DEVELOPING CHEMICAL PROTEOMIC TOOLS---CONNECTING PROTEINS AND SMALL MOLECULES

LIU KAI

(B.Sc. Huazhong University of Science & Technology)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOLOGICAL SCIENCES

NATIONAL UNIVERSITY OF SINGAPORE

2011

Acknowledgements

Summary

The completion of the human genome sequencing project has provided a wealth of new information about the genomic blueprint of the cell. The promise of this information is likely to re-define the way researchers approach the study of complex biological systems and drug development. But genes do not tell the entire story of life and living processes until proteins are translationally produced and post-translationally modified. Proteins are not only integral part of life but also are required for its regulation and diversification. Diseases can be caused by minor changes in protein dysfunction. Although there are roughly 20,000 genes in the human genome, only a few proteins have known functions. Little is understood about the physiological roles, substrate specificity, and downstream targets of the vast majority of these important proteins. The major challenge for fighting human disease lies in translating genomic information into understanding of the cellular functions of these proteins in both normal and pathological process. A key step toward the biological characterization of proteins, as well as their adoption as drug targets, is the development of global solutions that bridge the gap in understanding these proteins and their interactions. Recently developed chemical proteomics approaches are alternative and complementary approaches for gene expression analysis and thus are ideal utensils in decoding this flood of genomic information. This approach makes use of synthetic small molecules that can be used to covalently modify a set of related proteins and subsequently allow their purification and/or identification as valid drug targets. Furthermore, such methods enable rapid biochemical analysis and small molecule screening of targets there by accelerating the often difficult process of target validation and drug discovery.

This thesis examines and addresses these challenges by introducing a series of chemical proteomics tools that span various analytical modes, effectively expanding the chemical proteomics labelling's application on both specificity and scope. These include chemical (small molecules inhibitor) labelling (Chapter 2, 3 and 4) and metabolites (endogenous small molecules) analogue labelling (Chapters 5) platforms, for which I demonstrate with examples, novel strategies to garner implicit understanding of protein functions, enzyme-substrate interactions, protein-drug interactions, protein localizations and protein's post-translational modifications. Cohesively, these methodologies are applied (but not limited) to different phases of drug development--- protein targets identification, lead discovery and drug efficacy assessment.

Table of Contents

List of Publications

(2008 - 2011)

1. Liu, K., Yang, P.-Y., Na, Z., Yao, S.Q.* Dynamic Monitoring of Newly Synthesized Proteomes: Up-Regulation of Myristoylated Protein Kinase A During Butyric Acid Induced Apoptosis Angew. Chem. Int. Ed. **2011**, in press.

2. Yang, P.-Y., Wang, M., Liu, K., Ngai, M.H., Sheriff, O., Lear, M.J., Sze, S.K., He, C.Y.*; Yao, S.Q.* Parasite-Based Screening and Proteomic Profiling Reveal Orlistat™, an FDA-Approved Drug, as a Potential Anti-Trypanosoma brucei Agent Angew. Chem. Int. Ed. (2011), submitted.

3. Yang, P.-Y., Liu, K., Zhang, C., Chen, G. Y. J., Shen, Y., Ngai, M. H., Lear, M. J., Yao, S. Q.* Chemical Modification and Organelle-Specific Localization of Orlistat-Like Natural Product-Based Probes. Chem. -Asian. J., (2011) in press.

4. Liu, K., Shi, H., Xiao, H., Chong, A.G.L., Bi, X., Chang, Y.T., Tan, K., Yada, R.Y., Yao, S.Q.* Functional Profiling, Identification and Inhibition of Plasmepsins in Intraerythrocytic Malaria Parasites. *Angew. Chem. Intl. Ed.*, **2009** 48, 8293-8297.

5. Yang, P-Y., Liu, K., Ngai, M.H., Lear, M.J.., Wenk, M., Yao, S.Q.* Activity-Based Proteome Profiling of Potential Cellular Targets of Orlistat - An FDA-Approved Drug with Anti-Tumor Activities. *J. Am. Chem. Soc.* **2010**, 132, 656-666.

6. Kalesh, K.A., Sim, S. B. D.., Wang, J., Liu, K., Lin, Q., Yao, S.Q.* Small molecule probes that target Abl kinase. *Chem. Commun.* **2010**, 46, 1118-1120.

7. Kalesh, K.A,; Tan, L.P., Liu, K., Gao, L., Wang, J., Yao, S.Q Peptide-based Activity-Based Probes (ABPs) for Target-Specific Profiling of Protein Tyrosine Phosphatases (PTPs). *Chem. Commun.* **2010**, 46, 589-591.

8. Kalesh, K.A., Liu, K., Yao, S.Q.* Rapid Synthesis of Abelson Tyrosine Kinase Inhibitors Using Click Chemistry. *Org. Biomol. Chem.*, **2009**, 7, 5129-5136.

9. Shi, H., Liu, K., Xu, A., Yao, S.Q.* Small Molecule Microarray (SMM)-Facilitated Screening of Affinity-Based Probes (AfBPs) for g-Secretase. *Chem. Commun.* **2009**, 5030-5032.

10. Shi, H., Liu, K., Leong, W.Y.W.,Yao, S.Q.* Expedient Solid-Phase Synthesis of Both Symmetric and Asymmetric Diol Libraries Targeting Aspartic Proteases. *Bioorg. Med. Chem. Lett.* **2009**, 19, 3945-3948.

11. Liu, K., Kalesh, K.A,; Ong L.B., Yao, S.Q.* An Improved Mechanism-Based Cross-Linker for Multiplexed Kinase Detection and Inhibition in A Complex Proteome. *ChemBioChem,* **2008**, 9, 1883-1888.

12. Uttamchandani, M., Wang, J., Li, J., Hu, M., Sun, H., Chen, K. Y. -T., Liu, K., Yao, S. Q. "Inhibitor Fingerprinting of Matrix Metalloproteases using a Combinatorial Peptide Hydroxamate Library" *J. Am. Chem. Soc.* **2007**, *129*, 7848-7858.

13. Uttamchandani, M; Liu, K., Panicker, R.C., Yao, S.Q.* Activity-Based Fingerprinting and Inhibitor Discovery of Cysteine Proteases in a Microarray. *Chem. Commun.*, **2007** , 1518-1520.

List of Figures

List of Tables

List of Abbreviations

2DE	two-dimensional electrophoresis
6AzFuc	6-azido fucose
AA	amino acid
ABP	Activity Based Probe
AfBP	Affinity Based Probe
AHA	Azidohomoalanine
AOMK	acyloxymethyl ketone
BA	butyric acid
BP	benzophenone
BSA	bovine serum albumin
cAMP	cyclic adenosine monophosphate
$CaCl_2$	calcium chloride
CHX	cycloheximide
CID	collision-induced dissociation
C-terminus	carboxy terminus
Cu	copper
Cy	cyanine
CYP	cytochrome P450
DEVD	Asp-Glu-Val-Asp
DIGE	difference in gel electrophoresis
DMSO	dimethylsulfoxide
E. coli	*Escherichia coli*
EDTA	ethylenediamine tetraacetic acid
EN	ethynylnaphthalene

ESI	electrospray ionization
FAS	fatty acid synthase
GPCR	G-protein coupled receptors
GST	glutathione-S-transferase
DNA	deoxyribonucleic acid
FP	fluorophosphonates
Gal(N)Az	N-azidoacetylgalactosamine
Glc(N)Az	N-azidoacetylglucosamine
HCl	hydrochloric acid
HAP	Histoaspartic Protease
HDAC	histone deacetylases
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HGP	Human Genome Project
HIV	human immunodeficiency virus
HPG	homopropargylglycine
HPLC	high performance liquid chromatography
HTS	high-throughput screening
IC_{50}	half the maximal inhibitory concentration
IPG	immobilized pH gradients
IEF	isoelectric focusing
K_D	dissociation constant
K_i	inhibition constant
LB	luria-bertani
LC	liquid chromatography
MALDI	matrix-assisted laser desorption ionization

Man(N)Az	N-azidoacetylmannosamine
MAPK	mitogen-activated protein kinase
MP	metalloproteases
MMP	matrix metalloprotease
MS	mass spectrometry
MudPIT	multidimensional protein identification technology
N-terminus	amino terminus
NaCl	sodium chloride
$NaHCO_3$	sodium bicarbonate
NMR	nuclear magnetic resonance
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PC	phosphatidyl choline
pH	negative logarithm of the hydroxonium ion concentration
pI	isoelectric point
PM	plasmepsin
PMSF	phenylmethanesulfonylfluoride or phenylmethylsulfonyl fluoride
RBC	red blood cells
RNA	ribonucleic acid
RFU	relative fluorescence units
RP	reverse-phase
SAHA	suberoylanilide hydroxamic acid
SH	serine hydrolases
SiaNAz	N-azidoacetyl sialic acid
SDS	sodium dodecyl sulfate

SGI   silicon graphics

SPR   surface plasmon resonance

THL   tetrahydrolipstatin

TMR   tetramethylrhodamine

TOF   time of flight

Tris   Trishydroxymethyl amino methane

WH   warhead

UV   ultra-violet

List of Symbols

Å	angstrom
^{o}C	degree celsius
cal	calorie
g	gram
h	hour
k	kilo
l	litre
λ	wavelength
m	milli or meter
μ	micro
M	molar
min	minutes
mol	mole
n	nano
p	pico
s	seconds

List of 20 Natural Amino Acids

Single Letter	Three Letter	Full Name
A	Ala	Alanine
C	Cys	Cysteine
D	Asp	aspartic acid
E	Glu	glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Praline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine

Chapter 1.

Introduction

1.1 Summary

The complete sequence of the human genome (Lander *et al.*, 2001; Venter *et al*., 2001), in addition to the larger framework of other model organisms have established a firm foundation for modern biological investigations to unveil the blueprint of life. Rapid functional assignment and global characterization of every protein from the sequence data of tens of thousands of genes (so called functional genomics) is the next major step for the human genome project. Proteomics aims to accelerate this process by developing methods for the parallel analysis of large numbers of proteins. (Saghetalian *et al.*, 2005) By large-scale studying the dynamic description of gene's product, proteomics offers powerful utensils to decipher gene functions (Abersold *et al.*, 2001; Mann *et al.*, 2003). It holds promises to impact our understanding of the molecular composition and function of cells significantly. Moreover, by studying the entire complement of proteins, including the modifications made to

1

a particular set of proteins, it provides a global, integrated and comprehensive view of disease states and cellular processes (Hanash, *et al.* 2003). The alteration of the protein functional profiles, upon perturbation by extracellular and intracellular cues, may also be monitored and quantified, which provides a powerful tool to identify proteins that are potential targets for therapeutics and interventions (Lindsay, M.A. *et al.* 2003). Researchers studying proteomics aims to accelerate this process by developing state-of-the-art methods for the parallel analysis of large numbers of proteins. The following section will discuss the impact of proteomics in the post-genomic era and summarize the its advances and applications in various phases of drug development.

1.2 Proteomics in Post-Genomics Era

Proteins are involved in almost all biological activities and they also have diverse properties, which collectively contribute greatly to our understanding of biological systems. Proteomics systematically study such diverse properties of proteins in a high-throughput manner and aim to provide detailed descriptions of the structure, function, interaction, modification of proteins in health and disease state. It deals with the large-scale determination of gene and cellular function directly at the protein level, thus it is an extension of genomics study. To highlight the importance of proteomics study in post-genomic era, we will discuss the significance of both genomics and proteomics, the two field deeply connected with each other at different levels.

1.2.1 Genomics and Human Genome Sequencing

Genomics was firstly coined by Thomas H. Roderick in 1986 as a term to define the study of the complete set of genetic information of an organism (Mckusick,

et al. 1997). It mainly dealt with mapping, sequencing and analysis of the whole genome of an organism. The significance of genomics was highlighted by the initiation of the Human Genome Project (HGP) in 1985 with the primary goal of determining the sequence of chemical base pairs which make up DNA and to identify and map the approximately 20,000-25,000 genes of the human genome from both a physical and functional standpoint (Watson and Cook-Deegan, *et al.* 1991). In 2001, after more than a decade of efforts from international collaboration, the draft of the human genome sequence was produced and made available (Lander *et al.*, 2001; Venter *et al.*, 2001). The complete sequence of the human genome holds an extraordinary trove of information waiting to be further analyzed. However, the major challenge is how to rapidly annotate all the function and characterize the roles of genes from the overgrowing sequence data.

1.2.2 Challenges in Deciphering the Human Genome

To overcome this issue, the term genomics rapidly expanded from focused efforts on mapping and sequencing of complete genomes to a nearly all-inclusive combination of genome scale experimental and computational enterprises. Comparative genomics, one of the most insightful approaches to interpret genomic data, was born almost immediately upon landmark reports of completely sequenced bacterial genomes of Haemophilus ifluenzae and Mycoplasma genitalium in 1995, followed by the first archaeal (*Methanococcus jannaschii*) and eukaryotic (*Saccharomyces cerevisiae*) genomes in 1996. This is attempted by the analysis of linear sequence motifs or higher order structural motifs that indicate a statistically significant similarity of a sequence to a family of sequences with known function or by other means such as comparison of homologous protein functions across species. These efforts will lead to a more richly annotated sequence database and, not by

themselves, to an explanation of the structure, function, and control of biological processes.

With the development of high-throughput platform technologies, functional genomics is developed to make use of the vast wealth of data produced by genomic projects to describe gene functions (Hieter and Boguski, *et al.* 1997). Such functional analysis includes the large-scale characterization of genes and their derivatives dynamic in aspects of gene transcription, translation, and protein-protein interactions (Eisenberg *et al.*, 2000). Functional genomics as a means of assessing phenotype differs from more classical approaches primarily with respect to the scale and automation of biological investigations. A classical investigation of gene expression might examine how the expression of a single gene varies with the development of an organism in vivo. Functional genomics approaches, however, using a high-throughput tool like DNA microarray, functional genomics would examine the profiling of 1,000 to 10,000 genes at the transcriptional level as a function of development (Lockhart and Winzeler, *et al.* 2000). Up to date, unprecedented amounts of genome-wide data on gene expression patterns have been generated and the gene's transcript profiles from different cellular states have been compared on a genome-wide scale from DNA microarray experiments. These extension and variation of genomics did provide a first clue about gene function at different cellular states and thus a more complete picture of how biological function arises from the information encoded in an organism's genome. These -omics have changed the way we plan and perform our research in the most profound way. However, the intrinsic limitations of the study of gene functions at the prediction or transcriptional level should be noticed. Firstly, in the time scale of most biological processes, except evolution, the genomic DNA sequence can be viewed as static and a genomic

sequence database therefore represents only a stationary information resource. Secondly, even in the most state-of-the-art transcriptome study, there is accumulating evidence indicating that the data of mRNA abundance gathered do not corroborate well with the protein expression level (Pandey and Mann, *et al.* 2000). It has been reported that variation between certain protein abundance and the corresponding mRNA transcripts could be as high as 30 folds (Gygi and Rochon *et al.*, 1999; Tian *et al.*, 2004). Besides this poor correlation between mRNA levels and protein abundance, characterization of gene products in a sophisticated biological network is inevitably complicated by the un-expected numbers of gene products from a single gene as a result of alternative splicing and posttranslational modifications. Moreover, localization change and protein-protein interaction change, representing another level of protein function profiling, is also an obstacle that cannot be overcome through traditional genomics study of genetic sequence.

1.2.3 the Promise of Proteomics

Unlike static genomic sequences, proteins inside the cell are perpetually being created and discarded. Besides, proteins, as highly diverse entities inside the cells and key structural scaffolds, signal transducers, functional executors, reaction catalysts, are dynamic in every aspects of view like the types of expressed proteins, abundance, state of modification, sub-cellular location, and etc., being dependent on the physiological state of the cell or tissue (Hanash, *et al.* 2003). Understanding these properties of proteins is one of the grand goals of the post-genomic era, and has been given a disciplinary title of its own: proteomics. As a consequence, the large-scale parallel analysis of proteins, termed proteome profiling, is a more accurate and comprehensive description which reflects the cellular state or the external conditions encountered by a cell than genomics information alone. Therefore, direct

examination of the properties of proteins, such as expression level and protein activity, localization and etc., with the aid of DNA sequence information, will be more meaningful for our comprehensive understanding of cellular processes. This multidisciplinary field relies on a collection of various technologies, all of which contribute to proteomics. These include cell imaging by light and electron microscopy, array and chip experiments, and genetic readout experiments, as exemplified by the yeast two-hybrid assay. Such studies typically challenge the high complexity of cellular proteomes and the low abundance of many of the proteins, which require highly sensitive analytical techniques. This advance has major implications for our understanding of cellular organisation in health and disease, and for pharmaceutical biotechnology. Indeed, proteomics is already yielding important findings that will accelerate the process of drug discovery. The following section discusses new concepts, innovative technologies and biological applications in proteome research.

1.3 Core Technologies of Proteomics

Most proteomics research is aiming the goal of investigating protein expression and function under specified physiological conditions. Many proteomics strategies emerged and aimed to monitor the expression of large number of proteins in a specified cell or tissue and quantify the expression pattern changes under different cellular conditions like in the presence of drug or in the diseased tissue. Such study makes it possible to identify disease-specific proteins, drug targets and markers of drug efficacy and toxicity. Today, all proteomics studies reply on two core technologies---two-dimentional electrophoresis (2DE) and Mass Spectrometry (MS). Both these core technologies have been significantly advanced during the past

years.

1.3.1 Two Dimentional Gel Electrophoresis

More than three decades ago, O'Farrell *et al.* (1975) and Klose *et al.* (1975) first demonstrated large-scale protein separation by 2-DE. In their works, proteins were separated by isoelectric focusing (IEF) in the first dimension, followed by separation on SDS-PAGE according to the molecular weight of the protein in the second dimension. *E. coli*, a simple model organism, was chosen in O'Farrell's work and more than 1000 proteins from *E. coli* were resolved in 2DE. Since then, it has been the technique of choice for separation of complex protein mixtures into their individual polypeptide components and for analyzing protein composition. One could expect to visualize around 1000 proteins using 2D-polyacrylamide gel electrophorasis. However, this fraction is still short of the total number of proteins that may be present in a eukaryotic cell. Missing polypeptides do not enter the gel due to not being resolved by the pH gradient (too basic or too acidic) or simply are not detected due to limitations in the sensitivity of the current procedures. Ideally, every protein would be resolved as an isolated and detectable spot by 2D-PAGE. But it is estimated that at least 1000 copies of a protein have to be present in a cell for them to be detected by 2D and as much as 90% of the total protein of a typical cell is made up of only 10% of the 10 000-20 000 different protein species (Gygi, S.P., *et al.* 2000). Thus, many low-abundance proteins might not be detectable by 2D-PAGE Therefore these low abundance proteins are unintentionally omitted from the subsequent analysis.

More recently, the resolution and reproducibility of IEF was further increased by the introduction of immobilized pH gradients (IPG), which also enable researchers to tune the pH separation to any desired range. Using narrow-range IPG

strips allowed a larger number of proteins to be separated than had been possible with standard 2D-PAGE because a narrower pH range was spread out over a greater physical distance. This spread allowed proteins with similar isoelectric point (pI) values to be separated with higher resolution. In the initial study, Hoving *et al.* (2000) developed a 2D-PAGE method in which they applied narrow range IPG strips in the first dimension. The IPG strips were typically 1-3 pH U wide and overlapped one another by at least 0.5 pH U. Six IPG strips covering the pH range of 3.5 to 10 were used. Proteins from a B-lymphoma cell line were applied to each strip and separated using IEF. Each strip was then applied to an individual SDS-PAGE gel plate and proteins were separated in the second dimension based on their molecular weight. Approximately 5000 distinct spots were detected using the six IPG strips, compared with 1500 spots detected using a single IPG strip with a pH range of 3-10 and a single standard 2D-PAGE gel plate.

Another important technological advance in 2DE to improve the detection limit is the development of sensitive protein stains, including the ammoniacal silver stain, which permits detection of proteins at or below nanogram quantities. Although, silver staining, which is far more sensitive than Coomassie staining methods, has been widely used for high sensitivity visualisation on 2D-PAGE, it is not suitable for quantitative analysis with a limited dynamic range. Moreover, recently developed Sypro post-electrophoretic fluorescent stains (Molecular Probes, Eugene, Oregon, USA) have emerged as alternatives, offering a better detection limit and dynamic range and ease of use (Yan, J.X. *et al.* 2000; Malone, J.P. *et al.* 2001). Sypro Ruby has been shown to be more sensitive than silver, and is compatible for subsequent peptide mass mapping (Lopez, M.F. *et al.* 2000).

With improved detection limit and resolving power, 2D-PAGE has also been

used for comparison of protein expression profiles of sample pairs (normal versus transformed cells, cells at different stages of growth or differentiation, etc.). Moreover, 2D-PAGE also urged researchers to analyze the global pattern changes of protein expression in response to the addition of a cytokine or a drug in a given cell type or tissue.

However, when 2D-PAGE is more and more used for comparison, it was found that the application of conventional 2D-PAGE in quantitative proteomics has been largely hampered by its poor reproducibility from lab to lab. This is typically caused by the discrepancy of proteins absorbed by the IEF strips, protein transfer from IEF to PAGE gels and in-homogeneities of the gel composition and pH gradients during manufacture (Van den Bergh, G. & Arckens, 2004). Although the in-homogeneities of the gel composition and pH gradients could be improved with more strict quality control, it is still accepted that any subtle changes in experimental conditions may render the quantities of two aliquots of proteins analyzed in separate 2-D gels unequal, making it difficult to overlay and compare proteins with altered expression level and especially quantify them on different gels. This makes it difficult to distinguish between system errors and real changes in the proteome arising from biological perturbations. Since the amount of proteins transferred from the first dimension to the second dimension is usually inconstant, the poor reproducibility thus makes the comparison less significant and necessitates the running of replicate gels for the same protein Apart from being waste of samples and a tedious process, the accuracy of this method is highly in doubt, particularly when quantificatoin of protein expression by 2-DE is required.

To address this issue, a relatively new technique (i.e. Difference in gel electrophoresis, refered as DIGE) was developed by Unlu *et al.* (1997), where two

different protein samples can be labeled with two structurally similar fluorescent dyes respectively prior to 2D-PAGE and co-separated in the same gel. 2D DIGE technology is based on the specific properties of two sets of spectrally resolvable dyes— Cyanine- 2 (Cy2), Cyanine-3 (Cy3) or Cyanine-5 (Cy5), which have been designed to be both mass- and charge-matched. Therefore, identical proteins labeled with each of the dyes will migrate to the same position on a 2D-PAGE gel. Resolving two or even up to three different samples on single gel separates system variation from biological variation and thus allows the ratio of relative abundance of the same protein in different samples to be compared directly. Even small differences in expression levels can be determined by comparing the ratio obtained from one fluorescently labeled sample directly with another. As a result, it is possible to see small differences in protein abundance between samples, with high statistical confidence, enabling accurate analysis of differences in protein abundance between samples. In addition, only biological replicates are required rather than replicates of the same sample. This is an advantage over conventional 2-D electrophoresis. Using 2D-DIGE approach, minimal effort is required to gain meaningful statistical data. Moreover, labeling with Cy-Dye is very sensitive with a detection limit of around 500 pg of a single protein. It offers greater sensitivity and a broader linear dynamic range, comparing to fluorescent staining of 2-D gels (Patton, *et al.* 2002).

Although these important technological advances gives 2D-PAGE considerable resolving power, this technology still cannot fulfill the ultimate goal of displaying in one experiment an entire cell or tissue proteome. Several types of proteins have proven especially understated by 2DE, including low and high molecular mass proteins, membrane proteins, proteins with extreme isoelectric points (pIs) and low abundance proteins (Corthals, G.L. *et al.* 2000; Gygi, S.P. *et al.* 2000)

In one example, Gygi et al. found that the products of six genes migrated to the same faint spot on a silver-stained, narrow pH range gel of a yeast cell extract. . In Gygi, S.P.'s study, capacity and sensitivity of 2DE have been pushed to their limits, however, this inadequacy of resolution also confounds further accurate quantification and mass spectrometric identification. Unfortunately, the conclusion of the study is that 2DE as a separation method for the comprehensive analysis of the yeast proteome has limitations and without sample pre-enrichment, 2DE is not suitable for the detection of lower abundance classes of proteins, which collectively comprise at least one-half of the predicted proteome. Besides, the inadequacy in detecting low abundance proteins, membrane proteins is another un-resolved issue. About 30% of proteins are recently estimated to be membrane proteins (Paulsen, I.T. *et al.* 1998). It is well know that these membrane proteins are difficult to work with. There has been a report finding that only about 1% of integral membrane proteins are actually resolved on current 2DGE gels (even when thio-urea is used in the lysis buffer) **(**Garrels, J.I. *et al.* 1997**)** Moreover, even with continued development of sample preparation protocols using various detergents, there has still been an important and un-resolved problem of isolating highly hydrophobic membrane proteins using 2DE (Santoni, V., *et al.* 1999). Considering those limitations of 2D-PAGE, alternative and/or complementary separation strategies must be developed in order to permit a global characterization of the protein expression.

1.3.2 Mass Spectrometry

Apart from 2D-PAGE, MS is another technology overwhelmingly utilized in proteomics study. It emerged as a base for protein identification from 2D-PAGE gels in the late 1980s. Traditionally, protein identification involves de novo sequencing, most frequently by the automated, stepwise chemical degradation (Edman

degradation) of proteins or isolated peptide fragments.(Hewick, R. M. *et al.* 1981; Aebersold, R. H. *et al.*1987). As the size of sequence databases grow, even relatively short and otherwise imperfect sequences (gaps, ambiguous residues) became useful for the identification of proteins. This was realized by correlating sequences obtained experimentally from the peptides analysis with sequence databases. And these sequence information required for mapping sequence database could be easily generated by mass spectrometry when two technical breakthroughs made in the late 1980s. These breakthroughs were the development of the two ionization methods electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) (Fenn, J. B. *et al.*1989; Cole, R.B. *et al.*1997; Karas, M. *et al.*1995).

ESI gained immediate popularity because of the ease with which it could be interfaced with popular chromatographic and electrophoretic liquid-phase separation techniques and quickly supplanted fast atom bombardment as the ionization method of choice for protein and peptide samples dissolved in a liquid phase. Furthermore, due to the propensity of ESI to produce multiply charged analytes, simple quadrupole instruments and other types of mass analyzers with limited m/z range could be used to detect analytes with masses exceeding the nominal m/z range of the instrument.

For different but no less compelling reasons, MALDI also rapidly gained popularity. The time-of-flight (TOF) mass analyzer most commonly used with MALDI is robust, simple, and sensitive and has a large mass range. MALDI mass spectra are simple to interpret due to the propensity of the method to generate predominantly singly charged ions. The method is relatively resistant to interference with matrixes commonly used in protein chemistry. For MALDI, analytes are spotted onto a metal plate either one at a time or, in a higher throughput format, multiple samples on the same plate. The samples are usually tryptic digests from proteins

separated by 2DE, although proteins purified by other separation methods are also compatible with the method. Before deposition of the analytes, the matrix is placed on the plate or mixed in with the sample. The matrix will absorb energy from the laser causing the analytes to be ionized by MALDI. The m/z ratio of the ions is then typically measured based on the flight time in a field-free drift tube (as opposed to ion mobility MS where a field pushes ions through a gas) that constitutes the heart of the time-of-flight mass (TOF) analyzer. Using internal calibration on monoisotopic masses, a mass accuracy of 5 ppm at 1000 u can be achieved. An additional bonus for samples isolated from biological sources is that MALDI is compatible with biological buffers such as phosphate and Tris and low concentrations of urea, nonionic detergents, and some alkali metal salts. Peptide m/z ratios are calculated based on the energy equation ion $E = 1/2mv^2$ that accounts for contributions from kinetic energy, mass, and velocity. At a constant energy, low molecular weight ions will travel faster than high molecular weight ionsflight times of ions are inversely proportional to the square root of their molecular mass.

These methods solved the difficulty for generating ions from large, nonvolatile analytes such as proteins and peptides without analyte fragmentation. Since then, identifying proteins by correlating information extracted from a protein or peptide with sequence databases rather than by de novo sequencing gradually became a concept accepted. Moreover, the implementation of mass spectrometric methods for protein identification was further accelerated, because the methods initially developed for Edman sequencing to isolate small amounts of proteins and peptides were directly compatible with peptide analysis by LC-MS and LC-MS/MS (Aebersold, R.H. *et al.* 1987; Larive, C.K. *et al.* 1999). Correlating of sequence information generated from MS with the databases also rely on the development of

novel search algorithms. Such algorithms use available constraints in a decision-making process that distinguishes the correct match from all other sequences in the database. As the sequence databases were complete, breakthroughs in mass spectrometric methods were made, and the search algorithms were set up, a rapid and sensitive technology for protein identification became dramatically advanced and mature. All these converged as the basis of protein identification by MS and also basis of the emerging field of proteomics.

Today, there are in general two unveiling methods for protein identification, either using multiple related peptide (peptide mass mapping) or single peptide sequencing. Protein identification by peptide mass mapping was conceptually simple and reported by several groups independently at approximately the same time (Yates, J. R., III *et al.* 1993; James, P. *et al.* 1993). The principle is that, after proteolysis with a specific protease, proteins of different amino acid sequence will produce groups of peptides. And the masses of these peptides constitute mass fingerprints unique for a specific protein. Therefore, if these selected masses (the observed peptide mass fingerprint) are searched in a sequence database containing the specific protein sequence, then it is expected to correctly identify the protein within the database.

These methods implemented almost at the same time by different groups vary in specific details but share the following sequence of steps. Firstly using sequence-specific proteases, peptides can be generated from the sample protein by digestion leaving the carboxyl- or amino-terminus fixed for the search. For instance, the most widely used protease trypsin leaves arginine (R) or lysine (K) at the carboxyl-terminus after cleavage. Secondly, accuracy of mass spectrometer is in need to generate peptide as accurately as possible. Accuracy in mass will decrease the

number of peptides for any given mass in a sequence database and therefore increase the stringency of the search. Thirdly, the sequences of the protein available in the database are then in-silico digested under the rules of the proteolytic method used in the experiment. This will generate a theoretical list of masses which will be compared with the set of measured masses. Lastly, an algorithm matches the measured peptide masses with those sets of masses predicted for each protein in the database and ranks the quality of the matches by assigning a score to each match. It is obvious that, the sequence of the protein to be identified has to exist in the sequence database being used for comparison. If DNA sequence databases are being used, the DNA sequences will be translated into protein sequences prior to the in-silico digestion. Thus, genetically well-characterized organisms with genome sequenced will be best suited for this approach. However, since the protein identification by this approach depends on the correlation of multiple peptide masses generated from the same protein, it is neither suited for EST databases nor for identification of proteins in complex mixtures if the proteolyzed sample consists of un-separated mixtures. The problem with EST database is that it only represents a fragment of a gene's coding sequence and such fragment may not be long enough to cover a sufficient number of peptides observed in the mass spectrometry analysis to allow an unambiguous identification. Digests of unseparated protein mixtures is also problematic because it is not obvious which peptides in the complex peptide mixture come from the same protein. The peptide mass mapping approach is therefore most compatible with the identification of proteins from species of which complete genome sequences have been determined and for those sample proteins purified by 2D gel electrophoresis whereby protein molecular weight and isoelectric point information is available to aid unambiguous identification. When a pure protein is

digested and peptide masses are compared with the list of peptide masses calculated for the protein, typically there could be two observations. First, not all of the in-silico predicted peptides are detected in real experiment. Second, some of the measured peptide masses in mass analysis are not present in the list of masses predicted from the protein. The first observation, the missing masses are usually caused by a number of problems such as poor solubility, ion suppression, and selective ionization. There are many reasons that unassigned masses being observed. Firstly, changes in the expected peptide masses by post-translational modification (e.g., phosphorylation adds a net 80 u to an amino acid mass), and artificial modifications from sample handling (such as oxidation of methionine) could lead to unassigned masses. Some of these changes can be anticipated and incorporated into the search algorithm but some cannot. Secondly, un-expected proteolysis due to the presence of contaminating proteases can sometimes produce peptides unanticipated by the search algorithm (e.g., the presence of chymotryptic activity in a trypsin preparation). Thirdly, the presence of more than one protein in the sample is a frequent problem for protein bands in SDS gels and spots in 2D.

Other than peptide mass mapping, protein identification can also be achieved by single peptide sequencing. Not like peptide mass mapping, where several peptide masses from the same protein are required for unambiguous identification, using peptide sequencing, amino acid sequence of even a relatively small peptide can uniquely identify a protein, because the amino acid sequence of a peptide is more constraining than its mass for protein identification by sequence database searching. This approach largely depends on tandem mass spectrometry for the generation of sequence-specific spectra for peptides. In a tandem mass spectrometer, peptide ions are fragmented and the resulting fragment ion spectra are recorded. In tandem mass

spectrometers such as triple quadrupole, ion trap, or quadrupole/TOF instruments, fragment ion spectra are generated by cleavage at the amide bonds in a process called collision-induced dissociation (CID) where the peptide ion to be analyzed is isolated and fragmented in a collision cell. Such spectra generated are much less complex than the high collision energy spectra generated in magnetic sector or TOF/TOF instruments. Therefore these spectra are relatively simple to interpret To interpret and annotate the spectra generated from CID, a nomenclature has been adapted to differentiate fragment ions according to the amide bond that fragments and the end of the peptide that retains a charge after fragmentation. For example, if the positive charge associated with the parent peptide ion remains on the amino-terminal side of the fragmented amide bond, then this fragment ion is referred to as a b ion. On the contrary, if the charge remains on the carboxyl-terminal side of the broken amide bond, then the fragment ion is referred to as a y ion. These individual fragment ion m/z values contain redundant pieces of information from the same peptide and this redundancy makes fragment ion spectra a rich source of sequence-specific information. Currently, either for single isolated proteins or on a proteome wide scale, peptide sequencing is most often carried out by CID in a triple quadrupole (TQ), ion-trap (IT) or quadrupole time-of-flight (QTOF) mass spectrometer. Besides these core instruments, a particularly successful implementation of the peptide sequencing approach uses a variation in ESI called nanospray. In nanospray, a peptide sample is introduced at very low flow rates. Such low sample consumption afforded by the this technique allows for extended observation and accumulation of the ion signals and generally yields CID spectra of excellent quality.

