrolysis-dependent translocation of peptidyl-tRNA during translation elongation(27). Our results showed that [S]GTP binds to elongation factors 1-alpha, 1-alpha2, and 2, which may affect the GTP-related regulatory mechanism of protein synthesis. Interestingly, we also observed the over-representation of "cell cycle progression" process (Figure 4.4C), where more than 20 proteins involved in this process, including multiple CDKs, were found to be [S]GTP-binding proteins. These results offer a wealth of information for the further examination of [S]GTP-protein interaction and the mechanism of thiopurine drug toxicity.

3. Competitive binding of [S]GTP/GTP to known GTPases

As mentioned above, our nucleotide affinity profiling strategy differs from other nucleotide-binding assays by virtue of its capability in site-specific differentiation of nucleotide binding. Benefited from this high-throughput proteome-wide quantitative analysis, we sought to analyze the sequence context surrounding the desthiobiotin-modified lysine, which is considered as the [S]GTP-binding site in proteins. In this regard, we included all the desthiobiotin-labeled peptides with $R_{SGTP10/1}$ ratios being smaller than 2 for motif search using motif-X(28). This led to the discovery of the well-known P-loop

sequence motif of GxxxxGKS(29), with a 91-fold enrichment relative to the occurrence frequency of this motif in the entire proteome (Figure 4.5A). This finding suggests that [S]GTP directly competes with GTP in binding towards protein targets at exactly the same nucleotide-binding site at the P-loop region. Therefore, our [S]GTP affinity profiling assay and [S]GTP/GTP competition assay could allow for site-specific determination of the [S]GTP binding affinity and selectivity towards the unique P-loop motif sequence, which is known to be directly involved in GTP-binding for these GTPases. Based on the results from [S]GTP affinity profiling assay and [S]GTP/GTP competition assay, we constructed a heatmap to better visualize the [S]GTP binding affinity and selectivity for 34 quantified GTPases towards the unique P-loop binding motif (Figure 4.5B). The heatmap showed that the GxxxxGKS motif in most known GTP-binding proteins, including small GTPases, heterotrimeric G proteins and several other GTP-binding proteins (i.e., elongation factor 1-alpha 2, GTP-binding proteins 1 and 2) exhibited significant binding preference toward [S]GTP.

Small GTPases are structurally classified into at least five families, namely, the Ras, Rho, Rab, Sar1/Arf, and Ran families(12). Like other G proteins, small GTPases exist in the inter-convertible GDP-bound inactive and GTP-bound active forms. The regulation of the activity of small GTPases plays vital roles in cell signaling and loss of such regulation is believed to be closely associated with the development of various types of cancer(30). Our results led to the quantification of most small GTPases, including 14 Rab proteins, with their GxxxxGKS motif displaying significant binding preference

Figure 4.5 (A) Unique binding motifs found for the identified [S]GTP-binding proteins in [S]GTP affinity profiling experiment. (B) A heatmap displaying the [S]GTP-binding affinity and [S]GTP/GTP binding selectivity toward GxxxxGKS binding motif of known GTPases. For [S]GTP affinity profiling heatmap, dark blue indicates strong [S]GTP-binding affinity with small $R_{SGTP10/1}$ ratios and white represents low [S]GTP-binding affinity with larger $R_{SGTP10/1}$ ratios; for [S]GTP/GTP competition binding heatmap, dark red and dark green indicate significant GTP-binding preference with low $R_{SGTP/GTP}$ ratio and similar [S]GTP/GTP-binding preference (i.e., with $R_{SGTP/GTP}$ ratio close to unity), respectively (See scale bar below the heatmap). Grey represents unquantified GTPases.

(A)

(B)
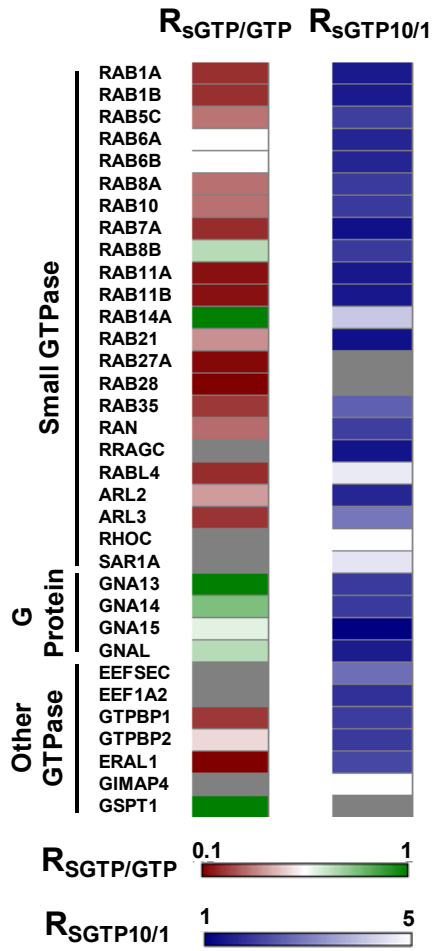
toward ^{S}GTP. For instance, the desthiobiotin-labeled peptide from the P-loop motif of Rab11b, namely, VVLIGDSGVGK$_{24}$#SNLLSR, was detected in our ^{S}GTP affinity profiling experiment with a significant ^{S}GTP binding affinity (Rs$_{GTP10/1}$ ratio= 1.39).

It is worth noting that several small GTPases, i.e., Ras-related protein Rab-14, RAS oncogene family-like 4, Rho-related GTP-binding protein RhoC, and GTP-binding protein SAR1a, displayed low ^{S}GTP binding affinity. Although P-loop motif sequence from these small GTPases can still be labeled with the ^{S}GTP probe, relatively large R$_{SGTP10/1}$ ratios were observed for these peptides, suggesting their low binding affinity toward ^{S}GTP. Moreover, our ^{S}GTP/GTP competition assay revealed that small GTPase-related pathways may not be the major targets for ^{S}GTP. As shown in Figure 4.5B, most small GTPases exhibit stronger binding to GTP than ^{S}GTP. This result suggests that, owing to the competition from binding to endogenous GTP, ^{S}GTP may not affect substantially the small GTPase-related cell function *in vivo*. This finding is reminiscent of a previous observation that although ^{S}GTP is capable of binding to eight small GTPases like Ran protein, much higher concentrations of ^{S}GTP were required to displace the bound GTP from these small GTPases(10).

Four heterotrimeric G proteins show significant binding affinity toward ^{S}GTP. For instance, one desthiobiotin-labeled peptide with P-loop motif from guanine nucleotide-binding protein G(olf) subunit alpha (GNAL), i.e., LLLLGAGESGK$_{55}$#STIVK, was successfully quantified in both forward and reverse ^{S}GTP affinity profiling assay, where K55 consistently displays a low R$_{SGTP10/1}$ of 1.46

167

(Figure 4.6A, B). This result suggests that GNAL may bind to ^{S}GTP, with K55 being directly involved in nucleotide binding. Different from what we found for small GTPases, P-loop regions of all heterotrimeric G proteins quantified in our experiments exhibited similar binding selectivities toward ^{S}GTP and GTP. For instance, the aforementioned K55 from GNAL can be effectively labeled by GTP and ^{S}GTP probes at similar efficiency, suggesting the similar binding affinity of this protein to these two nucleotides. The finding that ^{S}GTP competes with GTP in binding multiple heterotrimeric G proteins indicates that ^{S}GTP may affect heterotrimeric G protein-mediated signaling. Signal amplitude of G-protein-related pathway is a dynamic interplay of GDP/GTP exchange (activation) and GTP hydrolysis (deactivation), where the GDP-bound inactive form of heterotrimeric G proteins interact with membrane-bound G protein-coupled receptors (31). The competitive binding of ^{S}GTP instead of GTP to Gα subunit may block G protein activation by inhibiting Gα release. In addition, the loading of ^{S}GTP to Gα may lead to the accumulation of ^{S}GTP-bound, active form or ^{S}GDP-bound, inactive form of G proteins over time; either scenario may perturb G-protein mediated signaling.

4. Competitive binding of ^{S}GTP to multiple CDKs

Aside from the various GTPases, we also made an interesting discovery that ^{S}GTP can bind to multiple CDKs. CDKs are serine/threonine kinases that become active only when coupled with specific types of cyclins. These CDKs and their activating cyclins (A, B, D, and E) are key regulators in mammalian cell cycle progression(32). For

Figure 4.6 Light- and heavy-labeled peptides of selective GTPases from forward and reverse ^{S}GTP-affinity profiling and ^{S}GTP/GTP competition binding experiments. (A, B) Peptide VVLIGDSGVGK$_{24}$#SNLLSR with a low $R_{SGTP10/1}$ ratio from Rab-11B; (C, D) Peptide VVLIGDSGVGK$_{24}$#SNLLSR with a low $R_{SGTP/GTP}$ ratio from Rab-11B; (E, F) Peptide LLLLGAGESGK$_{55}$#STIVK with a low $R_{SGTP10/1}$ ratio from guanine nucleotide-binding protein G(olf) subunit alpha (GNAL). (G, H) Peptide LLLLGAGESGK$_{55}$#STIVK with a $R_{SGTP/GTP}$ ratio close to unity from GNAL. "#" indicates the desthiobiotin-labeling site.
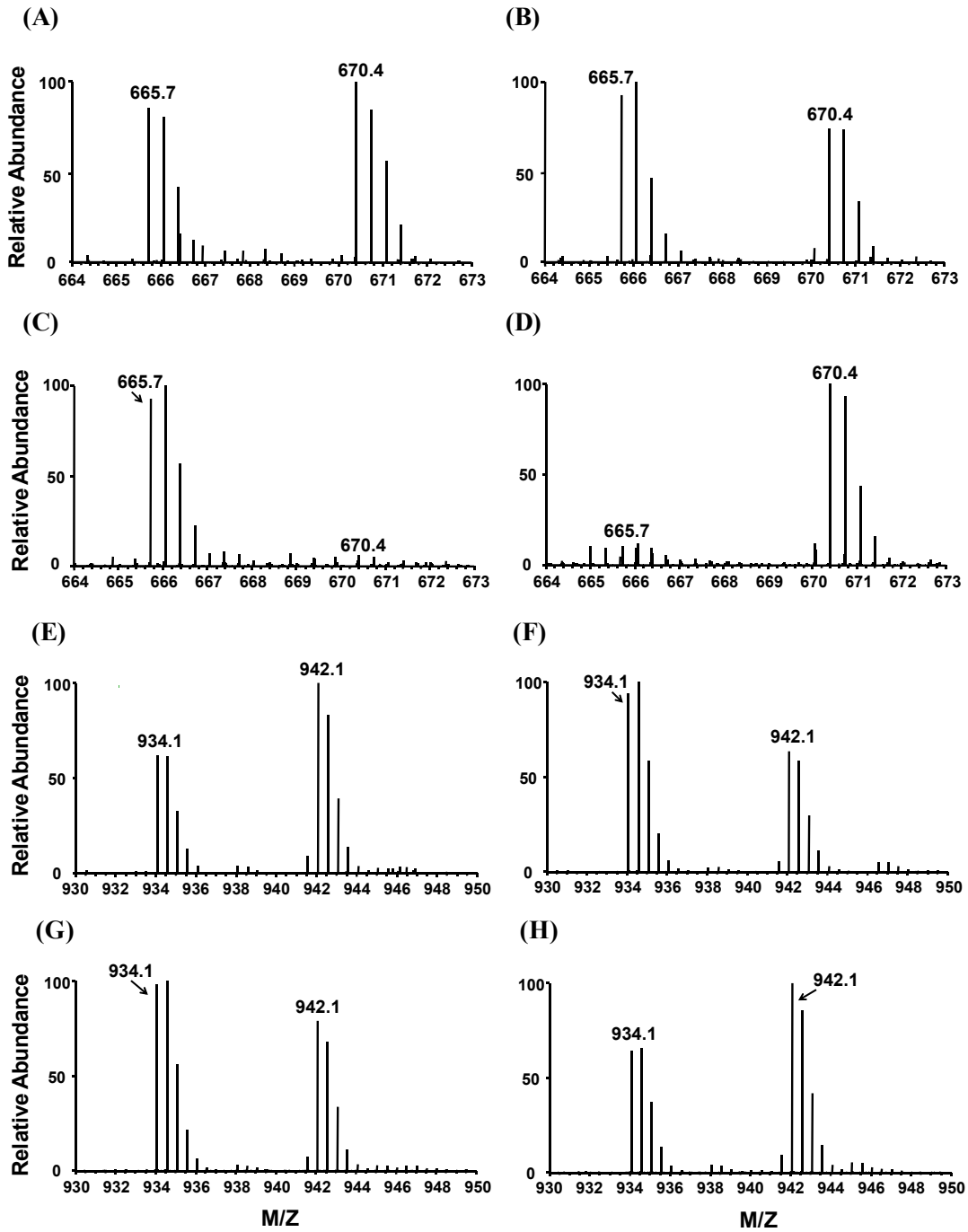
instance, CDK4/cyclin D and CDK2/cyclin E/A promote the passage through G1 and S phases, whereas CDK1/cyclin B, as the only non-redundant cell cycle driver, regulates the transition through late G2 and mitosis(33). Malfunction of CDKs, especially their hyperactivation may induce unregulated proliferation of cancer cells. Thus, CDKs are attractive targets for cancer therapy and multiple CDK inhibitors have been introduced as anticancer drugs for preclinical and clinical evaluation(34).

Intriguingly, our ^{S}GTP affinity profiling results showed that CDK1, CDK2, CDK4, CDK5, CDK6 can bind strongly to ^{S}GTP, with all $R_{SGTP10/1}$ being < 2. Moreover, we found that ^{S}GTP binds to the same binding motif (i.e., the HRD motif) in these CDKs as ATP(15). For instance, we successfully detected probe-labeled K80 located in the HRD motif of CDK6. In addition, results from ^{S}GTP/GTP competition experiment towards this unique ATP-binding HRD motif sequence indicate that CDK1 ($R_{SGTP/GTP}$=0.51), CDK2 ($R_{SGTP/GTP}$=0.85), CDK5 ($R_{SGTP/GTP}$=0.88) have similar nucleotide-binding preference between GTP and ^{S}GTP, whereas CDK4 ($R_{SGTP/GTP}$=3.33) and CDK6 ($R_{SGTP/GTP}$=3.85) display much stronger binding toward ^{S}GTP than GTP (Figure 4.7B). It is worth noting that CDK4 and CDK6 belong to the same D-Type cyclin-dependent kinase group, which bind to and are activated by cyclin D(35). The structural and functional similarities of CDK4 and CDK6 are consistent with their comparable ^{S}GTP binding property as revealed from our ^{S}GTP-binding profiling assay.

Although ^{S}GTP has been observed to bind to a broad spectrum of CDKs, it is possible that CDKs may be able to employ ^{S}GTP, in a similar manner as ATP, as

171

Figure 4.7 Light- and heavy-labeled peptides of CDK6 from forward and reverse ^{S}GTP-affinity profiling and ^{S}GTP/GTP competition binding experiments. (A, B) Peptide DLK$_{147}$#PQNILVTSSGQIK with a low R$_{SGTP10/1}$ ratio from CDK6; (C, D) Peptide DLK$_{147}$#PQNILVTSSGQIK with a high R$_{SGTP/GTP}$ ratio from CDK6; "#" indicates the desthiobiotin-labeling site.
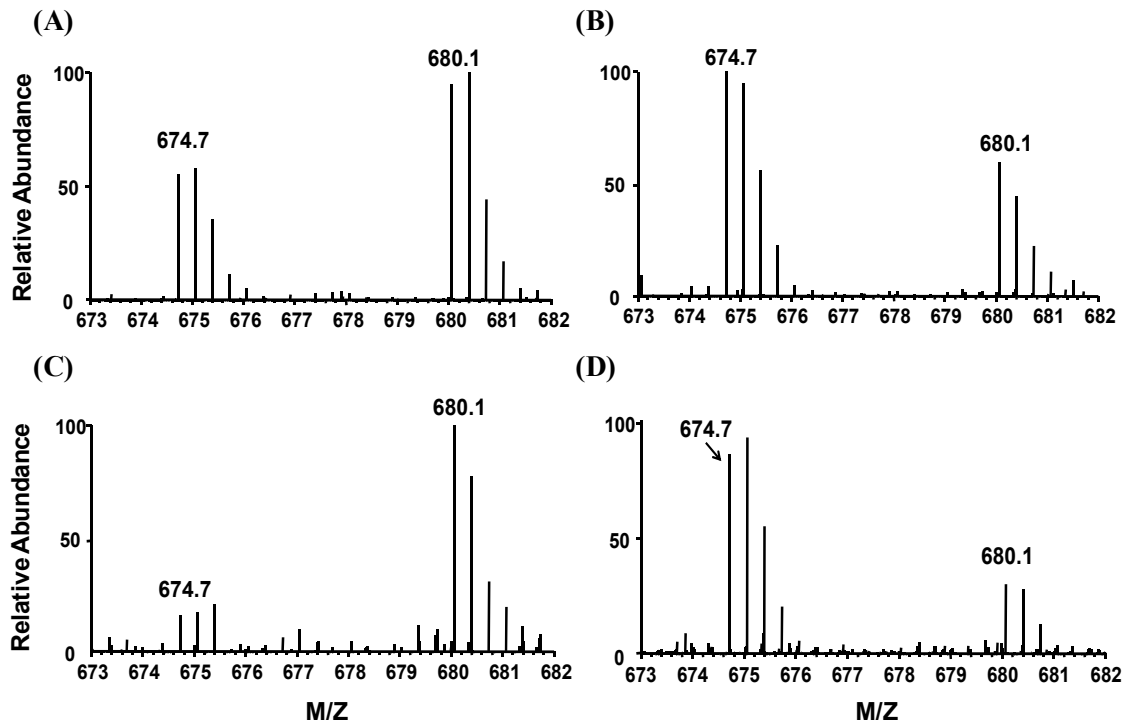
phosphate donor to phosphorylate its substrate proteins, which may not perturb the functions of CDKs. To examine this possibility, we performed an *in vitro* kinase activity assay using purified CDK6 and a synthetic peptide EGLPT$_{821}$PTKMTPPFR, derived from retinoblastoma-associated protein (Rb), which is a known CDK6 substrate(36). Mass spectrometry was employed to detect the phosphorylated substrate peptide. Our results showed that, in the presence of ATP, CDK6/cyclin D can successfully phosphorylate the substrate peptide with the signal ratio for phosphorylated/unmodified peptide being ~ 7% (Figure 4.8A). Furthermore, the MS/MS supports the phosphorylation of Thr821 in the substrate peptide (Figure 4.9), which is consistent with previous finding(37). However, when the same *in vitro* kinase activity assay of CDK6 was conducted using GTP or ^{S}GTP to replace ATP as the phosphate donor, no phosphorylated peptide was observed. There results suggest that, although ^{S}GTP displays robust binding toward CDK6, CDK6 can only employ ATP as the phosphate donor to phosphorylate the Rb protein. Therefore, we further assessed the inhibitory effect of ^{S}GTP for CDK6 phosphorylation by conducting the same *in vitro* phosphorylation reaction with constant ATP concentration along with increasing amount of ^{S}GTP. Significant inhibitory effect of CDK6 phosphorylation by ^{S}GTP binding was observed. As shown in Figure 4.8B, around 50% and 80% inhibition in CDK6 phosphorylation was observed when ^{S}GTP/ATP concentration ratio reached 1/2 and 1/1, respectively, and no phosphorylation was observed when the ratio was 2/1. Therefore, our results suggest that the competitive binding of ^{S}GTP to multiple CDKs may not only affect their binding

173

Figure 4.8 (A) *In vitro* CDK6 kinase activity assay employing ATP, GTP, and ^{S}GTP as phosphate donor. (B) Inhibitory effect of ^{S}GTP binding on *in vitro* CDK6 phosphorylation reaction**.**

(A)
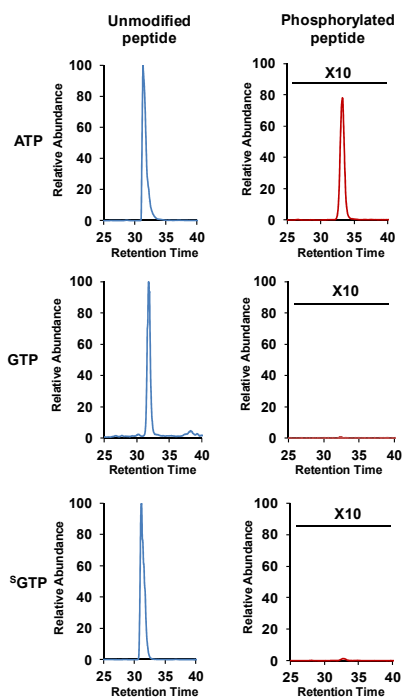
(B)
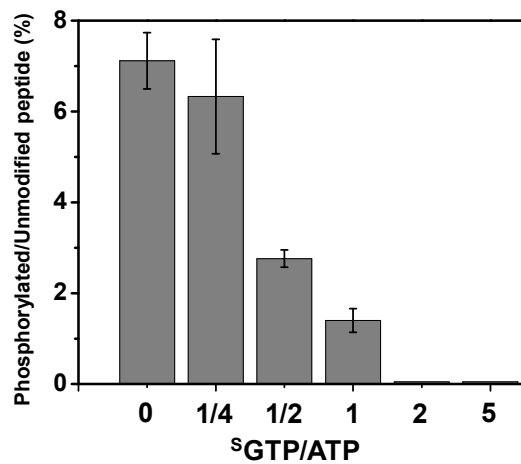
Figure 4.9 MS/MS of the [M+2H]$^{2+}$ ion of the phosphorylated CDK6 substrate peptide

EGLPT$_P$PTKMTPPFR (*m/z*= 826.3993)
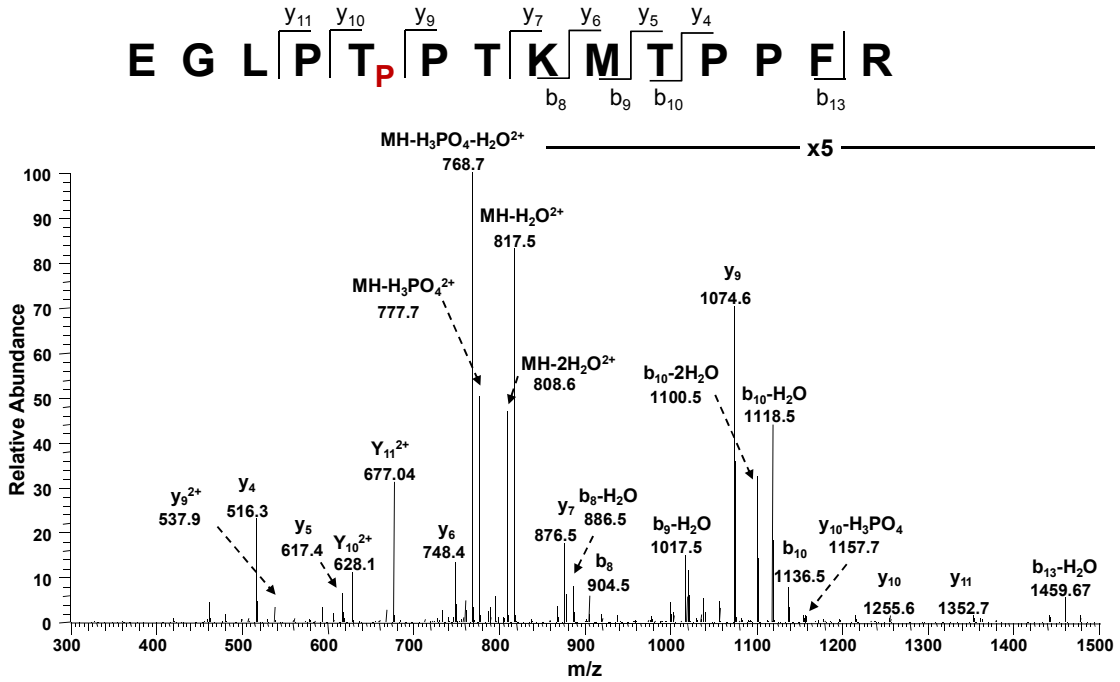
toward endogenous ATP but also greatly affect the phosphorylation efficiency of the kinases. Thus, the ^{S}GTP introduced by thiopurine drug treatment may exert a profound effect on CDK activity, which may ultimately affect cell cycle progression.

Conclusions

Here, we introduced an orthogonal strategy encompassing the nucleotide affinity profiling assay and nucleotide binding competition assay to comprehensively characterize, at the entire proteome scale, ^{S}GTP-binding proteins. With the simultaneous use of ^{S}GTP and GTP affinity probes, 165 proteins involved in different biological processes were determined to be ^{S}GTP-binding targets. In addition, the selectivity between ^{S}GTP- and GTP-binding for these ^{S}GTP targets was further characterized. Unlike traditional binding competition assay, our ^{S}GTP/GTP binding selectivity was determined from the binding preference of ^{S}GTP/GTP towards the verified nucleotide-binding motif sequence, which provides superior specificity and accuracy. Our results suggest that ^{S}GTP mainly targets GTPases and affects various GTPase-mediated signaling pathways. Especially, ^{S}GTP exhibits strong binding affinity towards multiple heterotrimeric G proteins. Furthermore, we demonstrated for the first time that ^{S}GTP binds to multiple CDKs, which may perturb the CDK-mediated phosphorylation and cell cycle progression.

Nucleotides are susceptible to damage through exposure to various genotoxic agents, including reactive oxygen species generated from normal metabolism or from

exposure to ionizing radiation and environmental chemicals(38). For example, 8-oxo-7,8-dihydroguanosine triphosphate (8-oxoGTP) could be produced at appreciable levels in the cytoplasm(39). Some damaged nucleotides may be incorporated into RNA or DNA, thereby perturbing the flow of genetic information. Alternatively, damaged nucleotides may be recognized by certain cellular proteins to perturb relevant cellular functions. The analytical strategy presented here should be applicable for quantitative studies of proteins that can bind to 8-oxoGTP and any other damaged nucleotides at the whole proteome scale. Such studies should also result in the discovery of specific proteins that can bind to these damaged nucleotides and how these proteins recognize the damaged nucleotides versus their endogenous undamaged counterparts.

References

1.	Elion G. The purine path to chemotherapy. Science. 1989;244:41-7.

2.	Evans WE, Relling MV. Mercaptopurine vs thioguanine for the treatment of acute lymphoblastic leukemia. Leuk Res. 1994;18:811-4.

3.	Swann PF, Waters TR, Moulton DC, Xu Y-Z, Zheng Q, Edwards M, et al. Role of Postreplicative DNA Mismatch Repair in the Cytotoxic Action of Thioguanine. Science. 1996;273:1109-11.

4.	Karran P, Offman J, Bignami M. Human mismatch repair, drug-induced DNA damage, and secondary cancer. Biochimie. 2003;85:1149-60.

5.	Krynetski EY, Krynetskaia NF, Gallo AE, Murti KG, Evans WE. A Novel Protein Complex Distinct from Mismatch Repair Binds Thioguanylated DNA. Mol Pharmacol. 2001;59:367-74.

6.	Yuan B, Zhang J, Wang H, Xiong L, Cai Q, Wang T, et al. 6-Thioguanine reactivates epigenetically silenced genes in acute lymphoblastic leukemia cells by facilitating proteasome-mediated degradation of DNMT1. Cancer Res. 2011;71:1904-11.

7.	Brem R, Karran P. Oxidation-mediated DNA cross-linking contributes to the toxicity of 6-thioguanine in human cells. Cancer Res. 2012;72:4787-95.

8.	Zhang F, Fu L, Wang Y. 6-Thioguanine induces mitochondrial dysfunction and oxidative DNA damage in acute lymphoblastic leukemia cells. Mol Cell Proteomics. 2013;12:3803-11.

9.	Tiede I, Fritz G, Strand S, Poppe D, Dvorsky R, Strand D, et al. CD28-dependent Rac1 activation is the molecular target of azathioprine in primary human CD4+ T lymphocytes. The Journal of Clinical Investigation. 2003;111:1133-45.

10.	Poppe D, Tiede I, Fritz G, Becker C, Bartsch B, Wirtz S, et al. Azathioprine Suppresses Ezrin-Radixin-Moesin-Dependent T Cell-APC Conjugation through Inhibition of Vav Guanosine Exchange Activity on Rac Proteins. The Journal of Immunology. 2006;176:640-51.

11.	Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. Science. 2002;298:1912-34.

12.   Takai Y, Sasaki T, Matozaki T. Small GTP-binding proteins. Physiol Rev. 2001;81:153-208.

13.   Ormö M, Sjöberg B-M. An ultrafiltration assay for nucleotide binding to ribonucleotide reductase. Anal Biochem. 1990;189:138-41.

14.   Guarnieri MT, Blagg BS, Zhao R. A high-throughput TNP-ATP displacement assay for screening inhibitors of ATP-binding in bacterial histidine kinases. Assay Drug Dev Technol. 2011;9:174-83.

15.   Xiao Y, Guo L, Wang Y. Isotope-coded ATP probe for quantitative affinity profiling of ATP-binding proteins. Anal Chem. 2013;85:7478-86.

16.   Yoshikawa M, Kato T, Takenishi T. A novel method for phosphorylation of nucleosides to 5′-nucleotides. Tetrahedron Lett. 1967;8:5065-8.

17.   Abramova TV, Vasileva SV, Serpokrylova IY, Kless H, Silnikov VN. A facile and effective synthesis of dinucleotide 5′-triphosphates. Bioorg Med Chem. 2007;15:6549-55.

18.   Qiu H, Wang Y. Probing adenosine nucleotide-binding proteins with an affinity-labeled nucleotide probe and mass spectrometry. Anal Chem. 2007;79:5547-56.

19.   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;85:3198-206.

20.   Olsen JV, de Godoy LMF, Li G, Macek B, Mortensen P, Pesch R, et al. Parts per Million Mass Accuracy on an Orbitrap Mass Spectrometer via Lock Mass Injection into a C-trap. Mol Cell Proteomics. 2005;4:2010-21.

21.   Taylor P, Nielsen PA, Trelle MB, Andersen MB, Vorm O, Moran MF, et al. Automated 2D peptide separation on a 1D Nano-LC-MS system. J Proteome Res. 2009;8:1610-6.

22.   Tabb DL, McDonald WH, Yates JR. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. J Proteome Res. 2002;1:21-6.

23.   Park SK, Venable JD, Xu T, Yates JR. A quantitative analysis software tool for mass spectrometry-based proteomics. Nat Methods. 2008;5:319-22.

24.   Ong S-E, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, et al. Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. Mol Cell Proteomics. 2002;1:376-86.

25.   Li X, Liu DR. DNA-Templated Organic Synthesis: Nature's Strategy for Controlling Chemical Reactivity Applied to Synthetic Molecules. Angewandte Chemie International Edition. 2004;43:4848-70.

26.   Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protocols. 2008;4:44-57.

27.   Negrutskii BS, El'skaya AV. Eukaryotic Translation Elongation Factor 1α: Structure, Expression, Functions, and Possible Role in Aminoacyl-tRNA Channeling. In: Kivie M, editor. Prog Nucleic Acid Res Mol Biol: Academic Press; 1998. p. 47-78.

28.   Schwartz D, Gygi SP. An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. Nat Biotech. 2005;23:1391-8.

29.   Saraste M, Sibbald PR, Wittinghofer A. The P-loop--a common motif in ATP- and GTP-binding proteins. Trends Biochem Sci. 1990;15:430-4.

30.   Cox AD, Der CJ. Ras Family Signaling: Therapeutic Targeting. Cancer Biol Ther. 2002;1:599-606.

31.   Preininger AM, Hamm HE. G Protein Signaling: Insights from New Structures. Sci STKE. 2004;2004:re3-.

32.   Malumbres M, Harlow E, Hunt T, Hunter T, Lahti JM, Manning G, et al. Cyclin-dependent kinases: a family portrait. Nat Cell Biol. 2009;11:1275-6.
33.   Vassilev LT, Tovar C, Chen S, Knezevic D, Zhao X, Sun H, et al. Selective small-molecule inhibitor reveals critical mitotic functions of human CDK1. Proceedings of the National Academy of Sciences. 2006;103:10660-5.

34.   Dai Y, Grant S. Cyclin-dependent kinase inhibitors. Curr Opin Pharmacol. 2003;3:362-70.

35.   Malumbres M, Sotillo Ro, Santamar1 a D, Galán J, Cerezo A, Ortega S, et al. Mammalian Cells Cycle without the D-Type Cyclin-Dependent Kinases Cdk4 and Cdk6. Cell. 2004;118:493-504.

36.   Kubota K, Anjum R, Yu Y, Kunz RC, Andersen JN, Kraus M, et al. Sensitive multiplexed analysis of kinase activities and activity-based kinase identification. Nat Biotech. 2009;27:933-40.