These sequencing instrumentations, when coupled with gel-free, chromatography-based separation methods, can even overcome the limitation of the

resolving power of 2D-PAGE approach and accomplish identification of the components in protein mixtures without the need for separation of the mixture into the individual protein components. It is realized by the digestion of the un-separated proteins and the analysis of the resulting complex peptide mixture by LC-MS/MS.

Andrew, J.L. described a rapid and sensitive method for comprehensively identifying proteins in macromolecular complexes that uses multidimensional liquid chromatography (LC) and tandem mass spectrometry (MS/MS) to separate and fragment peptides. This combination is able to identify individual proteins in even the most complex macromolecular complex in the cell without first purifying each protein component to homogeneity. With this method, they have identified more than 100 proteins in a single run and a novel protein component of the yeast and human 40S subunit in Saccharomyces cerevisiae ribosome (Link, A. J. *et al.* 1999).

In addition to protein identification from protein mixture, these newly developed MS methods can be further tuned to determine the protein expression levels (relative quantity) between two different pools of proteins when the stable isotope labeling technique has been adapted. The method involves the addition to a sample of chemically identical but stable isotope (e.g., 2H, 13C, 15N, etc.) labeled internal standards. Quantitative protein profiling is therefore accomplished when a protein mixture (reference sample) is compared to a second sample containing the same proteins at different abundances and labeled with heavy stable isotopes. Therefore all the peptides in the sample form pairs of identical sequence but different mass because of the heavy stable isotope. The same physicochemical properties of these the peptide pairs allow them to behave identically the same during isolation, separation, and ionization. Thus, an accurate measure of the relative abundance of the peptides could be achieved by comparing the the ratio of intensities of the lower

and higher mass components. This ratio of peptides to some extent reflects the ratio of proteins in the original protein mixtures. Three groups initially and independently reported measuring quantitative protein profiles based on stable isotopes (Pasa-Tolic, L., *et al.*1999; Oda, Y., *et al.*1999; Gygi, S.P. *et al.* 1999).

Gygi S.P. has recently developed a novel method for quantitative protein profiling based on isotope-coded affinity tags (ICAT) (Gygi S.P. *et al.* 1999). An ICAT reagent is specific toward sulfhydryl groups and has an eightfold deuterated linker, and a biotin affinity tag. In this strategy, the side chains of cysteinyl residues in a reduced protein sample representing one cell state are derivatized with the isotopically light form of the ICAT reagent. And equivalent groups in a sample representing a second cell state are derivatized with the isotopically heavy reagent. Afterwards, two groups are combined and enzymatically cleaved to generate peptide fragments while only cysteine containing peptides are tagged and isolated by avidin affinity chromatography to reduce the complexity. Finally, isolated csyteine containing peptides are analyzed by LC-MS/MS. During LC-MSMS analysis, both the quantity of the tagged peptides and identity of the sequence of the peptide are determined by automated multistage MS. The sequence information was then used to deduct the protein from which these peptides originate. This method is compatible with any amount of protein harvested from bodily fluids, cells, or tissues under any growth conditions and any type of biochemical, immunological, or physical fractionation for low abundance proteins. However, the size of the ICAT label (500 Da) is a relatively large modification remaining throughout the MS analysis which can complicate the database searching algorithms, especially for small peptides.

Similarly, stable isotope labeling with amino acids in cell culture (SILAC) is an in-vivo approach developed for incorporation of a label into proteins for mass

spectrometry (MS)-based quantitative proteomics. This method involves metabolic incorporation of a given 'light' or 'heavy' form of the amino acid into the proteins. In an experiment, two cell populations are grown in culture media that are identical except that one of them contains a 'light' and the other a 'heavy' form of a particular amino acid (e.g. 12C and 13C labeled L-lysine, respectively). When the labeled analog of an amino acid is supplied to cells in culture instead of the natural amino acid, it is incorporated into all newly synthesized proteins. After a number of cell divisions, each instance of this particular amino acid will be replaced by its isotope labeled analog. Since there is hardly any chemical difference between the labeled amino acid and the natural amino acid isotopes, the cells behave exactly like the control cell population grown in the presence of normal amino acid. After harvest, the proteins from both cell populations can be combined and analyzed together by mass spectrometry. Pairs of chemically identical peptides of different stable-isotope composition can be differentiated in a mass spectrometer owing to their mass difference. The ratio of peak intensities in the mass spectrum for such peptide pairs reflects the abundance ratio for the two proteins.

Everley, P.A. *et al.* (2004) applied SILAC to the study of metastatic prostate cancer. They added 12C- and 13C-labeled amino acids to the growth media of different prostate cancer cell lines with various metastatic potential, giving rise to cells containing either "light" or "heavy" proteins, respectively. By mixing lysates harvested from these cells, proteins can be identified by tandem mass spectrometry at the same time allowing a quantitative comparison between the two samples. The expression levels for more than 440 proteins in the microsomal fractions of prostate cancer cells were compared and 60 of them were found to be elevated greater than 3-fold in the highly metastatic cells, while 22 were reduced. Further validation was

carried out with western blotting to confirm the mass spectrometry-based quantification.

For proteomics in general, the separation sciences continue to make great strides in analyzing complex mixtures and offer the potential for circumventing gel electrophoresis as a preparative tool for MS. Over the past decade, gel electrophoresis followed by proteolysis of individual stained protein bands has been the most common method for separating proteins prior to MS identification (Patterson, S.D. *et al.* 1994). However, a number of laboratories have been investigating the use of chromatography only based approaches that bypass the electrophoretic based preparative gel methods altogether, except for diagnostic purposes. Two-dimensional or orthogonal chromatography approaches such as cation exchange followed by reverse-phase on-line with tandem MS have been successfully used to identify proteins in complex mixtures after proteolysis (Link, A. J. *et al.* 1999; Opiteck, G. J. *et al.*1997; Opiteck, G. J. *et al.* 1997). Even more complex approaches have used computer-controlled setup with an autosampler, five columns, and three 10-port switching valves to allow a series of steps to be performed on-line, obviating the need for any manual transfers of materials. The strategy included the following: immunoaffinity chromatography, desalting and buffer exchange on a mixed-bed strong ion-exchange absorbent, enzymatic digestion on an immobilized trypsin column, capture of peptides on a short perfusion capillary reversed-phase column, and final separation on an analytical reversed-phase column with on-line MS/MS analysis (Hsieh *et al.*1996). These orthogonal chromatography techniques have as a common goal the circumvention of the weakness of data-dependent analysis of complex mixtures, namely, that very complex mixtures of peptides exceed the capacity of these computer routines to carry out CID on all of the peptides present in

a given full-scan mass spectrum. Thus, by fractionating complex peptide mixtures on-line, these two-dimensional chromatography methods help extend the dynamic range of the overall analysis.

1.4 Emerging field of Chemical Proteomics

Although 2DE and MS methods allow researchers to analysis the relative levels of parallel proteins across multiple proteomic samples, these methods only focus on measuring protein abundance changes. Therefore these methods are limited in providing direct information on protein function and a series of protein post-translational regulation, including protein-protein, protein-small-molecule interactions and etc. Such information will increase the quality of the data obtained from 2D and MS experiment and therefore bring our study more steps towards the fundamental goal of biological research to understand complex physiological and pathological processes at the different level.

To facilitate the analysis of protein function, protein-protein interaction and protein-small-molecule interactions, which is complementary to abundance analysis, several proteomic methods have been developed including large-scale yeast two-hybrid screens (Uetz, P. *et al.* 2000; Ito, T. *et al.* 2001) and epitope-tagging immunoprecipitation experiments (Gavin, A.C. *et al.* 2002；Ho, Y. *et al.* 2002) aiming to delineate a comprehensive profile of protein-protein interactions. Besides, high-throughput platforms like protein microarrays (Macbeath G. *et al.* 2000) were also developed to facilitate the rapid analysis of protein activities. Although these methods have the advantage in annotating specific molecular functions of individual protein products, one of their limitation is that they all rely on the recombinant expression of proteins in artificial environments. Therefore, these approaches do not

directly assess the functional state of these biomolecules in their native environment.

To address this issue, the multidisciplinary science of chemical proteomics has emerged. These approaches utilize synthetic small molecules that can be used to covalently modify or being installed on a set of related proteins in-vitro or in-vivo and subsequently allow their purification, identification, and analysis of their function. The following sections will highlight the development of such chemical proteomic tools and their application to functional proteomic studies.

The field of chemical proteomics makes use of small molecules probes to tag and monitor distinct sets of proteins within a complex proteome. These chemical probes offers a means to systematically analyze proteins at various levels (activity, localization, interaction, co- or post-translaitional modifications and etc.) other than simple protein abundance .

Depending on the means through which these covalent chemical labels are introduced into proteins, vast majority of these strategies can be classified into three group: activity based (**Section 1.4.2**), affinity based (**Section 1.4.3**) or metabolic incorporation based (**Section 1.4.4**). In the following sections, we will discuss these probe families developed in the past decades. But before that, we will firstly summarize various reporters or chemical handles used to report probes once they are covalently linked to their target proteins (**Section 1.4.1**).

1.4.1 Tagging and Detection Strategies for Chemical Proteomics

By allowing quick and simple detection, identification and purification of probe-modified proteins from a complex proteome, reporters or chemical handles are one of the key component of a chemical probe and it distinguish a chemical probe from any other small molecule that can be covalently linked to a protein. Commonly used reporters include fluorophores and affinity tags.

1.4.1.1 Fluorophores

The use of small molecule fluorophores is probably the greatest advance in tagging methods for chemical proteomics. Fluorescent tags can be visualized by direct in-gel scanning with a fluorescent scanner such as the Typhoon scanner from Amersham Biosciences. Fluorescent tag is advantageous in several aspects over affinity tags like biotin for the purpose of detection and quantification, such as minimal handling, higher sensitivity and a greater dynamic range. Especially when using recently developed fluorescent tags with non-overlapping excitation/emission spectra, researchers can carry out multiple chemical proteomics experiments with different colors and obtain all results on a single gel (Greenbaum, D. *et al.* 2002; Patricelli, M.P. *et al.*2001). Thus far, diverse types of fluorophores with broad range of structural properties and absorbance and emission spectra have been developed as fluorescent tags for chemical proteomics strategies. These fluorophores includes fluorescein and rhodamine (Patricelli, M.P. *et al.*2001), dansyl (Berkers, C.R. *et al.*2002), NBD (nitrobenz-2-oxa-1,3-diazole)(Schmidinger, H., *et al.*2005), BODIPY (dipyrromethene boron difluoride) (Greenbaum, D. *et al.* 2002) and the cyanine (Cy)-dyes (Chan EW, *et al.* 2004). Fluorescein and rhodamine is relatively inexpensive, but the disadvantage of these fluorophores is that they photobleach rapidly making them less suitable for most imaging applications. BODIPY and Cy-dyes display high absorption coefficients, high quantum yields, narrow absorption peaks and relatively large stoke shifts. In addition, BODIPY and Cy-dyes are hydrophobic and can freely penetrate cell membranes. Therefore, these fluorophores are widely used for a variety of biological applications.

1.4.1.2 Affinity tags

Regardless of the advantages of fluorescent tags for detection, affinity tag is

still the most commonly used tag for chemical proteomics strategy because of its ability to provide a method for purification of labeled proteins. Among various affinity tags, biotin is the most commonly used affinity tag since biotin-streptavidin bond is one of the strongest known non-covalent interactions allowing for tight binding of low abundance biotinylated proteins. Such purification and enrichment of probe-bound proteins is essential for the identification of probe's targets in a complex proteome.

However, there are still limitations using biotin as a tag in chemical proteomics strategies. One of the limitations is that once a biotin-labeled protein is bound to an avidin resin, elution of the bound protein usually requires harsh conditions such as detergents or denaturants that release nonspecific, unlabeled background proteins. Moreover, a significant number of endogenously biotinylated proteins exist in most cells and tissue extracts therefore they are also co-purified by the avidin affinity resin. Recently cleavable linkers were developed to overcome these problems by facilitating probe specific release from the beads without using harsh conditions.

Another limitation is the low cell permeability properties, of biotin preventing in situ and in vivo applications. Besides, sometimes, reporter tags such as biotin and even fluorophores can influence the specificity of chemical probes and also limit their utility in living cells. These shortcomings have recently been addressed with the advent of tandem bioorthogonal labeling methods that enable biotin/fluorophore tag to be added after the probe finds its target protein. The following section discusses the development of such bio-orthogonal ligation methods.

1.4.1.3 Tandem Bio-Orthogonal Tagging

The main advantage of tandem bio-orthogonal labeling is the ability to use

small chemical functionality that has minimal effects on target binding and cell permeability and that can be later used for chemical modification with any various reporter tags (Figure 1.1). Therefore, highly specific bioorthogonal ligation chemistry that can function under physiological environments will be highly desirable to install the reporter tag after chemical probes have reacted with target proteins.
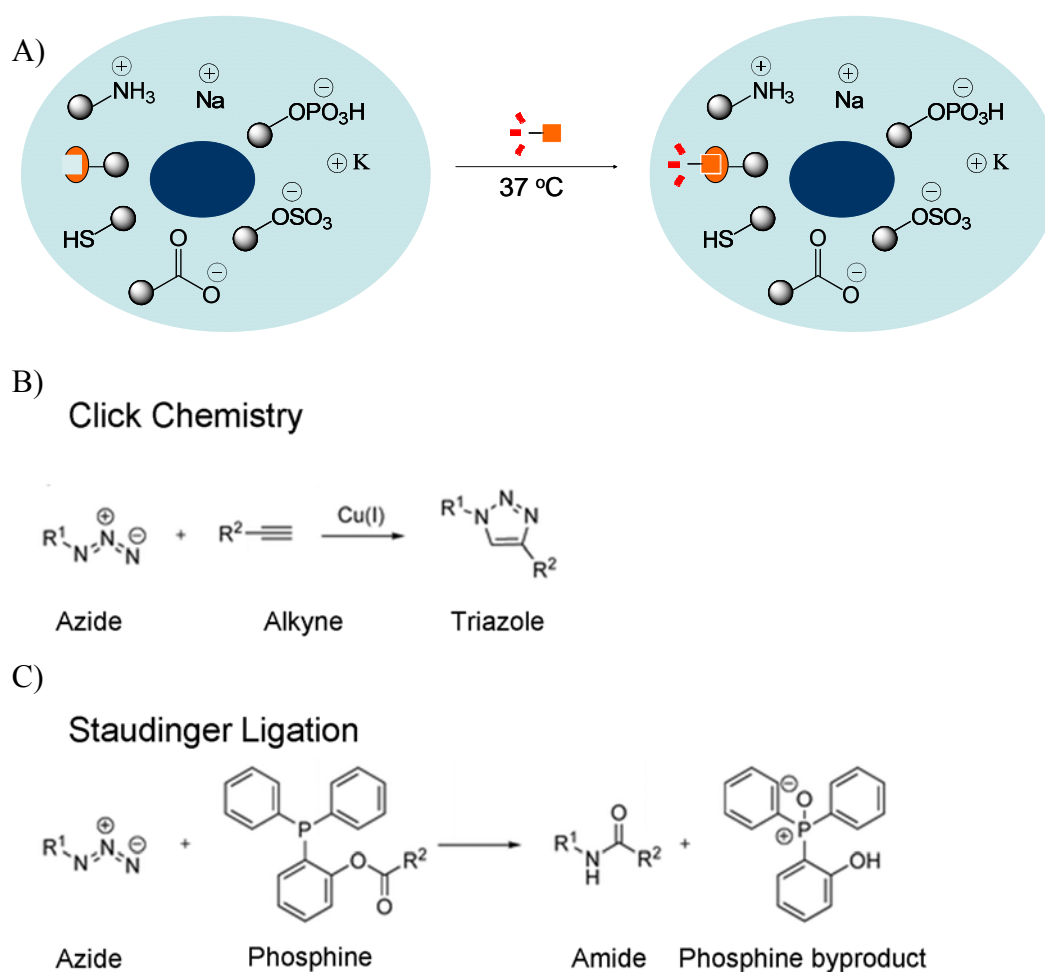
Figure 1.1 Bioorthogonal reactions commonly employed for labeling of biological molecules. A) The bioorthogonal functionality (highlighted in oval orange), reacts with its counterpart (orange square), to label a biomolecule in cell lysates or live cells or organisms. B) Click Chemistry: copper-catalyzed azide−alkyne cycloaddition C) Staudinger ligation utilizing azide and phosphine reagent.

The development of two chemoselective reactions, the Staudinger ligation (Saxon E. *et al.* 2000) and the Huisgen [3 + 2] cycloaddition or 'click-chemistry' (Rostovtsev V.V., *et al.* 2002; Tornoe C.W., *et al.* 2002; Wang Q., *et al.* 2003), has

enabled such installation. The Staudinger ligation involves the reaction between alkyl/aryl azides with methylester-modified triphenylphosphines, which results in the formation of a covalent amide bond between the two reactants Alternatively, alkyl/aryl azides can undergo Cu(I)-catalyzed [3 + 2] cycloadditions with terminal alkynes to yield triazole products Importantly, these reactions proceed with high rate in aqueous environments and exhibit extreme selectivity and biocompatibility with chemical functionality in biomolecules (nucleic acids, proteins and metabolites) and have consequently been termed 'bioorthogonal' (Prescher J.A., *et al.* 2005). The invention of these bioorthogonal reactions enables the use of alkyl azides and alkynes as small chemical reporters on chemical probes that can be subsequently converted into detection tags. This strategy has realized new *in vivo* applications for chemical proteomic methods.

1.4.2 Affinity/Activity Based Chemical Proteomic Tools

A popular approach of chemical proteomics relies on design and synthesis of chemical probes that can react with proteins specifically utilizing the distinct enzymatic mechanism of the target proteins (Activity or Mechanism Based Probes refered as ABP) or depending on its affinity to a certain class of proteins (Affinity Based Probe, refered as AfBP).

1.4.2.1 General design of an Affinity/Activity based chemical probe

In their most basic form, these chemical probes, either ABP or AfBP consist of three fundamental building blocks with distinct function (Figure 1.2): a reactive group for targeting to the conserved mechanistic or structural feature of a set of proteins, a tag or handle for identification and purification of modified proteins which we have discussed previously and a linker region that can modulate reactivity and specificity of the reactive group by connecting reporter and the reactive group

and providing enough space between reporter to the reactive group, the reactive group. In this section, we will mainly discuss the design and development of reactive groups for ABPs and AfBPs.

For ABPs, the reactive groups of many of the successful examples have been developed based on covalent, mechanism-based inhibitors (suicide inhibitors) of various enzyme families. The selective targeting of these inhibitors depends largely on the conserved mechanistic differences of each protein classes. Therefore, many of the best examples of ABPs have been designed to target hydrolases, like serine hydrolases and cysteine proteases, which have a reactive nucleophilic residue in the active site and distinct a catalytic mechanism. It may be extrapolated that, the design of ABPs for enzyme classes with known covalent inhibitors is, at least in concept, straightforward. However, not like serine hydrolases and cysteine proteases, the majority of proteins do not yet have a known covalent inhibitor attacking a conserved residue. Therefore, AfBPs adopting an alternative strategy utilizing photo reactive variants of reversible inhibitors, have been developed for such proteins, (Sieber S.A., *et al.* 2006, Saghatelian A, *et al.* 2004, Chan EW, *et al.* 2004). The major difference is that these photoreactive AfBPs achieve target selectivity through binding affinity, rather than "mechanism-based" reactivity. Covalent labeling is accomplished by exposure to UV light, which induces the photoreactive group (e.g. diazirine, benzophenone and tosyl functionality) to modify proteins in close spatial proximity to the probes. Probes for metallohydrolases—the metalloproteases (MPs) and histone deacetylases (HDACs) are standard examples of AfBPs. In addition, alternative strategy aiming to discover labeling reagents for proteins lacking known suicide inhibitor is in need. For this purpose, a combinatorial or nondirected strategy has been introduced in which libraries of candidate probes are synthesized and screened

28

against complex

Activity (Affinity) Based Probe
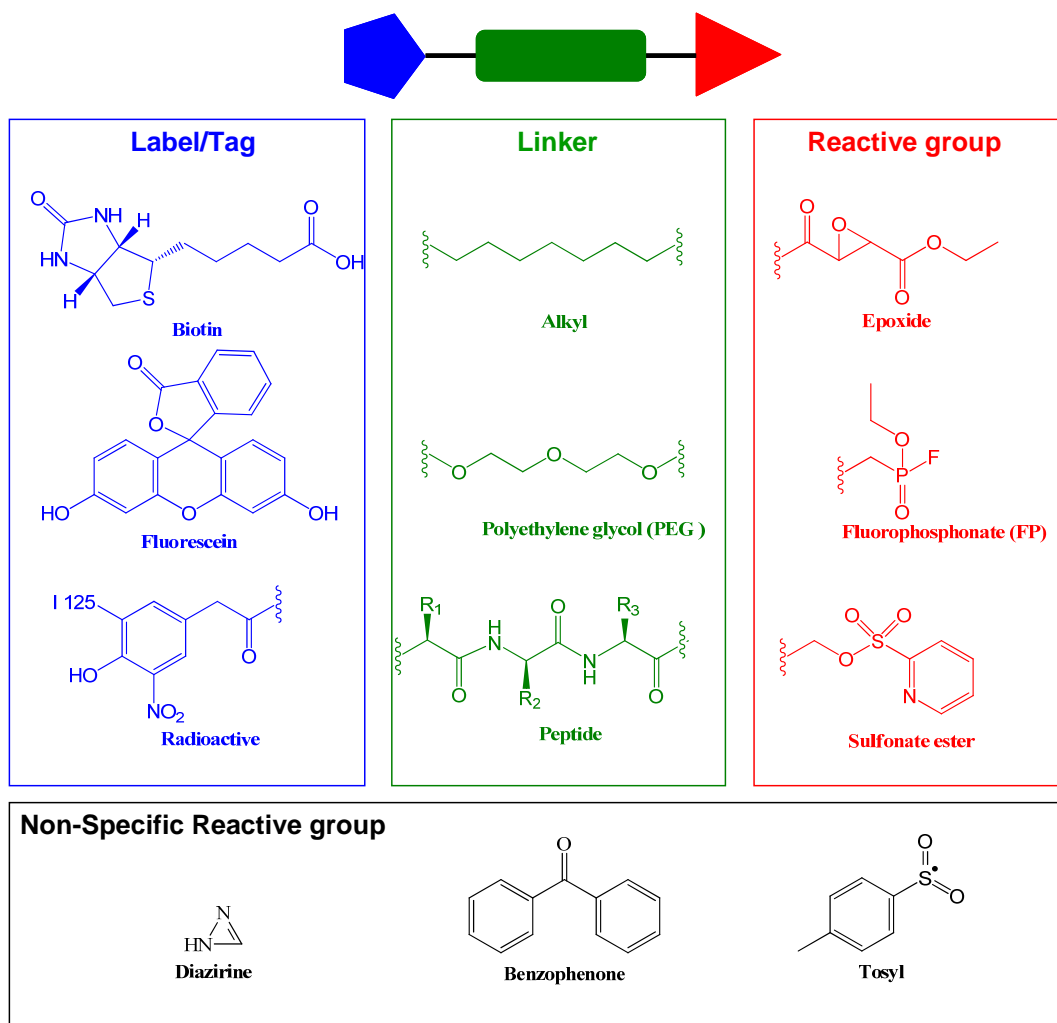
Figure 1.2 Structure of an Activity (Affinity) based chemical probe. A chemical probe has three basic components: a reactive group for covalent attachment to the protein of interest; a linker region to provide spacing and specificity; and a tag to allow for identification and/or purification.

proteomes to identify "specific" protein-labeling events, which were defined as those that occurred in native proteomes (Adam G.C., *et al.* 2001; Adam G.C., *et al.* 2002; Barglow, K. T. *et al.* 2004 Evans, M.J. *et al.* 2006).

Since the concept of chemical proteomics profiling brought up, research efforts over the past decade have bred the development of chemical probes for numerous protein classes. These probes have been widely used to offer impressive

29

insights into enzyme function in a various biological systems. Up to date, chemical probes for more than a dozen protein classes have been developed including serine dydrolases, cysteine proteases, kinases, phosphatases, glycosidases, and oxidoreductases. The following discussions summarize the development of chemical probes, either ABPs or AfBPs, for different enzyme classes.

1.4.2.2 Activity based Chemical probes

1.4.2.2.1 Activity based probes for Serine hydrolases

Serine hydrolases is a large and diverse enzyme class constitutes approximately 1% of the predicted protein products encoded by most eukaryotic genomes. This enzyme class includes proteases, peptidases, lipases, esterases, and amidases. These enzymes utilize a common catalytic mechanism involving the activation of a conserved serine nucleophile and they are most susceptible to covalent modification by many types of electrophiles. Among these electrophiles, fluorophosphonates (FPs) has emerged as a particularly powerful class of reactive group for the design of ABPs targeting serine hydrolases.

In 1997, Cravatt's group described the first method for monitoring dynamics in the expression and function of an entire enzyme family. In this study, they synthesized an active-site directed probe based on the general and irreversible serine hydrolases inhibitor FP. FP was chemically modified and a small molecule reporter group was added to make the FP probe. The key concept here is that the reactivity of FPs with serine hydrolases requires that the enzymes be in a catalytically active state. As expected, this probe is able to covalently label serine hydrolases in cell extract and report the presence of serine hydrolases using in-gel fluorescence scan. Additionally, FP probe is able to differentiate free (active) from inhibitor-bound (inactive) proteases in these samples, meaning that it covalently labels these proteins

in an activity-dependent manner**.** This probe was shown to be capable of simultaneously monitoring the activities of multiple serine hydrolases on a systems level.

1.4.2.2.2 Activity based probes for Cysteine proteases

Like serine hydrolases (SHs), cysteine protease is another huge enzyme class in the genome of both prokaryotic and eukaryotic organism. This class of enzymes shares a common catalytic mechanism that involves a nucleophilic cysteine thiol in a catalytic triad. This distinct mechanism makes them susceptible to inactivation by different electrophiles like epoxides, vinyl sulfones, diazomethyl ketones, α-halo ketones, and acyloxymethyl ketones. The first activity based chemical probes for cysteine proteases is developed by Bogyo and colleagues based on the cathepsins inhibitor E64 with an activated epoxide functionality. These reagents were shown to be able to covalently label various cathepsins and several unidentified polypeptides likely to be proteases. Among these probes, both highly selective probe reporting cathepsin B activity and broad spectrum probes simultaneously targeting multiple cathepsins were identified. These probes were then used to monitor different cysteine protease activity in primary human tumor tissue and cells derived from human placenta. These probes appended with a range of reporter tags including radioisotopes, fluorophores, and biotin were widely utilized for the functional proteomic analysis of cathepsins.

Besides for Cathepsins, chemical probes have also been developed for Clan CD cysteine proteases including caspases which play key roles in apoptosis-mediated cell death. The most popular probes targeting caspases are developed by chemically modifying the peptide acyloxymethyl ketone (AOMK) with various reporter tags. Peptide AOMK was reported for its high activity against cysteine proteases and low

reactivity with other weak nucleophiles. It inhibits cysteine proteases by alkylating the active site cysteine residue and a wide variety of peptides can be incorporated into AOMK structures allowing control of the selectivity and reactivity toward different cysteine proteases. Bogyo and colleagues chemically synthesized a library of probes with varying P1 position linked to AOMK structure and found that Asp-AOMK probe efficiently labeled caspase-3, caspase-6, caspase-7 and caspase-8 but not caspase-9. Starting from this initial library, probes with varying P2 and P3 positions were synthesized and the lead to the discovery of the Asp-Glu-Val-Asp (DEVD) substrate sequence which was most commonly used for these probes. Although this probe is optimal for caspase-3 and caspase-7, it is also efficiently recognized by several other cysteine proteases like the cathepsins and legumain. Further, to decrease the potency of these probes toward legumain, a screen for P3 amino acids that direct selectivity away from legumain was conducted. This screen thus identified the optimized the probe AB50-Cy5, containing a nonnatural amino acids biphenylalanine at P3, showing most selectivity against caspases-3 and caspases-7. Not only used for in gel analysis, this probe, producing a maximum fluorescent signal allowing non-invasive imaging, can be used to monitor the kinetics of apoptosis in live mice, whole organs and tissue extracts.

1.4.2.3.3 Activity based probes for protein kinases

Following these pioneering studies, research efforts over the past decade have bred the development of chemical probes for numerous enzyme classes. These probes have been widely used to offer impressive insights into enzyme function in a various biological systems. This section summarizes the development chemical probes for individual enzyme classes. Up to date, chemical probes for more than a dozen enzyme classes have been developed including proteases, kinases,

phosphatases, glycosidases, and oxidoreductases. Like chemical probes previously discussed for serine hydrolases and cysteine proteases, up to now at least 518 protein kinases has been determined in human genome. These protein kinases are clustered into 17 groups and 134 families. Although these kinases are diverse in their sequence and structure, they all recognize a common substrate ATP, which is the key phosphate donor in protein or peptide substrate's phosphorylation. Therefore, these kinases all share highly homologous consensus motif in their ATP binding sites. Protein kinases are main modulators of various cellular signaling pathways and have emerged as major drug targets for therapeutics. Not unexpected, the development of chemical tools to analyze the kinases has been an important focus. Not like previously discussed serine and cysteine hydrolases, protein kinases catalyze phosphate transfer from ATP to substrate by a direct transfer mechanism that does not involve covalent enzyme intermediates. Therefore, the kinase active site does not contain any unusually nucleophilic residues and the design is not as straight forward as for serine or cysteine hydrolases. Kinases do share a common conserved active-site lysines residue and this lysine residue has been the major target residue for the design of ABP probes. Therefore many kinase probes have been developed based on an appropriately placed electrophile that could efficiently react with the equilibrating deprotonated lysine.

Among these methods, Patricelli and colleagues reported a kinase probe relying on a reactive acyl phosphate-containing nucleotides, prepared from a biotin derivative and ATP or ADP (Patricelli MP, *et al.* 2007). A 6 carbon chain linker was inserted between the acyl phosphate and the biotin to distance the biotin from the nucleotide. The probe design has been proved to be successful to bind selectively and react covalently at the ATP-binding sites of at least 75% (400) of the known human

protein kinases in cell lysates. With this probe, the biotinylated kinases could then be analyzed by LC-MS/MS to determine the identity of the labeled protein and the functional state of many kinases could be profiled in parallel directly in native proteomes. Potency and selectivity of kinase inhibitors could be analyzed with competitive assay using this probe s to determine inhibitor potency and selectivity against native protein kinases as well as against hundreds of other ATPases.

Yee MC *et al.* (2005) reported a targeted kinase probe labeling phosphoinositide 3-kinase (PI 3-kinase) and the PI 3-kinase-related kinase (PIKK) families (Yee MC, 2005). This probe is based on potent and covalent inhibitor Wortmannin, which targets members of the PI3K and the PIKK families and covalently modify the lysine in the active site. In this study, biotin- and fluorophore-based probes of wortmannin were synthesized and were used to label PI3K and PIKK family members, consistent with the presence of the conserved active-site lysine. The high sensitivity of these probe to detect PI3K and PIKK family in complex proteomes allow accurate quantitation of known protein targets, as well as the identification of new targets of Wortmannin like polo-like kinase 1 (PLK1) and polo-like kinase 3 (PLK3).

1.4.2.2.4 Activity based probes for cytochrome P450

The cytochrome P450 superfamily (CYP) is a large and diverse group of enzymes. The function of most CYP enzymes is to catalyze the oxidation of organic substances. The substrates of CYP enzymes include metabolic intermediates such as lipids and steroidal hormones, as well as xenobiotic substances such as drugs and other toxic chemicals. CYPs are the major enzymes involved in drug metabolism and bioactivation, accounting for ∼75% of the total metabolism. The human genome encodes 57 distinct P450 enzymes (Guengerich F.P., *et al.* 2005) and many of these

enzymes' function are uncharacterized. Wright & Cravatt have developed ABP probes for P450 enzymes on the basis of mechanism-based inhibitor 2-ethynylnaphthalene (2EN). They modified 2EN with an alkyl (unconjugated) acetylene group which enables tagging of probe-modified enzymes with reporter tags after the proteome labeling step via the Click Chemistry. This probe has been shown to be capable of labeling multiple P450 enzymes in vitro and in vivo In addition the labeling was dependent on the presence of NADPH and was blocked by P450 inhibitors indicating its activity based nature of labeling.

1.4.2.3 Affinity based Chemical Probes

1.4.2.3.1 Affinity based Chemical probes for Metalloproteases (MPs)

Function of MPs is modulated by multiple posttranslational mechanisms in vivo, including zymogen activation and inhibition by endogenous protein-binding partners (e.g., TIMPs). These post-translaitional events complicate the analysis of MPs' function using traditional genomic or proteomic methods. Therefore the development of chemical probes to analyze its function in a complex proteome is highly desired and has been the focus of intense study. Saghatelian and colleagues reported the first chemical probe for Matrix Metalloproteases (MMPs) based on several broad-spectrum, tight-binding reversible inhibitors of MMPs (Saghatelian A, *et al.* 2004). Most of these inhibitors contain a hydroxamic acid (hydroxamate) group that tightly binds the conserved active-site zinc atom of MMPs. They replaced the hydrophobic P2 group of the hydroxamate-based inhibitor GM6001 with a photoreactive benzophenone and appended a rhodamine reporter tag onto the molecule. The resulting probe has been shown to be able to selectively label active, but not in-active forms, such as zymogen or inhibitor-bound, forms of MMPs. Later on the same group developed an advanced version of probes containing an alkyne

group in place of the bulky rhodamine reporter tag (Sieber SA, *et al.* 2006). These probes were found to display higher-labeling selectivity and efficiencies than the original probe for several MMPs.

1.4.2.3.2 Affinity based Chemical probes for HDAC

Lysine residues' acetylation and deacetylation on histone proteins is a critical event modulating transcriptional activation and repression. (Bolden JE, *et al.* 2006 ssThe deacetylation event is regulated by a class of enzyme named histone deacetylases (HDAC) which functions as parts of larger protein complexes and removes acetyl groups from the lysine residue (Marks PA, *et al.* 2007). Depending on sequence identity and motif organization, HDACs are classified into four groups. Among these Classes, Class I and II HDACs are zinc-dependent metallohydrolases and their inhibitor suberoylanilide hydroxamic acid (SAHA) has been shown to be potent and selective inhibitor in vivo and are used to induce differentiation in cancer cell lines and reduce tumor volume (Michael S. *et al.*1999). Using SAHA as scaffold, Salisbury & Cravatt developed the first AfBP for class I/II HDACs by introducing benzophenone photo-reactive group into SAHA structure to covalently label HDACs via UV irradiation-induced reaction (Salisbury C.M., *et al.* 2007). Alkyne groups were also introduced into the probe thus making it a clickable probe through click chemistry-based tagging. The resulting probe, named SAHA-BPyne, was shown to target multiple class I and II HDACs complex in native proteomic preparations. Further optimization demonstrated that SAHA-BPyne was markedly superior for profiling HDAC activities in live cells. (Salisbury C.M., *et al.*2008)

1.4.2.4 Applications of Affinity/Activity based chemical probes

Affinity/Activity based chemical probes can be used to study many aspects of proteomics from protein expression, activity to identification and even inhibitor

discovery. This section will discuss the applications of chemical probes.

1.4.2.4.1 Comparative target discovery

The most common application of chemical probes is for "target discovery" in biological systems (Figure 1.3). Target discovery is one of the primary challenges faced in drug discovery in pharmaceutical companies. One of such target identification experiment would comparatively analyze two or more proteomes by chemical probes to identify targets with differing levels of activity. The concept behind is that, if the proteomes under comparison display distinct biological properties (e.g., normal versus disease, drug treated versus un-treated), then any enzyme activities differences identified can potentially be a target that regulating these phenotypes. Such identification requires further experimentation to confirm its validity.

For example, chemical probes have been used for comparative study of enzyme activities among different cancer cell lines for target identification. This is
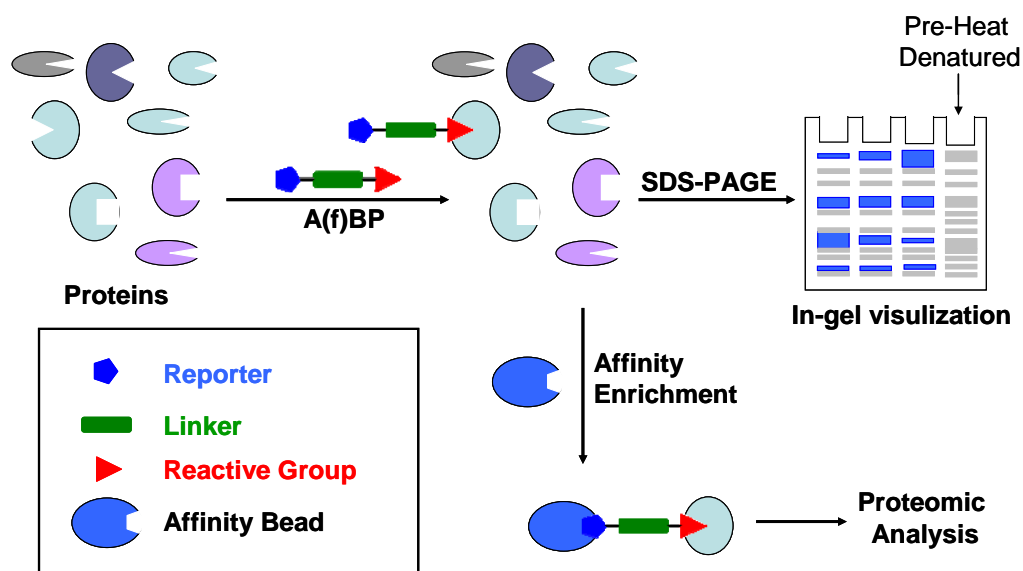
Figure 1.3 Proteomic analysis with activity (affinity) based probes. Small molecules modified with reactive electrophiles or photocrosslinkers enable covalent labeling of specific protein targets depending on the chemical scaffolds of the probes. In-gel comparative analysis of different proteome will lead to identification of proteins with different levels of activity. Alternatively, the probe-labeled proteins can be identified

by mass spectrometry-based proteomics after affinity enrichment.

accomplished by identification of enzyme activities selectively expressed by tumor cells or tissues. Such target identification provides a rich source of potential new biomarkers and targets for the diagnosis and treatment of cancer. Jessani and colleagues selected a panel of human cancer cell lines for comparative analysis using FP probes (rhodamine-tagged and biotin tagged) targeting the serine hydrolase superfamily of enzymes developed by Cravatt's group (Jessani, N., *et al.* 2002). They functionally profiled serine hydrolases activities of different subcellular localization across a panel of human breast and melanoma cancer cell lines. Specifically, secreted, membrane, and soluble fractions of cell extracts from different cancer cell lines were prepared accordingly before reaction with a rhodamine-tagged FP probe. Fluorescently labeled proteins were then separated by SDS/PAGE and visualized by in-gel scanning with a fluorescence scanner. To identify these protein targets, biotin-tagged FP probes were used to affinity enrich the active enzymes, which are further analyzed by mass spectrometry methods. Based on the serine hydrolase activity profiles of the secreted, membrane, and soluble proteomes obtained, they further hierachically clustered different cancer cell lines and identified a cluster of proteases, lipases, and esterases down-regulated in the most invasive cancer lines examined. Most interestingly, the activity of KIAA1363, an integral membrane serine hydrolase without any link to cancer made, was found to be consistently up-regulated in invasive cancer lines origin from several different tumor types. Together, this finding suggests that this enzyme may support the progression of tumors from a variety of origins and thus represent attractive targets for the diagnosis and treatment of cancer.

Another example highlights the use of such activity based chemical probes for target identification from plasmodium falciparum proteome (Greenbaum, D.C. *et*

al. 2002). In this study, Greenbaum and colleagues reported comparative study of cysteine protease activities across different malaria parasite life cycles. Firstly, with cysteine protease specific probes, they detected four cysteine protease activities in whole-cell lysates from mixed blood stages of P. falciparum parasites with the radiolabeled cysteine protease probe 125I-DCG-04. These cysteine protease activities were identified as a member of the papain family of cysteine proteases, refered as falcipains. This was accomplished by a single-step purification of all labeled proteins through the biotin affinity tag of DCG-04 followed by MS-based sequencing. Further, they used highly synchronized parasite populations to profile protease activities throughout the multiple developmental stages of the parasite and found that each of these falcipains was differentially regulated throughout different life cycles of the parasite. This distinguished activity profile suggests a primary role for falcipain 1 either in red blood cell rupture or during re infection of new host red blood cells. Further, the role of falcipain 1 in live cultures of P. falciparum were determined by treating parasites with the falcipain 1-specific inhibitor YA29-Eps(S,S). Dose dependent effect of this inhibitor at the late schizont stage suggests that falcipain 1 has a specific role in the invasion of red blood cells by extracellular merozoites and further validated its legitimate drug targets for treating the infection of malaria pasasite.

Comparing to classical expression-based genomics and proteomics, chemical proteomics strategies for target identification have multiple advantages. Firstly, it measures the activity of proteins which is a final output of all post-translational events like localization, interaction and post-translational modifications rather than the expression level. Secondly, chemical probes with biotin tag allow these reagents to affinity enrich low-abundance proteins in samples of high complexity and

therefore molecule identification of the targets is accelerated.

1.4.2.4.2 Competitive inhibitor discovery

Not only served as powerful tools for the discovering protein targets associated with distinct physiological and pathological processes, chemical probes have also been applied to discover irreversible inhibitors for certain enzyme classes such as cysteine proteases, serine hydrolases, metalloproteases and even HDACs.

This application involves chemical probe labeling of protein targets in a competitive manner (Figure 1.4), in other words, inhibitors of enzymes are identified by their ability to block probe labeling in the complex proteome (Greenbaum DC, *et al.* 2002; Kidd D, *et al.* 2001; Leung, D. *et al.* 2003).

Such competitive proteome labeling approach offers several advantages over traditional inhibitor screening assay. Firstly, enzymes are tested in their native environment, alleviating the requirement for recombinant expression and purification.
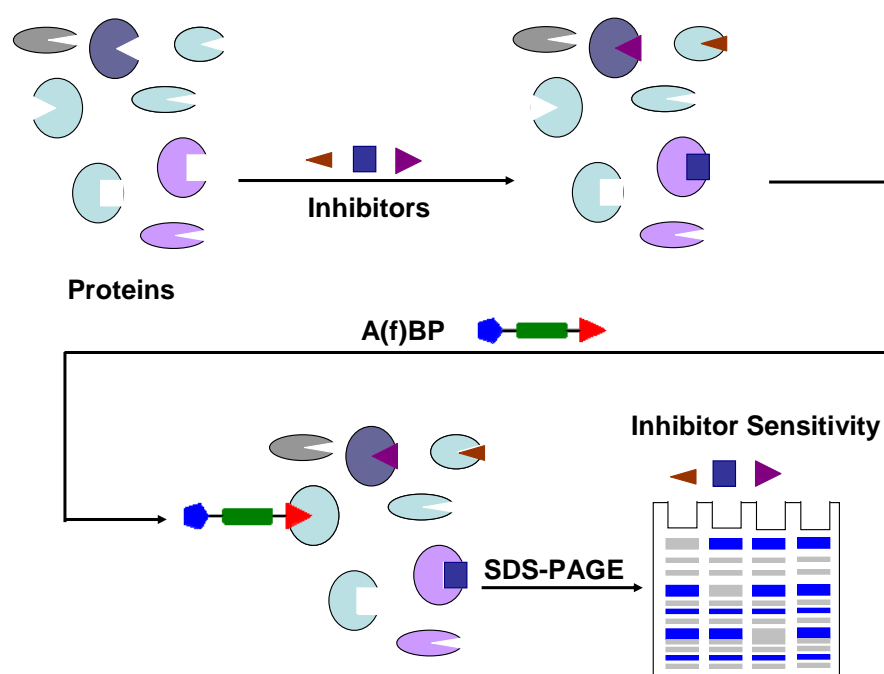
Figure 1.4 Inhibitor discovery by competitive activity-based protein profiling. The selectivity and potency of inhibitors can be determined by initial incubation of a proteome with inhibitors followed by probe treatment. Inhibitor-bound proteins are

detected by a reduction in probe labeling intensity.

Secondly, probe labeling acts as a substitution for substrate assays therefore providing a convenient activity assessment for novel enzymes lacking known substrates. Finally, potency and selectivity of the inhibitor can be simultaneously determined when broad spectrum chemical probes labels many enzymes in parallel from the complex proteome.

For instance, Leung and colleagues adapted chemical proteomic labeling to competitive screening to assess the activity of a library of candidate reversible inhibitors against serine hydrolase activities expressed in mouse tissue proteomes (Leung, D., *et al.* 2003). This approach was based on the well developed, serine hydrolase specific FP-rhodamine (rhodamine-tagged fluorophosphonate). Firstly, they reacted FP-rhodamine with mouse tissue proteomes and identified a condition that allows measurement of the extent of probe labeling for the majority of enzymes. Under this condition and at this single, kinetically relevant time point, the probe labeling of all different targets is not completed. Therefore the rate of reaction between probe and protein should be slowed down by the binding of competitive reversible inhibitors to the protein, and this competitive effect should be detected as a decrease in fluorescence signal intensity. Then each inhibitor was initially tested for its competitive effect against FP-rhodamine's labeling of multiple targets in the mouse brain membrane proteome at varied concentrations. Further, IC50 values were generated from these data for each inhibitor with corresponding target(s) in the proteome. As reported, these calculated IC50 values measured by chemical proteomics labeling, matched well with Ki values determined by conventional substrate assays. With this approach, they are able to identify nanomolar reversible inhibitors of several enzymes simultaneously, including the endocannabinoid-

degrading enzyme fatty acid amide hydrolase (FAAH), triacylglycerol hydrolase (TGH) and an uncharacterized membrane-associated hydrolase that lacks known substrates.

In addition, Greenbaum and colleagues reported another screening platform utilizing chemical proteomic labeling approach (Greenbaum DC, *et al.* 2002). In their study, this screening platform was developed with chemical probes targeting cysteine proteases in plasmodium falciparum (Falcipains). Similarly, the parasite extracts were pre-incubated with each inhibitor, followed by reaction with the general cysteine protease activity-based probe 125I-DCG-04 with a radioactive tag. The radioactive readout serves as a measurement for the potency of specific inhibitor scaffolds by comparing the ratio of labeled proteases after inhibitor treatment to an untreated control. Additionally, the specificity element of the chemical probe in this study also serves as a starting point for small-molecule inhibitor design. The inhibitor libraries was generated with a fixed single amino acid residue on a tripeptide inhibitor scaffold and variations on the remaining two positions based on the activity based probe 125I-DCG-04. This method produces a series of sublibraries made up of several hundred compounds, each having a single different fixed amino acid residue. From these sub-libraries, several falcipain 1-specific P2 amino acid residues were identified and were used to design a series of specific inhibitors. Screening of these optimized inhibitors in parasite extracts over a range of concentrations give rise to the discovery of the most selective inhibitors, YA29-Eps(S,S), showing greater than 25-fold selectivity for falcipain 1 over all other cysteine proteases. Such chemical approach with a selective inhibitor that renders a specific target protease inactive serves as an alternative of traditional gene ablation or knockout study which usually results in lethal phenotype. Therefore Such a "chemical genetic" approach allows

dissection of the protease's biochemical function. and phenotypic evaluation at specific stages within the life cycle of the parasite.

However, such competitive approaches of inhibitor discovery target only a portion of the whole proteome, thus inhibitor's selectivity determined by this approach does not represent proteome-wide specificity. Therefore, approaches that assessing inhibitor selectivity in a whole proteome level is of significant value.

1.4.3 Metabolic Incorporation Based Chemical Proteomic Tools

Other popular approaches involves metabolic incorporation of the reagent (usually chemically modified unnatural metabolites with an analyzable chemical handle) utilizing the tolerant selectivity of endogenous enzymes during the process of protein synthesis or post-translational modifications in-vivo.

Instead of utilizing the reactivity between a chemical probe and a conserved site of a protein, several groups have adopted a strategy to re-engineer metabolites to tolerate the selectivity of the protein synthesis machinery or in some cases the endogenous translational/post-translational machinery without any manipulations to incorporate a unique chemical handle during translation or post-translational modification into proteins.

This strategy involves two steps. The first step is metabolic labeling, which is very similar to traditional labeling with radioactive methionine or cysteine. In this first step, chemically synthesized, Azide/alkyne bearing metabolites are fed to cells or organisms and integrated by cellular machinery into various proteins. This azide or alkyne functionality serves to distinguish metabolically labeled protein from other proteins. In the second step, the azide/alkyne bearing unnatural metabolites are then covalently tagged with fluorescent tags or affinity tags, employing bioorthogonal reactions. either ex vivo or in vivo (Figure 1.5).
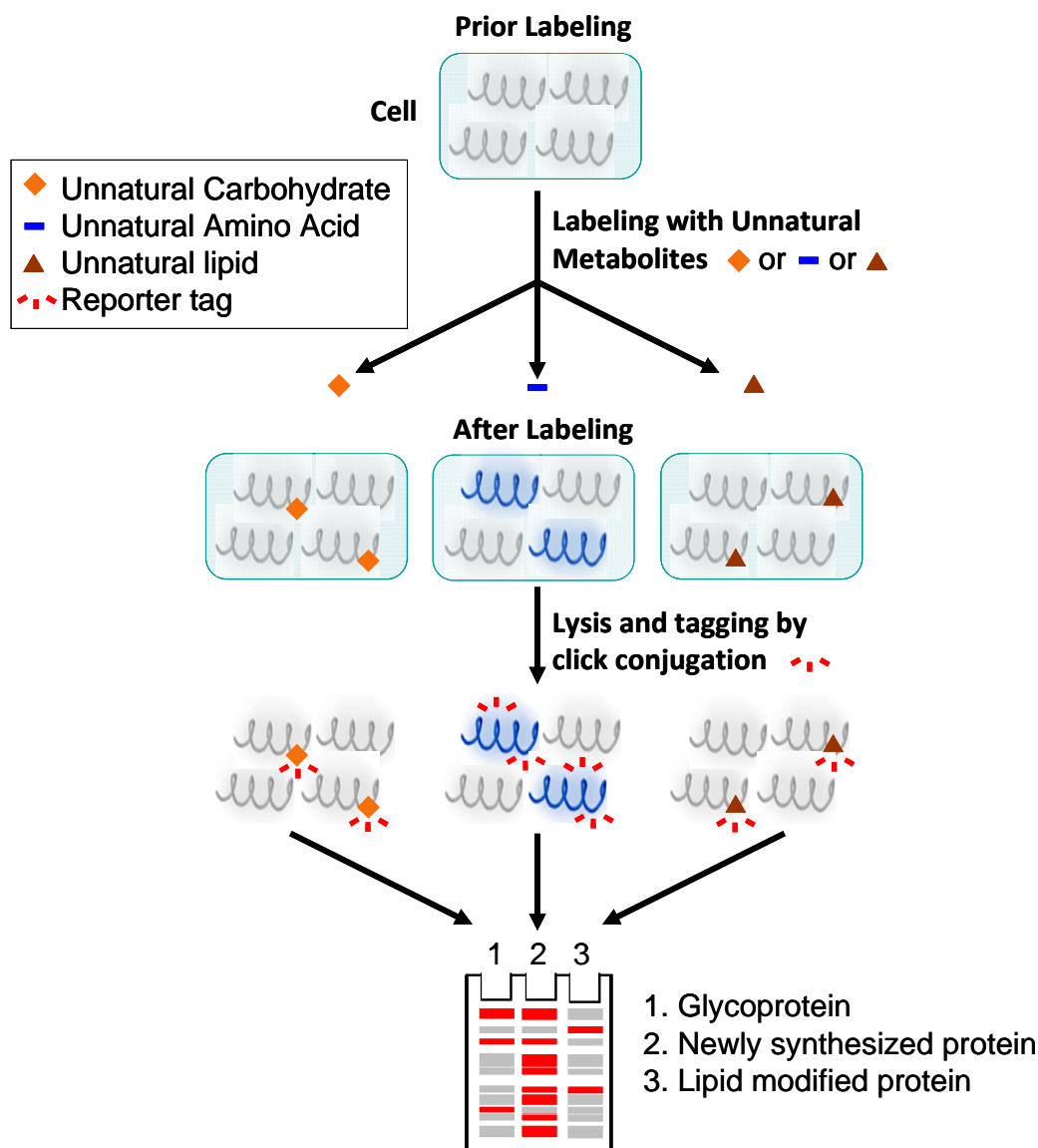
Figure 1.5 Metabolic based incorporation of chemically synthesized probes. Cells are incubated with unnatural metabolites bearing azide or alkyne functional groups. After incubation, cells are lysed and then coupled to an alkyne (or azide)-bearing reporter tag, followed by SDS-PAGE and visualization by in-gel fluorescence or avidin blot.

There are two general reaction classes that have been developed qualifying for bioorthogonal reactions: the Staudinger ligation with phosphines (Saxon, E., *et al.* 2000) and the [3+2] cycloaddition with alkynes. Two forms of the latter reaction have been described: the copper-catalyzed reaction with terminal alkynes, also called 'click chemistry'(Speers, A.E., *et al.* 2003), and the strain-promoted cycloaddition

with cyclooctynes(Agard, N.J., *et al.* 2004). Among these three reactions, click chemistry is most utilized for tagging from cell or tissue lysates because of its fast reaction rate (Agard, N.J., *et al.* 2006). However, this reaction can not be used for applications with live cells or in living animal owing to the cytotoxicity of the copper catalyst used in this reaction. Instead, the Staudinger ligation and strain-promoted cycloaddition are therefore a better and preferred choic for in vivo applications.

The following section discusses the development and application of such metabolic labeling based chemical proteomic strategies.

1.4.3.1 Metabolic Incorporation of unnatural amino acids

Tirrell and colleagues developed the two of the most widely used unnatural amino acids, Azidohomoalanine (AHA) and homopropargylglycine (HPG), both of which are susceptible to selective modification via the click chemistry reaction. They have been able to shown that the methionyl tRNA synthetase (MetRS)'s promiscuity can tolerate the chemically modified amino acids AHA and HPG in place of methionine. AHA-charged or HPG-charged methionyl tRNAs can be used by ribosomes for protein translation/synthesis in the absence of the competing natural counterpart methionine**.** With this metabolic incorporation technology, Tirrell and coworkers successfully incorporated the methionine surrogate into overexpressed proteins in Escherichia coli (Kiick, K.L. *et al.* 2002). Later, Schuman and Tirrell reported an extension of this metabolic incorporation technology introducing BONCAT (bioorthogonal non-canonical amino acid tagging) to monitor global protein synthesis in different cell types and culture systems with high specificity (Dieterich, D.C., *et al.* 2006). As reported, the presence and incorporation of AHA is non-toxic to the cell. Besides, the global rates of protein synthesis and degradation are not affected by the presence of AHA. This metabolic incorporation of AHA

endows newly synthesized protein with the azide functionalities which are covalently coupled to an alkyne-bearing affinity tag in the subsequent click chemistry tagging reaction. After affinity purification with avidin beads and on-resin digestion with trypsin, the resulting peptide mixture is subjected to tandem mass spectrometry for identification.

Moreover, this technique has been applied to a broad range of biological study. Recently, Letourneau and colleagues used AHA to visualize local protein synthesis in axonal growth cones (Roche, F.K. *et al.*2009), and the Schuman lab used the technology (which is named FUNCAT for fluorescent non-canonical amino acid tagging) to monitor basal and neurotrophin-induced protein dynamics both in time and space in hippocampal neurons. The authors also studied newly synthesized protein's dynamics by combining this technology with single particle tracking using Quantum Dots (Dieterich, D.C. *et al.* 2010). In a recent study, Tirrell and Schuman's group explore the possibility to exclusively target this residue-specific incorporation of non-natural methionine surrogates to specified cells in a complex cellular mixture. To achieve this aim, they developed a new non-canonical amino acid Azidonorleucine (ANL), which is rejected by the endogenous MetRS due to its long side-chain. And they over-expressed a mutant MetRS in a specified cell line therefore ANL is only recognized and utilized in the cell line over-expressing mutant MetRS. Therefore, de novo synthesized proteins are exclusively visualized in this specified cell in a cell selective manner (Ngo, J.T. *et al.* 2010). This cell-selective metabolic labeling approach could be applied to a mammalian context in combination with live-tagging methods using strain-promoted cycloaddition. It would be interesting to track proteome dynamics of distinct types of neurons or astrocytes in co-culture systems or even in living animals.

1.4.3.2 Metabolic Incorporation of Unnatural Monosaccharide

Apart from chemical modified, unnatural amino acids restricted to monitor global de novo protein synthesis, metabolic labeling of proteins with other classes of unnatural metabolites like carbohydrates and lipids is also well explored. Even before AHA and HPG were developed for monitoring de novo protein synthesis, Bertozzi's group implemented the metabolic labeling strategy to track glycoproteins with different, chemically engineered unnatural monosaccharides.

Metabolic incorporation of chemically modified, unnatural oligosaccharide is conceptually similar with incorporation of unnatural amino acid. It introduces subtle modifications into monosaccharide residues within cellular glycans. By taking advantage of the tolerance of a cell's biosynthetic machinery, unnatural monosaccharides can be activated with nucleotide sugars and further transported into the Golgi compartment. In the Golgi compartment, activated monosaccharides is transferred to glycoconjugates, like glycoproteins which are processed for secretion, delivery to various cellular compartments, or to the cell surface. The site of modification of the unnatural monosaccharide is represented in its biosynthetic product, permitting subsequent tagging, purification and identification.

The first example of metabolic incorporation of oligosaccharide followed by chemoselective ligation was reported by Mahal et al. They demonstrated that the modified oligosaccharide, ManLev, when fed to cells, is transformed to the corresponding sialic acid within cell surface glycoproteins. The ketone functionality is unique in the context of the cell surface and can be further tagged with bio-orthogonal reactions using hydrazine. Following the first study, Hang reported that the unnatural ketone analogue of GalNAc, in which a ketone group substituted for the N-acetyl group, is accepted by the GalNAc salvage pathway. This study

demonstrated that unnatural GalNAz (N-azidoacetylgalactosamine, refered as GalNAz)analogue can be incorporated into cell-surface glycans.

To overcome the limitation of ketone functionality that ketone condensation reactions is relatively slow at physiological pH, alternative chemoselective reactions like Staudinger ligation or copper catalyzed 3+2 cyclo-addition (click chemistry) have been used for covalent tagging of unnatural sugars incorporated biosynthetically into cell surface glycans. One of these examples is by using chemically modified ManNAc bearing an azido functional group on the N-acyl substituent, (N-azidoacetylmannosamine, refered as ManNAz), which can be converted to the corresponding sialic acid on human cells (Saxon, E. *et al.* 2000; Saxon, E. *et al.* 2002). The azide is then available for tagging with various reporters based on bio-orthogonal reactions. Afterwards, more and more unnatural oligosaccharide bearing azido groups, like GalNAz and GlcNAz, have been shown to be tolerated by corresponding salvage pathways. As demonstrated by Vocadlo, and Pamela's study, GalNAz is able to be incorporated into O-linked glycoproteins and GlcNAz is able to be incorporated into nuclear and cytosolic proteins as an analog of O-GlcNAc (Vocadlo, D.J. *et al.* 2003; Pamela V. *et al.* 2007). Rabuka and colleagues synthesized azido fucose analogues and exploits the fucose salvage pathway for the incorporation of azido analogues into fucosylated glycans in the human T lymphoma cell line Jurkat. These techniques provide a new means for proteomics analysis of glycosylation (discussed in later section).

Up to now, there are five azido sugars described for metabolic labeling of glycoproteins: N-azidoacetylmannosamine (ManNAz)1, 5, N-azidoacetyl sialic acid (SiaNAz)9, N-azidoacetylgalactosamine (GalNAz)2, 10, N-azidoacetylglucosamine (GlcNAz)11, 12 and 6-azido fucose (6AzFuc)13. ManNAz is metabolically

converted to SiaNAz.and thus used for profiling of sialic acid modified proteins. This could also be achieved through a salvage pathway using SiaNAz. GalNAz is utilized by cells as a substitute for GalNAc as the core residue of mucin-type O-linked glycans and therefore is mainly used for profiling of mucin-type O-linked glycoproteins. As GalNAc is also found in chondroitin sulfate proteoglycans and as a terminal modification of rare N-linked glycans, these glycoproteins might also be analyzed with GalNAz. In mammalian cells, GlcNAz was used to selectively label nuclear and cytosolic glycoproteins modified by O-GlcNAc. This subset of glycoproteins can also be labeled robustly with GalNAz, which is partially metabolically converted to GlcNAz in vivo. Finally, fucosylated glycoproteins can be labeled with 6AzFuc through a salvage pathway. In all cases, the azido sugar is typically administered in peracetylated form, which ensures entry into cells by passive diffusion and is followed by deacetylation by cytosolic esterases.

Gurcel, C. and coworkers have demonstrated the application of GlcNAz probe for the proteomic analysis of protein O-GlcNAcylation, a form of glycosylation that is unique to cytosolic and nuclear proteins (Gurcel, C., *et al.* 2008). This dynamic post-translational modification occurs on serine and threonine residues of nuclear proteins or cytoplasmic proteins. It is thought to modulate numerous cellular processes including transcription and translation. However, the precise functions of the 'O-GlcNAc' modification and the complete list of O-GlcNAc modified proteins are not fully uncovered. In their study, researchers used GlcNAz to analyze and identify proteins modified by O-GlcNAc in a proteomic scale. In detail, cells were cultured with GlcNAz , which was incorporated onto cytoplasmic and nuclear proteins in place of O-GlcNAc. The nucleocytoplasmic fraction of these cells was allowed to react with a biotin-phosphine capture reagent,

followed by affinity purification of labeled glycoproteins on streptavidin beads. The affinity-purified glycoproteins were trypsinized and the fragments identified by nano-high-performance LC-MS/MS analysis. O-GlcNAc-azido-tagged proteins have been identified by this approach with LC-MS/MS analysis in an MCF-7 cellular model and 14 of these proteins were previously unreported.

To achieve the same purpose, Zhao and co-workers, exploited this discovery using similar approach to capture O-GlcNAcylated proteins for a comprehensive proteomic analysis of cell lysates (Nandi, A. *et al.* 2006; Sprung, R., *et al.*2005). This global identification study confirmed 21 known, and discovered 178 novel, O-GlcNAc-modified proteins present in HeLa cells. The identification of 21 previously known O-GlcNAc-modified proteins represents almost 20% coverage of previously reported O-GlcNAc-modified proteins. This coverage serves as a good positive control and this technology enable the identification of most previously reported O-GlcNAc-modified proteins in a single study. In addition, this study nearly tripled the list of putative O-GlcNAc-modified proteins, extending the list to >200.

Not only used for cell culture, unnatural saccharides have also been used to profile glycoproteins in mice (Prescher, J.A., *et al.* 2004; Dube, D.H. *et al.* 2006). Bertozzi's group have successfully shown that both the metabolic labeling and subsequent Staudinger ligation could proceed in living mice. The obvious advantages of performing metabolic labeling in living animal is that cellular glycosylation pattern could be profiled within organs. Specifically, they administered living mice with ManNAz to probe sialic acid modified proteins in various organs. Interestingly, after lysis, tagging and separation of proteins from various organs, discrete protein bands in SDS-PAGE analysis were observed from heart, kidney and liver lysates, but not from brain or thymus homogenates. The liver is known to secrete numerous

sialylated glycoproteins and is a target of first-pass metabolism, factors that may explain the robust labelling observed in that organ. Notably, the kidney and heart lack significant levels of UDP-GlcNAc 2-epimerase (Stasche, R. *et al.* 1997), the enzyme that produces endogenous ManNAc.

Further, researchers tested the ability to tag cells coated with SiaNAz in vivo within their native physiological environment. Therefore, phosphine probes were injected into these living animals and the results demonstrated that these injected probes undergo the Staudinger ligation and accumulate on cell surfaces in an azide-dependent manner, indicating Staudinger ligation successfully function on cells within living animals. This ligation reaction displayed extraordinary chemical selectivity and the reactants displayed exquisite biological compatibility, no harmful side effects were produced and the product is not prone to metabolic breakdown on the timescale of the reaction. Successful tagging of cell-surface glycoproteins in living animals offers a method to monitor these protein modifications in a physiologically relevant system. Potentially this could be used for non-invasive imaging of cells with various glycosylation pattern. Changes in glycosylation pattern associated with different stages of organ development or disease progression could be visualized by this approach.

Further, imaging reagents that target azides are developed by different groups, reporting the development of phosphine- or alkyne-functionalized dyes that fluoresce only upon reaction with azides and thereby label cell surface glycans with low background staining (Lemieux, G. A. *et al.* 2003; Sawa, M. *et al.* 2006; Hsu, T. L. *et al.* 2007). However, these reagents are limited in its slow kinetics in living animals and toxicity caused by Copper catalyst. To overcome these problems, Bertozzi's group developed a difluorinated cyclooctyne reagent, termed DIFO. The advantage

of DIFO is that its reaction rate with azides in living animals is good and it is nontoxic in mice (Baskin, J. M. *et al.* 2007; Codelli, J. A. *et al.* 2008). This DIFO−azide reaction thus was applied to monitor the dynamics of cellular glycosylation during zebrafish embryo development due to its fast reaction kinetics. Laughlin, S. T. (2008) used Zebrafish embryos as a model system and metabolically labeled the embryo with GalNAz followed by tagging with fluorescently labeled DIFO reagents. This approach allowed them to image the fish's total O-linked glycosylation. Further, to extend this technique, they performed spatiotemporal analysis of glycan expression and trafficking by quenching unreacted azide groups on labeled cells with TCEP, pulsing the embryos with additional GalNAz, and probing the embryos with a multicolor detection strategy. Areas of rapid O-linked glycan biosynthesis including the fins, jaw, and olfactory organs showed increased labeling with the blue-shifted conjugate. Low or no toxicity of these reagents was observed with zebrafish after labeling which continues normal development. The further application of these metabolic labeling reagents might be extended to mammalian systems enabling imaging of tumors and sites of microbial infection.

1.4.3.3 Metabolic Incorporation of Unnatural Lipids

Apart from glycans, metabolic incorporation of bioorthogonal reporters, in combination with azide-alkyne ligation has also been applied to analyze lipid modified proteins. Major forms of lipid modification include N-myristoylation, S-palmitoylation or farnesylation. The identification of lipid modified proteins and characterization of their biosynthetic pathways have significant implication for our understanding of protein's role in many physiological processes.