37.   Takaki T, Fukasawa K, Suzuki-Takahashi I, Semba K, Kitagawa M, Taya Y, et al. Preferences for Phosphorylation Sites in the Retinoblastoma Protein of D-Type Cyclin–Dependent Kinases, Cdk4 and Cdk6, In Vitro. J Biochem (Tokyo). 2005;137:381-6.

38.   Tsuchimoto D, Iyama T, Nonaka M, Abolhassani N, Ohta E, Sakumi K, et al. A comprehensive screening system for damaged nucleotide-binding proteins. Mutation Research/Genetic Toxicology and Environmental Mutagenesis. 2010;703:37-42.

39.   Yoon S-H, Hyun J-W, Choi J, Choi E-Y, Kim H-J, Lee S-J, et al. In vitro evidence for the recognition of 8-oxoGTP by Ras, a small GTP-binding protein. Biochem Biophys Res Commun. 2005;327:342-8.

Chapter 5

A Targeted Quantitative Proteomics Approach for Global Kinome Profiling of Cancer Cells and Tissues

Introduction

Protein phosphorylation, one of the most important types of post-translational modifications (PTMs), is catalyzed by protein kinases (collectively referred to as the kinome), which are encoded by over 500 genes in higher eukaryotes (1). Aberrant expression and/or activation/deactivation of kinases have been implicated as among the major mechanisms through which cancer cells escape normal physiological constraints of cell growth and survival (2). Additionally, dynamic kinome reprogramming has been found to be closely associated with resistance toward cancer chemotherapy (3). Owing to their crucial roles in cancer development, kinases have become one of the most intensively pursued enzyme superfamilies as drug targets for cancer chemotherapy and more than 130 distinct kinase inhibitors have been developed for phase 1-3 clinical trials (4). Recently, inhibitor potency and selectivity for more than 400 kinases have been reported, which provided a comprehensive target-inhibition profile for the majority of the human kinome (5-7). Therefore, the kinome-inhibitor interaction networks coupled with comprehensive profiling of global kinome expression and activity associated with certain

types of cancer could be invaluable for understanding the mechanisms of carcinogenesis and for designing rationally novel kinase-directed anti-cancer chemotherapies.

Unfortunately, currently there is no optimal strategy to profile the expression levels of the entire kinome at the protein level. Traditional methods for measuring kinase expression rely primarily on antibody-based immunoassays due to their high specificity and sensitivity (8). The immunoassays, however, are limited by the availability of high-quality antibodies; therefore, these methods are only useful for assessing a small number of kinases in low-throughput. The development of mass spectrometry (MS) instrumentation and bioinformatic tools enables the identification and quantification of a significant portion of the human proteome from complex samples (9). However, proteomic studies of global kinome by MS are still very challenging, which is largely attributed to the fact that, similar as other regulatory enzymes, protein kinases are generally expressed at low levels in cells (10). This analytical challenge is further aggravated in shotgun proteomics approach where even more complex mixtures of peptides instead of proteins from whole cell or tissue extracts are analyzed (11). Therefore, selective enrichment of protein kinases from cellular extracts is essential for the comprehensive identification and quantification of the global kinome.

Affinity columns immobilized with kinase inhibitors have been employed as capture ligands for the enrichment of kinases, and approximately 200 protein kinases could be identified and quantified by subsequent LC-MS/MS analyses (3, 10, 12). Recently, we and others reported the application of biotin-conjugated acyl nucleotide

probes for the enrichment and identification of kinases from complex protein mixtures (13-17). This enrichment technique, in combination with multi-dimensional LC-MS platform, facilitates the identification of approximately 200 protein kinases (15). Despite these advances, such large-scale kinome studies are often performed in the data-dependent acquisition (DDA) mode, where typically 10-20 most abundant ions found in MS are selected for fragmentation in MS/MS to enable peptide identification (18). Although this discovery-mode (or shotgun) proteomic approach provides the potential to uncover novel protein targets, sample complexity, together with inherent variation in automated peak selection, results in compromised sensitivity and reproducibility for protein quantification. As a result, only partially overlapping sets of proteins can be identified even from substantially similar samples (11). The inadequate sensitivity and reproducibility of these kinome detection strategies hamper their utility in biomarker discovery and clinical studies.

Targeted proteomics technique, which relies on multiple-reaction monitoring (MRM) on triple quadrupole mass spectrometers, has become increasingly used in quantitative proteomics studies (19). In the MRM mode, mass filtering of both the precursor and product ions is employed to provide high specificity for the quantification of target proteins. Additionally, this MRM-based targeted MS analysis permits rapid and continuous monitoring of specific ions of interest, which enhances the sensitivity for peptide detection by up to 100-fold as compared with MS analysis in DDA-based discovery mode (20). Therefore, the MRM-based targeted proteomic approach may

enable global kinome profiling with high specificity, sensitivity, throughput and reproducibility.

Here, we developed the first MRM-based platform to support the multiplexed, reproducible, and sensitive quantification of approximately 300 protein kinases in the human kinome. Aside from conventional MRM-based assay design, we selectively label and enrich kinases from complex human proteome prior to MRM analysis with the use of desthiobiotin-based Isotope-Coded ATP-affinity Probe (ICAP) (21) to attain high specificity and sensitivity. We demonstrated that this MRM-based kinome detection strategy coupled with ICAP reagent is applicable for clinical samples that are not amenable to metabolic labeling. Additionally, this MRM-based kinome assay is easily transferable between instruments and laboratories, rendering it a facile and universal strategy for global kinome detection.

Materials and Methods

Preparation of the Desthiobiotinylated Nucleotide Affinity Probe

The desthiobiotinylated nucleotide affinity probes were prepared according to previously published procedures with minor modifications (1, 2). In this regard, we first prepared light or heavy form of desthiobiotinyl-aminobutyric acid (desthiobiotin-C4) following previously published procedures. Isobutyl chloroformate (0.19 mL) was added to a solution containing 150 mg desthiobiotin, 15 mL DMF, and 0.38 mL tri-*n*-butylamine. After being incubated at room temperature for 10 min, the mixture was

185

slowly added to a suspension of 230 mg light or heavy γ-aminobutyric acid (Sigma) in DMF (7.5 mL) at 5 °C. After stirring at 5 °C for 2 h, the solvent was removed under reduced pressure and the crude product was dissolved in 5 mL water at 40 °C. The solution pH was adjusted to 2 with 2.0 N HCl, and the mixture was kept at 0 °C for 12 h. The desired desthiobiotin-C4 was precipitated out of solution. The precipitate was filtered, washed with water, and dried under vacuum.

To render nucleotide soluble in organic solvent, the commercially available sodium salt form of ATP was first converted to the tributylammonium form by passing the nucleotide through a cation-exchange column packed with Spectra/Gel IE 50×8 resin (40-75 μm) at 4 °C once. Fractions containing the tributylammonium form of ATP were collected and lyophilized. Desthiobiotin-C4 (8 mg), dissolved in a 1-mL solvent mixture of ice-cold dry CH_2Cl_2 and DMF (4:1, v/v), was mixed with tri-*n*-butylamine (11 μL) and ethyl chloroformate (5 μL). After stirring at 0 °C for 5 min, the mixture was stirred at room temperature under argon atmosphere for another 60 min. Tributylammonium form of ATP (50 mg), dissolved in a 1.25 mL solution of CH_2Cl_2 and DMF (4:1, v/v), was then added to the above reaction mixture. The reaction was continued at room temperature and under argon atmosphere for 18 h. The CH_2Cl_2 was then removed by argon purging for 10 min and the remaining 200 μL solution was directly subjected to HPLC purification using a YMC ODS-AQ column (4.8×250 mm, 120 Å in pore size, 5 μm in particle size, Waters). The flow rate was 0.8 mL/min, and a 45-min linear gradient of 0-30% acetonitrile in 50 mM triethylammonium acetate (pH 6.8) was used for the purification. A

UV detector was set at 265 nm to monitor the effluents. Appropriate HPLC fractions were pooled, lyophilized, and stored at -80 °C. The structures of the products were confirmed by ESI-MS analysis (3).

Protein lysate preparation from cultured cells and human lung tissues

HeLa-S3 cells were purchased from the National Cell Culture Center (Minneapolis, MN). IMR-90, WM-115 and WM-266-4 cells (ATCC; Manassas, VA) were cultured in Eagle's Minimum Essential Medium supplemented with 10% fetal bovine serum (FBS, Invitrogen, Carlsbad, CA) and penicillin (100 IU/mL). K562 cells (ATCC) were cultured in Iscove's Modified Dulbecco's Medium supplemented with 10% fetal bovine serum and penicillin. Cells were maintained in a humidified atmosphere with 5% CO_2 at 37 °C. Approximately 2×10^7 cells were harvested, washed with cold PBS for three times, and lysed in a 1-mL lysis buffer, which contained 0.7% CHAPS, 50 mM HEPES (pH 7.4), 0.5 mM EDTA, 100 mM NaCl, and 10 μL (1:100) protease inhibitor cocktail on ice for 30 min. The cell lysates were centrifuged at 16,000g at 4 °C for 30 min and the resulting supernatants were collected.

Human lung tumor tissues and paired adjacent normal tissues were purchased from National Disease Research Interchange (NDRI, Philadelphia, PA). Frozen tissue samples were thawed, minced with razor blades, and weighed to obtain about 0.2 g of tissue prior to protein extraction. In the presence of liquid nitrogen, tissue samples were grounded into fine powders and then homogenized on ice with the same cell lysis buffer described above in a Dounce homogenizer for 2-3 min. The same experimental

procedures and conditions were applied for the protein extracts of all the examined tissue samples to achieve comparable extraction efficiency. Cell debris was removed by centrifugation at 16,000 g at 4 °C for 30 min.

Approximately 1 mL of the resulting protein extracts from either cultured cells or human tissues were subjected to gel filtration separation using NAP-25 columns (Amersham Biosciences) to remove free endogenous nucleotides. Cell lysates were eluted into a 2 mL buffer, containing 50 mM HEPES (pH 7.4), 75 mM NaCl, and 5% glycerol. The amounts of proteins in the lysates were quantified using Quick Start Bradford Protein Assay (Bio-Rad, Hercules, CA) and stored at –80 °C. Prior to the labeling reaction, $MgCl_2$, $MnCl_2$, and $CaCl_2$ were added to the concentrated cell lysate until their final concentrations reached 50, 5, and 5 mM, respectively.

Kinase labeling with nucleotide affinity probe followed by affinity purification

Approximately 1 mg cell lysate was treated separately with light and heavy labeled desthiobiotin-ATP affinity probes at a final concentration of 100 μM. Labeling reactions were carried out with gentle shaking at room temperature for 1.5 h. After the reaction, the remaining probes in the cell lysates were removed by buffer exchange with 25 mM NH_4HCO_3 (pH 8.5) using Amicon Ultra-4 filters (10,000 NMWL, Millipore).

After addition of 8 M urea for protein denaturation, and dithiothreitol and iodoacetamide for cysteine reduction and alkylation, the labeled proteins were digested with modified sequencing-grade trypsin (Roche Applied Science) at an enzyme/substrate ratio of 1:100 in 25 mM NH_4HCO_3 (pH 8.5) at 37 °C for overnight. The peptide mixture

was subsequently dried in a Speed-vac and redissolved in 1 mL PBS buffer (100 mM potassium phosphate and 0.15 M NaCl, pH 7.5), to which solution was subsequently added 200 μL avidin-agarose resin (Sigma-Aldrich). The mixture was then incubated at 25 °C for 1 hr with gentle shaking. The agarose resin was washed sequentially with 3 mL PBS buffer and 3 mL H_2O to remove unbound peptides, and the desthiobiotin-conjugated peptides were subsequently eluted with 1% TFA in CH_3CN/H_2O (7:3, v/v) at 65°C. Enriched peptides were first analyzed on an LTQ-Orbitrap Velos mass spectrometer for discovery-mode proteomic analysis.

On-line 2D-LC-MS/MS analysis on an LTQ-Orbitrap Velos for discovery mode analysis

On-line 2D-LC-MS/MS analysis was performed on an LTQ-Orbitrap Velos mass spectrometer equipped with a nanoelectrospray ionization source (Thermo Fisher Scientific, San Jose, CA). The fully automated 7-cycle on-line two-dimensional LC-MS/MS was set up as described (4). Briefly, samples were automatically loaded from a 48-well microplate autosampler using an EASY-nLC II system (Thermo) at 3 μL/min onto a biphasic precolumn (150 μm i.d.) comprised of a 3.5-cm column packed with 5 μm C18 120 Å reversed-phase material (ReproSil-Pur 120 C18-AQ, Dr. Maisch) and 3.5-cm column packed with Luna 5 μm SCX 100 Å strong cation exchange resin (Phenomenex, Torrance, CA). The biphasic trapping column was connected to a 20 cm fused silica analytical column (PicoTip Emitter, New Objective, 75 μm i.d.) with 3 μm C18 beads (ReproSil-Pur 120 C18-AQ, Dr. Maisch). Ammonium acetate at

189

concentrations of 0, 25, 50, 75, 125, 200 and 500 mM were then sequentially injected using a 48-well autosampler from the sample vial to elute bound peptides from precolumn to the analytical column with reversed-phase separation. The peptides were then separated with a 180-min linear gradient of 2-35% acetonitrile in 0.1% formic acid and at a flow rate of 250 nL/min. The LTQ-Orbitrap Velos was operated in a data-dependent scan mode. Full-scan mass spectra were acquired using the Orbitrap analyzer with a resolution of 60,000 with lock mass option enabled for the ion of *m/z* 445.120025 (5). Up to 20 most abundant ions found in MS with charge state ≥2 were sequentially isolated and sequenced in the linear ion trap with a normalized collision energy of 35, an activation Q value of 0.25 and an activation time of 10 ms.

Data processing and analysis for discovery mode results.

The raw data were first converted to mzXML files and DTA files using ReAdW (http://sourceforge.net/projects/sashimi/files/) and MzXML2Search (http://tools.proteomecenter.org/wiki/index.php?title=Software:MzXML2Search) programs, respectively. Bioworks 3.2 was used for protein identification by searching the DTA files against the human IPI protein database version 3.68 (87,062 entries) and its reversed complement. Initial precursor mass tolerance of 10 ppm and fragment mass deviation of 0.8 Th were set as the search criteria. The maximum number of miss-cleavages for trypsin was set at two per peptide. Cysteine carbamidomethylation was considered as a fixed modification, whereas methionine oxidation and light or heavy desthiobiotinylation of lysine (+281.17394 Da and +287.21160 Da, respectively) were

included as variable modifications. The search results were then filtered with DTASelect (6) to achieve a peptide false discovery rate of 1%. Census was employed for peptide and protein quantification (7). Extracted-ion chromatograms were first generated for peptide ions based on their *m/z* values and peptide intensity ratios were subsequently calculated in Census from peak areas found in each pair of extracted-ion chromatograms.

MRM spectral library generation for the kinome peptides

Skyline (version 1.4.0.4421) (8) was used to generate the spectral libraries for kinome peptides. All raw files generated in discovery-mode proteomic analysis on an LTQ-Orbitrap Velos were searched by Maxquant (Version 1.2.2.5) (9) against the human IPI protein database version 3.68 (87,062 entries) using the same search parameters as described above. The resulting MS/MS files were then imported into Skyline and filtered with a threshold score of 0.9. An interactive Skyline spectral library file containing tandem mass spectra of all desthiobiotin-labeled kinome peptides was then generated.

Retention time (RT) extraction and iRT calculation for targeted peptides

To calculate the iRT score (10) for each kinome peptide in the MRM kinome library, 10 peptides derived from the tryptic digestion mixture of bovine serum albumin (BSA) were selected to constitute the reference peptides for a new iRT scale. iRT of these 10 BSA standard peptides were calculated using the empirically measured RT from shotgun LC-MS/MS analysis by setting iRT scores of peptides AEFVEVTK and DAFLGSFLYEYSR as 0 and 100, respectively. The BSA peptide mixture was then added to desthiobiotin-labeled kinase peptide mixture from IMR-90 and K562 cell lysates

and measured by LC-MS/MS on the Orbitrap system with a regular 130-min linear gradient. RTs were extracted for all BSA standard peptides as well as kinase peptides using the Skyline MS1 filtering workflow (11). The transformed iRT for all the newly identified kinase peptides were calculated based on linear regression of iRT and empirically measured RT of peptides with previously determined iRT score. With the accumulation of iRT score for targeted kinase peptides, it is not necessary to include the BSA standard mixture for RT extraction and iRT calculation. Any common kinase peptides that yield a regression with correlation of 0.99 or higher were used as standard peptides to calculate the new iRT score for newly identified kinase peptides by Skyline.