Traditionally, lipid modified proteins in cells has been detected by metabolic labeling with radioactive (3H, 14C) lipids followed by autoradiography of labeled

proteins (Resh, M.D., *et al.* 2006). However, this approach suffers extremely long exposure times and its hazardous (Buglino, J.A., *et al.* 2008). The development of metabolic incorporation of chemical probes (azide/alkyne-modified lipids) in combination with bioorthogonal labeling reactions has provided improved methods for profiling lipid modified proteins. Conceptually similar with chemical probes based on amino acids and oligo-saccharide, an lipid analogue bearing an azide/alkyne handle can be utilized by mammalian cells lipid transferring machinary, installed onto proteins' lipid modification site and readily visualized by tagging with fluorescent or affinity reporter tag after bio-orthoghonal reactions. (Hang, H.C., *et al.* 2007; Charron, G., *et al.* 2009; Martin, B.R., *et al.* 2009)

Hang and colleagues in 2007 reported the synthesis of ω-azido-fatty acids as non-radioactive chemical probes for the rapid detection of fatty-acylated proteins in mammalian cells (Hang, H.C. *et al.* 2007). This method is based on metabolic incorporation of the fatty acid analogues bearing azide functional groups onto target proteins by cellular machinery. Following incorporation, fatty-acylated proteins are selectively tagged with biotin using a phosphine based biotin tag via the Staudinger ligation. The detection was achieved by streptavidin blot. By varing the fatty acid chain length of the azide bearing analogue, this method could be applied to the detection of both N-myristoylated and S-palmitoylated proteins in cell lysates. Based on this work, the same group further explored this technology by synthesizing alkyne bearing analogues of fatty acids with various chain lengths which is compatible with tagging by click chemistry reaction (Charron, G. *et al.* 2009). By comparing this set of probes with previous set in Staudinger ligation and click chemistry, this study reported significantly improved detection of fatty acid modified proteins using either biotin (alkyne-biotin, azide-biotin) or fluorescence (alkyne-rhodamine, azide-

rhodamine) detection tags. In addition, fatty acid modification of protein exhibited chain length-dependent protein activity and competitive incorporation with naturally occurring fatty acids. Further by comparing the labeling profile of different mammalian cell types (HeLa, 3T3, DC2.4 and Jurkat), the comparative analysis displayed significantly diverse profiles of fatty-acylated proteins among different cell types.

With this technology developed for detecting lipid modified proteins, Martin, D.O. *et al.* (2008) exploited this new technique to demonstrate the existence of several post-translationally myristoylated proteins in Jurkat T cells undergoing apoptosis. They investigated if post-translational myristoylation is a widely used modification process during the onset of apoptosis. To do this, they induced Jurkat T cell apoptosis with staurosporine (STS) and cycloheximide (CHX) and metabolically labeled cell with azido-myristate to detect post-translational myristoylation of caspase-cleaved proteins. After metabolic labeling, phosphine-Flag tagging and PAGE separation, results demonstrate that at least 15 proteins appear to be post-translationally myristoylated. Among these 15 myristoylation targets, 5 of them are not reported before and are identified as new post-translationally myristoylatable proteins (PKCε, CD-IC2, Bap31, MST3, and the catalytic subunit of glutamate cysteine ligase). In addition, detection with phosphine-Flag and in-gel fluorescence scan represents over a million-fold signal amplification in comparison to using radioactive labeling methods. These results indicate that the nonradioactive chemical detection method is useful for detection of myristoylated proteins in a proteome scale. There might be additional proteins that are post-translationally myristoylated during apoptosis and that this phenomenon might play an important role in the regulation of apoptosis.

Cravatt's group used similar chemical proteomic method to detect protein S-palmitoylation, which is a pervasive post-translational modification required for the trafficking, compartmentalization and membrane tethering of many proteins (Martin, B.R., 2009). Their method uses the palmitic analogue 17-octadecynoic acid (17-ODYA) in a bioorthogonal metabolic labeling approach. As reported, 17-ODYA can be used as a bioorthogonal probe which is metabolically incorporated on proteins' S-palmitoylation sites for profiling endogenous protein targets of S-palmitoylation. After incorporation, proteins covalently modified by 17-ODYA can be further detected via the Cu(I)-catalyzed azide-alkyne [3 + 2] cycloaddition reaction (click chemistry) to commercially available rhodamine-azide or biotin-azide reporter groups. To identify the targets of palmitoylation, the membrane fractions of Jurkat T cell proteome were reacted with biotin-azide, enriched with avidin beads and further ananlyzed by the multidimensional protein identification technology (MudPIT). From this analysis, a total of 125 palmitoylated proteins were identified in Jurkat T cells. Among these 125 proteins identified were known GTPases, G-protein-subunits, receptors as well as many new candidate fatty-acylated proteins. And many of these identified palmitoylated proteins are homologs of known palmitoylated proteins discovered in yeast. To further validate the efficacy of this method, 18 proteins, including 12 high confidence targets and 6 medium confidence targets, were overexpressed in 293 T cell and membrane fraction of the cell lysates were directly labeled with rhodamine-azide, analyzed by SDS-PAGE . Of the 18 proteins analyzed, 16 show evidence of palmitoylation in 293T cells suggesting the low false positive rate of this method in identification of S-palmitoylation modified protein in a proteomic scale. Corresponding approaches have also been developed for analysis of proteins farnesylation and geranylgeranylation as well (Kho, Y., 2004).

Although metabolic incorporation of chemical probes has lead to the discovery of a great numbers of glycoproteins or lipid modified proteins, it requires long incubation time to detect proteins of low level modification, transient modification or reversible modification, therefore, its ability to discover regulated PTM events has been severely impeded. Chemical approaches capturing the modification states in a temporal regulated time window will be of great importance.

1.5 Objectives

Chemical proteomic tools discussed in the previous section offered methods that are complimentary to 2D and MS analysis by labeling, detecting and separating various subsets of low abundance proteins, for example serine hydrolases subproteome, kinase subproteome, newly-synthesized protein sub-proteome, glycoprotein sub proteome, lipid modified sub-proteome and etc. These tools have already proved their usefulness and provided important insight for our understanding of protein activity, localization, protein de-novo synthesis and even protein's co- or post-translational modificaitions (glycosylation, lipid acylation, farnesylation and etc.). However, chemical proteomics is still a new discipline and novel chemical tools addressing more different subproteomes need be developed. Besides, there are limitations for the current chemical proteomic tools. Firstly, proteomic wide analysis with ABP/AfBP are currently not available for many protein classes. Secondly, many of ABP/AfBP's approaches target only a portion of the whole proteome, thus inhibitor's selectivity determined by this approach does not represent proteome-wide specificity. Thirdly, many PTMs are transient and sometimes even reversible preventing them to be detected by standard metabolic incorporation approaches which require long incubation time. It is the aim of this thesis to improve existing

and develop novel chemical proteomic tools targeting different aspects/components of the proteome and to expand the types of experiment could be performed by chemical proteomics. We hope this improvement and expansion of repertoire will enrich our understanding of the biology of life and the pathology of disease and offer rapid methodology for accelerating target identification, drug design and assessment.

Chapter 2.

Developing Mechanism Based Cross-Linker for Functional Profiling, Identification and Inhibition of Protein Kinases

2.1 Summary

To offer valuable information from drug target discovery, different aspects of the proteome need be addressed other than simply the level of protein abundance and activity. For instance, with the high degree of substrate promiscuity of kinases, protein phosphorylation level is not linear with only one executioner kinase's activity but many. When one protein kinase's activity is inhibited, other kinases will become a back up executioner responsible for that phosphorylation event and thus compromises the analysis. Therefore, searching for upstream kinase responsible for a known substrate in a proteomic level will provide complimentary information to kinase activity level addressed by ABPs and global phosphorylation site identification through MSMS.

In this chapter, we describe an improved mechanism-based crosslinker for functional profiling, identification and inhibition of protein kinases-substrate pairs. Based on the mechanism-based crosslinker OPA-AD reported by Shokat recently, we designed and synthesized NDA-AD. We demonstrated for the first time a general chemical approach for the identification of multiple kinase activities directly from the whole proteome. With both purified enzymes as well as kinases present in a crude lysate, we showed unequivocally that the strategy is compatible with not only

serine/threonine kinases, but also tyrosine kinases. Preliminary results indicated that the method is robust enough to crosslink endogenous kinases in mammalian cells. Our results also indicated that the approach was useful for multiplexed detection and kinase activities present in a proteome, and is amenable for potential screenings of potent and selective inhibitors of kinases in their native environments. The establishment of the highly specific and sensitive NDA-adenosine guided crosslinking reactions with desired kinase-substrate pairs in their native states represents a step forward towards the creation of novel chemical tools in cell signaling and drug discovery. Studies are in progress to extend the use of this method for the cross-linking of other kinases and their protein substrates.

2.2 Introduction

Protein phosphorylation is the most prevalent event in cell signaling (Manning, G., *et al.* 2002). Despite decades of intensive research, there exist several major challenges in the field of phosphoproteomics: 1) how does one identify new kinase-substrate pairs in the ever-expanding phosphosignaling cascades, 2) how can one detect multiple kinases in their native environment, and subsequently 3) look for potent and selective small-molecule inhibitors (Ubersax, J. A. *et al.* 2007)? Existing biological and chemical methods have offered invaluable tools for the identification of phosphorylated proteins, as well as sites of phosphorylation (McLachlin, D. T. *et al.* 2001; McLachlin, D.T. *et al.* 2003). Shokat *et al.* recently developed a mechanism-based cross-linker, *o*-phthalaldehyde adenosine (OPA-AD), that potentially allows researchers to use known phosphoproteins/phosphopeptides to identifytheir upstream kinases (Maly, D. J. *et al.* 2004). To detect kinases in their native environments, protein-based biosensors have routinelybeen used but offer

limited success (Zhang, J. *et al.* 2002). Imperiali et al. recently developed a homogeneous fluorescence-based assaythat enables multiplexed kinase detection in cell lysates (Shults, M. D. *et al.* 2005), Activity-based probes (ABPs), based on either reversible small-molecule kinase inhibitors or irreversible ATP analogues, have also demonstrated good utilities in large-scale kinase detection and identification (Evans, M.J. *et al.* 2006; Hagenstein, M.C. *et al.* 2003; Patricelli,M. P. *et al.* 2007; Blair, J. A. *et al.* 2007). None of these kinase-detecting methods, however, has thus far been expanded to the screening and identification of inhibitors against specific kinase-substrate pairs in their native states. In order to expand these strategies for multiplexed detection and inhibition of kinase-substrate pairs in a complex proteome, we aimed to improve the mechanism-based cross-linker OPA-AD reported by Shokat.

2.3 Results and Discussion

2.3.1 Design and Synthesis of NDA based Cross-linker

Based on the originally reported OPA-AD, When it was initially used in a complex proteome, we un-expectedly discovered that it produces a large number of nonspecific cross-linking bands (Figure 2.1 and 2.2) in addition to the one corresponding to the desired kinase pseudosubstrate pair (i.e., a kinase peptide substrate in which the S/T/Y phosphorylation site was replaced by cysteine). This severely limits its potential applications. Thus we aimed to modify the highly reactive o-phthaldialdehyde (OPA) moiety in OPA-AD and make it compatible with proteomic experiments. Consequently, naphthalene-2,3-dicarboxaldehyde adenosine (NDA-AD) was designed to covalently trap the transient kinase-substrate-ATP

ternarycomplex formed during the phosphorylation (Maly, D. J. *et al.* 2004). With

an adenosine moiety guiding NDA-AD to the ATP-binding
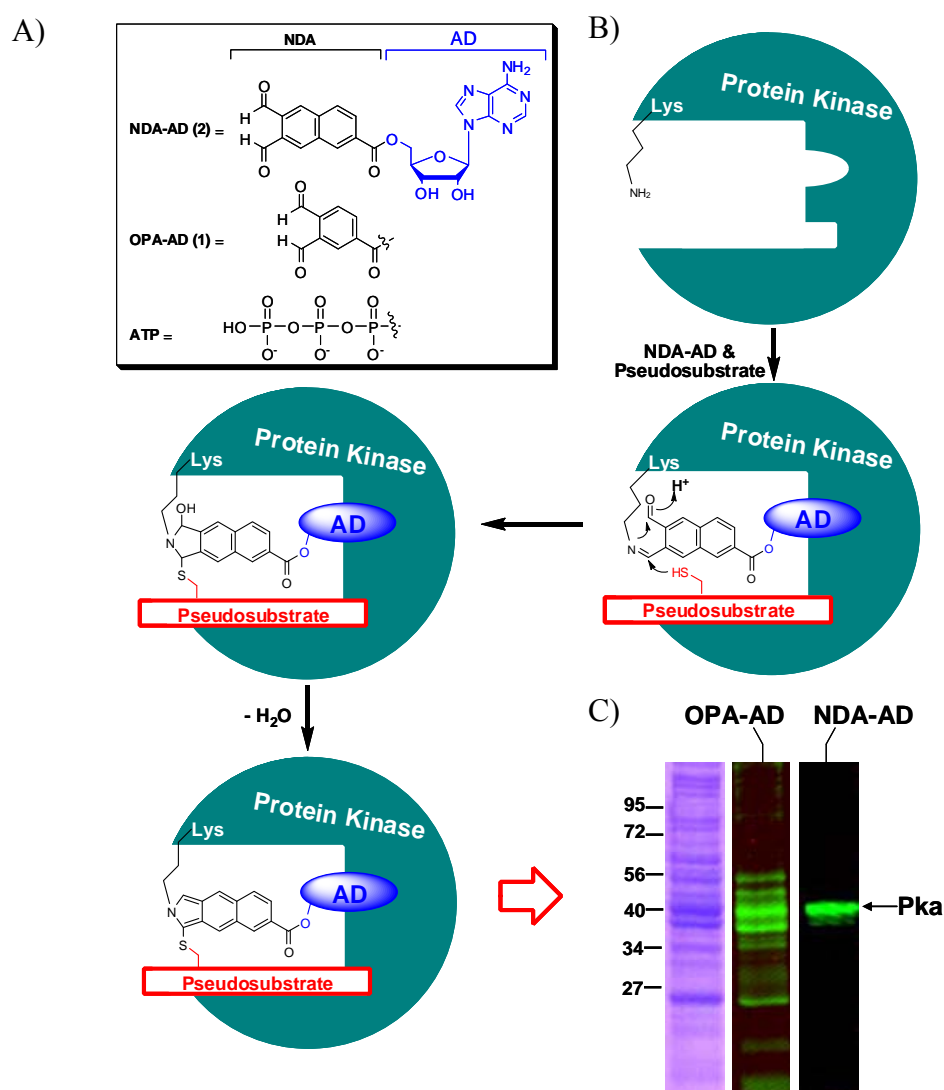
Figure 2.1. Schematic representation of the substrate−kinase cross-linking reaction. A) Structures of ATP, OPA-AD and NDA-AD. B) Scheme showing the cross-linking reaction of the kinase/pseudosubstrate/2 ternary complex. C) Improved cross-linking specificity of kinase/substrate pair in a crude proteome by 2 vs 1. The reactions were done with Pka/PKAtide under identical conditions (see Supporting Information for details). (left) Coomassie gel.

pocket of a kinase, the naphthalene-2,3-dicarboxaldehyde (NDA) group serves as a

bifunctional chemical that cross-links the proximal catalytic lysine residue (from the

kinase) and the cysteine residues (from the pseudosubstrate); this generates a stable

isoindole linkage between the kinase-substrate pair (Figure 2.1). The chemical

61

synthesis of the NDA-AD and pseudosubstrates was performed by my collaborators (see Chapter 8.2) We reasoned that NDA would cross-link kinase-substrate pairs more specifically than OPA, because of its more desirable chemical properties and better structural fit in the kinase active site.
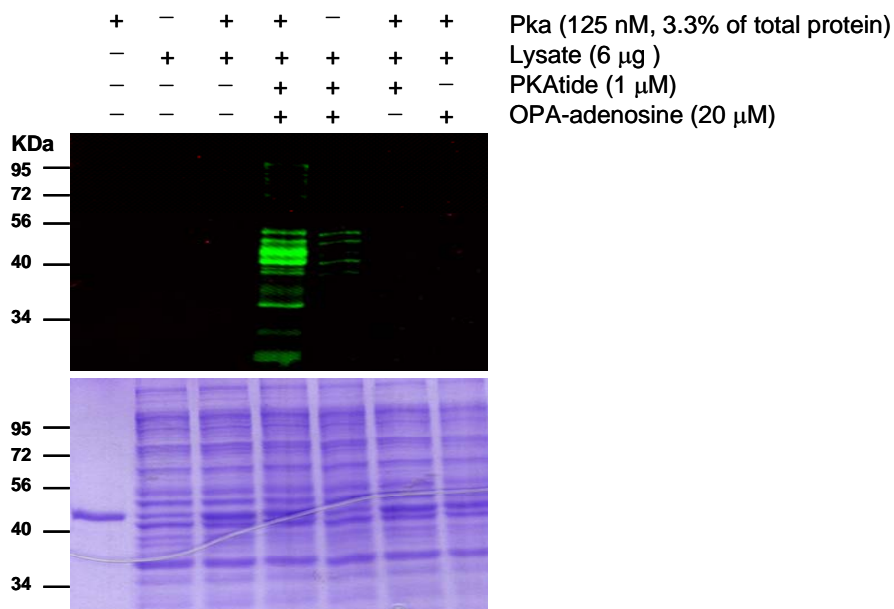
Figure 2.2 OPA-adenosine guided cross-linking of Pka/PKAtide in the crude bacterial proteome.

2.3.2 NDA-AD as a general mechanism based cross linker

We first assessed whether NDA-AD could serve as a general mechanism-based cross-linker for both Tyrosine and Serine/Threonine kinases. The cross-linking reactions were tested with a set of 6 purified kinases, of which three are Tyr kinases (Csk, Src & Abl) and the other three Ser/Thr kinases (Erk1, Erk2 & Pka).

All kinases were recombinantly expressed and tested to ensure their purity as well as enzymatic activities. Fluorescein-labelled, cysteine-containing kinase pseudosubstrates were chemically synthesized based on their known peptide substrate sequences (Table 2.1) (Casnellie,J.E. *et al.* 1991; Cheng, H.C. *et al.* 1992;

Gonzalez,F.A. *et al.* 1991; Kemp,B.E. *et al.* 1997; Sekimoto, H. *et al.* 2003). As shown in Figure 2a, incubation of each the six kinases, regardless of whether they are Tyr or Ser/Thr kinases, with their cognate pseudosubstrates in the presence of NDA-AD led to the successful cross-linking of kinase-substrate complex, as indicated by a fluorescence band on the SDS-PAGE. All three components (i.e. kinase, pseudosubstrate and NDA-AD) were necessary, as labelling was not observed in the absence of any of them. No cross-linking was seen with heat-denatured kinases, indicating the cross-linking was dependent upon the active conformation of kinases.

Table 2.1 Sequences of kinase Pseudosubstrates used in our studies.

Peptide [a]	Pseudosubstrate sequence[b]
CSKtide	Fluorescein-GG-KKKKEEI**C**FFF
SRCtide	Fluorescein-GG-KVEKIGEGT**C**GVVYK
ABLtide	Fluorescein-GG-EAI**C**AAPFAKKK
PKAtide	Fluorescein-GG-LRRA**C**LG
ERKtide	Fluorescein-GG-ELVEPL**C**PSGEAPNQ

[a] The Fluorescein-labelled, cysteine-containing pseudosubstrate for each kinase was denoted "XXXtide" after the name of each kinase. [b] The sequences of the pseudosubstrates were based on reported optimal peptide substrates for the kinases, with a cysteine mutation in the phosphorylation site (bold and in red).
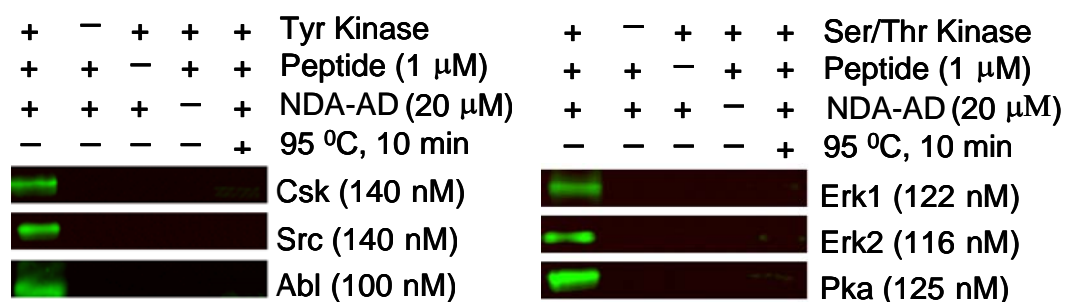
Figure 2.3 NDA-AD Cross-linking of purified kinases. Fluorescence-scanned gels showing cross-linking profiles of NDA-AD against three different Tyr kinases (left) and three different Ser/Thr kinases (right).

2.3.3 Tolerance of NDA-AD guilded cross-linking towards active-site non-specific or specific competitors

To test the tolerance of the cross-linking towards non-specific competitors like exogenous thiols or amines, Pka-PKAtide pair was incubated with NDA-AD with increasing concentrations of β-mercaptoethanol (BME) or lysine (Figure 2.4A&B). Similar to previous reports with OPA-AD, NDA-AD guided cross-linking reactions were not affected by 200-fold excess of exogenous thiols or 1000-fold excess of exogenous amines. Competition experiments were performed with ATP and LRRASLG-OH (a Pka peptide substrate); an 1000-fold excess of either ATP or LRRASLG-OH was necessary to completely block the cross-linking (Figure 2.4C&D).
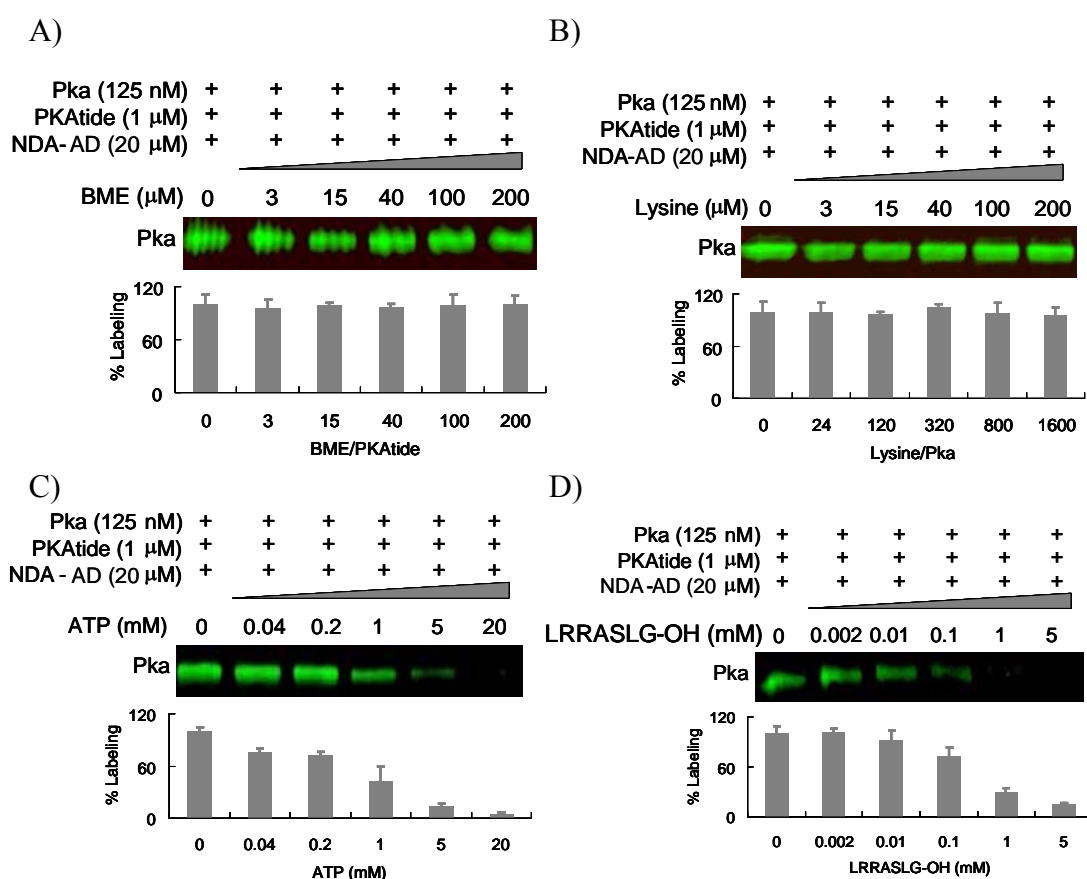
Figure 2.4. Effects of an exogenous thiol/amine/ATP/peptide susbstate on cross-linking. % Labeling = (fluorescence band intensity in the presence of competitor/fluorescence band intensity without competitor) x 100. BME/PKAtide indicates the folds of exogenous BME over pseudosubstrate. Lysine/Pka indicates the folds of exogenous lysine over kinase's active-site lysine.

2.3.4 Dose dependent and active kinase dependent nature of NDA-AD guided cross-linking

To further optimize the cross-linking reaction, dose-dependent experiments were carried out by varying the concentration of each component in the reaction (kinase, pseudosubstrate and NDA-AD). Two kinases, Pka and Csk, were chosen as model proteins. Firstly, the concentration of the pseudosubstrate in the cross-linking reaction was optimized. Varied concentrations of the pseudosubstrate peptide (0 to 1 μM) were used in cross-linking reaction. Results showed that the cross-link reaction using peptide pseudosubstrate (1 μM) gave the strongest labeling signal with minimum background (Figure 2.5A&B). Next, NDA-AD concentration was optimized. Varied concentrations of NDA-AD (0 to 20 μM) were used in cross-linking reaction. Results showed that the cross-link reaction using NDA-AD (20 μM) gave the strongest labeling signal with minimum background (Figure 2.5C&D).
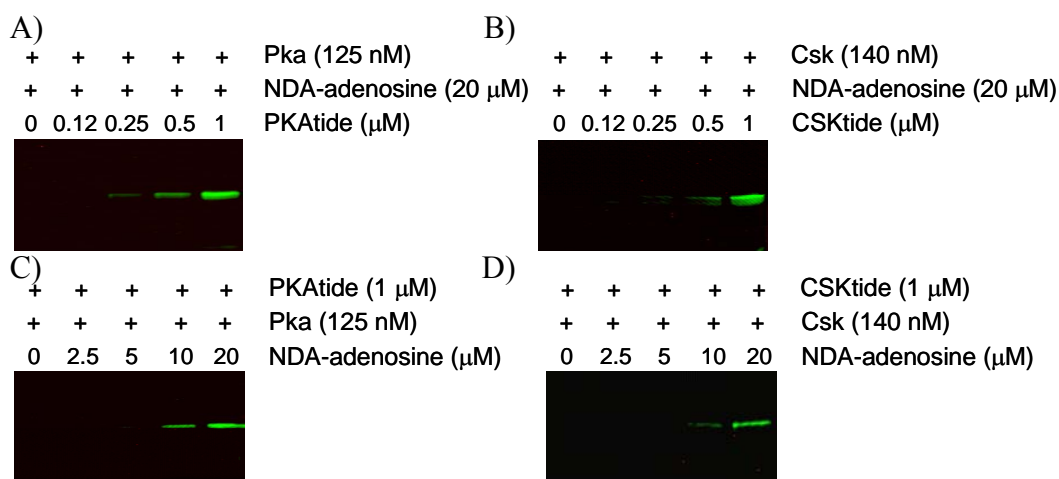
Figure 2.5 Cross-linking experiments with varied amounts of peptide pseudosubstrate, or NDA-adenosine.

To assess whether the degree of cross-linking depends proportionally on the availability of the kinase active site, different amounts of the kinase were cross-linked. Pka and Csk were taken as models. As expected (Figure 2.6), a kinase dependent increase of fluorescence band intensity was observed with increased

amounts of kinases in the cross-linking reactions, indicating that the fluorescence band intensity observed is proportional to the amount of kinases used in the reactions. As little as 300 fmol (15 nM x 20 μL) of Pka or 1400 fmol (70 nM x 20 μL) of Csk, respectively, could be easily detected. A higher amount of Csk was needed for detection was probably due to the lower enzymatic activity of this kinase.
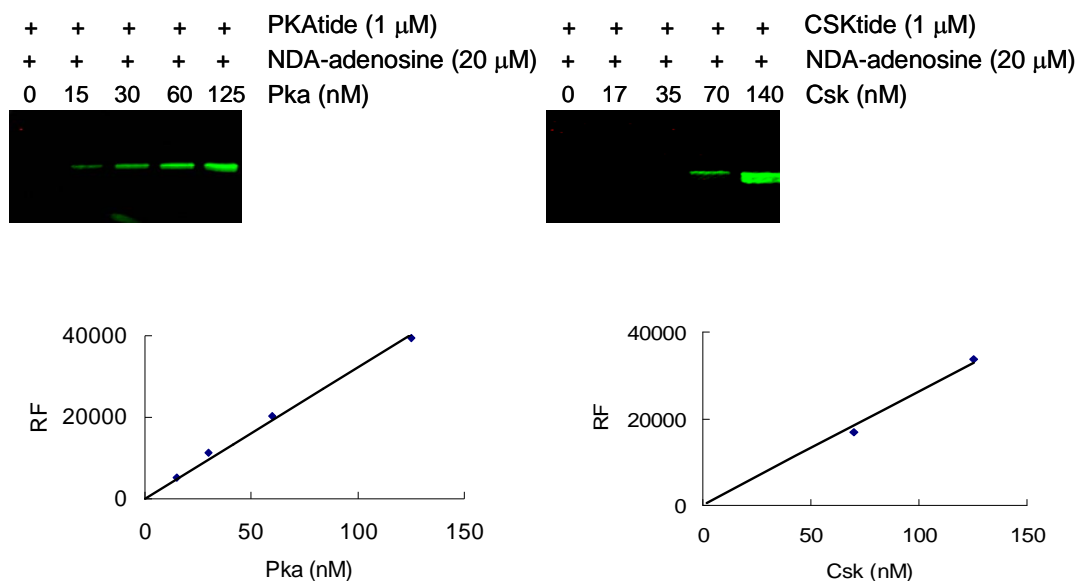
Figure 2.6 Cross-linking experiments with varied amounts of kinase.

In addition, the same amount of a kinase was cross-linked in the presence of different amounts of Staurosporine, a general kinase inhibitor (Meggio, F. *et al.* 1995). Both results confirmed the dose-dependent nature of the cross-linking against the active kinase (Figure 2.7). Also shown in Figure 2.7, the gel-based inhibition results could be conveniently plotted to generate the corresponding IC50 curves so as to obtain quantitative data of the tested kinases against ST; results gave an IC50 value of 1016 nM and 19.9 nM for Csk and Pka, respectively, which is in close agreement with previous literature values (Meggio, F. *et al.* 1995). All these lines of evidence indicate the robustness of the cross-linking reaction by NDA-AD and its

potential applications to study kinase-substrate interactions and inhibition in a complex proteome.
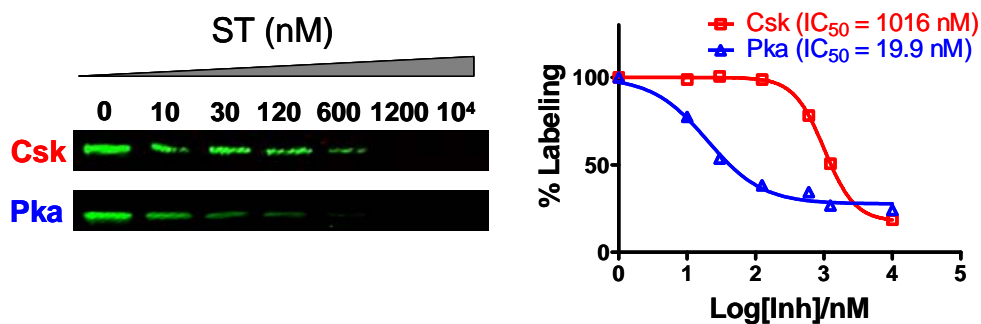
Figure 2.7 Cross-linking of NDA-AD with purified kinases in the presence of varied amounts of Staurosporine. "ST" indicates Staurosporine. "% Labeling" indicates the relative fluorescence of cross-linked kinase in the presence of the inhibitor (100% = no inhibitor). All experiments were done in duplicate.