LC-MRM analysis

All LC-MRM experiments were carried out on a TSQ Vantage triple quadrupole mass spectrometer equipped with a nanoelectrospray ionization source coupled to an Accela HPLC with customized split-flow configuration or an EASY-nLC II system (Thermo Scientific). Samples were automatically loaded onto a 4-cm trapping column (150 μm i.d.) packed with 5 μm 120 Å reversed-phase C18 material (ReproSil-Pur 120 C18-AQ, Dr. Maisch) at 3 μL/min. The trapping column was coupled to a 20-cm fused silica analytical column (PicoTip Emitter, New Objective, 75 μm i.d.) with 3 μm C18 beads (ReproSil-Pur 120 C18-AQ, Dr. Maisch). The peptides were then separated with a 130-min linear gradient of 2-35% acetonitrile in 0.1% formic acid and at a flow rate of 250 nL/min. The spray voltage was 1.9 kV. Q1 and Q3 resolutions were 0.7 Da and the cycle time was 4-5 s depending on the maximum number of transitions per cycle.

To enable automated, multiplexed MRM analyses, we first generated an interactive Skyline spectral library file containing tandem mass spectra of all desthiobiotin-conjugated kinase peptides along with their normalized retention time score (iRT) (22), which were acquired from discovery mode analysis on an LTQ Orbitrap Velos mass spectrometer, using Skyline (version 1.4.0.4421, see Supplementary Materials) (23). Collision energies were calculated using a linear equation that was specific to the TSQ Vantage instrument according to the Skyline default setting. BSA standard mixtures were analyzed in unscheduled MRM-mode prior to the analysis of the enriched desthiobiotin-conjugated peptides. The linear predictor of empirical RT from iRT for targeted kinase peptides was then determined by the linear regression of RTs of BSA standard peptides obtained for the current chromatography setup. This linear predictor was rechecked between every four MRM sample analyses by injecting another BSA standard mixture. In identification experiment, four transitions were monitored for each light desthiobiotin-labeled peptide, whereas three pairs of transitions were monitored for each light/heavy desthiobiotin-labeled peptide pair in quantification experiments. These targeted transitions were monitored in four separate injections for each sample in scheduled MRM mode with a retention time window of 10 min.

All raw files were processed using Skyline (version 1.4.0.4421) for generation of extracted ion chromatograms and peak integration. The targeted peptides were first manually checked to ensure the overlaid chromatographic profiles of multiple fragment ions derived from light and heavy forms of the same peptide. The data were then

processed to ensure that the distribution of the relative intensities of multiple transitions associated with the same precursor ion correlates with the theoretical distribution from kinome MS/MS spectral library entry. Finally, the remaining transitions from both forward and reverse ICAP labeling experiments were exported and filtered using AuDIT algorithm (24), where T-test was performed to evaluate whether the relative product ion intensities of the light-labeled peptides to the heavy labeled peptides are similar. The significant difference (P value <0.1) of the relative product ion intensities indicates the presence of interference or imprecision of the corresponding transition and these transitions were therefore discarded (24). The sum of peak area from all transitions of light or heavy labeled peptides was used for quantification.

Western blot

WM-115 and WM-266-4 cells were cultured in 75 cm^2 cell culture flasks until 70-80% confluency and lysed as described above. The concentrations of the resulting protein lysates were determined by Bradford Assay (Bio-Rad). The whole cell lysate for each sample (20 μg) was denatured by boiling in Laemmli loading buffer and then electrophoresed on a 12% SDS-PAGE gel with a 4% stacking gel. Subsequently, proteins were transferred onto the nitrocellulose membrane at 4 °C overnight. The resulting membrane was blocked with PBS-T (PBS with 0.1% Tween 20) containing 5% milk (Bio-Rad) at 4 °C for 6 h. Next, the membrane was incubated with primary antibody at 4 °C overnight and then secondary antibody at room temperature for 1 h.

194

After thorough washing with PBS-T buffer, the HRP signals were detected with Pierce ECL Western Blotting Substrate (Thermo).

EphA2 (D4A2) XP® Rabbit mAb (Cell Signaling # 6997, with dilution ratio 1:5000) and CDK2 (78B2) Rabbit mAb (Cell Signaling #2546, with dilution ratio 1:1000) were employed as primary antibodies, and horseradish peroxidase-conjugated anti-rabbit IgG were used as secondary antibodies. Membranes were also probed with anti-actin antibody (1:10,000) to confirm equal protein loading.

MTT cell proliferation assay

For MTT assay, WM-115 and WM-266-4 cells were placed in a 96 well plate at a density of 4×10^3 cells per well with 100 μL EMEM medium. After 24 h of pre-incubation, cells were treated with dasatinib or flavopiridol (Santa Cruz Biotechnology) at concentrations ranging from 32 nM to 1 μM in serum-free EMEM medium for 24 h. The medium was then supplemented with 10 μL of MTT [3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyltetrazolium bromide, Roche, Basel, Switzerland] reagent (0.5 g/L) to each well and incubated at 37 °C for 4 h. After this incubation period, purple formazan salt crystals were formed. Thereafter 100 μL of solubilizing solution was added to each well. The plates were incubated at 37 °C overnight to dissolve the formazan crystals. The solubilized formazan product was spectrophotometrically quantified using the Victor 2 plate reader (Perkin Elmer, Waltham, MA). Absorbance readings were performed at 570 nm with reference of 630 nm.

Results and Discussions

1. Global kinome profiling based on isotope-coded ATP-affinity probe (ICAP) and MRM

The MRM-based kinome assay is built upon our previously described ICAP-based strategy for the simultaneous enrichment and isotopic labeling of ATP-binding proteins in cell lysates (21). The ICAP reagent harbors three components (Figure 5.1A), namely, a binding moiety (ATP), an isotope-coded linker present in light (contains six hydrogens) or heavy form (contains six deuterons), and an enrichment moiety (i.e. desthiobiotin), where the former two are conjugated through an acyl phosphate linkage. Upon binding to kinases, the acyl-phosphate component of the ICAP reacts with the ε-amino group of the P-loop lysine residue in kinases to yield a stable amide bond, which results in the covalent attachment of desthiobiotin together with light or heavy isotope-coded linker to the lysine residue (Figure 5.2). Therefore, the ICAP labeling of a kinase relies on its expression level and ATP binding affinity, which renders the method useful for profiling the expression and ATP-binding affinities of kinases at the global proteome scale.

The general experimental procedures for the MRM-based kinome assay with the use of ICAP encompass the following steps (Figure 5.1B): (1) Kinases from protein samples representing two experimental states are labeled individually with the isotopically light and heavy forms of the ATP affinity probe. (2) The two protein samples are combined and digested with trypsin, and the resultant light/heavy

Figure 5.1 (A) The structure of the isotope-coded ATP affinity probe (ICAP). (B) A schematic diagram showing the general workflow for multiple-reaction monitoring (MRM) analysis of global kinome using isotope-coded ATP affinity probe (ICAP).

(A)
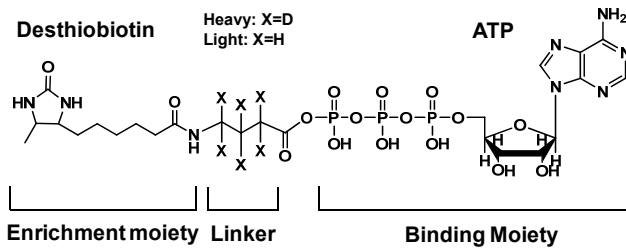
(B)
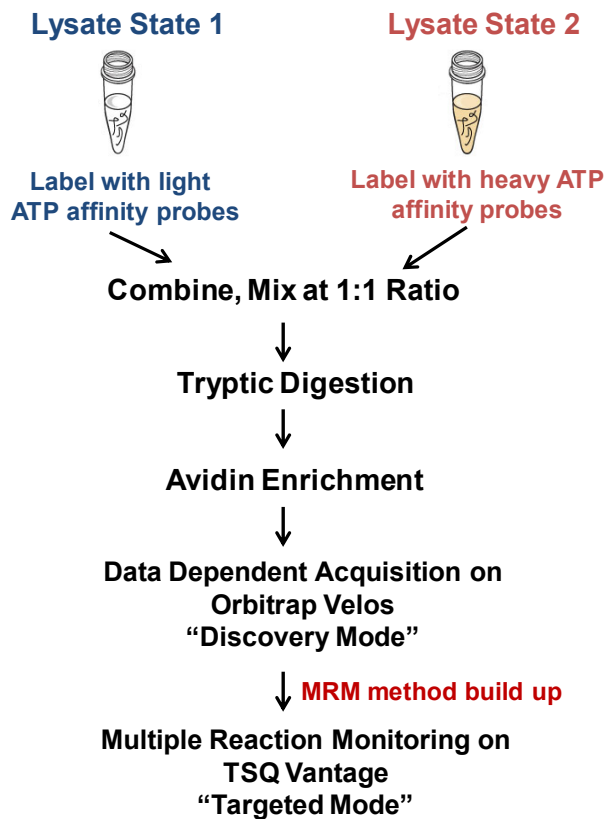
Figure 5.2 A schematic diagram showing the reaction between isotope-coded ATP affinity probe with an ATP-binding protein.
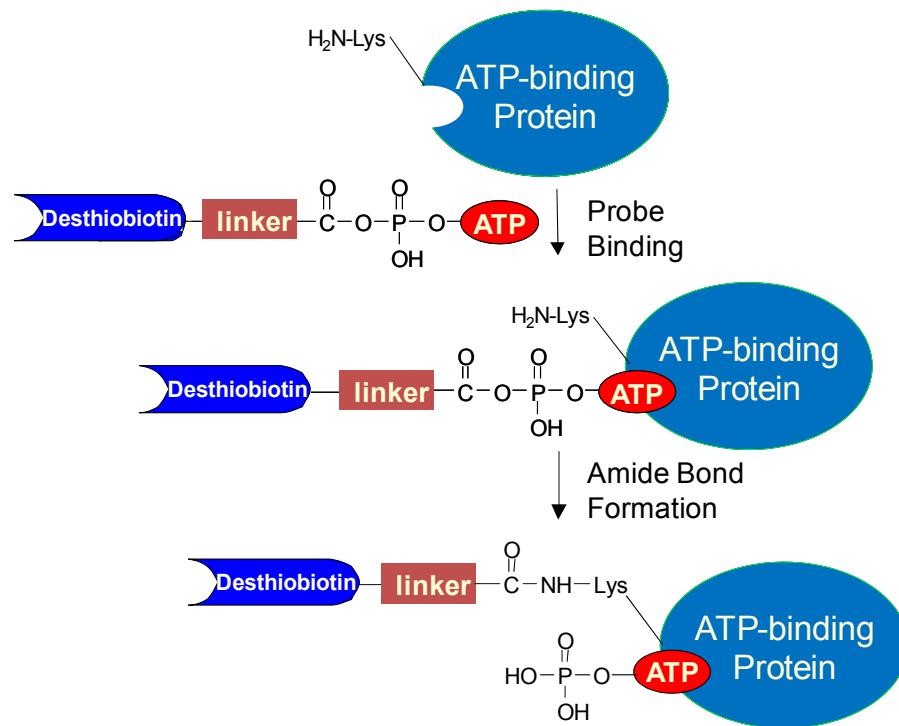
desthiobiotin-labeled peptide pairs are enriched with avidin agarose. (3) The affinity-purified peptides with the desthiobiotin tag are analyzed by LC-MS/MS in DDA mode to generate tandem mass spectra of interested kinase peptides, which allowed for the construction of MRM kinome library. (4) Quantitative analysis of global kinome was performed on a triple quadruple mass spectrometer in MRM mode using the established MRM kinome library.

2. Development of an MRM assay for human kinome profiling

The design of MRM-based kinome assay requires *a priori* tandem mass spectral information for the interested desthiobiotin-modified kinase peptides (20). The primary sources for building the MRM library of human kinome were large-scale discovery-based shotgun MS experiments conducted in our laboratory. Human kinases from whole cell lysates of K562, IMR-90, HeLa-S3, Jurkat-T, WM-115 and WM-266-4 cells were enriched using either light or heavy labeled ICAP reagents and the resulting desthiobiotin-labeled peptides were analyzed on an LTQ Orbitrap Velos mass spectrometer in DDA mode, which together included more than 100 LC-MS/MS runs. The tandem mass spectra of all desthiobiotin-modified peptides from human kinases together with their retention time information were imported into Skyline to establish a primary kinome library. A maximum of five peptides were used for each protein. Transitions were chosen from the 3 or 4 most abundant y-ions based on reference spectra acquired from the shotgun approach.

Peptide selection. Owing to its relatively high reactivity, the ICAP probe may react with, aside from the lysine residue(s) located at the ATP-binding site, other lysine residues in kinases or other proteins through electrostatic interactions (15, 21). The probe labeling emanating from such non-specific interactions does not reflect the ATP-binding affinities of the kinases. Moreover, the probe-labeled peptides stemming from the non-specific interaction are generally less abundant than those arising from specific nucleotide binding, which may compromise the sensitivity for kinase quantification. Thus, we reason that the quantification of kinome based on the pre-selected representative peptides would generate more reliable results than the quantification based on all the peptides identified in the DDA mode. To improve the specificity of our MRM-based kinome assay, these targeted peptides in our final kinome library were further divided into two categories depending on the local amino acid sequences surrounding the desthiobiotin-modified lysine. In the category with the highest confidence (class I), target peptides represent the confirmed ATP-binding sites on kinases and must satisfy one of the following two criteria: (A) Peptides cover at least one of the three ATP-binding motifs discovered in our previous studies, i.e., HRDxKxxN, VAxK or GxxxxGK (15, 21); (B) Peptides contain the ATP-binding sites discovered in our previous ATP-affinity profiling assay even though they do not reside in any of the conserved motif sequences (21). For instance, although FLSGLELVK#QGAEAR (K# represents the probe-labeled lysine) from TP53-regulating kinase, does not belong to any of the aforementioned conserved motifs, it displays strong ATP binding affinity in our previous

report (21). The class-II peptides include those that were identified at least five times in our shotgun proteomics experiment. Although these peptides do not reside in any of the conserved ATP binding sites and their involvement in ATP binding remains undefined, their high frequencies of identification in shotgun proteomic studies render these peptides good candidates for determining the quantities of their corresponding kinases. Our current MRM peptide list includes 386 peptides, with 265 and 121 being class-I and class-II peptides, respectively.

Peptide uniqueness. In our enrichment strategy for human kinases using ICAP reagents, a large portion of protein kinases were labeled on the reactive lysine close to the ATP-binding sites, which share highly conserved sequence (1). Therefore, some targeted peptides in the final MRM peptide list may be assigned to multiple protein kinases. We manually inspected the uniqueness of each targeted peptide using Skyline. Among the 386 peptides in the library, 320 (83%) are unique and belong only to single protein sequences, whereas the rest 66 (17%) can be assigned to multiple kinases. Accordingly, these 386 peptides can be mapped to 313 human kinases and 242 (77%) of these kinases contained at least one unique targeted peptide to unambiguously identify the targeted protein isoform. Unlike the cross-activity of antibody-based immunoblots (25), the advantage of this MRM kinome assay lies in its capability in identifying unequivocally the potential cross-talk targets. Our current MRM kinome library contains a total of 386 peptides from 313 human kinases, including 292 protein kinases, 14 lipid kinases, and 7 metabolic kinases bearing GxxxxGK motif. Placement of these protein kinases to the

201

human kinome map (1) revealed that our kinome library spreads over all major categories of the human kinome (Figure 5.3A).

Retention time prediction by iRT. Our kinome peptide list encompasses 386 peptides, which involves the monitoring of more than 2000 MRM transitions in quantitative measurements. To achieve this level of multiplexed detection, it is essential to perform scheduled MRM analysis in which the mass spectrometer is programmed to detect only a limited number of peptides in pre-defined retention time windows. Therefore, accurate retention time (RT) prediction for these kinase peptides becomes necessary for our multiplexed MRM-based kinome assays. To this end, we calculated the normalized retention time (iRT) for each peptide on our target list following a previously described method (22). Based on previous retention time information of targeted peptides on Orbitrap Velos coupled with EASY-nLC II system, we used 10 BSA peptides as standards to successfully convert empirically determined retention times of 94% (362 out of 386) of targeted peptides into normalized iRT scores, which reflect their conserved elution order. As shown in Figure 5.3B, peptides from human kinases detected in one of our LC-MS/MS runs on the Orbitrap system show linear correlation ($R^2 > 0.999$) between the assigned iRT values and empirically measured RT values following the Skyline MS1 filter workflow. As reported previously (22), we found that this iRT value is very stable, and the iRT score and empirically measured RT exhibit a linear relationship across different LC configurations, including online 2D-LC and 1D-LC configurations with different column lengths (Figure 5.4). Moreover, since the iRT value represents the

Figure 5.3 Targeted protein kinases mapped in the dendrogram of the human kinome (A) and linearity of iRT vs.measured RT on different instruments and experiment platforms (B-C). (B) iRT values predict measured RT in on-line 2D experiment on Orbitrap Velos (180 min linear gradient) with a very high correlation ($R^2 = 0.999$). (C) iRT values predict measured RT in MRM experiment on TSQ vantage (130 min linear gradient) with a very high correlation ($R^2 = 0.999$).
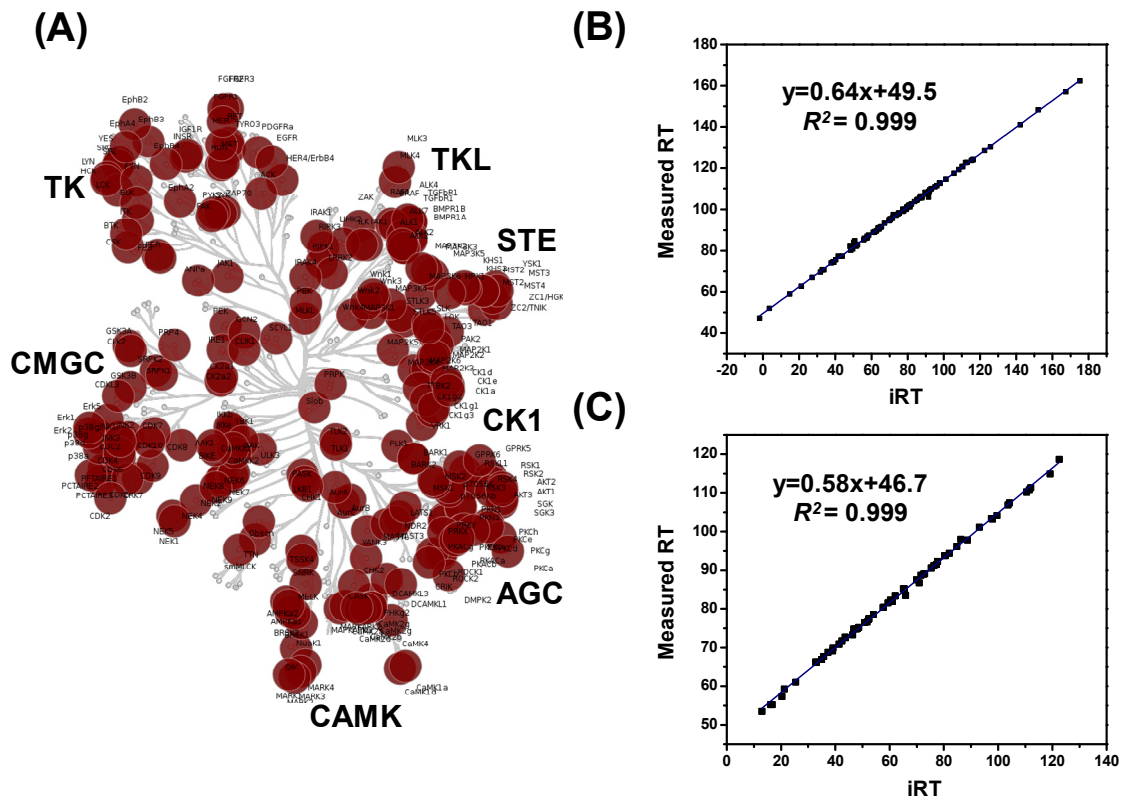
Figure 5.4 Linearity of iRT and measured RT on different instrument and experiment platforms. (A) iRT values predict measured RT in 1D LC-MS/MS experiment using Jurkat-T cell lysates on an Orbitrap Velos (180 min linear gradient) with a very high correlation coefficient (R^2=0.998); (B) iRT values predict measured RT in online 2D LC-MS/MS experiment using K562 cell lysates on an Orbitrap Velos (180 min linear gradient) with a very high correlation coefficient (R^2=0.998) (C) iRT values predict measured RT in 1D LC-MS/MS experiment using HeLa-S3 cell lysates after offline HPLC separation on Orbitrap Velos (130 min linear gradient) with a very high correlation coefficient (R^2=0.997).

(A)
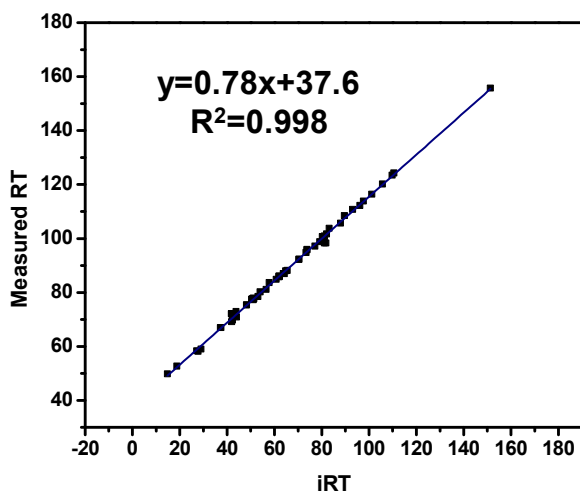
(B)
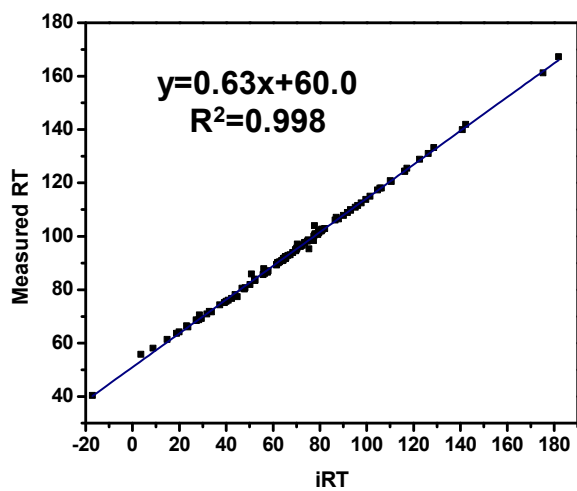
(C)
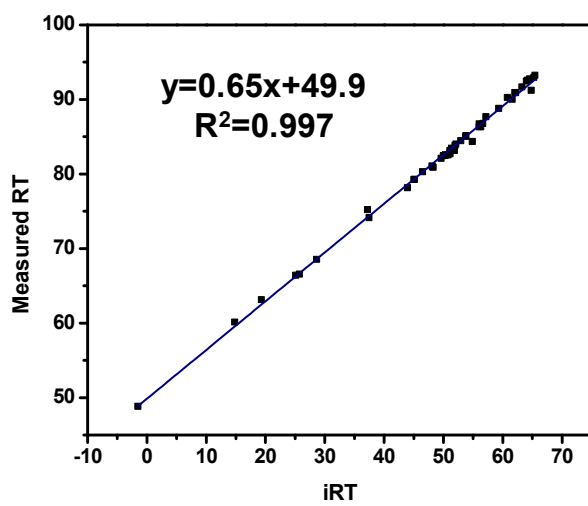
inherent attribute of the hydrophobicity of a peptide, it can also serve as another parameter for the validation of the quantification results from MRM assay, where any outliners deviated from the linear plot of iRT vs. measured RT could be attributed to false-positive detection.

Retention time calibration by on-the-fly correction. Although iRT provide a excellent solution for accurate retention time prediction, the actual RT deviation is also affected by the variance that occurs as systematic shifts from run to run. This variation is mainly attributed to differences in sample composition, nonreproducible sample loading, unstable column temperature or random air bubble generated by the liquid chromatographic pump. Any of these shifts in retention time can result in missing targeted peaks from the scheduled time window and lead to lower overall analytical precision. Additionally, this technical variance cannot be corrected by any pre-run RT prediction including iRT technology and therefore real-time RT calibration methods during the LC runs are highly desirable to minimize the effect of this RT shift from run to run.

To address these challenges, we incorporated on-the-fly RT calibration technique developed by Thermo Fisher into our MRM analysis workflow. In on-the-fly RT calibration process, any spiked-in standard peptide can be chosen as a reference compound to predict additional variance during acquisition. With changes in chromatographic conditions, the retention time shift for a reference compound between the expected retention time and the observed retention time is then captured. If a single

reference peak is used, the captured time shift is used as an offset. If multiple reference peaks are used, then a linear fit is generated during the LC-run using the last two observed reference peaks. For example, if three reference compounds are used that elute at time 1, time 2, and time 3, the correction is generated first by using the reference compounds at time 1 and time 2. As soon as the reference compound at time 3 is eluted, the instrument ignores the reference peak of time 1 and instead uses the reference compounds of time 2 and time 3 for the correction. The software then automatically calculates an offset and a slope correction to the running retention time windows for the later eluting compounds.

We test if on-the-fly RT calibration technique can help to minimize the effect of LC technical variance by measure a series of BSA standard peptides. We first predicted the retention time of BSA standard peptides and set RT windows of 6 min according to previous unscheduled BSA analysis runs. To introduce the RT shift, we manually delay the chromatographic gradient by 7 min. Then we setup the MRM method by using the same scheduled RT windows of 6min for all the target BSA peptides except using the K.DLGEEHFK.G as the reference peptide with RT windows of 18 min for on-the-fly RT calibration. All BSA target peptides were successfully detected although there were generally 5-7 min RT shifts. For example, as shown in Figure 5.5, although around 6 min RT shifts for peptides K.YLYEIAR.R and K.LGEYGFQNALIVR.Y were induced by manually change of gradient, the MS analyzer still successfully capture these two peptides by real-time adjustment of schedule detection window. All of target BSA

Figure 5.5 MRM-based kinome assay exhibits better sensitivity, reproducibility compared to data-dependant shotgun proteomics. (A) The Venn diagrams showing the overlap of quantified kinases from cell lysates of two melanoma cells obtained from MRM analysis and shotgun proteomics. (B) The Venn diagrams showing the overlap of quantified kinase peptides from cell lysates of two melanoma cells by MRM analysis and shotgun proteomics. (C) The Venn diagrams showing the overlap of quantified kinase peptides from cell lysates of two melanoma cells by replicates of MRM analysis. (D) The Venn diagrams showing the overlap of quantified kinase peptides from cell lysates of two melanoma cells by replicates of shotgun proteomics.
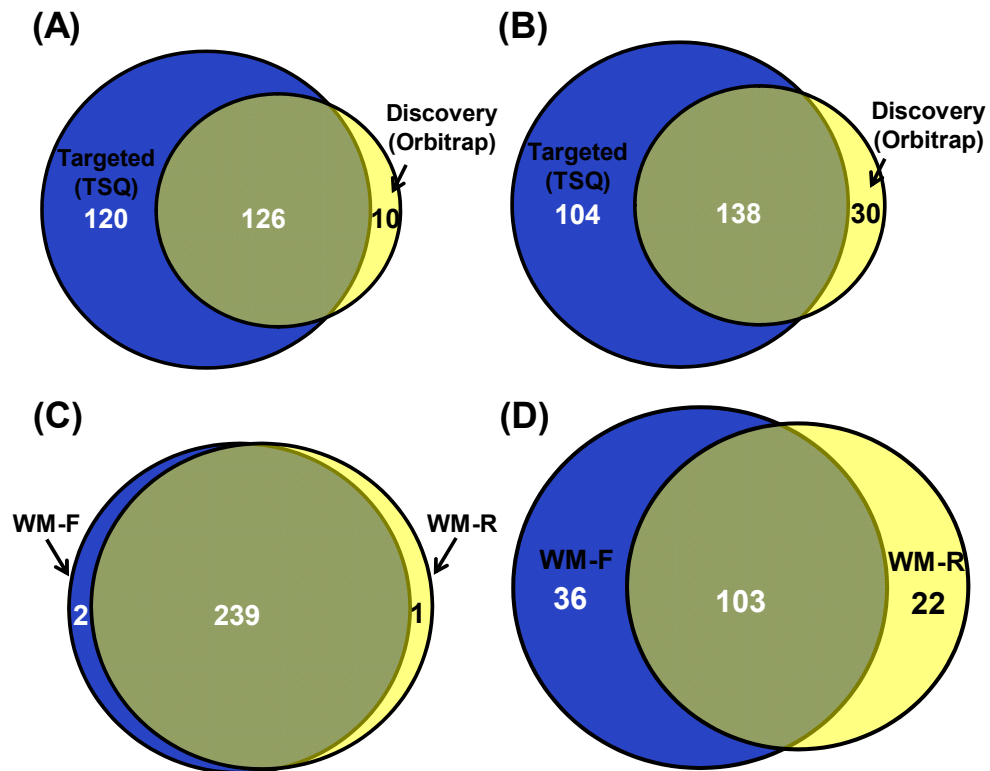
peptides would be missed using the pre-set scheduled MRM windows without the on-the-fly RT calibration. Therefore, we demonstrated that the on-the-fly calibration method can effectively minimize the peak missing induced by technical variance of chromatographic system. This allows shorter scheduled time windows than iRT alone, resulting in less temporal overlap and greater dwell times for each SRM transition.

Based on the above MRM-based assay design, we set out to test the sensitivity and reproducibility of this kinome detection assay using lysates of IMR-90 cells, where four transitions were monitored for each of the 386 peptides on the target list. This MRM-based kinome analysis led to the detection of 242 peptides from the kinase list corresponding to 227 human kinases, which include 210 protein kinases covering all major categories of the human kinome. This result demonstrates the superior sensitivity of the MRM-based kinome assay. Moreover, we found that the linearity between the iRT value and measured RT remains for each peptide detected in this MRM assay, indicating that our MRM-based kinome assay can be easily transferred among different instrument configurations and laboratories (Figure 5.3C).

3. **Quantitative profiling of the global kinomes of WM-115 and WM-266-4 human melanoma cells**

We next applied the MRM-based kinome assay to assess the differential expression of kinases in a pair of human melanoma cell lines, WM-115 and WM-266-4, which were initially derived from the primary and metastatic melanoma sites of the same patient (26). These two cell lines are particularly interesting because dasatinib, a tyrosine

kinase inhibitor, was found to inhibit the growth and reduce the migration and invasion of WM-115, but not WM-266-4 cells (27). However, conventional immunoblot assay for a limited number of kinases did not reveal the target kinase(s) conferring the distinct sensitivities of these two cell lines (27). We reason that quantitative global kinome profiling of these two cell lines may uncover the target kinases of dasatinib in WM-115 cells. In addition, such a study may provide important new knowledge about kinome reprogramming during cancer progression. To this end, we employed our ICAP probes, along with the aforementioned MRM-based assay, and assessed the differential expression of kinases in these two cell lines (Figure 5.1). For comparison, we also injected the same samples into the Orbitrap Velos instrument for shotgun proteomics analysis.

A total of 246 kinases in the library, including 228 protein kinases, 14 lipid kinases, and 4 other metabolic kinases, were quantified in MRM-based targeted kinome assay, whereas only 136 protein kinases were quantified by shotgun proteomics strategy even with multi-dimensional LC separation (Figure 5.5A). In addition, 10 and 120 protein kinases were only quantified in DDA-based discovery mode and MRM-based targeted mode, respectively. Accordingly, from the total list of 386 kinase peptides, 242 were successfully quantified by the MRM method, whereas only 168 were quantified from shotgun proteomics analysis (Figure 5.5B). The much better kinome coverage provided by the MRM assay than the shotgun proteomic approach is attributed to the superior sensitivity of the former method. In this vein, our ATP probe enriches kinases

along with other ATP-binding proteins. Therefore, the precursor ion intensities for the probe-labeled peptides from some low-abundance kinases are likely to be too weak to trigger collisional activation, thereby preventing their identification and quantification, in the DDA-based shotgun proteomics approach. Monitoring constantly the target peptides in the MRM mode, however, provided much improved sensitivity that is needed for the quantification of kinases of low abundance.

Aside from inferior sensitivity, inadequate reproducibility is another major disadvantage for shotgun proteomics. Along this line, 99% (239 out of 242) of the kinase peptides were successfully quantified in both forward and reverse labeling experiments in MRM-based targeted analysis, whereas only 64% (103 out of 161) was quantified with the same samples using the DDA-based shotgun proteomics approach (Figures 5.5C&D). Thus, the MRM-based targeted approach provided much better reproducibility than the DDA-based shotgun approach for kinome profiling analysis in which multiple experimental replicates are generally required.