2.3.5 Specificity of NDA-AD assisted cross-linking

To determine the specificity of the cross-linking, Csk and Pka were taken as model kinases and tested against the full set of five fluorescein-labelled pseudosubstrates. As shown in Figure 2.8, the strongest labelling was observed for a kinase with its cognate pseudosubstrate. Some noticeable "cross-talks", however, were observed between the kinase and other pseudosubstrates. For example, as much as 30-40 % cross-linking was observed between Csk and SRCtide (assume 100 % cross-linking for Csk/CSKtide). Similarly, up to 20% cross-linking was observed between Pka and SRCtide (assume 100% cross-linking for Pka/PKAtide). We wondered if this was due the well-documented promiscuous nature of the kinases in their substrate recognition.(Maly, D. J. *et al*. 2004). We therefore performed in situ phosphorylation assay using the commercially available Kinase-Glo™ Plus Kit. Both results from the cross-linking experiments and the phosphorylation assay were

compared (Figure 2.8 and Table 2.2). In general, the substrate preferences of the two kinases showed good consistency across the set of substrates (or pseudosubstrates) from the two independent kinase screening platforms. The high fidelity of 2-guided cross-linking of specific kinase-substrate pairs again indicates the potential of its application in real proteomic experiments.
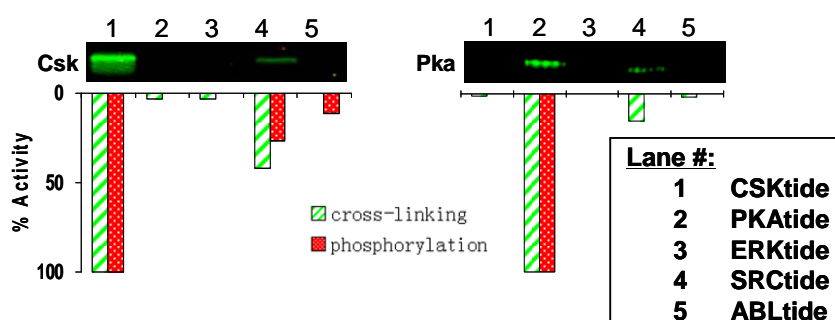
Figure 2.8 Specificity of NDA-AD guided cross-linking. (top) Fluorescence gel images of Csk (left) and Pka (right) cross-linked with 5 pseudosubstrates. (bottom) Comparison of kinase substrate preferences as determined by cross-linking and phosphorylation assay. "% Activity" indicates the relative extent of cross-linking or phosphorylation of a kinase vs different substrates (100% = kinase with its cognate substrate).

Table 2.2. *In vitro* kinase phosphorylation towards different substrates

substrate specificity determined by in vitro phosphorylation assay						
Kinase		Peptide				
		CSKtide	PKAtide	ERKtide	SRCtide	ABLtide
Csk	Phosophate transfered [a] (nmol)	0.9	0	0	0.239	0.096
	% Activity [b]	**100%**	0%	0%	27%	11%
Pka	Phosophate transfered (nmol)	0	1.84	0	0	0
	% Activity	0%	**100%**	0%	0%	0%
[a] the amount of phosphate transferred to peptide substrate in 20 min, at rt, in 20 µL reaction [b] % Activity indicates the reactivity obtained with certain kinase/substrate pair divided by the reactivity obtained with target kinase/ substrate pair.						

2.3.6 NDA-AD assisted cross-linking in crude proteome

We next assessed the NDA-AD guided cross-linking of specific kinase-substrate pairs in a crude proteome. Lysates from the E. coli DE3 strain were used as the crude proteome and were spiked with the target kinase (i.e. Csk or Pka). Subsequently, the corresponding fluorescein-labelled pseudosubstrate together with NDA-AD was added. For comparison, identical experiments were performed with 1,
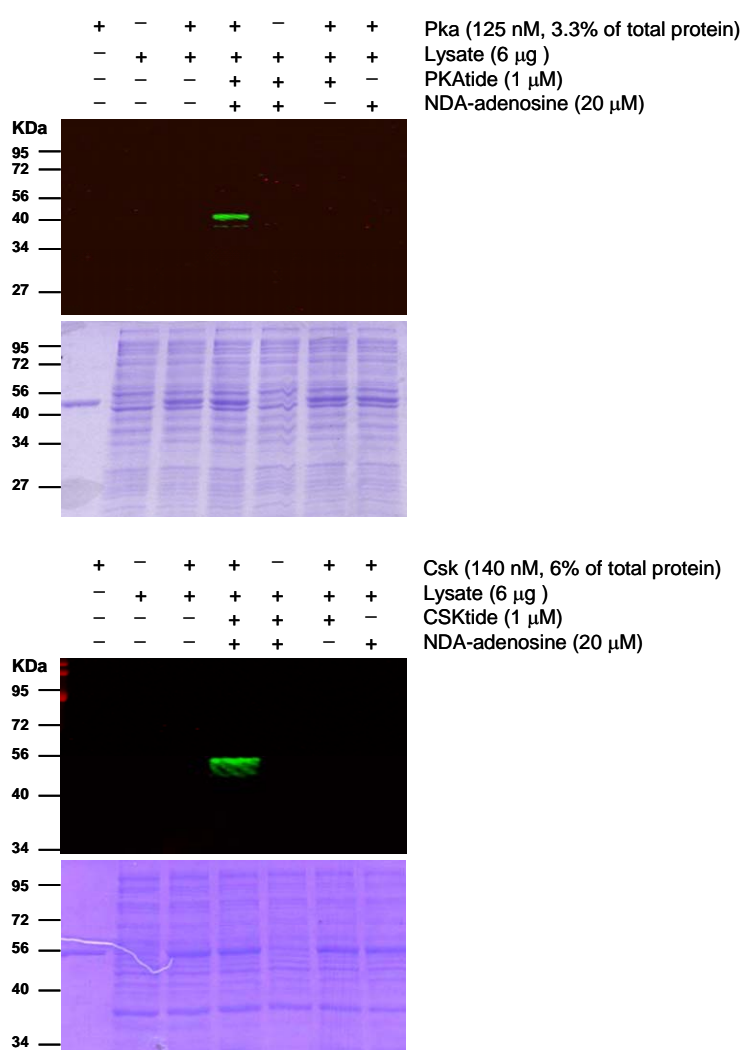
Figure 2.9 NDA-AD guided cross-linking of kinase-substrate pair in a bacterial proteome (A) Pka/pseudosubstrate pair, and (B) Csk/pseudosubstrate pair. Both fluorescence (top) and Coomassie gels (bottom) were shown. The only fluorescent band observed in each gel was that of the desired cross-linked pair.

OPA-AD. As shown in Figure 2.9, highly specific cross-linking between Csk/CSKtide and Pka/PKAtide pairs was observed with NDA-AD, but not with OPA-AD (Figure 2.2), in the presence of the bacterial proteome. It should be noted that no known kinases are present in the E coli DE3 proteome.

Further, to eliminate gel-to-gel variation of labeling results with OPA-adenosine and NDA-adenosine, the cross-linking reactions were set up side by side, under identical conditions, and the samples were resolved on a single gel (Figure 2.10). Results unambiguously confirmed that, compared to its OPA-adenosine counterpart, NDA-adenosine consistently produced much higher quality labeling profiles (in terms of specificity, labeling intensity and background labeling).
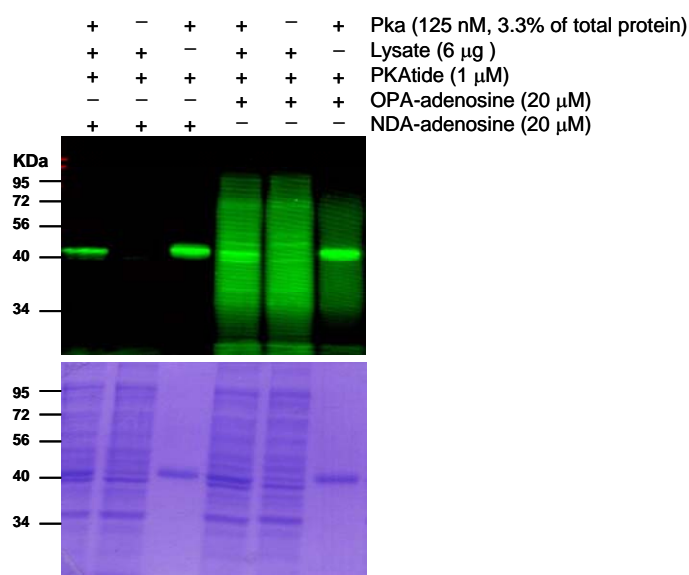
Figure 2.10 Comparison NDA-AD and OPA-AD guided crosslinking in bacterial proteome on the same gel.

2.3.7 Detection limit of NDA-AD assisted cross-linking in crude proteome

The detection limit of the cross-linking in the crude proteome was further determined by incubating NDA-AD, the pseudosubstrate and serial dilutions of the

kinase; as little as 20 ng (0.3 % of total proteome) for Pka and 200 ng (3 % of total proteome) for Csk were detectable (Figure 2.11). The 10 folds higher in the detection limit of Csk could be attributed to its intrinsically lower enzymatic activity (Sondhi, D. *et al.* 1998). Our results thus provide the first documented example of cross-linking kinase-substrate pairs in a proteomic experiment.
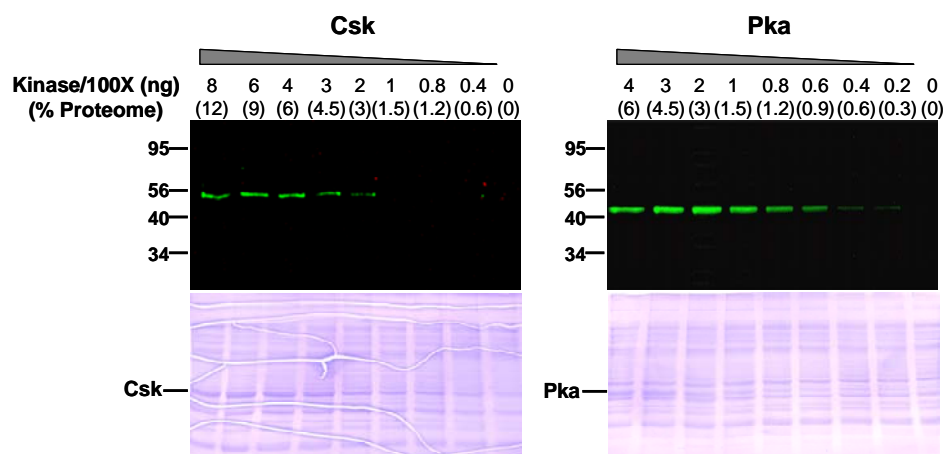
Figure 2.11 Detection limit of NDA-AD guided cross-linking in bacterial proteome Upper panel: Fluorescence gels profile showing cross-linking in bacterial proteome; Lower panel: commassie stain of mammalian proteome spiked with different % of a kinase.

2.3.8 NDA-AD assisted cross-linking for detecting potential kinase inhibitors

Many kinases are proven therapeutic targets. Yet highly potent and specific inhibitors against select kinases are relatively scarce due to their propensity to inhibit multiple kinases (Baselga, J. *et al.* 2006; Knight,Z.A. *et al.* 2005). Therefore, much effort in kinase research has been spent on developing strategies capable of rapidly screening potential kinase inhibitors that address issues related to efficacy, selectivity and safety (Karaman, M.W. *et al.* 2008). Most kinase inhibitors developed thus far have been identified from in situ assays involving the use of recombinant enzymes. As a result, off-targets of the inhibitors often escape unnoticed. Recent advances in

activity-based protein profiling (ABPP) have shown that, with suitably designed activity-based probes targeting specific enzymes, one can screen inhibitors against multiple enzymes in their native environment, thus ensuring both the potency and selectivity of these inhibitors be concurrently evaluated (Leung, D. *et al.* 2003). In our study, the highly efficient and specific cross-linking ability of NDA-AD has presented a good opportunity for the strategy to be used for inhibition studies of kinases in the crude proteome. To confirm this, we performed dose-dependent inhibition of Csk/CSKtide and Pka/PKAtide pairs in the bacterial proteome. As shown in Figure 2.12, a concomitant decrease in the fluorescent band was observed with increasing amounts of Staurophorine. The cross-linking results were further quantified and plotted to generate the corresponding inhibition curves. The obtained IC50 values (6.49 μM and 40.24 nM for Csk and Pka, respectively) were only 2-5 times higher than that obtained with pure enzymes (Figure 2.7).
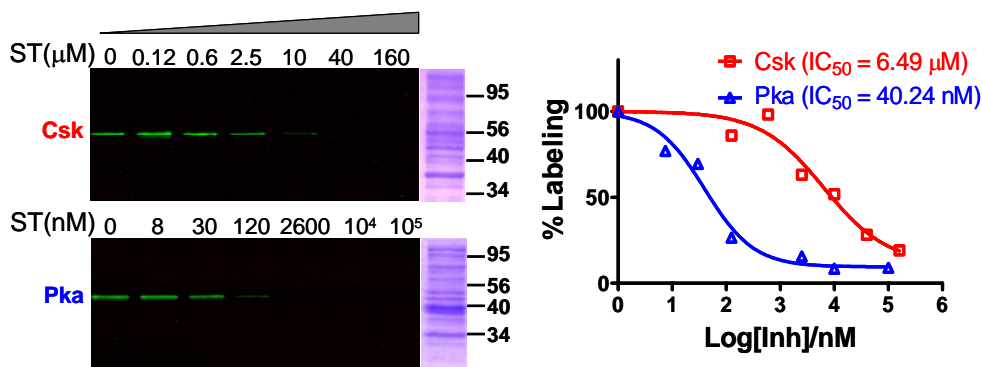
Figure 2.12 Detection of kinase inhibition in a complex proteome assisted by NDA-AD guided cross-linking. Left: Kinase Inhibition by ST in a complex proteome; right; the corresponding IC50 curves (right). "% Labeling" indicates the relative fluorescence of cross-linked kinase in the presence of the inhibitor (100 % = no inhibitor).

2.3.9 Multiple kinase detection assisted by NDA-AD guided cross-linking

Next, we assessed whether the cross-linking strategy could be used for multiplexed detection of kinase activities, as well as their inhibition, under their native environment. Specifically, we asked whether this approach could selectively identify the activity of one kinase in the presence of another, and whether the quantification of the activity of one kinase would be affected by another more abundant kinase. As shown in Figure 2.13, by spiking increasing amounts of Pka (0 to 200 ng) and a fixed amount of Csk (400 ng) into the same bacterial proteome, followed by cross-linking with 2 and either PKAtide or CSKtide individually, or together, we were able to detect highly specific cross-linking of Pka/PKAtide and Csk/CSKtide pairs separately and/or in a multiplex manner. Furthermore, the presence of the more abundant Csk (400 ng in this case) did not appear to have any noticeable effect on the detection of the lowly abundant Pka (40 ng in lanes 2). This is a crucial feature of multiplexed experiments in a crude proteome where multiple endogenous kinases may be inevitably present at different expression levels (Shults, M. D. *et al.* 2005). Next, to assess how the multiplexed kinase detection strategy could be extended for inhibitor discovery, the above experiments were repeated with increasing amounts of Staurosporine (Figure 2.13); between 10 to 500 nM of ST was sufficient to completely abolish the Pka/PKAtide cross-linking, consistent with its earlier determined IC50 of ~ 40 nM against Pka. Partial cross-linking of Csk/CSKtide pair was still observed even with 25 μM of ST, further confirming the poor potency of ST against Csk. Taken together, results herein indicate multiplexed kinase detection and inhibition in a complex proteome is possible with NDA-AD and

our reported strategy. Studies are underway to further validate these findings with more kinases and inhibitors.

A)

B)
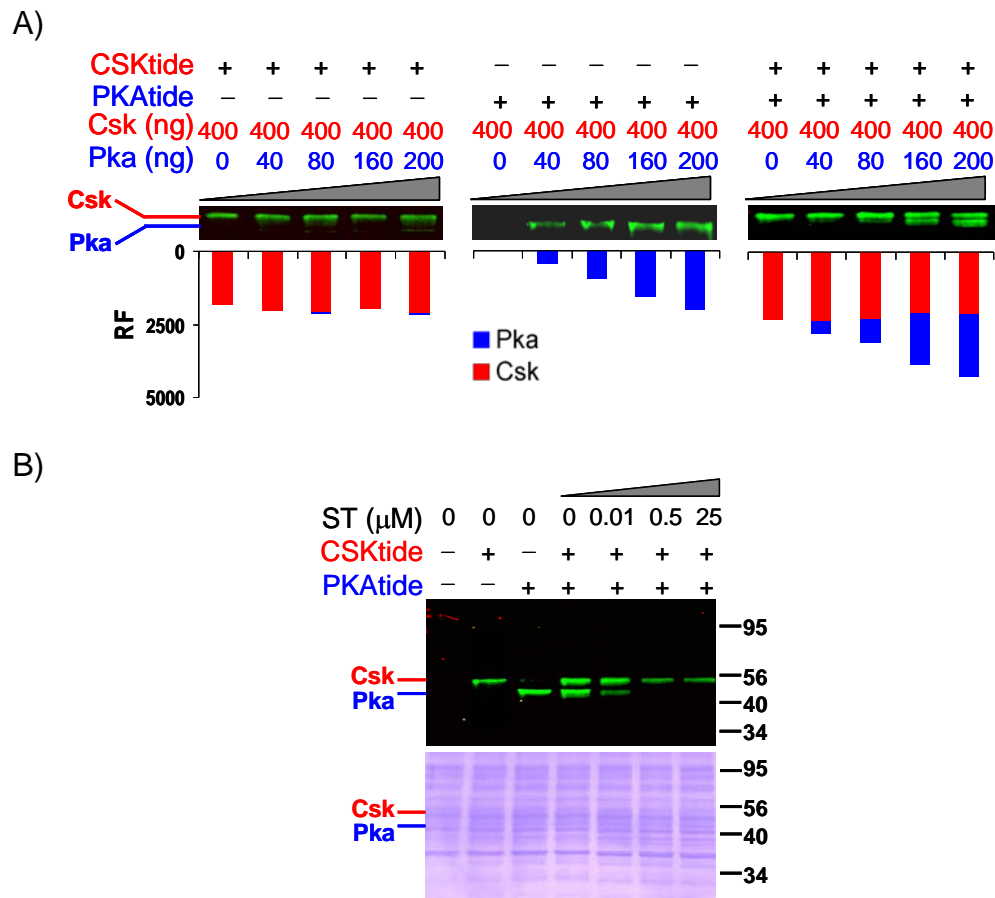
Figure 2.13 NDA-AD-guided multiplexed kinase detection and inhibition A) Multiplexed kinase detection in a complex proteome. An increasing amount of Pka (0 to 200 ng) and a fixed amount of Csk (400 ng) were used. Gels (left to right) represent cross-linking with CSKtide, PKAtide or a CSKtide/PKAtide mixture, respectively. Bottom graphs: quantification of fluorescent bands in lanes above. B) Multiplexed kinase inhibition in a complex proteome with increasing amounts of Staurosporine (ST).

2.3.10 NDA-AD assisted cross-linking of endogenous kinase in mammalian proteome

Finally, we tested whether NDA-AD guided cross-linking experiments could be extended to the detection of kinases expressed endogenouslyin cells. Pka was again used as our model kinase as it is ubiquitously expressed in most mammalian

cells. To do this, CHOK1 lysate was treated with NDA-AD and PKAtide, followed by SD S-PAGE analysis, fluorescence scanning, and Western blotting with an anti-Pka antibody. As shown in Figure 6 (left), a number of cross-linked bands were detected, one of which was attributed to the endogenous Pka and was unambiguously confirmed by both Western blot (right) and Pka-spiked experiments. Other fluorescent bands were likely cross-linked products between other endogenous kinases and PKAtide. This is possible because PKAtide might be the substrate of multiple kinases in addition to Pka).(Casnellie,J.E. *et al.* 1991; Cheng, H.C. *et al.* 1992; Gonzalez, F.A. *et al.* 1991; Kemp,B.E. *et al.* 1997; Sekimoto, H. *et al.* 2003). Work is underway to further characterize the identity of these unknown bands and results will be reported in due course.
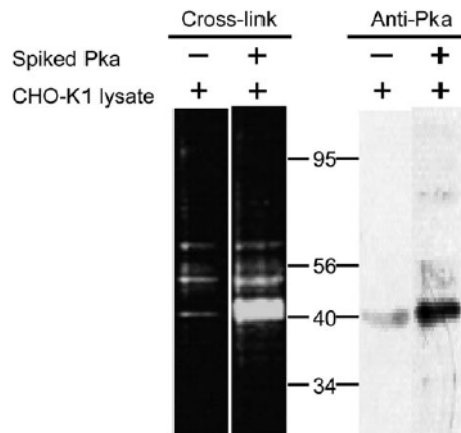
Figure 2.14 Detection of endogenous Pka expression in CHO-K1 cell lysate by NDA-AD guided cross-linking. CHO-K1 cell lysates (6 mg), with or without spiked Pka (400 ng), pseudosubstrate (2 mm), NDA-adenosine (20 mm) in the reaction buffer were incubated for 40 min at room temperature before SDSPAGE analysis. After fluorescence scanning, proteins were transferred to PVDF membrane and probed with anti-Pka antibody.

2.4 Conclusion

To conclude, by using an improved mechanism based cross-linker, we demonstrated here a general approach for identifying multiple kinase activities directly from the whole proteome. Given the fact that peptide pseudosubstrates of any sequences could be synthesized easily, and the increasing information available in phosphor-proteome database, this approach would be compatible for detecting activities of any diverse sets of kinases. This platform is also useful in a competitively inhibition profiling assay for the discovery of potent and selective inhibitors for kinases. The establishment of the NDA-adenosine guided kinase/peptide cross-linking reaction and the initial application represents a new step toward the creation of novel chemical tools targeting phosphorylation signalling and drug discovery. Study is still in progress to extend the use of this new cross-linker for the cross-linking of kinases and their protein substrates.

Chapter 3.

Functional Profiling, Identification and Inhibition of Plasmepsins in Intraerythrocytic Malaria Parasites

3.1 Summary

Another obvious challenge is that, more classes of proteins should be covered by expanding the proteome coverage of ABPs or AfBPs. For example, aspartic proteases do not have chemical probes that are selective enough for comparative proteomic analysis and competitive inhibitor analysis. Small molecules known to interrogate with this family of proteins need be converted to chemical probes. And this process need not only the synthesis effort of these probes but more effort in optimization in probe design, candidate probe screening and even tuning of compatible analysis platform. This chapter describes the development of first affinity-based probes for functional profiling of all 4 plasmepsins (PMs) in intraerythrocytic malaria parasites. Subsequent in situ screening of parasites with these probes has led to the identification of a compound, G16, which show good inhibition against all 4 PMs and parasite growth in infected red blood cells (RBCs). Our finding indicates that feasibility of using ABP approaches for identification of inhibitors against less-characterized enzymes (i.e. Histoaspartic Protease, or HAP). We anticipate that these new chemical tools should facilitate discovery of new parasite biology and new anti-malaria drugs.

3.2 Introduction

Malaria is a global disease, affecting 300-500 million people annually and killing 1-2 million. The most deadly form of the disease is caused by the pathogen Plasmodium falciparum. Currently, quinolines and antifolates are the most common anti-malaria drugs (Snow, R.W. *et al.* 2005). The cost of the drugs, as well as the emergence of multidrug resistance, is however a major problem, highlighting the need for new drugs against this devastating disease. P. falciparum has two sexual and asexual stages of growth. The human asexual erythrocytic phase (blood stage) is the cause of most malaria-associated pathology. Upon invasion of RBCs, the parasites differentiate (ring stage), metabolize hemoglobin (trophozoite stage), and replicate (schizont stage) over the next 48 hours before being released (by rupture of the host cell) into the blood stream. Proteases, including cysteine (i.e. falcipains) and aspartic proteases (i.e. plasmepsins, or PMs), are required for parasite growth through digestion of human hemoglobin and delivery of necessary nutrients. They are promising anti-malaria targets (Boss, C. *et al.* 2006). Genomic data obtained for P. falciparum predict at least 10 genes encoding aspartic proteases, four of which (PM-I, -II, -IV and HAP) have been validated and are found in the food vacuole of the parasite. The functional redundancy of these PMs calls for inhibition of all four of them in order to effectively kill the parasite (Fidock, D. A. *et al.* 2008; Rosenthal, P. J., *et al.* 2004; Goldberg, D. E. *et al.* 2005). At present, most inhibitors developed are only effective against selected PMs (Liu, J. *et al.* 2006; Drew, M. E. *et al.* 2008). This is due to difficulties associated with recombinant expression and insufficient biochemical characterizations of certain PMs (i.e. PM-I and HAP) in vitro, and the lack of methods that allow simultaneous screening of PMs' activity in situ (Hof, F. *et al.* 2006; Nezami, A. *et al.* 2003). So we aimed to develop chemical proteomic tools

for functional profiling of activities of multiple PMs in the crude proteome of intraerythrocytic malaria parasites.

3.3 Results and Discussion

3.3.1 Design and synthesis of AfBPs and inhibitors library

Previously, activity-based probes (ABPs) were successfully used for in situ screening of malaria cysteine proteases (Ersmark, K. *et al.* 2006). We report herein the first chemical proteomic approach for functional profiling of all 4 PMs in intraerythrocytic malaria parasites. This could be made possible by the development
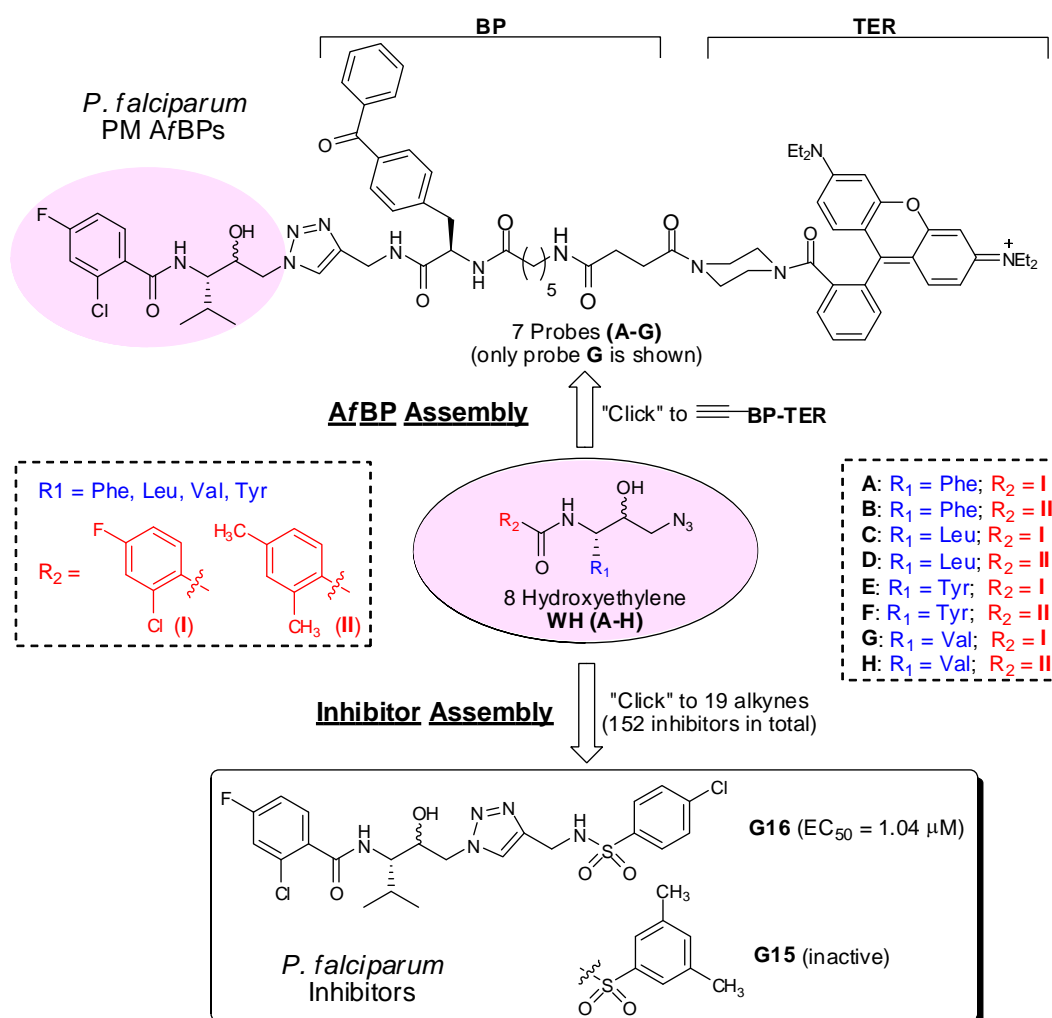
Figure 3.1. Assembly of affinity-based probes (AfBPs) and the 152-member inhibitors against all four plasmepsins in P. falciparum.

of affinity-based probes (AfBPs) against PMs (Figure 1) (Xiao, H. *et al.* 2006; Xiao, H. *et al.* 2007). In situ screening of PMs with these probes against a focused library of 152 hydroxyethyl-containing small molecules has led to the identification of G16 as an effective inhibitor against the parasite in infected RBC cultures. Unlike other known aspartic protease probes, which were based on specific inhibitors against their respective targets (e.g. presenilin and   β-secretase), we aimed to establish a general approach that could be applicable for a variety of aspartic proteases. As shown in Figure 3.1 (top), the 7 AfBP probes (A-G), each containing a hydroxyethyl-based warhead (WH) with a varied P1 and P2 group (assuming the orientation of probe-enzyme was as drawn and Figure 9.1), were assembled from the corresponding azide-containing WH and the alkyne, which contains a benzophenone (BP) photo-crossing unit and a tertraethylrhodamine (TER) reporter (Figure 9.3), using click chemistry. Hydroxyethyl-containing scaffolds are general transition state analogs of aspartic proteases. In probes A-G, aliphatic/aromatic groups were strategically chosen since they are preferred in three of the four PMs (PM-I, -II & -IV; HAP is not well-characterized) (Liu, J. *et al.* 2006; Drew, M. E. *et al.* 2008). Other aspartic proteases may be targeted by structurally tuning the WH in future. The use of click chemistry for efficient chemical assembly of complex ABPs was well documented (Chan, E.W.S. *et al.* 2004). In our case, it also provided rapid access to the 152 hydroxyethyl inhibitors against the PMs (A1-H19; see Figure 9.1, 9.2 for complete structures). The 8 hydroxyethyl WHs were chemically synthesized. Upon "click" assembly of the probes and the inhibitors, they were further characterized and purified (where necessary).

3.3.2 Labeling of recombinantly purified aspartic proteases by AfBPs

3.3.2.1 UV initiated labeling by AfBPs
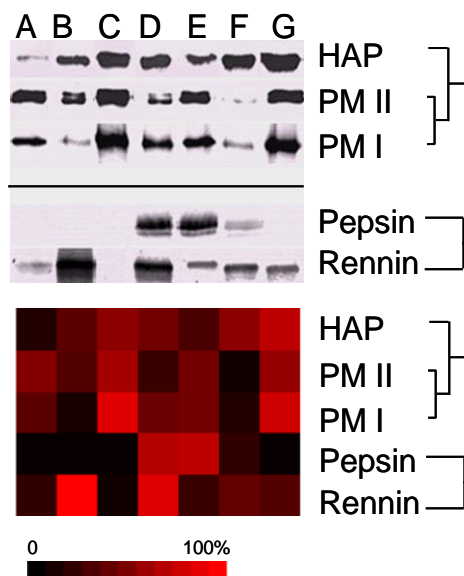
Figure 3.2. Distinct labeling profiles of different aspartic proteases against the 7 AfBPs. (top) Fluorescence gels of five different recombinant aspartyl proteases labeled by the 7 AfBPs; (bottom) The treeview representation of the labeled bands.

To demonstrate the capability of the probes for UV-initiated proteomic profiling of aspartic proteases, pepsin, rennin, recombinant PM-I, -II, -IV and HAP were initially used; highly distinct labeling profiles against different aspartic proteases were observed, indicating the variable P1/P2 group exerted a strong influence over specific enzyme/probe interactions.

3.3.2.2 Competitive labeling by AfBPs with known inhibitors

To confirm our AfBPs label plasmepsins in an activity-dependent manner, we carried out the in-gel fluorescence labeling experiments of the plasmepsins with probe **G** in the presence of different concentrations of each inhibitor. (Liu, K. *et al.* 2008; Leung, D. *et al.* 2003).
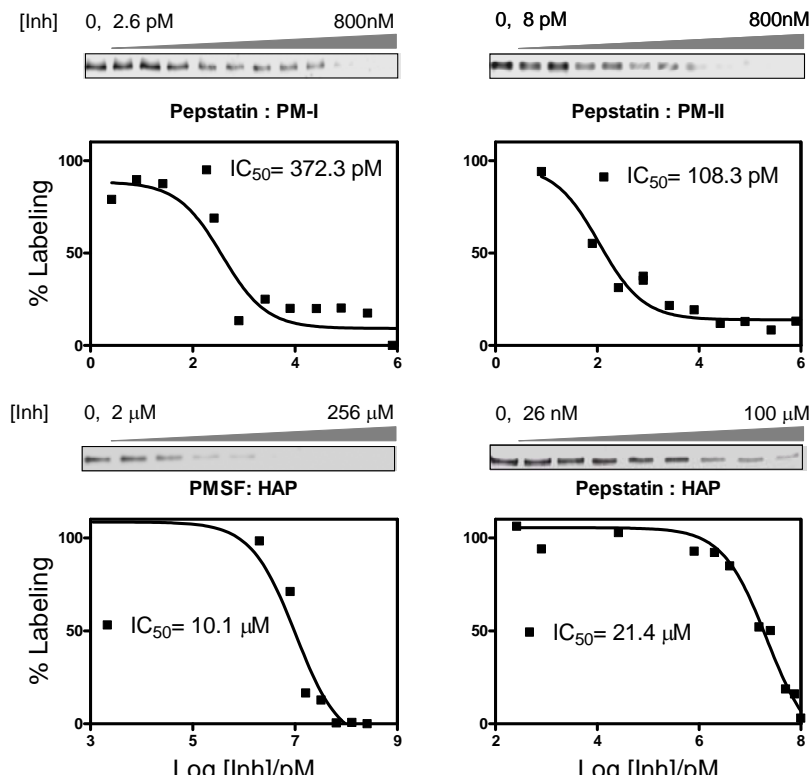
Figure 3.3. Characterizations of several well-known inhibitors against recombinant FV plasmepsins in gel-based competitive labeling assay (top in each panel). IC_{50} values were subsequently determined (bottom in each panel). HAP belongs to a unique class of aspartic proteases which require an actives-site serine residue.[1] Thus its enzymatic activity is sensitive towards PMSF.

3.3.2.3 Comparative Profiling of HAP Active Site Mutants

The exact catalytic mechanism of HAP is still debatable (Parr, C. L. *et al.* 2008, Xiao, H. *et al.* 2007, Xiao, H. *et al.* 2006). In order to further validate the activity-based nature of our AfBPs, and to determine the key catalytic residues of HAP, we compare the in-gel fluorescence labeling profile of wtHAP and four of its active site mutants (H34A, S37A, K78A, E278A, respectively) (Parr, C. L. *et al.* 2008, Xiao, H. *et al.* 2007, Xiao, H. *et al.* 2006) with probe AfBP-**G**. The results were compared with the k_{cat} of the proteins as determined using conventional enzymatic assays (Xiao, H., 2007,). We noticed that probe AfBP-**G** was able to accurately report the relative enzymatic activity of the different HAP mutants. Silver

stained protein bands indicate that the labeling intensity is not due to uneven protein amount loaded.

A) B)

Protein	k_{cat} (s^{-1})
Wild-Type	3.15E-03
H34A	9.03E-04
S37A	8.04E-04
E278A	6.70E-04
K78A	2.70E-03
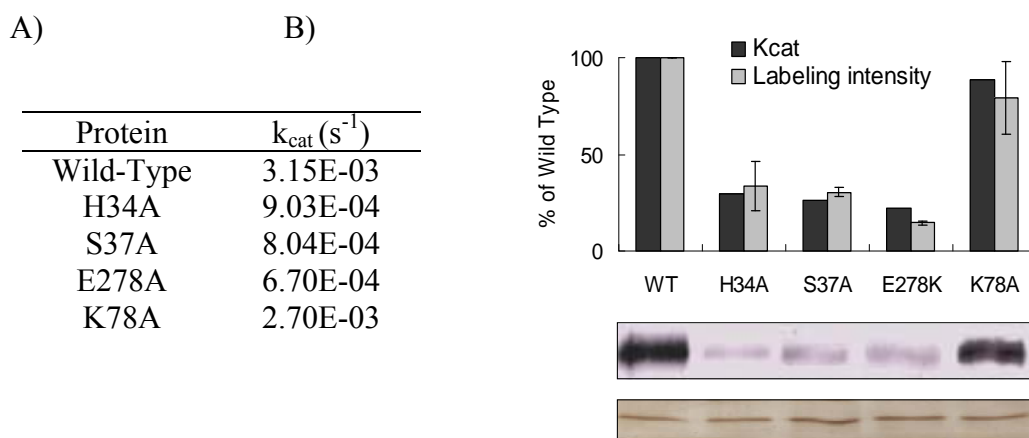
Figure 3.4. Comparative Profiling of HAP Active Site Mutants (A) k_{cat} obtained for the wtHAP and different mutants from traditional enzymatic assays. (B) (bottom) In-gel fluorescence labeling profiles of wtHAP and its mutants with probe AfBP-**G** and silver stained protein band; (top) graphical summary/comparision of k_{cat} and in-gel fluorescence labeling profiles of wt HAP and its mutants.

3.3.3 Labeling of Plasmepsins by AfBPs in Malaria Parasite lysates

3.3.3.1 Profiling of PM Activities throughout Different Blood Stages of *P. Falciparum*

The probes were subsequently used to label proteomes of highly synchronized parasites obtained at different development stages (ring, trophozoite and schizont) (Figure 3.5); a 37-KDa protein band, which corresponds to the molecular weight of all 4 PMs, was highly visible, albeit with varied labeling intensities. To determine whether the labeling was dependent on the activity of the labeled proteins, the same experiments were repeated in the prescence of Pepstatin (1 µM). Results show that the labeling was abolished by the addition of pepstatin demonstrating the activity dependent nature of the labeling. Across probes **A-G**, probe **G** consistently gave the

strongest labeling profile and thus was chosen for further studies. 3.3.3.2
Identification of PMs by 2DGE & MS, and Western Blotting.
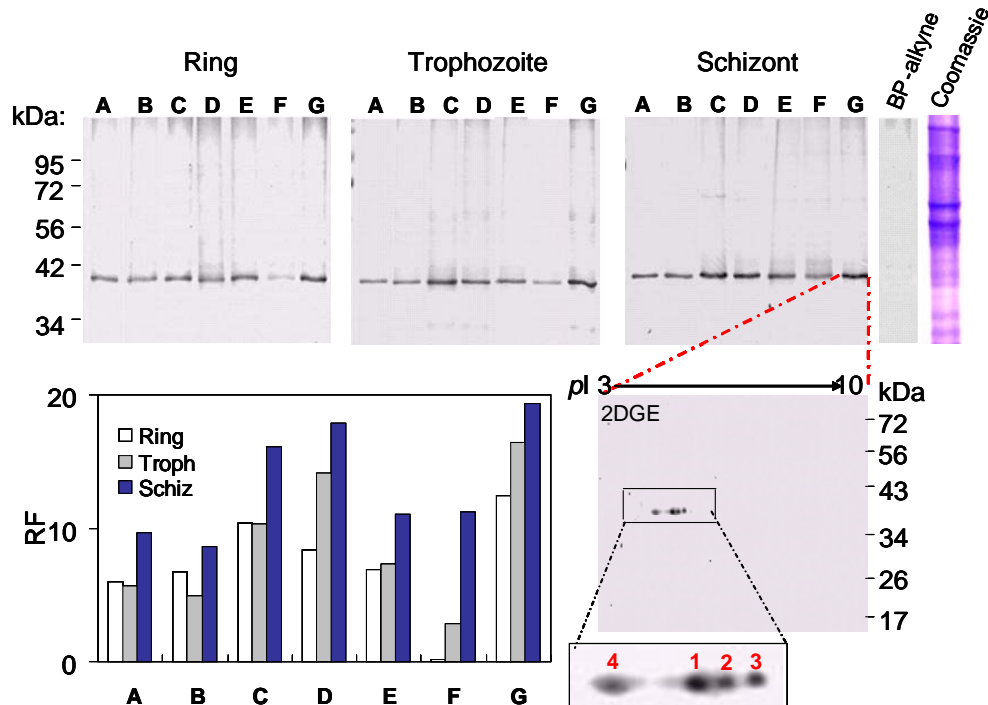
Figure 3.5 Labeling of Plasmepsins by 7 AfBPs in Malaria Parasite lysates; (Top) In-gel fluorescence scanning showing specific labeling of the 37-kDa band across different stages using all seven AfBPs, with spectral counts of the labeled bands (bottom left). (Bottom right) Schizont stage parasite extracts were labeled by probe G, followed by 2D-GE/MS analysis to identify the 4 PMs (spots 1-4 were identified as PM-II, -I, HAP & -IV, respectively in the next section).

3.3.3.2 Identification of PMs by 2DGE, MS, and Western Blotting

To positively confirm the 37-KDa fluorescent band in the schizont stage parasite extract corresponds to the four known plasmepsins (PM-I, PM-II, HAP and PM-IV), 2D-PAGE followed by MS analysis of the fluorescent spots and affinity pulldown followed by westernblot were carried out. Both 2D-PAGE/MS and immuno-pulldown analysis indicates that the 37-KDa band corresponds to all 4 labeled PMs (inset in Figure 3.5 and Figure 3.6, Table 3.1 & 3.2), thus confirming the probes could be used to profile all 4 PMs' activities from the parasite. Neither zymogens of

PMs nor other proteins were detected, indicating the specificity of the probes in targeting only active PMs.

Table 3.1. Summary of 2D-MS/MS Identification for Probe **G**-Labeled Plasempsins

Protein Name	Spot No.	Accession No.	Predicted *p*I	Observed *p*I	MW (kDa)
PM-I	2	gi|124808172	4.7	4~5	37
PM-II	1	gi|858754	4.6	4~5	37
HAP	3	gi|124808181	5.0	4~5	37
PM-IV	4	gi|21730846	4.4	4~5	37

Table 3.2. Mascot MS/MS searching match of PM-I, PM-II, HAP and PM-IV peptides

Protein Name	Peptide	Precursor	Score	Expect
PM-I	GYLTIGGIEDR	1192.61	55	0.00017
	LPTLEFR	874.49	50	0.00052
PM-II	HTGFLTIGGIEER	1428.74	114	2.5e-10
	FYEGPLTYEK	1246.58	38	0.0065
HAP	GYLTIGGIEER	1206.62	34	0.017
	FFDGPLNYEK	1228.5764	32	0.025
PM-IV	DGTKVEISYGSGTVR	1567.78	35	0.011
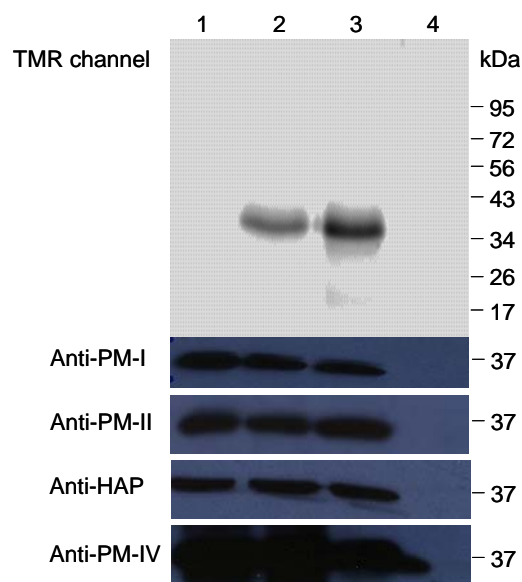
Figure 3.6. anti-TMR immuno pull-down of probe labeled protein; experiment showing that all four plasmepsins were labeled by probe **G**. Lane 1, unlabeled lysate; lane 2, labeled lysate; lane 3, anti-TMR pull-down fraction from labeled lysate; lane 4, anti-TMR pull-down from unlabeled lysate

3.3.3.3 Membrane/Soluble Sub-proteome Profiling of PM Activities

Subsequently, to gain insight into the functional roles of specific aspartic proteases, we used highly synchronized parasite populations to profile PM activities throughout the multiple developmental stages of the parasite. For each stage, PM activity was determined by labeling total lysates, detergent soluble and -insoluble lysates from each stage with probe **G**. As shown in Figure 3.7, results indicated that the activity profiles of FV plasmepsins in either soluble or in-soluble fractions are highly regulated. Interestingly, PM activity peaked at the trophozoite stage for the insoluble fraction. This is consistent with the role of these PMs in hemoglobin degradation which occurs in the food vacuole. The distinct profile of soluble-fraction activities peaked at the schizont stage; this peak probably indicates a change in the subcellular localization of the PMs (PM-II was previously shown to be released from
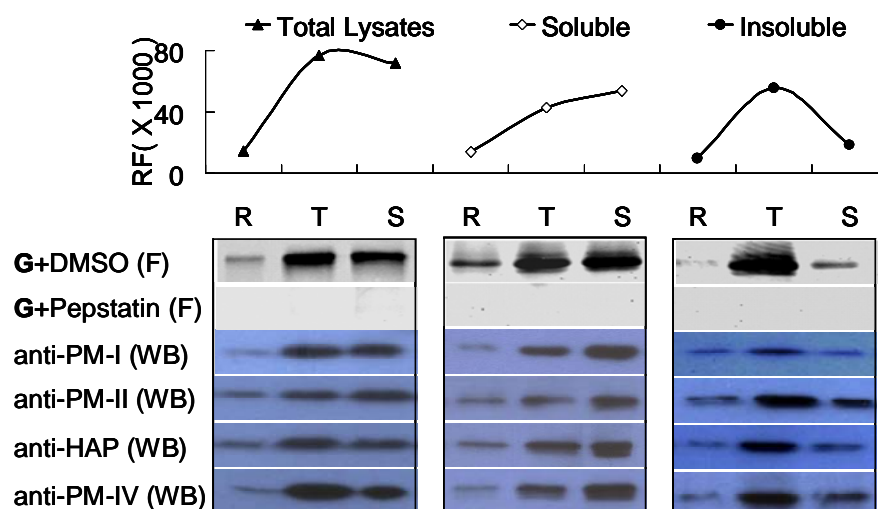
Figure 3.7 Characterizations of PMs activities from total lysates, NP-40 soluble and insoluble fractions across different intraerythrocytic stages of P. falciparum; R indicates Ring stage; T indicates Trophzoite stage; S indicates Schizonts stage. Upper panel shows the spectral counts of respective bands. lower panel shows the labeling profile of parasite lysate from different stages with probe G and respective PM level in different fractions detected by specific antibodies

the FV in the late schizont stage). Western blotting with antibodies against the four PMs also indicated that their absolute protein expression levels corroborated well.with labeling profiles observed with probe G. Thus, the utility of our probes for the accurate reporting of the activities of PMs from crude malaria proteomes was further confirmed

3.3.4 In-Situ Inhibitor library Screening and effect of selected compounds against PMs

3.3.4.1 In-situ library screening by competitive AfBP labeling

One of the key advantages of these AfBPs is the ability to use them for the simultaneous detection of the activities of multiple PMs in their native environment. This so-called in situ screening method, originally described by Cravatt and coworkers for other enzymatic systems (Leung, D. *et al.* 2003) enabled us to identify potential inhibitors against all four PMs directly from the malaria proteome without the recombinant production of every active PM. To identify the most potent inhibitors against all four PMs, we preincubated each compound of the 152-membered hydroxyethyl-based library with the parasite lysate and then added probe G and subjected the samples to UV irradiation. We determined the relative potency of each inhibitor by measuring the decrease in fluorescence intensity in the 37 kDa labeled band. The relative fluorescence intensity of the 37-Kda band labeled by G were quantified using ImageQuant™ software. The potency of specific inhibitor scaffolds was measured as a ratio of the percent residual labeled proteases after inhibitor treatment relative to an untreated control. For analysis, the inhibition data were displayed in a colorimetric format and clustered on the basis of similarities in inhibitor profiles using treeview software. Results are shown in Figure 3.8, from which representative hits potently inhibiting entire FV family plasmepsins in the
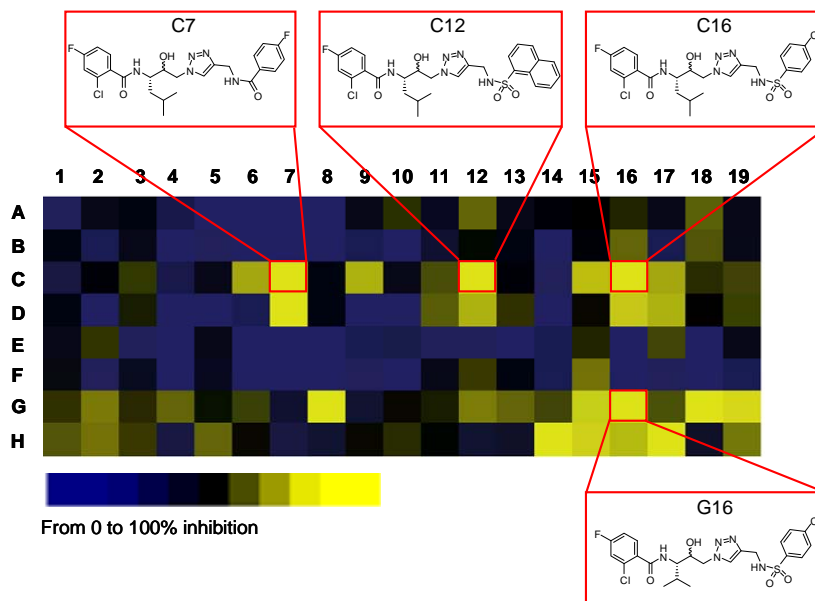
Figure 3.8. Single-point *In situ* screening of 152 member library and selected compounds
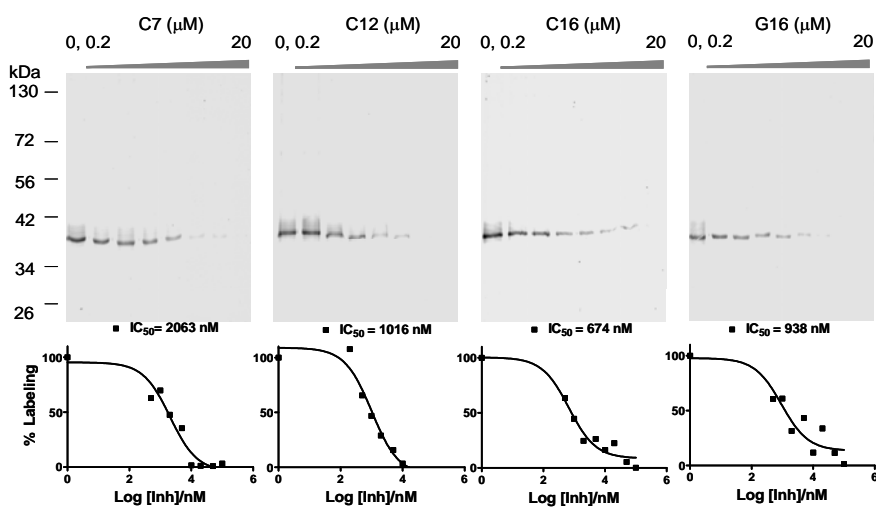
Figure 3.9. *In-situ* Dose dependent inhibition by selected compounds. In-gel fluorescence profiles of the parasite lysate labelled by probe G in the presence of different dosages of an inhibitor (top). The 37-KDa labeled band was quantified and plotted to generate the corresponding IC_{50} curve (bottom).

complex proteome were identified: C7, C12, C16 and G16. As controls, the 152-member library was also screened against each of the three recombinant plasmepsins (PM-I, PM-II, HAP) using a standard microplate-based enzymatic assay or

competitive labeling assay. From these screening results, four more candidate "hits" were identified: H3, H19, G15 and G19. Each of these hits was potent against at least one, but none was potent against all, of the three plasmepsins tested. These eight compounds were resynthesized and purified, and their potency was confirmed in a dose-dependent in situ screening assay (Figure 3.9).

One of the most potent inhibitors identified, G16, showed an IC50 value of 937.5 nM. In contrast, G15, a "false positive" identified from standard enzymatic assays by using selected recombinant PMs, showed significantly weaker inhibition (IC50=21.6 μM) (Figure 3.10). To show that the difference in labeling intensities were due to competitive inhibition of the PM labeling by G16 rather than loading error, specific antibodies were used to confirm consistent PMs level acrosss all lanes.
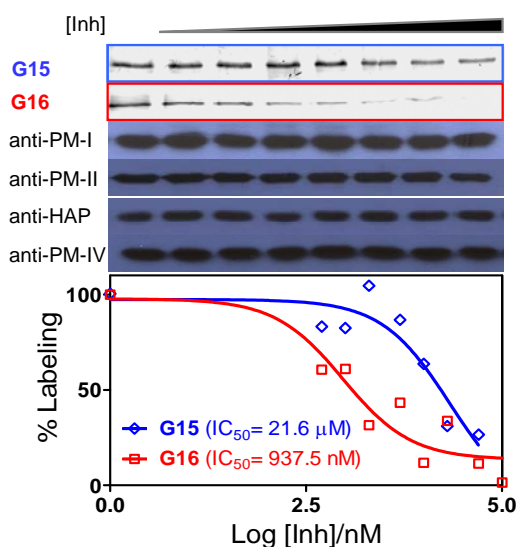
Figure 3.10 In situ screening assay and determination of the IC50 values of G15 and G16 against all four PMs in the parasite proteome

3.3.4.2 Effect of selected compounds in live culture of Parasite infected RBC

The inhibitory effect of these candidate compounds was tested with parasite-infected RBC cultures. The distinct activity/solubility profiles of PMs (Figure 3.7) prompted us to test the inhibitory effect of the inhibitors against schizontstage PMs, which showed the highest activity in the detergent nonyl phenoxylpolyethoxylethanol (NP-40). RBCs were treated with the inhibitors 40 h after parasite invasion (i.e. in the late schizont stage). Upon cell rupture, we measured the percentage of parasites in the ring and schizont stages. Respectively, **G16** gives the best potency in live parasite culture with an EC_{50} of 1.038 μM, followed by **C16** with EC_{50} of 3.191 μM, **C12** with EC_{50} of 8.924 μM, C7 with $EC_{50} > 20$ μM (Figure 3.11). Compound G16, but not G15, caused a marked decrease in the number of newly formed ring-stage parasites, and at the same time an increase in free extracellular merozoites (Figure 3.12).

Thus, as well as blocking parasite development at the trophozoite/schizont stage, as one might expect, the inhibition of PM activity also caused the blockage of either the escape of the parasites from RBCs, or their reinvasion of RBCs. This speculation is further supported by previous findings that PM-II was able to digest an RBC membrane-skeleton protein in the late schizont stage at neutral pH, and that the invasion of P. falciparum merozoites was affected by treatment with pepstatin (a general aspartic protease inhibitor) (Le Bonniec, 1999; Dejkriengkraikhul, P.,1983). The inhibition of parasites by G16 was dose-dependent, whereby the estimated EC50 value of 1.04 μM was similar to that obtained from the in situ proteomic screening (IC50=937.5 nM). In contrast, G15 showed little or no inhibition towards infected
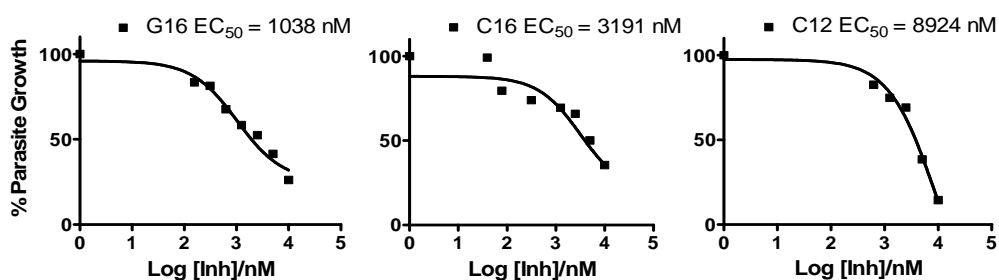
Figure 3.11. Inhibition effects of live parasite-infected RBC cultures treated with selected compounds. EC_{50} was plotted by calculating from the inhibition effect of newly formed ring stage parasites compared to DMSO control.
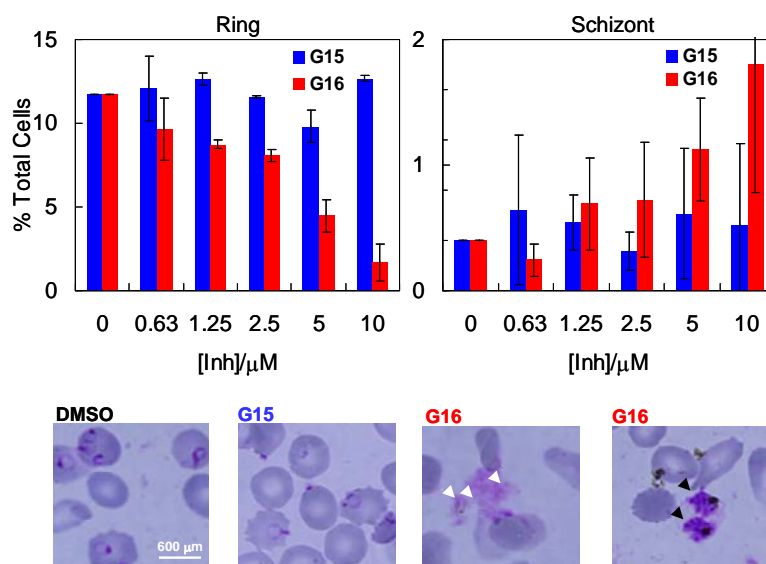
Figure 3.12 Inhibition of parasite growth in infected RBC at Schizont stage. Top, dose dependent inhibition effect of G15 and G16; Lower panel, representative images of parasite infected RBC treated with G15 and G16 with arrows showing abnormal development of parasite.

RBCs, even at the highest concentration tested. Finally, G16 showed no apparent cytotoxicity against common mammalian cell lines (Figure 3.13). These results underscore the importance of our in situ screening assay for the future discovery of general PM inhibitors (Table 3.3)

91
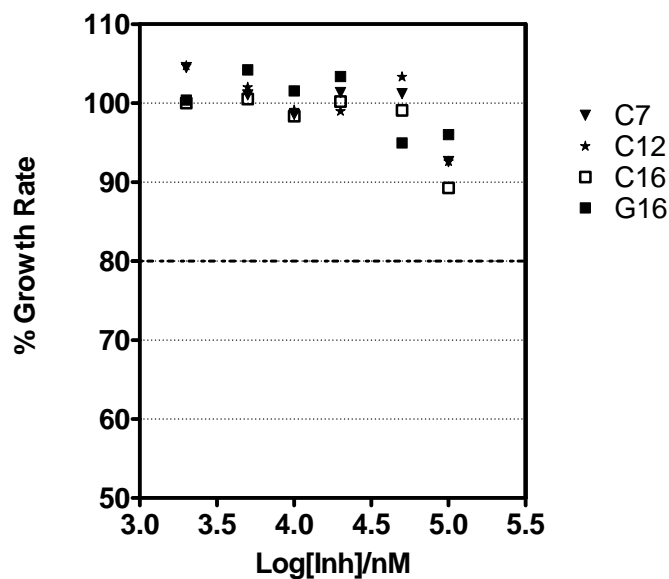
Figure 3.13 . Cytotoxicity analysis in mammalian cell cultures for selected hits identified from the proteomic/parasite screening.

Table 3.3 Summary of Inhibitors identified from proteomic competitive assay

Inhibitor	IC_{50} in competitive proteomic assay	EC_{50} in live RBC culture	CC_{50} against mammalian cells
C7	2.063 μM	> 20 μM	>100 μM
C12	1.016 μM	8.924 μM	>100 μM
C16	0.674 μM	3.191 μM	>100 μM
G16	0.938 μM	1.038 μM	>100 μM

3.3.4.3 Prediction of Binding Mode by Molecule Docking

To gain insight into the binding mode of G16 to FV PMs, molecular docking studies were performed. The results showed that the molecule binds nicely in the active site confined by D34 and D214 aspartic acid pair in the structure of PM-II and PM-IV or by the H32 and D214 in the structure of HAP (Figure 3.14 A, B, C). Detailed analysis of the top 100 orientations identified preferred conformations of

G16 in PM-II, PM-IV as well as the non-classical HAP (Figure 3.14 G, H, I). The results also show that the hydroxyl group of G16 has close interaction with D34 and D214 in PM-II and PM-IV just like proposed; while binding to HAP, G16 assumes a position with its hydroxyl group fitted between H32 and D214 which is observed for the statine hydroxyl group of pepstatin12 (Figure 3.14 D, E, F).
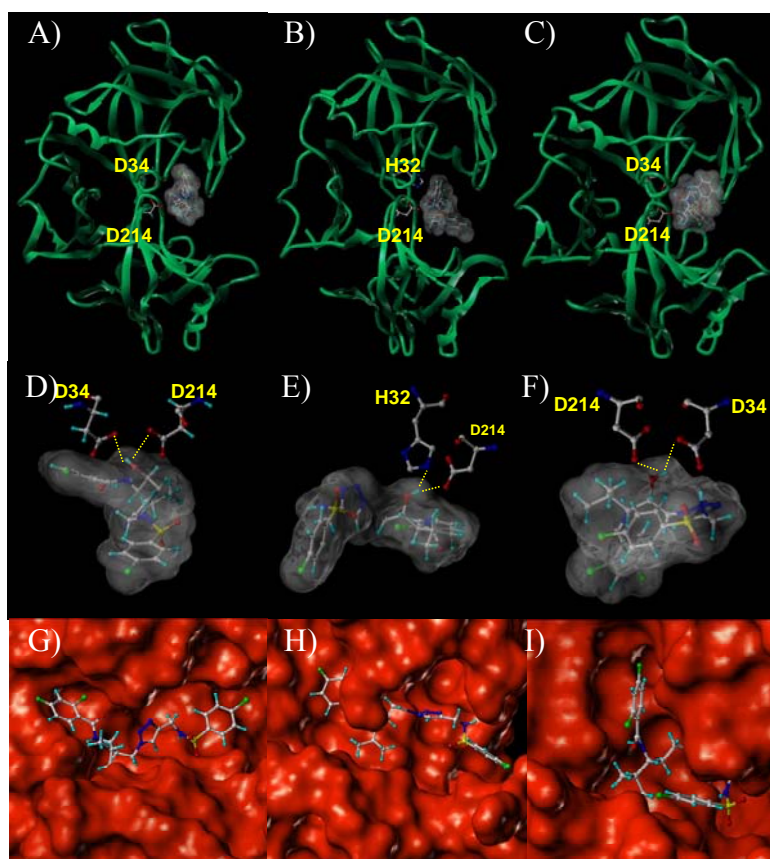
Figure 3.14 Molecular docking of G16 into the active site of PM-II, HAP and PM-IV A), G16 binds to the active site binding pocket of PM-II, B), HAP and C) PM-IV; D) Hydroxyl group of G16 closely interact with D34 and D214 in PM-II; E) D32 and D214 in HAP; F) D34 and D214 in PM-IV; G) Prefered conformation of G16 in the binding pocket of PM-II; H) of HAP; I) of PM-IV

3.4 Conclusion

In summary, we have developed the first affinity-based probes for the functional profiling of all four PMs in intraerythrocytic malaria parasites. Subsequent

in situ screening of parasites with these probes led to the identification of a compound, G16, which showed good inhibition against all four PMs and parasite growth in infected RBCs. Molecular modeling indicated that this inhibitor binds to the active site of the plasmepsins tested, as originally designed. Our results indicate the feasibility of using AfBP approaches for the identification of inhibitors of less-characterized enzymes (such as HAP) and inhibitors of multiple targets. We anticipate that these new chemical tools should facilitate the discovery of unknown parasite biology and new antimalarial drugs. In the current study, we were unable to detect previously predicted but unidentified aspartic proteases from the malaria proteome. Thus, G16 might have further targets, other than the four plasmepsins, in the malaria proteome. Future research will focus on the development of new chemical probes to address these issues.

Chapter 4.

Activity-Based Proteome Profiling of Potential Cellular Targets of Orlistat - An FDA-Approved Drug with Anti-Tumor Activities

4.1 Summary

Since small molecule inhibitor's selectivity addressed by ABP/AfBP approach only represents the selectivity within this class of enzyme targeted by the ABP/AfBP, profiling of a small molecule drug's targets in a whole proteome level is of significant value. This chapter describes an activity-based profiling of small molecule-protein interactions approach for unbiased, empirical identification of off-targets of development-stage or even marketed drugs. Tetrahydrolipstatin is a natural product-derived drug that is a lipase inhibitor used against obesity that may also be useful in cancer therapy. Its action is based on acylation of active-site nucleophiles by its beta-lactone unit. Cellular off-targets and potential side effects of Orlistat in cancer therapies, however, have not been extensively explored thus far. In this chapter, we describe the application of THL-like protein-reactive probes, in which extremely conservative modifications were introduced in the parental THL structure to maintain the native biological properties of Orlistat, while providing the necessary functionality for target identification via the bio-orthogonal click chemistry. With

these natural productlike, cell-permeable probes, we were able to demonstrate, for the first time, this chemical proteomic approach is suitable for the identification of previously unknown cellular targets of Orlistat. In addition to the expected fatty acid synthase (FAS), we identified a total of eight new targets, some of which were further validated by experiments including Western blotting, recombinant protein expression, and site-directed mutagenesis. Our findings have important implications in the consideration of Orlistat as a potential anticancer drug at its early stages of development for cancer therapy. Our strategy should be broadly useful for off-target identification against quite a number of existing drugs and/or candidates, which are also covalent modifiers of their biological targets.

4.2 Introduction

Drug discovery is a long and costly process, yet most drugs have side effects, ranging from simple nuisances to life-threatening complications (Giacomini, K. M. *et al.* 2007). Unanticipated effects of a drug, often revealed either during clinical trials or sometimes after the drug enters the market, could lead to termination of a drug development program/recall of the drug, or, in some rare cases where the effects are beneficial, new drug applications (Campillos, M. *et al.* 2008, Crunkhorn, S *et al.* 2008). Therefore one of the most critical steps in the drug discovery process is the effective identification of the so-called off-targets and anticipation of their potential side effects *a priori*.

Orlistat™, or tetrahydrolipstatin (THL), is an FDA-approved anti-obesity drug, which works primarily on pancreatic and gastric lipases within the gastrointestinal (GI) tract (Guerciolini, R *et al.* 1997). Recently, Orlistat was found to

inhibit the thioesterase domain of fatty acid synthase (FAS), an enzyme essential for the growth of cancer cells, but not normal cells (Kridel, S. J. *et al.* 2004；Menendez, J. A. *et al.* 2007). By effectively blocking the cellular FAS activity, Orlistat induces endoplasmic reticulum stress in tumor cells, inhibits endothelial cell proliferation and angiogenesis, and consequently delays tumor progression on a variety cancer cells, including prostate, breast, ovary, and melanoma cancer cells (Little, J. L. *et al.* 2007). As a result, this compound (as well as other Orlistat-like analogs with improved potency and bioavailability) is being pursued as a promising anti-cancer drug (Ma, G. *et al.* 2006；Richardson, R. D. *et al.* 2008). Cellular off-targets and potential side effects of Orlistat in cancer therapies, however, have not been extensively explored thus far (Filippatos, T. D. *et al.* 2008). Our long-term research goals focus on developing novel chemical proteomic strategies that enable large-scale studies of therapeutically relevant enzymes, as well as small molecules (i.e. potential drug candidates) that can modulate these enzymes' cellular activities (Uttamchandani, M. *et al.* 2009). In our study, we set out to look for new cellular targets, including off-targets, of Orlistat at its early stages of development for cancer therapy. In the next section, we describe for the first time, the identification and putative validation of several previously unknown cellular targets of Orlistat by using a natural product-based, chemical proteomic approach.