Our results also revealed that the quantification accuracy for the MRM method is better than the shotgun approach. As displayed in Figure 5.6A, we observed a very good correlation (R^2=0.92) for the quantification results of 71 peptides of high abundance (peak intensity in MRM assay >100000) commonly quantified by these two modes of analyses. Nevertheless, there are a few discrepant results obtained by these two methods for some low-abundance kinase peptides (peak intensity <100,000 in MRM assay). For example, a unique peptide ETSVLAAAK#VIDTK from SLK kinase exhibited

Figure 5.6 (A) Linear regression comparing quantification results of abundant kinase peptides from two melanoma cells obtained by MRM assay and shotgun proteomics analysis. (B) Quantitative results by MRM assay for peptide ETSVLAAAK#VIDTK from SLK kinase: Extracted-ion chromatograms for three transitions monitored for light-labeled (Red) and heavy-labeled (Blue) peptides ETSVLAAAK#VIDTK in both forward (Left) and reverse labeling reaction (Middle); The consistent distribution of the peak area observed for each monitored transition from light- and heavy- labeled peptides in both forward and reverse labeling reaction along with the theoretical distribution derived from MS/MS spectra store in MRM kinome library (Right).

212
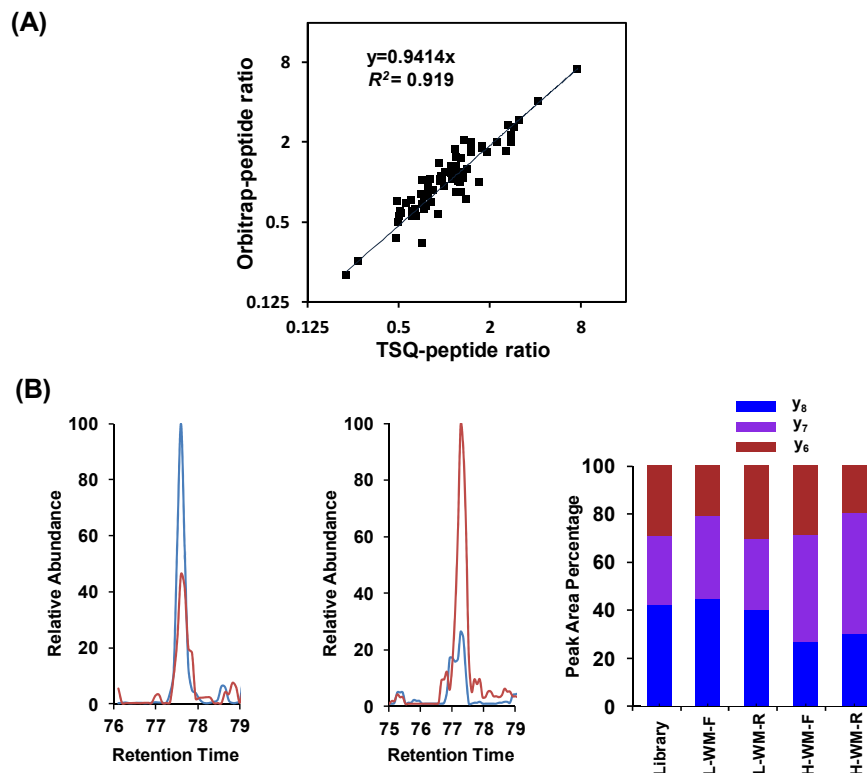
conflicting results based on these two modes of analysis ($Ratio_{MRM}$=0.47; $Ratio_{DDA}$=2.63). In MRM assay, the peptide shows consistent ratios in forward and reverse experiments (Figure 5.6B), which revealed the upregulation of the SLK kinase in WM-266-4 cells. Additionally, distribution of the peak area observed for each monitored transition from light- and heavy-labeled peptides was consistent with the theoretical distribution derived from MS/MS spectra stored in MRM kinome library, further validating the identity of this unique peptide. In DDA-based shotgun proteomic analysis, we only identified this peptide once in the reverse experiment, and the chromatogram displays very poor signal-to-noise ratio, suggesting the lack of reliability for the quantification of this peptide (Figure 5.7). Moreover, the consistent quantification results from both MRM and DDA analyses ($Ratio_{MRM}$=0.50; $Ratio_{DDA}$=0.56) for another peptide DLK#AGNILFTLDGDIK from the same kinase further substantiated our MRM quantification results for the peptide ETSVLAAAK#VIDTK. Likewise, the MRM method offers more robust quantification for the peptide FLEDDTSDPTYTSALGGK#IPIR from Ephrin type-B receptor 2 than DDA mode (Figure 5.8). Together, the MRM-based targeted analysis offers better sensitivity, reproducibility and accuracy than DDA-based shotgun proteomics for kinome profiling.

By combining MRM and DDA analyses, we quantified a total of 256 kinases in WM-115 and WM-266-4 cells. Kinases from all seven major kinome groups as well as other and atypical kinases were covered in our analysis. Although a large portion of the kinome is similar in WM-115 and WM-266-4 cells, the two lines of melanoma cells

Figure 5.7 Quantitative results by DDA shotgun proteomics analysis for peptide ETSVLAAAK#VIDTK from SLK kinase.
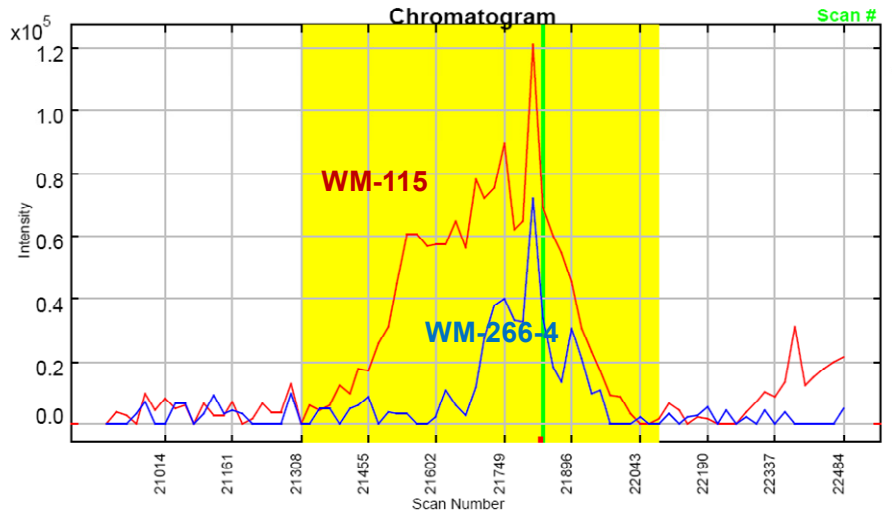
Figure 5.8 (A, B, C) Quantitative results obtained from MRM assay for peptide FLEDDTSDPTYTSALGGK#IPIR from Ephrin type-B receptor 2:   Extracted ion chromatograms for three transitions monitored for light- (Red) and heavy-labeled (Blue) peptides in forward (A) and reverse (B) labeling reactions; (C) The distribution of the peak areas observed for each monitored transition from light- and heavy- labeled peptides in forward and reverse labeling reactions along with the theoretical distribution derived from MS/MS acquired from discovery mode analysis and stored in the kinome MRM library; (D) Quantitative results by DDA shotgun proteomics analysis for peptide FLEDDTSDPTYTSALGGK#IPIR.
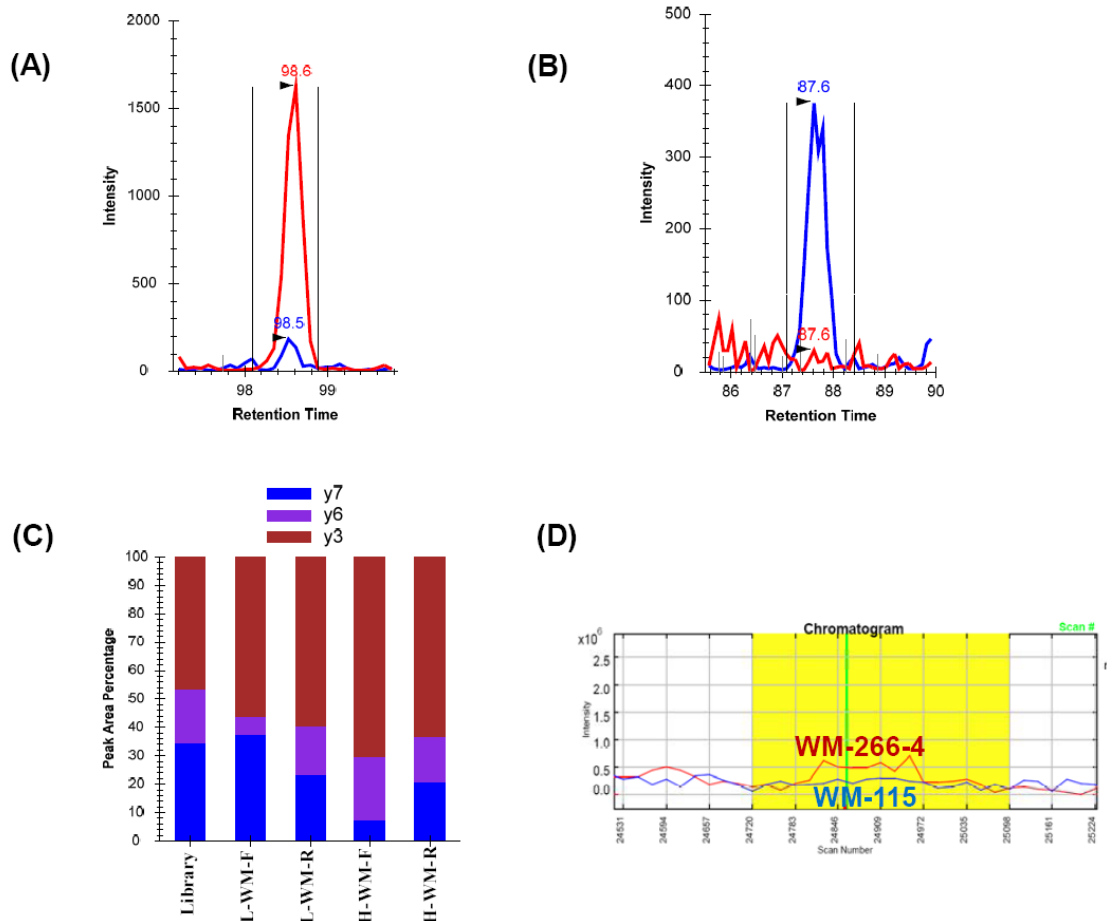
exhibit distinct kinome expression profiles (Figure 5.9). A total of 48 protein kinases display substantially different expression profiles ($R_{115/266-4} > 2$ or $R_{115/266-4} < 0.5$); except for PDK1, the rest 47 protein kinases show consistent expression ratios in MRM-based kinome assay for samples obtained from forward and reverse ICAP labeling experiments (Figure 5.10).

The tyrosine kinase group exhibits the most distinct expression profiles between these two melanoma cell lines and more than 30% of tyrosine kinases display different expression levels (Figure 5.11). Moreover, significantly more kinases in AGC, STE, and tyrosine kinase groups were upregulated in WM-115 cells than those that were upregulated in WM-266-4 cells. For example, in the AGC group, GPRK5, MAST3, PKCα, PKCβ, PKCγ, PDK1, and RSKL1 are upregulated in WM-115 cells, whereas only MSK1, MSK2, and PKN1 are upregulated in WM-266-4 cells. The distinct kinome expression profiles between these two cell lines suggest considerable kinome reprogramming during melanoma progression.

Among the differentially expressed tyrosine kinases, multiple receptor tyrosine kinases in the Eph family are upregulated in WM-115 cells. For instance, EphA2, EphB2 and EphB3 are expressed at much higher levels in WM-115 than WM-266-4 cells (Figure 5.12A). On the other hand, CMGC is the only kinase group where significantly more members are upregulated in WM-266-4 than WM-115 cells. In this vein, CDK1 and CDK2 are expressed at levels that are 4.5- and 3.6-fold higher in the former than latter