4.3 Results and Discussion

4.3.1 THL-R Design and Synthesis of Orlistat-like Probes.

Our strategy is based on the well-established activity-based protein profiling (ABPP) approach (Evans, M. J., 2006; Uttamchandani, M., 2008), by making use of

THL-like protein-reactive probes THL-R, -L and -T (structure 1, 2, 3 Figure 4.1). We took advantage of several key properties known to THL in the design of our activity-based probes: (1) THL (being derived from a natural product) is cell-permeable,
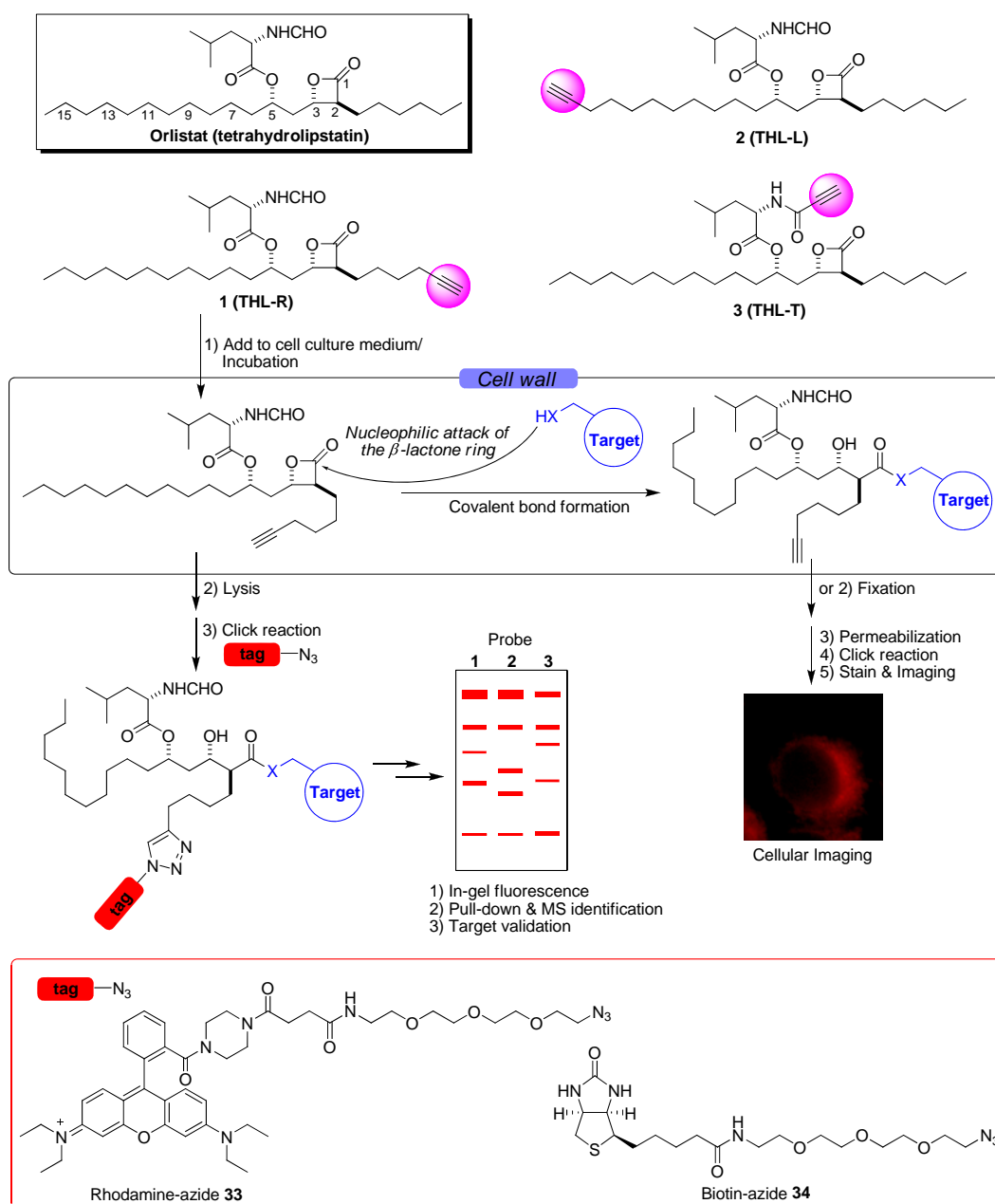
Figure 4.THL-R. Overall strategy for activity-based proteome profiling of potential cellular targets of Orlistat using alkyne-containing, cell-permeable THL analogs (**THL-R**, **-L & -T**). The alkyne handles in the probes are shaded (purple circle). The "Click"-based reporters used in the manuscript are shown (in red rectangle).

making our probes applicable for direct *in situ* cell-based screening; (2) THL reacts with its known cellular targets via a covalent reaction through its reactive β-lactone moiety and the nucleophilic active-site residue (typically Ser/Cys residues, e.g. Ser^{2308} in FAS (Pemble, C. W. *et al.* 2007) of the target protein, resulting in the formation of an isolatable protein/THL complex; (3) previous minor structural modifications at either the 16-carbon or 6-carbon aliphatic chains of THL did not significantly alter its native biological activities (Richardson, R. D. *et al.* 2008). These extremely conservative modifications of introducing an alkyne handle in the parental THL structure were aimed at maintaining the native biological properties of Orlistat, while providing the necessary functionality for identification and characterization (i.e. imaging) of previously unknown cellular targets by downstream conjugation of the protein/probe complex to reporter tags via the bio-orthogonal click chemistry (Kalesh, K. A. *et al.* 2007; Uttamchandani, M. *et al.* 2008).

4.3.2 Comparing the Cellular Effects of THL and THL-based probes

Next, the three probes were evaluated against Orlistat (as a positive control) for potential biological activities. Three different types of cellular assays based on previously established Orlistat biology were used (Kridel, S. J. *et al.* 2004; Knowel, L. M. *et al.* 2008; Little, J. L. *et al.* 2007). First, the anti-proliferation activity of the four compounds (THL-R, -L, -T & Orlistat) against HepG2 cells (a human hepatocellular liver carcinoma cell line) were evaluated using the XTT assay (Kridel, S. J. *et al.* 2004;); all four compounds showed a dose-dependent inhibition of tumor cell proliferation over a 72-h time period with comparable potency (Figure 4.2A). Secondly, we carried out comparative analysis of the compounds in their ability to induce phosphorylation of eIf2 in the prostate cancer PC-3 cells (Knowel, L. M. *et*

99
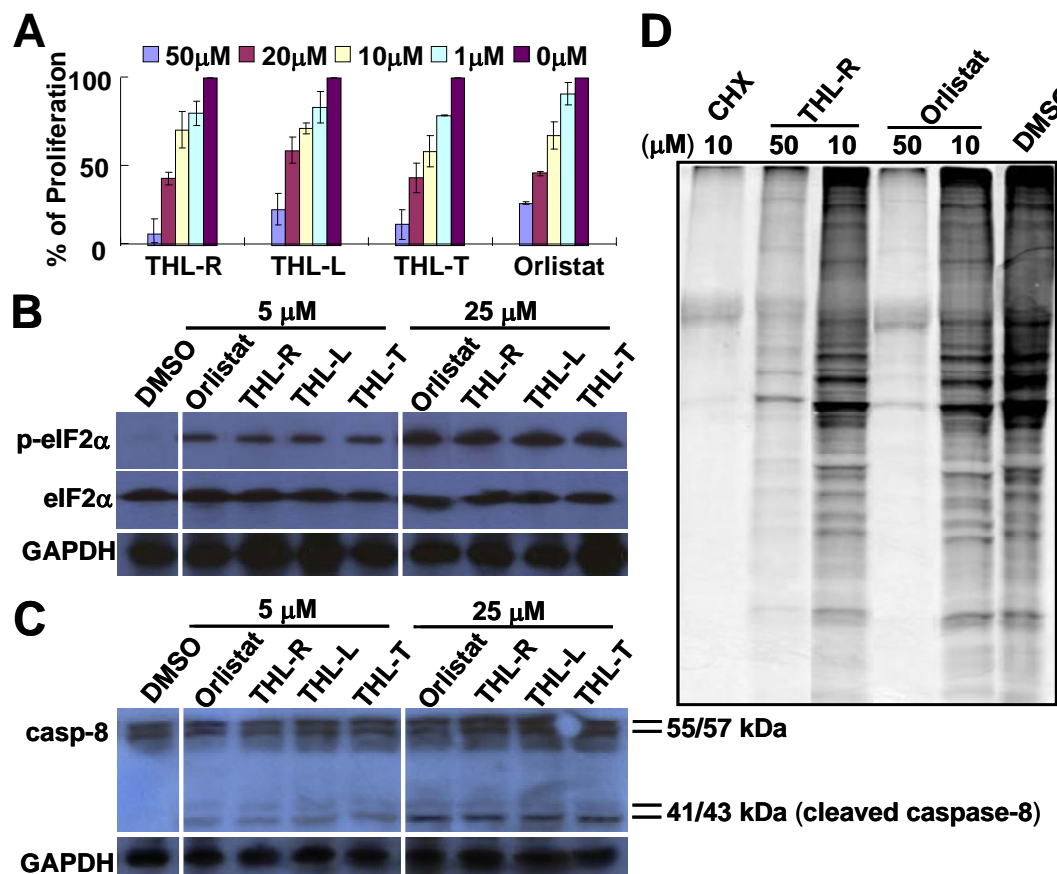
Figure 4.2 Cellular Effects of THL and THL-based probes (A) Dose-dependent inhibition of HepG2 cell proliferation by Orlistat and three analogues (1, 2, and 3) using XTT assay. Data represent the average (s.d. for two trials. (B) Western blot analysis of eIF2R phosphorylation in PC-3 cells upon treatment with the four compounds. GAPDH was used as a loading control. (C) Inhibition of protein synthesis in HepG2 cells treated with the indicated concentrations of Orlistat, THL-R, or CHX (cycloheximide, an inhibitor of protein biosynthesis) for 12 h and then pulsed with AHA (L-azidohomoalanine) for 4 h. Cell lysates were prepared and subjected to click chemistry with rhodamine-alkyne (provided with the AHA kit) following vendor's protocols, SDS-PAGE analysis, and in-gel fluorescence scanning (fluorescent gel is shown in grayscale). (D) Activation of caspase-8 in MCF-7 cells treated with the indicated concentrations of Orlistat/analogues.

al. 2008;). Previous studies had shown that inhibition of FAS by Orlistat induces endoplasmic reticulum (ER) stress and results in the phosphorylation of the translation initiation factor eIF2. As shown in Figure 4.2 B, PC-3 cells treated with different amounts of each compound showed similarly elevated

100

eIF2 phosphorylation. Because the phosphorylation of eIF2 is known to inhibit cellular protein synthesis (Little, J. L. *et al.* 2007), we carried out metabolic labeling experiments with azidohomoalanine (Dieterich et al. 2006) to measure the level of newly synthesized proteins in cells treated with Orlistat, THL-R or cycloheximide (a well-known protein synthesis inhibitor). As shown in Figure 4.2C, greatly reduced levels of protein synthesis were observed in cells treated with each of the three compounds. The inhibition of protein synthesis was dose-dependent, which is consistent with previous findings carried out using radiolabeled ^{35}S-methionine (Little, J. L. *et al.* 2007). Lastly, inhibition of FAS by Orlistat was previously shown to induce tumor cell apoptosis by activating caspase-8 (Knowel, L. M. *et al.* 2008;). We therefore tested the compounds against the invasive human breast cancer MCF-7 cells (Figure 4.2 D); similar degrees of caspase-8 activation (as evidenced by the appearance of the p41/43 bands corresponding to cleaved caspase-8) were observed against all four compounds (at either 5 or 25 μM). Collectively, these data show the introduction of a terminal alkyne handle at various designated locations in Orlistat scaffold did not noticeably affect its biological activities, and THL-R, -L, and -T were indeed suitable chemical probes for cell-based proteome profiling and identification of previously unknown cellular targets of Orlistat identification.

4.3.3 In Situ and in Vitro Proteome Profiling.

We next compared the in situ proteome reactivity profiles of the three probes to identify proteins which were covalently labeled by the probes in live HepG2 cells (Evans, M. J. *et al.* 2005). THL Probes (1-20 μM) were directly added to the cell culture medium, either alone or in the presence of 100 μM of the competing Orlistat. After two hours, the cells were washed (to remove excessive probes), homogenized, incubated with rhodamine-azide under click-chemistryconditions, separated by SDS-

PAGE gel, and analyzed by ingel fluorescence scanning. In addition to the expected FAS band (265 kDa, marked with red arrow; Figure 4.3A), confirmed by treatment with an anti-FAS antibody (Figure 4.4A), we also observed a number of Orlistat-sensitive targets, that is, those labeled bands that were competed away by treatments with excessive Orlistat (marked with blue arrows and asterisks in Figure 4.3A and B). Most of these labeled bands were clearly visible even at a low probe concentration (e.g., 1 μM of THL-R; lane 1 in Figure 4.3A). Both THL-R and -L gave similar proteome labeling profiles, whereas THL-T consistently produced weaker labeled bands (possibly caused by inefficient click-chemistry conjugation due to inaccessibility of the alkyne handle located at the N-formyl- L-leucine end, which, based on X-ray structure of FAS/Orlistat complex, was buried deep into the protein) (Pemble, C. W. *et al.* 2007). To assess which nucleophilic residue of the labeled proteins might have been covalently modified by our probes, SDS-PAGE gels from the in situ labeling were subjected to in-gel treatment with hydroxylamine (NH2OH), which preferentially cleaves thioesters, and esters to a lesser extent, under neutral pH conditions (Charron, G. *et al.* 2009). As shown in Figure 4.3D, the labeled FAS and most other Orlistat-sensitive bands showed a much reduced fluorescence signal upon treatment with NH2OH, suggesting likely involvemen of a cysteine/serine residue in the labeling between these proteins and THL (vide infra). It is interesting to note that the labeled FAS band was to some degree resistant to NH_2OH treatment, clearly indicating the formation of a more chemically stable ester linkage between Ser2308 in FAS and THL-R. In a related experiment, we also performed competitive ABPP with Cerulenin, a known FAS inhibitor which irreversibly inactivates the -ketoacyl-ACP synthase domain but not the thioesterase domain of FAS (Martin Moche, M. *et al.* 1999); as expected (Figure 4.3C), Cerulenin did not abolish the
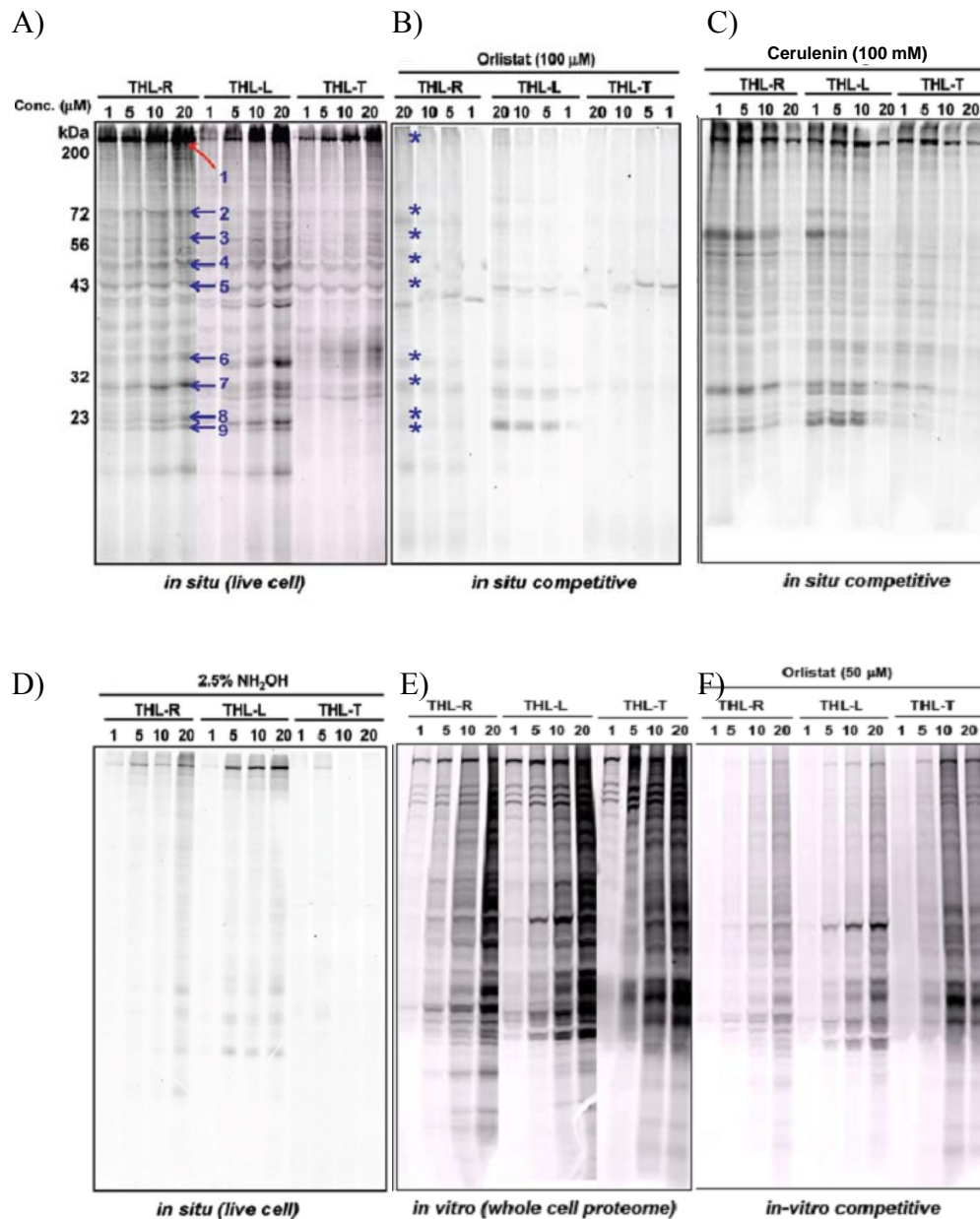
Figure 4.3 Proteome profiling of HepG2 cells using THL analogs. Both *in situ* (live cell) and *in vitro* (whole-cell lysates) labeling were carried out. (A) *in situ* proteome labeling of HepG2 cells, with or (B) without Orlistat (100 μM), (C) Cerulenin (100 μM), followed by click chemistry with rhodamine-azide SDS-PAGE analysis, in-gel fluorescence scanning , . The labeled FAS (marked with red arrow) and multiple Orlistat-sensitive targets (marked with blue arrows) were isolated and identified subsequently using biotin-azide, and summarized in Table 4.1. In (B), asterisks show the expected locations of FAS and other Orlistat-sensitive bands. (D) showed the fluorescence profile of the gel from Figure 3A upon treatment with hydroxylamine. The fluorescence labeling of most proteins (but not FAS) were reversed by hydroxylamine, indicating a thioester linkage. (E) and (F) *in vitro* labeling of HepG2 whole-cell proteome lysates with or without Orlistatlabeling of the THL probes toward FAS as well as most other Orlistat-sensitive proteins. This indicates that the labeling of our probes is both target- and domain-specific, and they may in the future be used to distinguish different domains of FAS.

As shown in Figure 4.3E and Figure 4.3F, when compared to in situ (live cell) labeling, similar proteome labeling profiles, albeit with significantly higher background labeling, were obtained when whole-cell lysate were used instead. This shows the importance of the in situ labeling (made possible by the cell-permeable property of the THL probes) as a prerequisite for accurate and specific target identification.

4.3.4 Target identification and validation

Subsequently, the labeled protein extract was enriched (following click-chemistry conjugation with biotin-azide) by avidin-agarose beads, separated by SDS-PAGE gel, confirmed by strepavidin blotting, subjected to in-gel trypsin digestion, and identified by MS/MS analysis (Figure 4.4A). In addition to FAS, eight new proteins were identified (numbered **1-8**; in Figure 4.3A, 4.4A and Table 4.1), of which one is an unnamed protein. Two of the proteins, GAPDH and β-Tubulin, are house-keeping genes constitutively expressed in most cell lines, but known to be expressed more highly in cancer cells (Mori, R. *et al.* 2008; Phadke, M. S. *et al.* 2009). Both GAPDH (a dehydrogenase) and β-tubulin (a hydrolase) possess nucleophilic active-site cysteine residues. It is therefore not surprising they were targets of Orlistat and their labeling was reversed by NH_2OH treatment. It should be noted that tubulins are well-known targets of anticancer drugs (Kingston, D.G. *et al.* 2007). Three other proteins identified, RPL7a, RPL14 and RPS9, are ribosomal proteins. They are known to be implicated in protein synthesis, control of cellular transformation, tumor growth, aggressiveness and metastasis. Overexpression of these proteins had previously been reported in colon, brain liver, breast and prostate carcinoma (Zhu, Y. *et al.* 2001; Liu, Y. *et al.* 2007; Wang, Y. *et al.* 2000; Vaarala, M. H. *et al.* 1998). In retrospect, our earlier findings that protein synthesis was greatly

inhibited in HEPG2 cells treated with Orlistat, as shown in Figure 4.2D, may also imply that these ribosomal proteins are probable cellular targets of the drug. The remaining two proteins, Annexin A2 and Hsp90AB1, are involved in cell proliferation/division and protein degradation, respectively.
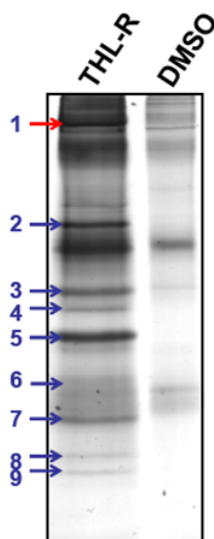
Figure 4.4 Silver staining of gels of pulled-down fractions from THL-R-labeled or DMSO-treated HepG2 live cell lysates. Only distinct bands appears in the pull-down fraction from THL-R-labeled sample were cut from the gel. None of the cut bands showed up in the DMSO-treated control sample (right lane), indicating they are specifically labeled proteins.

In an effort to further validate the MS results, we carried out pull-down and western blotting experiments with the respective antibodies; results confirmed all proteins tested (including FAS) were indeed positively labeled by THL-R, and thus are likely true cellular targets of Orlistat (Figure 4.5A). Four of the proteins (β-tubulin, GAPDH, RPL7a and RPL14) were taken for additional validation experiments. First, the c-Myc fusions of these proteins were transiently expressed in HEK-293T cells (Figure 4.5B), labeled (by THL-R *in situ*), immune-purified (with c-Myc agarose beads), subjected to click chemistry (with rhodamine-azide), and

105

analyzed by SDS-PAGE (Figure 4.5C); results unambiguously confirmed that all four proteins were fluorescently labeled by the probe. To confirm whether labeling of the probe is active site-directed, GAPDH, whose active site residue Cys[151] had previously been characterized (Phadke, M.S. *et al.* 2009), was taken as an example for further site-directed mutagenesis experiments. GAPDH is a multi-function protein and well-known for its primary role as a glycolytic enzyme. Recently, increasing evidence has suggested that this enzyme is also involved in a variety of activities which are unrelated to energy production, including membrane fusion, microtubule bundling, DNA repair, and apoptosis (Phadke, M. S *et al.* 2009). An active site (Cys[151] to Ala) mutant of GAPDH was therefore generated, transiently expressed and purified from HEK-293T cells, labeled with THL-R, reacted with rhodamine-azide, analyzed by SDS-PAGE and in-gel fluorescence scanning (Figure 4.5D); as expected, probe THL-R only labeled the wild-type GAPDH and not the Cys[151]Ala mutant, thus confirming Cys[151] as the residue in GAPDH being covalently labeled by THL-R, as well as the active site-directed nature of our probes. It should be noted that, in a series of recent reports (Bo¨ttcher, T. *et al.* 2008; Bo¨ttcher, T. *et al.* 2008; Bo¨ttcher, T. *et al.* 2009), Sieber *et al* had developed activity-based probes based on small molecule libraries containing a reactive β-lactone moiety. The authors concluded that β-lactones are promising privileged structures and could be used to identify a variety of mechanistically distinct enzymes. Our Orlistat-based probes, though structurally much more complex, appear to be capable of going after a number of other cellular targets, in addition to FAS. Even though the exact physiological roles of these proteins in connection with Orlistat and its pharmacological effects have not been established from this study, we believe these

putative cellular targets of Orlistat should be carefully evaluated when one considers

using Orlistat in cancer therapy.
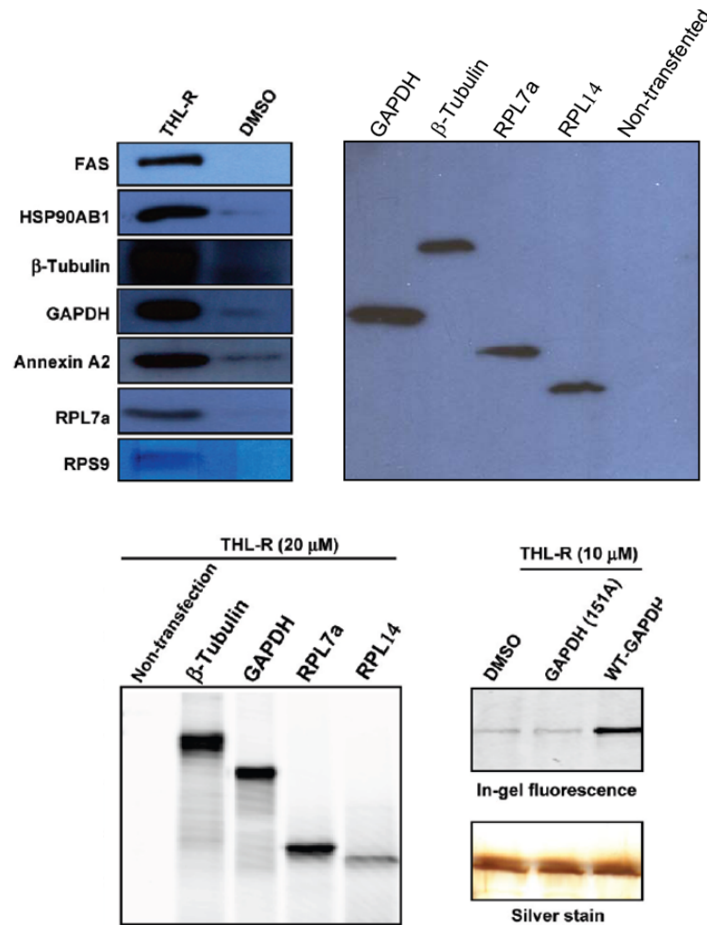
Figure 4.5 Target validation of the identified "hits". (A) Western blotting analysis of pulled-down fractions of HepG2 live cells treated with THL-R (or DMSO as negative controls; right lanes) with their respective antibodies. Biotin-azide 34 was used in the click chemistry with avidin-agarose beads for pull-down experiments. (B) Western blotting confirming the recombinantly expression of proteins with anti-c-Myc. (C) In situ labeling of recombinantly expressed -tubulin, GAPDH, RPL7a, and RPL14 by THL-R (fluorescent gel shown in grayscale). (D) Comparative labeling analysis of wild-type GAPDH and Cys151Ala mutant (upper panel, fluorescent gel shown in grayscale); (lower panel, silver-stained gel) comparable amounts of protein loading were demonstrated in all three lanes.

Table 4.1. Proteins identified by pull down and mass spectrometry

#	Protein Name	Calc/Obs Mass Da/kDa	Localization	Protein Function
2	Hsp90AB1	75088/72	Cytoplasm, Nucleus	Molecule Chaperone with ATPase acitivity; Stress Response
3	Unnamed	56156/56	Cytoplasm, Nucleus	Protein biosynthesis; Translational elongation
4	-Tubulin	49671/52	Nucleus	GTPase acitivity and cell division cell division
5	ANXA2	38808/39	Cytoplasm, Nucleus	Calcium binding; Cell division
6	GAPDH	36201/36	Cytoplasm, Nucleus	Glycolysis; Energy production
7	RPL7a	30148/30	Cytoplasm, Ribosome	Biogenesis; Protein synthesis
8	RPL14	23902/23	Cytoplasm, Ribosome	Biogenesis; Protein synthesis
9	RPS9	22635/22	Cytoplasm, Ribosome	Biogenesis; Protein synthesis

4.3.5 Fluorescence microscopy of Orlistat cellular targets

To demonstrate the utility of our cell-permeable probes for potential cellular imaging of Orlistat targets, we performed fluorescence microscopy to visualize probe-treated cells by colocalizing with ER (Figure 4.6) or FAS (Figure 4.7). Live HepG2 cells were first treated with THL-R, fixed with PFA, permeabilized with Triton X-100, conjugated to rhodamine-azide by click chemistry, and imaged (colored in red). Immunofluorescence was also carried out on the same cells to visualize the localization of endogenous FAS (colored in green in top panels). Minimal fluorescence was observed in samples treated with only DMSO, whereas, in

THL probes (-R, -L and -T) treated cells, fluorescence was mostly distributed in endoplasmic reticulum (Figure 4.6). Note that endogenous FAS was mostly cytosolic (which includes ER, top panels). Thus, our imaging results are consistent with previous findings that inhibition of FAS with Orlistat induces ER stress specifically in tumor cells (Little, J. L. *et al.* 2007). We take note, however, that, at its current state, THL-R might not be the most suitable chemical probe for bioimaging of FAS, as it also labels a number of other cellular proteins (as evidenced from our studies herein). Work is in progress to develop other Orlistat analogues, which may confer much greater specificities, and results will be reported in due course.
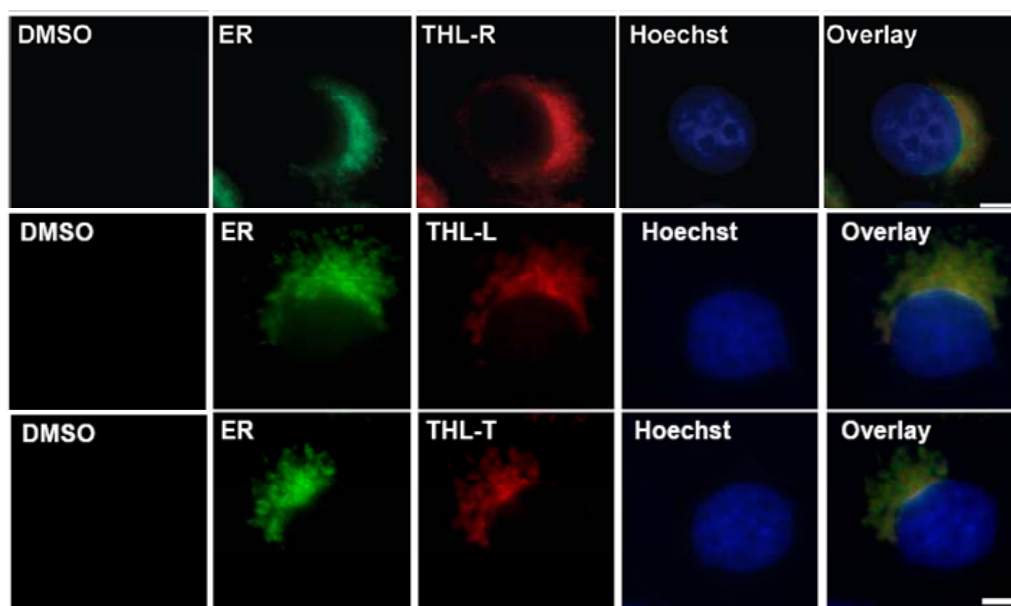
Figure 4.6 Cellular imaging of HepG2 cells with probes. Live HepG2 cells were treated with 1 (20 μM) for 2 h, fixed, permeabilized, reacted with rhodamine-azide (10 μM, false-colored in red) under click-chemistry conditions, stained with Hoechst (blue), anti-FASN primary antibodies, followed by FITC-conjugated antimouse IgG antibody (green), or ER-Tracker Green (glibenclamide BODIPY FL), mounted, and imaged. Samples were imaged with an Olympus IX71 inverted microscope, equipped with a 60X oil objective (NA 1.4, WD 0.13 mm) and CoolSNAP HQ CCD camera (Roper Scientific, Tucson, AZ, USA). Images were processed with MetaMorph software (version 7.1.2., Molecular Devices, PA, USA). Scale bar) 8 μm. All images were acquired similarly. Cells treated with DMSO (negative controls) are also shown
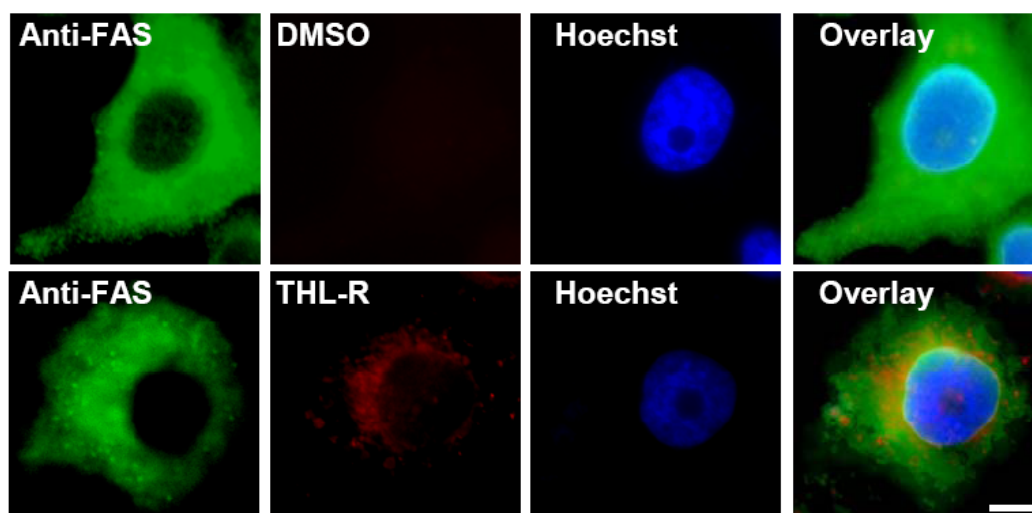
Figure, 4.7 Colocalizing in situ THL analogs target with FAS. HepG2 cells were incubated with THL-R, L, T or DMSO. Further, cells were tagged with Rhodamine-azide (Red), Anti-FAS and fluorecein tagged secondary Ab (Green) and Hoechst (Blue). Samples were imaged with an Olympus IX71 inverted microscope, equipped with a 60X oil objective (NA 1.4, WD 0.13 mm) and CoolSNAP HQ CCD camera (Roper Scientific, Tucson, AZ, USA). Images were processed with MetaMorph software (version 7.1.2., Molecular Devices, PA, USA). (Scale Bar: 8 mm). All images were acquired the same way

4.4 Conclusion

We have developed a novel chemical proteomic approach that enabled, for the first time, identification and putative validation of several previously unknown cellular targets of Orlistat. The potential of these cell-permeable probes to be used as future imaging probes has also been explored. Whereas further studies are needed to better understand the exact relevance of Orlistat and its pharmacological effects in relation to these newly identified cellular targets, our findings point to a likely scenario that these proteins might be potential off-targets of Orlistat. It is also possible that the antitumor activities of Orlistat might have originated from the drug's ability to inhibit both FAS (Knowel, L. M. *et al* 2008) and some of these newly identified targets. In either case, our findings have important implications in consideration of Orlistat as a potential anticancer drug. Finally, our strategy should

be broadly useful for off target identification against quite a number of existing drugs

and/or candidates, which are also covalent modifiers of their biological targets.