Figure 5.9 Quantitative comparison of kinome expression for WM-115 and WM-266-4 cells. Blue bar denotes kinase that is upregulated in WM-266-4 cells; red bar denotes kinase that is upregulated in WM-266-4 cells.
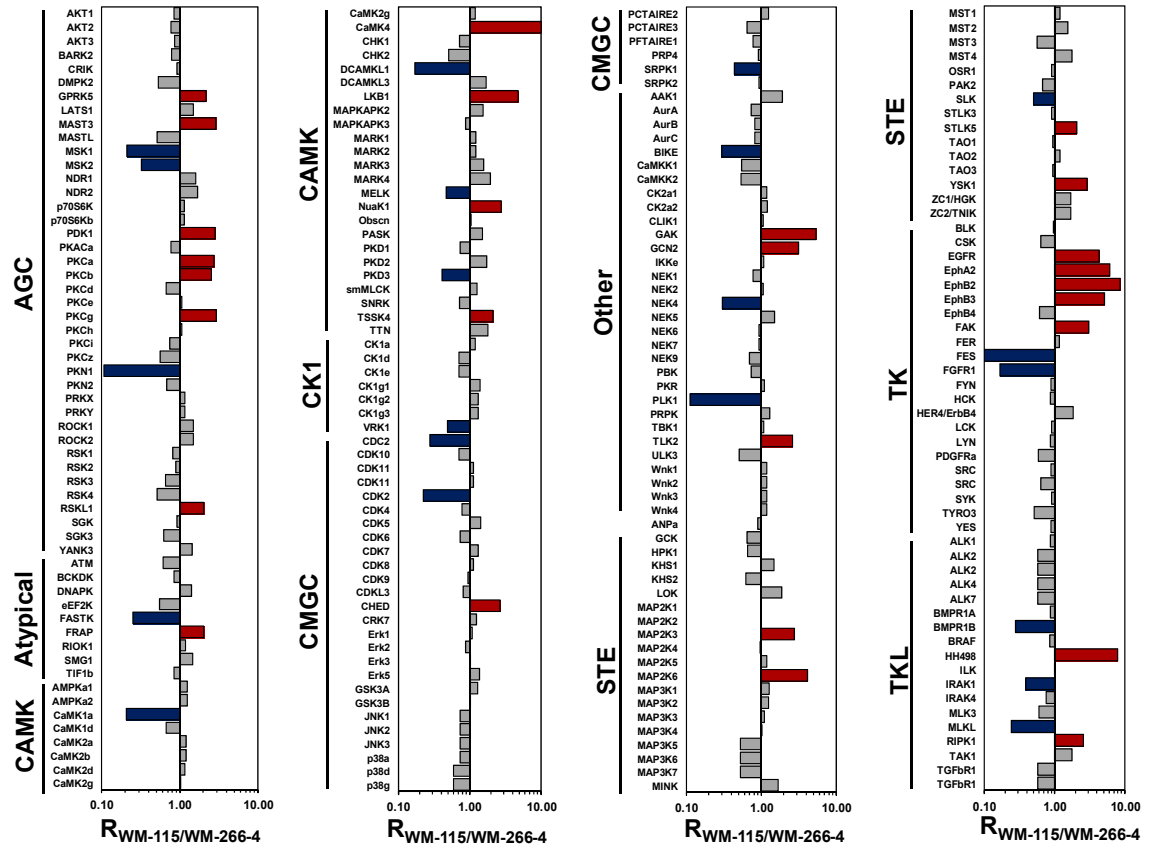
Figure 5.10 A heatmap showing the differential expression of kinases in WM-115 and WM-266-4 cells based on $R_{WM-115/WM-266-4}$ ratio. Dark red and white boxes represent kinases that are upregulated in WM-115 cells and WM-266-4 cells, respectively, as indicated by the scale bar above the heatmap.
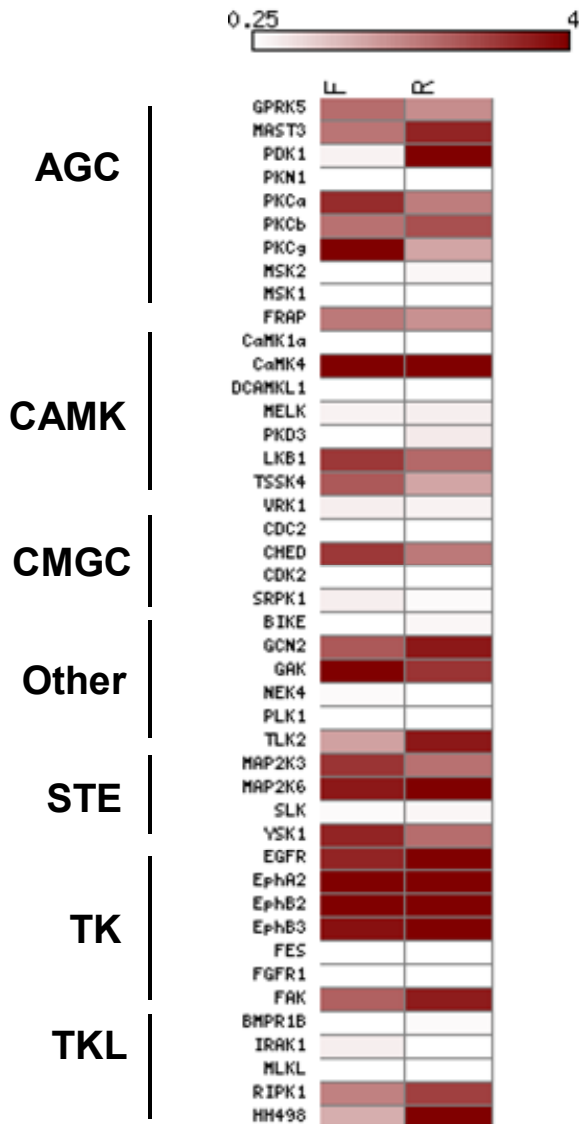
Figure 5.11 (A) Quantified protein kinases from WM-115 and WM-266-4 cells mapped in the dendrogram of the human kinome. Filled red circles indicate the kinases that are upregulated in WM-115 cells; filled blue circles indicate the kinase that are upregulated in WM-266-4 cells; Red shadow covers the targeted kinases for dasatinib; blue shadow covers the targeted kinases for flavopiridol. (B) Percentage of differentially expressed kinases from WM-115 and WM-266-4 cells for the seven major kinase groups as well as the atypical and other kinases.   White bar represents the percentage of kinases upregulated in WM-115 cells; grey bar represents the percentage of kinases upregulated in WM-266-4 cells; black bar represents the percentage of kinases differentially expressed in either of these two cells.
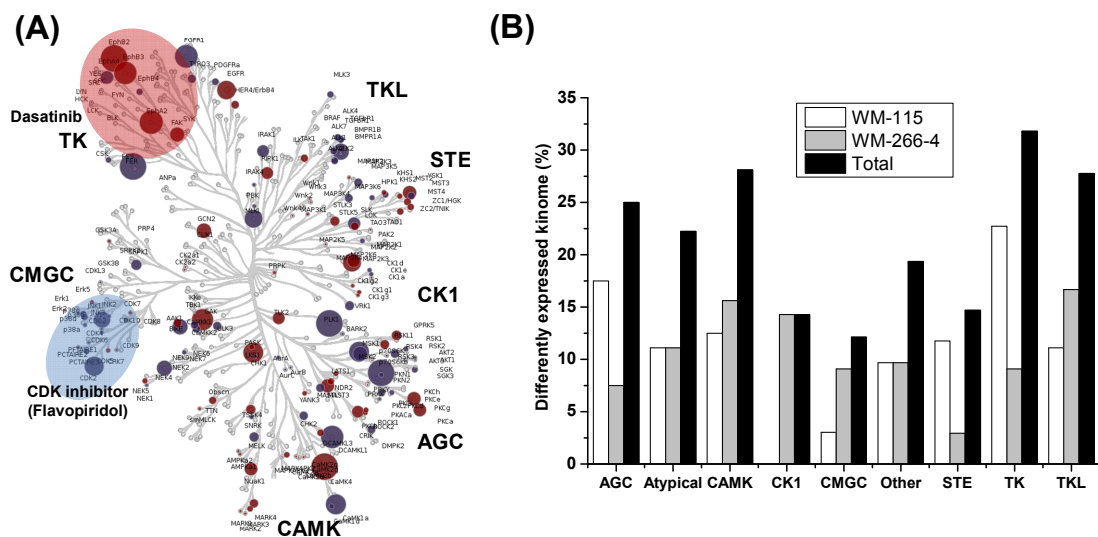
Figure 5.12 (A) Representative quantitative results by MRM assay in forward labeling reaction for peptide VLEDDPEATYTTSGGK#IPIR from EphA2, FLEDDTSDPTYTSALGGK#IPIR from EphB2, and FLEENSSDPTYTSSLGGK#IPIR from EphB3; peptide DLK#PQNLLIDDK from CDK1, and DLK#PQNLLINTEGAIK from CDK2. Extracted ion chromatograms of light-labeled peptides from WM-115 cells are depicted in red and extracted ion chromatograms of heavy-labeled peptides from WM-266-4 cells are depicted in blue. (B) Total cell lysates of WM-115 and WM-266-4 were immunoblotted with antibodies recognizing the kinases EphA2 (left) and CDK2 (right); actin served as a loading control. (C) Cell viability of WM-266-4 (blue line) and WM-115 cells (red line) when treated with dasatinib (left panel) and flavopiridol (right panel).

(A)
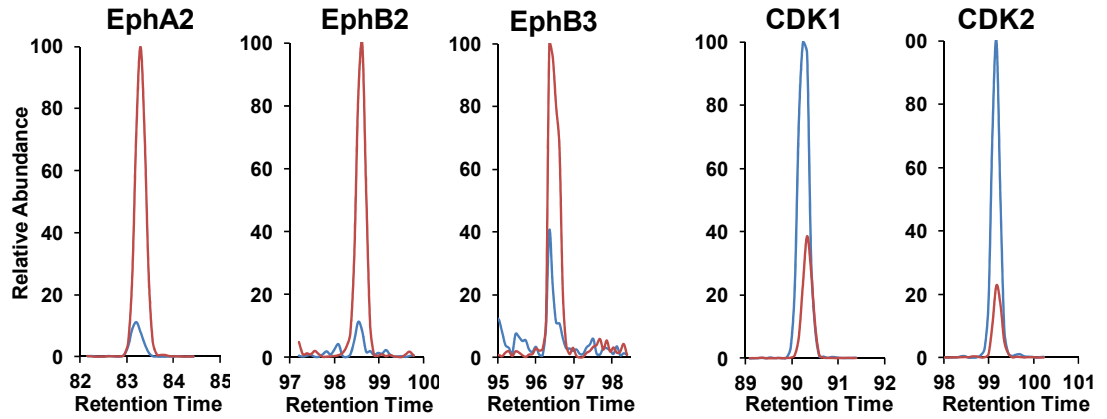
(B)
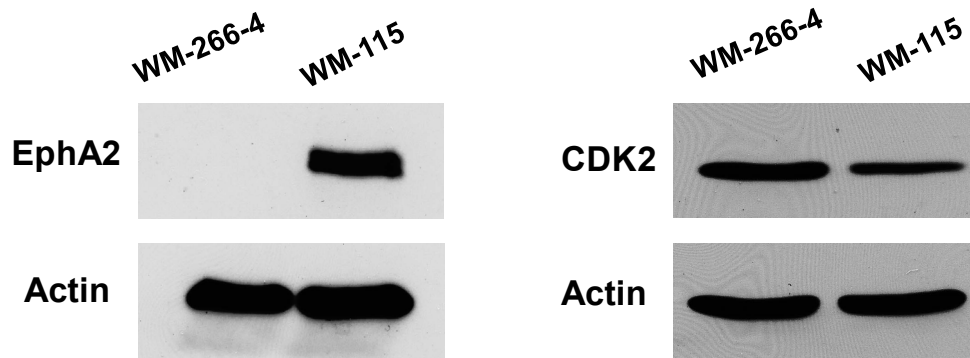
(C)

cells (Figure 5.12A). The differential expression of EphA2 and CDK2 in these two cell lines was also validated by Western analysis (Figure 5.12B).

The differential expression of the aforementioned kinases in these two melanoma cell lines may confer differential sensitivities of these two lines of cells toward the inhibitors for these kinases. In this regard, dasatinib is a pote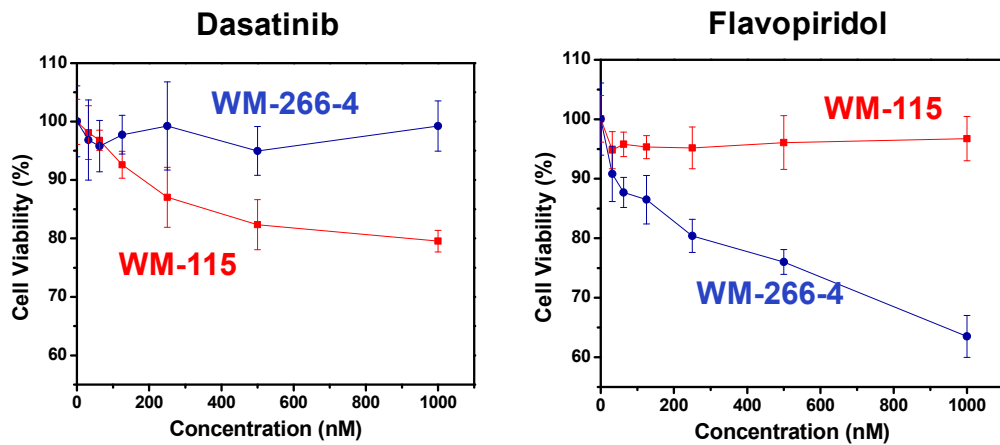nt inhibitor for EphA2, EphB3, EphB4, with dissociation constants (K_d) being < 1 nM; dasatinib, however, is a relatively poor inhibitor for those tyrosine kinases that are upregulated in WM-266-4 cells, which included FES (K_d > 10 µM) and FGFR1 (K_d > 3.7 µM) (5). We assessed the relative sensitivities of WM-115 and WM-266-4 cells toward dasatinib, and our results revealed that WM-115 cells are markedly more sensitive toward this tyrosine kinase inhibitor than WM-266-4 cells (Figure 5.12C). This finding is reminiscent of a previous study showing that dasatinib could inhibit the growth and reduce the migration and invasion of WM-115 cells, but not WM-266-4 cells (27).

We reason that the substantially higher levels of expression of CDK1 and CDK2 in the WM-266-4 than WM-115 cells may render WM-266-4 cells more sensitive toward CDK inhibitors than WM-115 cells. To test this, we examined whether flavopiridol, which is the first cyclin-dependent kinase inhibitor in human clinical trials and could potently inhibit CDKs 1, 2, and 4 (28), can preferentially inhibit the proliferation of WM-266-4 over WM-115 cells. It turned out that flavopiridol can effectively inhibit the growth of WM-266-4 cells at nM concentration range and approximately 40% cell death was observed at 1 µM (Figure 5.12C). In contrast, no appreciable sensitivity toward

flavopiridol was observed for WM-115 cells, even at the highest concentration tested (1 μM, Figure 5.12C). Together, our results demonstrate that the anti-proliferative effects of kinase inhibitors are correlated with the expression or ATP-binding affinity of their targeted kinases.

4. Application of MRM-based kinome profiling for human tissues

An advantage of the ICAP-based strategy lies in its potential application for kinome profiling for any biological samples, including clinical samples that are not amenable to metabolic labeling (29). To exploit this, we employed our MRM-based kinome assay to assess the differential kinase expression between human lung tumor and adjacent normal lung tissues from the same patient.

We successfully quantified 98 unique kinase peptides representing 124 kinases on our list (Figure 5.13). In addition, results from our MRM kinome assay for human tissue samples again exhibit excellent reproducibility among three independent labeling experiments. In this vein, 96 out of the 98 (98%) kinase peptides were quantified in all three labeling experiments (Figure 5.13), and most quantified kinases exhibit consistent trends in all three experiments (Figure 5.14A). Additionally, most kinases display elevated expression in lung tumor tissue than adjacent normal tissue (Figure 5.14A), which is a common phenomenon for tumor tissues (3). For instance, a unique peptide from MAPK3 (Erk1), i.e., DLK#PSNLLINTTCDLK, showed a higher level of expression in lung tumor than adjacent normal lung tissue in all three labeling experiments (Figure 5.14B). Although this proof-of-concept experiment for a single pair

Figure 5.13 The Venn diagrams showing the overlap of quantified kinase peptides from human tumor and normal lung tissues by triplicates of MRM analysis.
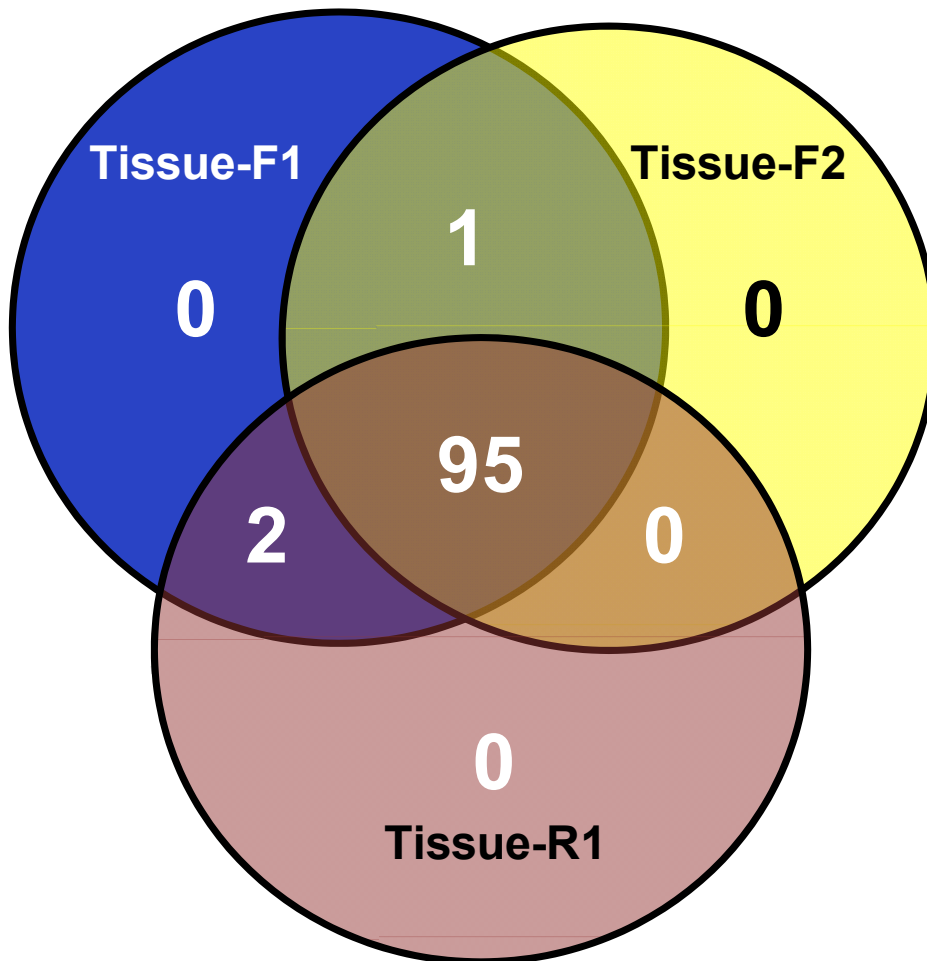
Figure 5.14 (A) A heatmap showing the differential expression of kinases from tumor and adjacent normal lung tissue based on $R_{tumor/normal}$ ratio in two forward and one reverse labeling reactions. Dark red and white boxes designate those kinases that are upregulated in tumor tissue and normal tissue, respectively, as indicated by the scale bar. (B) Quantitative results by MRM assay for peptide DLK#PSNLLINTTCDLK from MAPK3 kinase: (Left and Middle) Extracted-ion chromatograms for three transitions monitored for light-labeled (Red) and heavy-labeled (Blue) peptides in both forward and reverse labeling reaction; (Right) the consistent distribution of the peak area observed for each monitored transition from light- and heavy- labeled peptides in both forward and reverse labeling reaction along with the theoretical distribution derived from MS/MS spectra store in MRM kinome library.

(A)
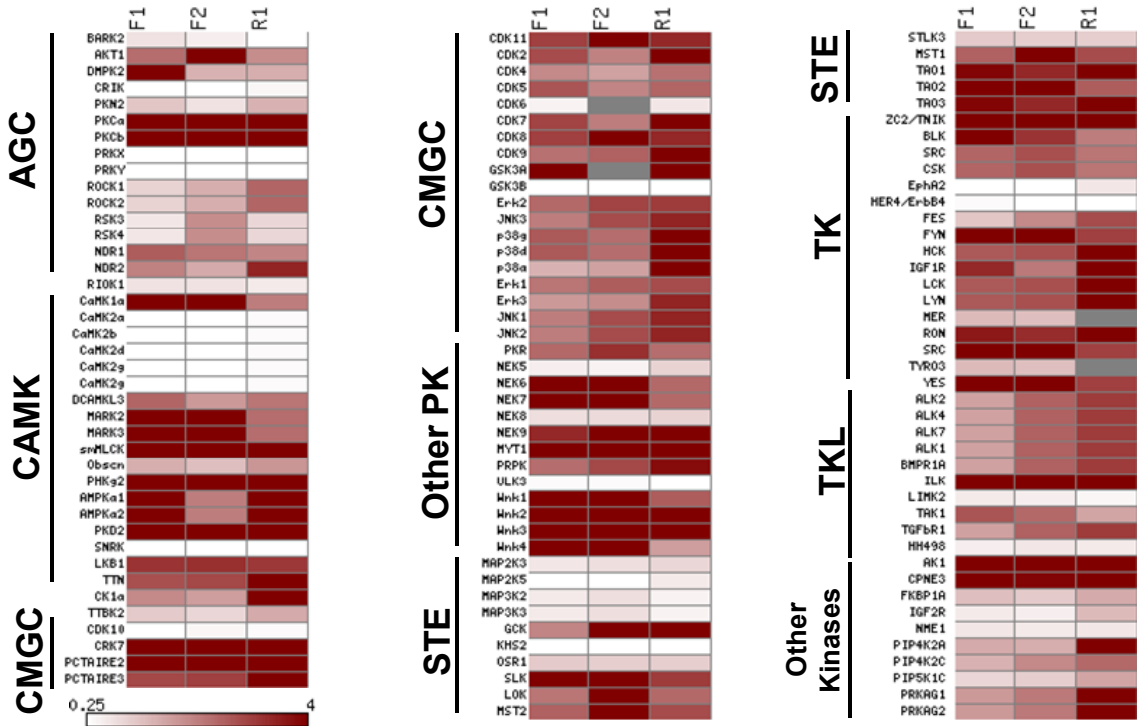
(B)
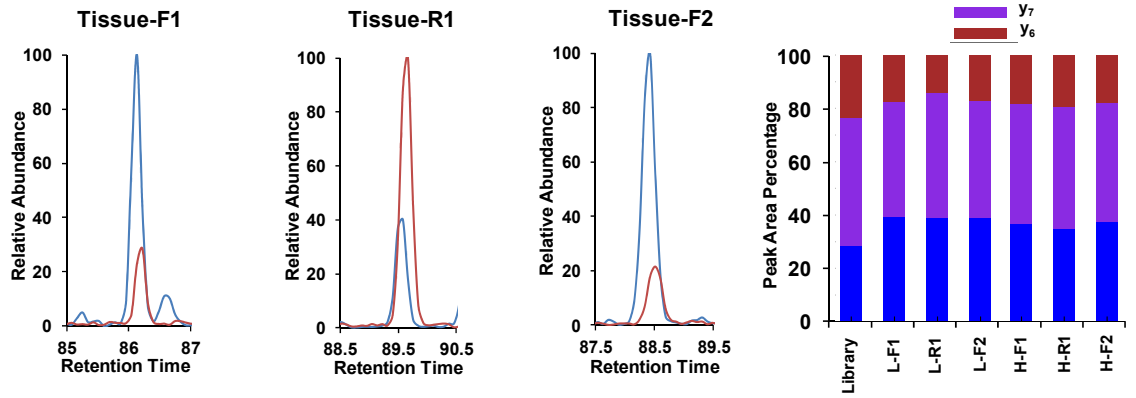
of human lung tissue samples does not allow us to draw a general conclusion about the implication of this and other aberrantly expressed kinases in lung cancer, our results clearly demonstrate the potential of our MRM kinome assay in quantitative study of tissue samples.

It is of note that we quantified much less kinases using the protein extracts from human tissue sample than using whole cell lysates. By analyzing the same tissue sample in DDA mode on the Orbitrap Velos, we found that the signal is dominated with highly abundant blood proteins, especially the different isoforms of human hemoglobin and albumin. The compromised performance of ICAP reagents for human tissue sample could be attributed to the inherent design of this probe. The ICAP reagents were developed to target any ATP-binding proteins instead of specific kinases. Unfortunately, the highly abundant proteins including albumin and hemoglobin in human tissue and plasma also possess strong ATP-binding affinity (30, 31). These proteins, therefore, compete favorably with low-abundance kinases in almost every step of the sample processing and analysis, such as probe binding and conjugation, avidin enrichment and MS detection. Nevertheless, we reason that depletion of highly abundant proteins from tissue samples, as has been widely used for improving the coverage of low-abundance proteins for biomarker discovery using human tissue or plasma samples (32), should enable the profiling of much more kinases from tissue lysates.

5. Application of MRM-based kinome profiling for mouse tissues

On the other hand, since the fragmentation pattern of peptides in SRM detection is the inherent attribute only related to its primary amino acid sequence, we expect that the kinome SRM library build up from human kinases can be readily applied to study biological sample from other organism. To this end, we map these 383 peptides in our human kinome library to mouse IPI database using skyline, generating a mouse kinome library including 292 kinase peptides. These 292 peptides can be mapped to 257 mouse kinases including 241 protein kinases and other 16 kinases. For concept of proof, we label protein extracts from mouse liver tissue with light form of ICAP reagent and analyzed by TSQ-vantage. As a result, 127 peptides corresponding to 146 kinases were detected. Carefully examination of the distribution of each transition for the detected peptides reveals that they share similar fragmentation as the peptides from human. For example, human and mouse CDK5 share the same sequence R.DLK#PQNLLINR.N. SRM assay reveal that this peptides share the very similar fragmentation distribution, in which y_8 ion is the dominant ion, no matter that the peptide is from human or mouse (Figure 5.15). These results indicate our kinase peptide liabrary, even though is generated from human samples, is not limited to study human kinome and can also be used for other organism.

Figure 5.15 Overlay of extracted ion chromatograms and distribution of peak area for four transitions monitored for peptide DLK#PQNLLINR from human and mouse CDK5.
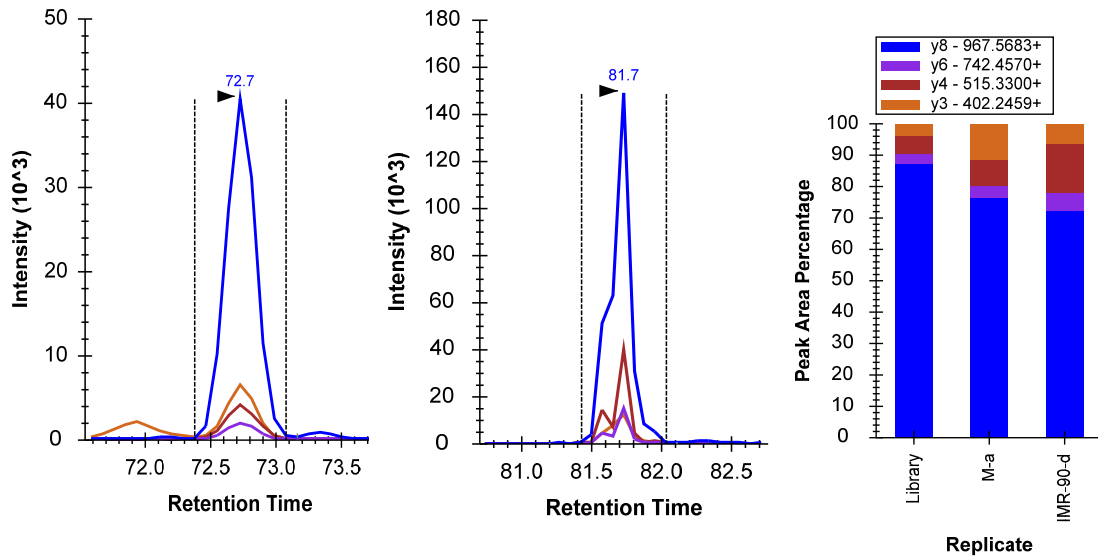
Conclusions

Kinases are among the most intensively pursued superfamilies of enzymes as targets for anti-cancer drugs. Recently large data sets on inhibitor potency and selectivity for more than 400 human kinases became available (5-7), which provided the opportunity to design rationally novel kinase-directed anti-cancer therapies. However, the kinase expression and activity are highly heterogeneous among different types of cancer and even among different stages of the same cancer, which point to the needs for global kinome expression profile in the development of kinase-targeted cancer therapy. Here, we introduced an MRM-based kinome profiling assay for more than 300 human protein and lipid kinases by monitoring specifically targeted peptides located at the ATP-binding sites of the kinases. In particular, we demonstrated that ~ 250 protein kinases (50% of the entire human kinome and more than 80% of human kinome in a single cell line) could be routinely quantified without extensive separation using multi-dimensional chromatography. Our on-going discovery-based analysis of kinases labeled using our ICAP reagents for different cell lines will further expand our kinome library. Moreover, our studies showed that the MRM-based kinome assay exhibits superior sensitivity, reproducibility and accuracy over discovery-based shotgun proteomics. Importantly, the incorporation of iRT into MRM kinome library renders our MRM kinome assay easily transferrable across different chromatographic systems and laboratories. This MRM-based approach for kinome expression profiling of two melanoma cell lines derived from the same patient revealed that the anti-proliferative effects of kinase inhibitor drugs are

correlated with the expression levels of their targeted kinases. Therefore, this facile and accurate kinome profiling assay, together with the kinome-inhibitor interaction map, may provide invaluable knowledge to predict the effectiveness of kinase inhibitor drugs and offer the opportunity for individualized therapy in cancer treatment.

We also demonstrated that this MRM-based kinome assay is applicable to clinical samples that are not amenable to metabolic isotope labeling. We believe that the depletion of highly abundant proteins from tissue samples prior to the ICAP labeling will further improve kinome coverage for tissue samples. It can be envisaged that our method could be used for future global kinome profiling of human tumor biopsy samples, which may guide the design of kinase-targeted, individualized anti-cancer therapy (33, 34).

References

1.	Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. Science. 2002;298:1912-34.

2.	Blume-Jensen P, Hunter T. Oncogenic kinase signalling. Nature. 2001;411:355-65.

3.	Duncan James S, Whittle Martin C, Nakamura K, Abell Amy N, Midland Alicia A, Zawistowski Jon S, et al. Dynamic Reprogramming of the Kinome in Response to Targeted MEK Inhibition in Triple-Negative Breast Cancer. Cell. 2012;149:307-21.

4.	Zhang J, Yang PL, Gray NS. Targeting cancer with small molecule kinase inhibitors. Nat. Rev. Cancer. 2009;9:28-39.

5.	Karaman MW, Herrgard S, Treiber DK, Gallant P, Atteridge CE, Campbell BT, et al. A quantitative analysis of kinase inhibitor selectivity. Nat. Biotech. 2008;26:127-32.

6.	Fabian MA, Biggs WH, Treiber DK, Atteridge CE, Azimioara MD, Benedetti MG, et al. A small molecule-kinase interaction map for clinical kinase inhibitors. Nat. Biotech. 2005;23:329-36.

7.	Davis MI, Hunt JP, Herrgard S, Ciceri P, Wodicka LM, Pallares G, et al. Comprehensive analysis of kinase inhibitor selectivity. Nat. Biotech. 2011;29:1046-51.

8.	Pulford K, Lamant L, Morris SW, Butler LH, Wood KM, Stroud D, et al. Detection of anaplastic lymphoma kinase (ALK) and nucleolar protein nucleophosmin (NPM)-ALK proteins in normal and neoplastic cells with the monoclonal antibody ALK1. Blood. 1997;89:1394-404.

9.	Cox J, Mann M. Quantitative, high-resolution proteomics for data-driven systems biology. Annu. Rev. Biochem. 2011;80:273-99.

10.	Oppermann FS, Gnad F, Olsen JV, Hornberger R, Mann M, Daub H. Large-scale proteomics analysis of the human kinome. Mol. Cell. Proteomics. 2009;8:1751-64.

11.	Marx V. Targeted proteomics. Nat. Meth. 2013;10:19-22.

12.	Daub H, Olsen JV, Bairlein M, Gnad F, Oppermann FS, Stemmann O, et al. Kinase-selective enrichment enables quantitative phosphoproteomics of the kinome across the cell cycle. Mol. Cell. 2008;31:438-48.

13.   Villamor JG, Kaschani F, Colby T, Oeljeklaus J, Zhao D, Kaiser M, et al. Profiling protein kinases and other ATP binding proteins in arabidopsis using Acyl-ATP Probes. Mol. Cell. Proteomics. 2013;12:2481-96.

14.   Ansong C, Ortega C, Payne Samuel H, Haft Daniel H, Chauvignè-Hines Lacie M, Lewis Michael P, et al. Identification of widespread adenosine nucleotide binding in mycobacterium tuberculosis. Chem. Biol. 2013;20:123-33.

15.   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;85:3198-206.

16.   Patricelli MP, Szardenings AK, Liyanage M, Nomanbhoy TK, Wu M, Weissig H, et al. Functional interrogation of the kinome using nucleotide acyl phosphates. Biochemistry (Mosc). 2006;46:350-8.

17.   Patricelli Matthew P, Nomanbhoy Tyzoon K, Wu J, Brown H, Zhou D, Zhang J, et al. In situ kinase profiling reveals functionally relevant properties of native kinases. Chem. Biol. 2011;18:699-710.

18.   Olsen JV, Schwartz JC, Griep-Raming J, Nielsen ML, Damoc E, Denisov E, et al. A dual pressure linear ion trap orbitrap instrument with very high sequencing speed. Mol. Cell. Proteomics. 2009;8:2759-69.

19.   Picotti P, Aebersold R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. Nat. Meth. 2012;9:555-66.

20.   Lange V, Picotti P, Domon B, Aebersold R. Selected reaction monitoring for quantitative proteomics: a tutorial. Mol. Syst. Biol. 2008;4.

21.   Xiao Y, Guo L, Wang Y. Isotope-coded ATP probe for quantitative affinity profiling of ATP-binding proteins. Anal. Chem. 2013;85:7478-86.

22.   Escher C, Reiter L, MacLean B, Ossola R, Herzog F, Chilton J, et al. Using iRT, a normalized retention time for more targeted measurement of peptides. Proteomics. 2012;12:1111-21.

23.   MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics. 2010;26:966-8.

24.   Abbatiello SE, Mani DR, Keshishian H, Carr SA. Automated detection of inaccurate and imprecise transitions in peptide quantification by multiple reaction monitoring mass spectrometry. Clin. Chem. 2010;56:291-305.

25.   James LC, Tawfik DS. The specificity of cross-reactivity: Promiscuous antibody binding involves specific hydrogen bonds rather than nonspecific hydrophobic stickiness. Protein. Sci. 2003;12:2183-93.

26.   Westermark B, Johnsson A, Paulsson Y, Betsholtz C, Heldin CH, Herlyn M, et al. Human melanoma cell lines of primary and metastatic origin express the genes encoding the chains of platelet-derived growth factor (PDGF) and produce a PDGF-like growth factor. Proc. Natl. Acad. Sci. USA. 1986;83:7197-200.

27.   Eustace AJ, Dowling P, Henry M, Doolan P, Meleady P, Clynes M, et al. 2D-DIGE analysis of phospho-enriched fractions from dasatinib-treated melanoma cell lines. J. Proteomics. 2011;74:490-501.

28.   Senderowicz A. Flavopiridol: the first cyclin-dependent kinase inhibitor in human clinical trials. Invest. New. Drugs. 1999;17:313-20.

29.   Mann M. Functional and quantitative proteomics using SILAC. Nat. Rev. Mol. Cell. Biol. 2006;7:952-8.

30.   Bauer M, Baumann J, Trommer WE. ATP binding to bovine serum albumin. FEBS Lett. 1992;313:288-90.

31.   Costello AJR, Marshall WE, Omachi A, Henderson TO. ATP binding to human hemoglobin in the presence and absence of magnesium ions investigated with 31P NMR spectroscopy and ultrafiltration. Biochim. Biophys. Acta. 1977;491:469-72.

32.   Keshishian H, Addona T, Burgess M, Kuhn E, Carr SA. Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. Mol. Cell. Proteomics. 2007;6:2212-29.

33.   Ocana A, Pandiella A. Personalized therapies in the cancer "omics" era. Mol. Cancer. 2010;9:202.

34.   Garman KS, Nevins JR, Potti A. Genomic strategies for personalized cancer therapy. Hum. Mol. Genet. 2007;16:R226-R32.