Chapter 5.

Dynamic Profiling of Post-Translational Modifications on Newly Synthesized Proteins Using a Double Metabolic Incorporation Strategy

5.1 Summary

Post-Translational Modification (PTM) occurs in most eukaryotic proteins and directly modulates their function, localization and interactions. The process is highly regulated and at the same time, dynamic. To promote the identification of transient PTM events, we present in this study a methodology that enables the proteome-wide profiling of PTMs on proteins synthesized in defined time window, as well as their dynamics, by using a double metabolic incorporation strategy. We validated the feasibility of this approach with a number of proteins covering a total of eight different types of PTMs which occur on different time scales (rapid and enduring). We further applied the strategy to monitor the myristoylation of newly synthesized proteins in apoptotic Jurkat cells, and successfully identified Protein Kinase A (PKA), a key signaling enzyme, whose myristoylation appeared to be up-regulated in response to butyric acid (BA)-induced

apoptosis. We anticipate this new chemical proteomic tool may facilitate the discovery of primary PTM changes associated with different extracellular and intracellular cues.

5.2 Introduction

Post-translational modification (PTM) is a highly dynamic yet precisely controlled process by which most eukaryotic proteins are diversified chemically (Walsh, C.T. et al. 2005). Many critical cellular responses are mediated through PTMs, leading to modulation of enzyme activity, protein conformation, protein-protein interaction and cellular localization. Analysis of these modifications at the proteome level could provide invaluable biological insights, but
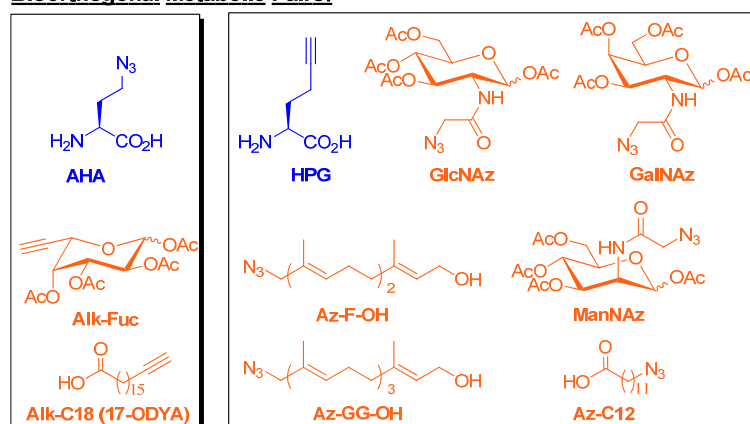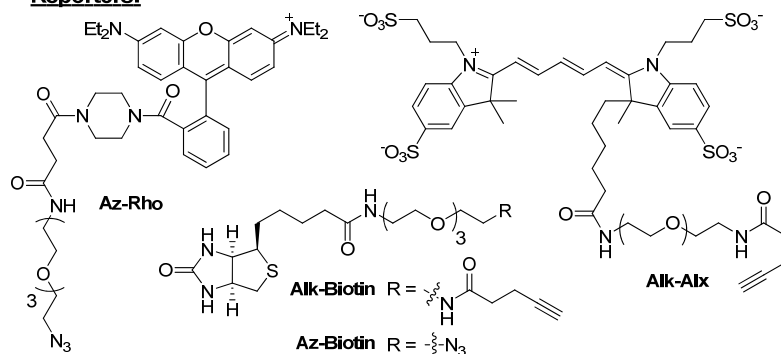
Figure 5.1 Panels of methionine surrogates (in blue) and unnatural metabolite PTM probes (in orange) form bio-orthogonal pairs for compatible double metabolic incorporation. Note that those forming pairs are boxed in the same group. Azide/alkyne-containing fluorophore/biotin reporters are also shown (bottom).
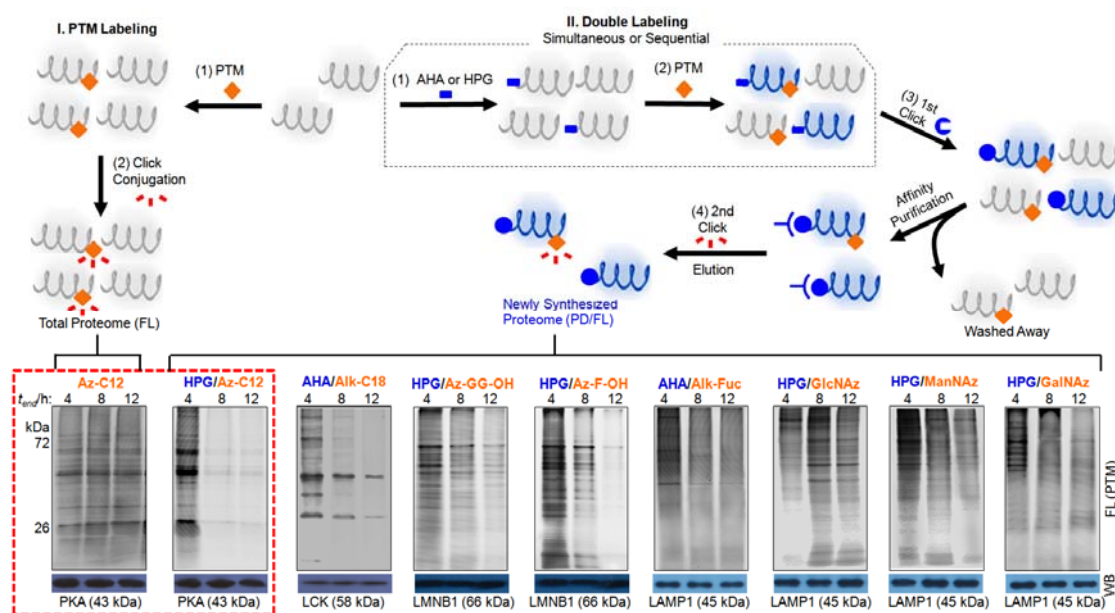
Figure 5.2 Double metabolic incorporation strategy for proteome-wide PTM profiling of newly synthesized proteomes. (a) Overall flow of the strategy. (b) Dynamic monitoring of eight PTMs on newly synthesized proteomes (AHA/HPG feeding: 1 h; PTM probe feeding: three 4-h windows with tend represents the end time (4: 0-4 h; 8: 4-8 h; 12: 8-12 h)). Each experiment was conducted in duplicate to ensure reproducibility. (Top) Fluorescence (FL) gels detecting PTM incorporation/profiles. (Bottom) Western blotting (WB) of the same gel was carried out to ensure newly synthesized proteins in three different lanes (0-4 h, 4-8 h, 8-12 h), upon isolation and loading on the gel, remained at a constant level. Each type of PTM was represented by a key protein (PKA for myristoylation, LCK for palmitoylation, LMNB1 for prenylation, LAMP1 for glycosylation). (Boxed in red) Representative examples of myristoylation showing significant differences in the PTM profiles between the "total proteome" and "newly synthesized proteome". PD = pull down. See **Figure 5.7** for full details.

remains a technically challenging undertaking. Traditionally, PTMs have been studied by

standard molecular biology techniques involving tedious isolation of individual proteins

followed by direct detection/analysis of amino acids bearing the modification. Recent

advances in mass spectrometry, when combined with stable isotope and/or metabolic

labeling approaches, has witnessed the successful examples of several large-scale studies

of PTMs and their dynamics (Mann, M. et al. 2003; Heal, W.P. et al. 2010; Hang, H.C. et

al. 2003; Hsu, T.L. et al. 2007; Hang, H.C. et al. 2007; Martin, B.R. et al. 2009; Kostiuk, M.A. et al. 2008; Davies, B.S. et al. 2008; Kratchmarova, I. et al. 2005; Matsuoka, S. et al. 2007; Daub, H. et al. 2008). These methods, however, analyze PTM changes of total proteins (old and new) present in the cell at the time of sampling, and thus are only able to evaluate PTM dynamics at the ensemble level (Kratchmarova, I. et al. 2005; Matsuoka, S. et al. 2007; Daub, H. et al. 2008; Pedersen, S. et al. 1978).With unnatural metabolic building blocks and in vivo-compatible conjugation chemistries becoming increasingly available (Figure 5.1) (Johnson, J.A. et al. 2010; Agard, N.J. et al. 2009; Hannoush, R.N. et al. 2010; Sletten, E.M. et al. 2009), we sought to develop a proteomic strategy for the detection and identification of newly synthesized proteomes and their PTMs (Figure 5.2). We envisioned several advantages to study the PTM dynamics of newly synthesized proteomes: (i) it decreases the complexity of the proteome and enables the identification of PTM changes that occur in a pre-defined protein synthesis window; (ii) it gives an accurate estimate on the time scale of different PTM events in transforming newly synthesized, modification-free proteins into mature functional entities; (iii) it permits PTM analysis of primary protein synthesis responses to internal and external cues. To isolate a newly synthesized proteome, we made use of BONCAT (bio-orthogonal noncanonical amino acid tagging) (Johnson, J.A. et al. 2010; Dieterich, D.C. et al. 2006; Beatty, K.E. et al. 2006; Kramer, R.R. et al. 2009; Deal, R.B. et al. 2010), which uses known methionine surrogates, azidohomoalanine (AHA) and homopropargylglycine (HPG), for metabolic incorporation into newly synthesized proteins (Figure 5.1; in blue), and the corresponding alkyne- or azide-modified biotin reporter (Figure 5.1; bottom) for subsequent proteome isolation. To follow dynamic changes of an PTM event, we fed growing cells with an azide- or alkyne-containing sugar, fatty acid or lipid building block

(Figure 5.1; in orange) (Agard, N.J. et al. 2009; Hannoush, R.N. et al. 2010). It should be noted that, while our work was in progress, Hang and coworkers recently reported a tandem labeling and detection method to monitor the dynamic acylation of LCK (a tyrosine kinase) and its turnover (Zhang, M.M. et al. 2010). Their work focused on the study of single PTM events (i.e. protein palmitoylation) of a specific protein (i.e. LCK) in a total proteome. Our work herein, while conceptually similar, greatly expands the scope of this double metabolic incorporation strategy by successfully demonstrating, for the first time, simultaneous monitoring of PTM dynamics on multiple newly synthesized proteins (at the proteome scale) and against different types of PTMs (8 in total). As a result, unique primary PTM changes caused by external stimuli may be discovered. By further applying this improved strategy to monitor PTM changes of newly synthesized proteomes in Jurkat T cells, we discovered, for the first time, the up-regulation of myristoylated protein kinase A (PKA; a key signalling enzyme) was intimitely linked to butyric acid (BA)-induced apoptosis.

In our double incorporation strategy (Figure 5.2), growing Jurkat cells were fed first with AHA/HPG at a pre-defined protein synthesis window (Step 1), then again with a desired "unnatural" PTM probe (orange structures shown in Figure 5.2a) (Step 2). Next, a copper-catalyzed azide-alkyne [3+2] cycloaddition reaction (i.e. 1st Click) was used to tag (with a biotin reporter), followed by affinity isolation (with avidin agarose beads) of, only the newly synthesized proteome (Step 3). "Old" proteins, modified or unmodified post-translationally, were removed at this stage. Subsequent on-bead, 2nd-Click reaction (Step 4; with a fluorophore reporter), followed by elution and SDS-PAGE analysis enabled the fluorescence visualization and quantitative analysis of any PTM event that might have

occurred on these newly synthesized proteins. The double incorporation of AHA/HPG and PTM probes may be carried out either simultaneously or sequentially.

5.3 Results and Discussion

5.3.1 Optimization of AHA/HPG incorporation

In order to obtain optimized protocols suitable for the double incorporation strategy and subsequent proteomic studies, we first determined the optimal concentration and time needed for metabolic incorporation of AHA/HPG in Jurkat cells (Figure 5.3 & 5.4); treatment of the cells with either 0.5 mM of AHA/HPG in 10 minutes or 50 μM in 2 hours, showed that sufficient and comparable amounts of newly synthesized proteins could be obtained. Concurrent addition of cycloheximide (CHX; a protein synthesis
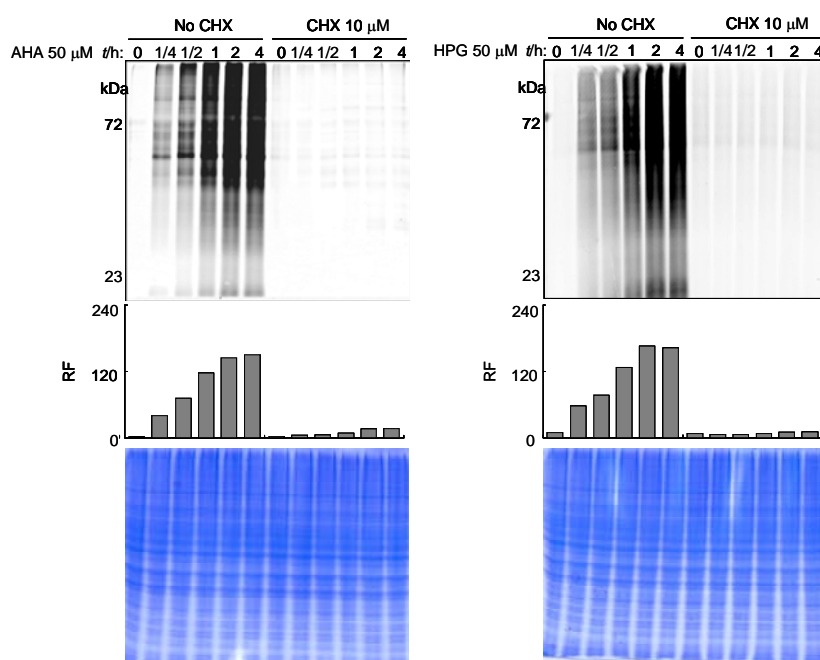
Figure 5.3 Time-dependent metabolic labeling of AHA/HPG in Jurkat T cells. Top: in-gel fluorescence scanning of time-dependent metabolic incorporation of AHA (left) or HPG (right) in the presence or absence of CHX. Bottom: quantitative analysis of each lane from the corresponding fluorescence gels (top). Y-axis: relative total fluorescence (RF) of all bands in each lane.
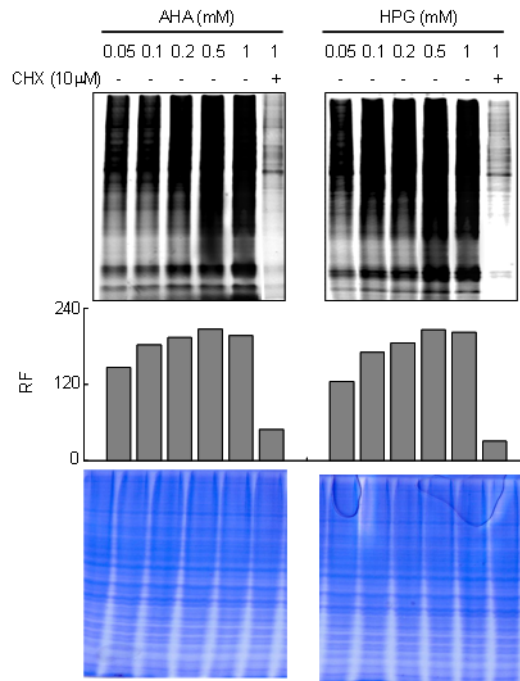
Figure 5.4 Concentration-dependent metabolic labeling of AHA/HPG in Jurkat cells. Top: fluorescence gels showing AHA or HPG's concentration-dependant incorporation (with 10 min AHA/HPG incorporation time). Bottom: quantitative analysis of each lane from the corresponding fluorescence gels (top). Y-axis: relative total fluorescence (RF) of all bands in each lane.
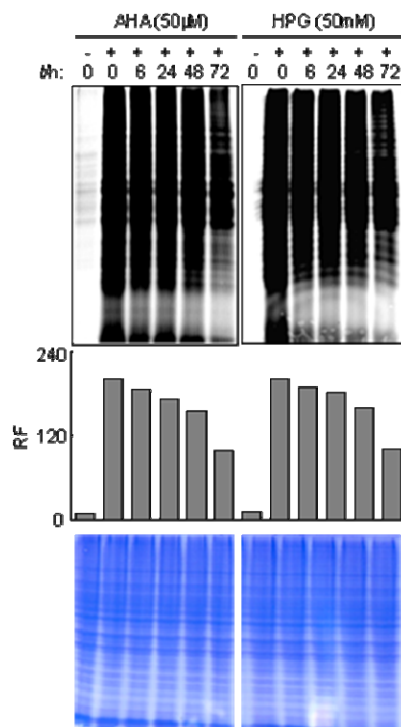
Figure 5.5 Recycling rate of AHA/HPG-labeled proteins in Jurkat cells. Top: fluorescence gels showing recycling rate of AHA/HPG-labeled proteins (50 μM x 2 hours); Bottom, quantitative analysis of each lane above.

inhibitor) during AHA/HPG feeding windows served as a convenient means to monitor/quantify newly synthesized proteomes from the resulting fluorescence gel, while the corresponding Coomassie-stained gel (which detects total proteomes) ensured that equal amounts of total proteomes across different gel lanes were loaded/compared. The AHA/HPG-labeled, newly synthesized proteomes were shown to retain at least 80% of their total original fluorescence intensities after 48 hours (Figure 5.5), indicating their overall recycling rate was insignificant under our assay windows.

5.3.2 Optimization of PTM probes incorporation

We also determined the incorporation efficiency of eight different PTM probes (structures shown in Figure 5.1) covering major protein PTM events - glycosylation, acylation (palmitoylation & myristoylation) and prenylation (farnesylation & geranylation); results showed that, with 50   M of a probe, as little as 4 hours was
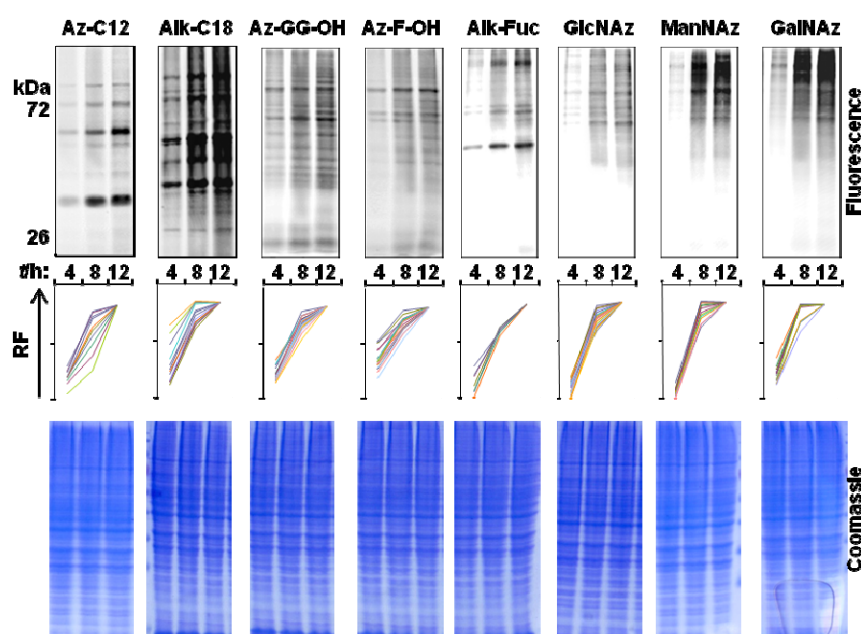
Figure 5.6 Time-dependent metabolic incorporation of PTM probes. Top: fluorescence gels showing metabolic labeling of PTMs (4, 24 and 72 h; 50   M). Center: quantitative analysis of each lane. All distinguishable bands were quantified and plots in the same graph, with each curve in the graph representing one band. Y axis: relative normalized fluorescence (RF) of each individual band throughout the duration. Bottom: coomassie gels showing equal protein loading in each lane.

sufficient to detect incorporation on major protein bands for most PTM events (Figure 5.6). PTM probe incorporation normally reached saturation after 1 day of feeding. Increasing probe dosage correspondingly shortened the time needed to achieve the same level of PTM incorporation.

5.3.3 Identification of PTM probe modified proteins

Large-scale pull-down and LC-MS/MS experiments were subsequently carried out to enrich these post-translationally modified proteins and unambiguously confirm their protein ID (Table S5.1); upon PTM probe incorporation, proteins were tagged with the corresponding biotin probe by Click chemistry, affinity-purified on avidin agarose beads, followed by SDS-PAGE separation. Subsequent LC-MS/MS analysis of gel slices led to the identification of a total of 177 post-translational modifications on 112 different protein entries (some proteins have more than one PTM). On average, the identification rate was about 22.1 modifications for each of the eight pull-down experiments. These proteins were then cross-checked with known literatures. At the end, we confirmed a total of 71 modifications on 44 unique proteins with high confidence (that is, their PTM had previously been documented in the literature). It should be noted that our main goal in this study was not to exhaustively identify/validate unknown PTMs. Therefore, only previously identified, high-confidence PTMs were listed (in Table S5.1) and selected protein targets (i.e. LCK & PKA) were followed up.

Table 5.1 Proteins identified by mass spectrometry

Entry	Protein Name	Accession No.	Calc/Obs Mass Da/kDa	# Pep.	Protein Score	E value	Ion Score	PTM Probe Used	Modification[1]
1	TFRC	IPI00022462	85274/85	6	180	5.2e-007	86	Alk-C18，GlcNAz, ManNAz,	A and G
2	CANX	IPI00941747	67982/68	21	280	2.8e-005	69	Alk-C18, Az-C12	A
3	LCK	IPI00394952	60878/58	7	103	2.5e-012	47	Alk-C18, Az-C12	A
4	SLC1A5	IPI00019472	56201/56	6	61	2e-023	42	Alk-C18, ManNAz	A and G
5	PSMC1	IPI00011126	49325/49	9	83	0.00034	56	Alk-C18, Az-C12	A
6	SCAMP3	IPI00306382	38661/38	1	62	0.00016	62	Alk-C18	A
7	MLEC	IPI00029046	32385/32	2	40	0.045	37	Alk-C18	A
8	SURF4	IPI00005737	30602/30	2	52	0.00088	52	Alk-C18	A
9	GNAI3	IPI00748145	40451/40	22	504	5.4e-008	97	Alk-C18, Az-C12	A
10	CYB5R3	IPI00328415	38544/38	3	56	0.0052	43	Az-C12	A
11	ARF1	IPI00215914	20741/21	28	345	1.3e-006	82	Alk-C18, Az-C12	A
12	ARF4	IPI00215918	20612/21	12	240	0.0004	58	Alk-C18, Az-C12	A
13	ARF5	IPI00215919	20631/21	11	163	0.00086	50	Alk-C18, Az-C12	A
14	PPM1G	IPI00006167	59919/60	8	103	0.00017	59	Az-C12	A
15	GORASP2	IPI00743931	48929/49	1	45	0.0056	45	Az-C12	A
16	FAM49B	IPI00303318	37010/37	20	352	1.9e-005	70	Alk-C18, Az-C12	A
17	CHCHD3	IPI00015833	26421/27	4	111	0.00037	57	Az-C12	A
18	CHMP6	IPI00305423	23527/24	2	58	0.0048	46	Az-C12	A
19	PKA	IPI00217960	40590/41	10	192	1.8e-005	70	Az-C12	A
20	MARCKSL1	IPI00641181	19574/20	7	92	0.00025	59	Az-C12	A
21	RAB39	IPI00001618	25007/25	2	47	0.0032	47	Az-F-OH	P
22	RAB33B	IPI00021475	25718/26	2	47	0.0032	47	Az-F-OH	P
23	RAB3A	IPI00023504	24984/25	2	47	0.0032	47	Az-F-OH	P
24	RAB35	IPI00300096	23296/23	3	41	0.0032	47	Az-F-OH	P
25	RAB7A	IPI00016342	23760/24	3	76	0.00014	61	Az-F-OH	P

26	RAB1A	IPI00005719	22891/22	6	41	0.0032	47	Az-F-OH	P
27	RAB15	IPI00383449	23731/23	3	41	0.0032	47	Az-F-OH	P
28	LMNB1	IPI00217975	66653/67	7	107	7.1e-005	63	Az-F-OH&Az-GG-OH	P
29	NAP1L1	IPI00023860	45631/45	2	47	0.038	36	Az-F-OH &Az-GG-OH	P
30	HSP90B1	IPI00027230	92696/90	1	51	0.0013	51	Alk-Fuc, GalNAz, ManNAz, GlcNAz	G
31	LAMP1	IPI00884105	45367/45	2	36	0.02	36	Alk-Fuc, GalNAz, ManNAz, GlcNAz	G
32	ATP1B3	IPI00008167	31834/32	2	56	0.02	39	Alk-Fuc GalNAz, ManNAz, GlcNAz	G
33	NUP214	IPI00183294	214230/200	1	36	0.027	36	GlcNAz	G
34	SLC3A2	IPI00027493	58023/58	4	101	3.7e-005	66	GalNAz, ManNAz, and GlcNAz	G
35	NUP54	IPI00172580	55643/56	11	277	9.3e-008	92	GlcNAz	G
36	NUPL1	IPI00107122	60974/60	3	101	0.00012	59	GlcNAz	G
37	NUP62	IPI00293533	53394/54	4	53	0.0069	43	GlcNAz	G
38	NUP153	IPI00292059	155440/150	8	138	0.0043	48	GlcNAz	G
39	HCFC1	IPI00019848	210598/200	16	114	0.00064	49	GlcNAz	G
40	M6PR	IPI00025049	31487/31	3	67	9.7e-005	63	GalNAz, ManNAz	G
41	PTPRC	IPI00155168	132530/130	7	84	0.00014	58	GlcNAz	G
42	BSG	IPI00019906	29431/29	10	118	2.2e-005	71	GalNAz, ManNAz	G
43	SPN	IPI00027430	40297/40	2	35	0.027	35	GalNAz, ManNAz	G
44	LAMP2	IPI00009030	45503/45	2	35	0.034	35	GalNAz, ManNAz	G

[1]Summary of type of PTM detected from the experiment. A = Acylation (including palmitoylation and myristoylation); G = Glycosylation; P = Prenylation.

5.3.4 Double Metabolic Incorporation of AHA/HPG-PTM probe pairs

We next carried out double metabolic incorporation using all eight pairs of AHA/HPG-PTM probe combinations and studied PTM dynamics on newly synthesized proteomes (Figure 5.2b). The most important feature of our double incorporation strategy is the ability to artificially "fix" a protein synthesis window while "varying" the timing of PTM. In doing so, we were able to quantitatively analyze the same pool of proteins at different intervals of a PTM event during the cellular process (e.g. comparing different lanes in each fluorescence gel in Figure 5.2b), thus a potentially more accurate picture of PTM dynamics could be depicted. Briefly, protein lysates were obtained from cells treated with an one-hour feeding of AHA/HPG (0.2 mM) and each of the three successive 4-hour feedings with each of the eight PTM probes (0.2 mM; 0-4, 4-8 and 8-12 h, where t = 0 at start of AHA/HPG feeding). Following sequential Click chemistry, affinity purification and gel separation, PTM profiles of these newly synthesized proteomes were rendered visible by in-gel fluorescence scanning (Figure 5.2b, top gels); each fluorescent band may be assigned to a unique protein undergoing a specific type of post-translational modification. Equal protein loading in each lane was assured by Western blotting of the same gel (Figure 5.2b, bottom gels). Striking differences were observed between PTM profiles of the "total proteome" and "newly synthesized proteome"; representative examples of myristoylation profiles were shown in Figure 5.2b (boxed in red). This highlights the key advantage of our strategy over existing pulse-chase, isotope-based labeling methods; its ability to isolate/amplify newly synthesized
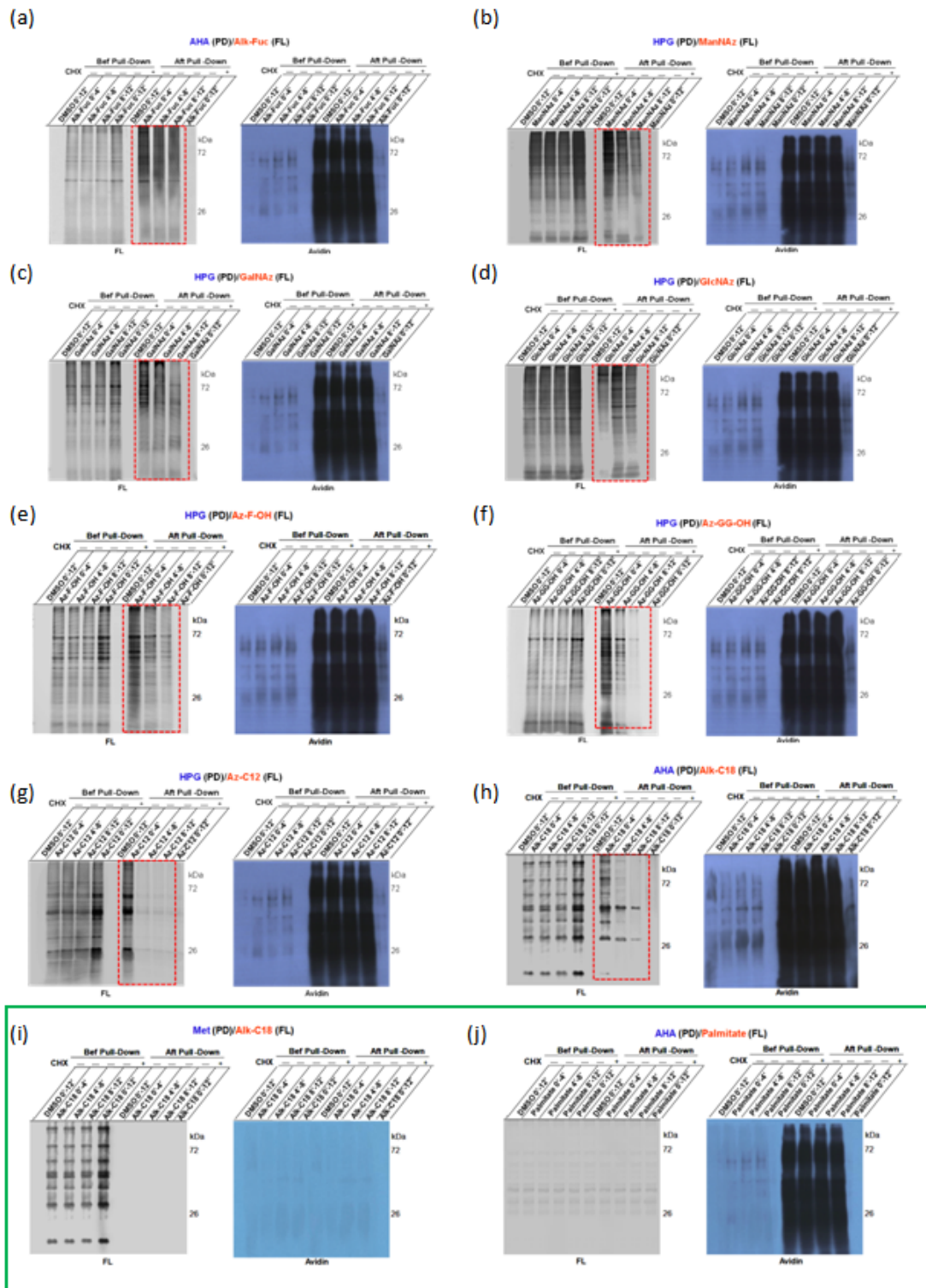
Figure 5.7 Dynamic profiling of PTMs on newly synthesized proteins by each of the 8 AHA/HPG-PTM probe pairs. Dynamic profilings of 8 PTMs on proteins newly synthesized in an 1-hour window, then chased in three 4-hour windows (**4**: 0-4 hour; **8**: 4-8 hour; **12**: 8-12 hour).AHA/HPG was tagged with biotin and PTM probes were tagged with a fluorophore, respectively. Left gel: fluorescence gel profile of the total proteome (before pull-down) and newly synthesized proteome (after pull-down); Right gel: avidin-blot of the same gel showing the total proteome (before pull-down) and newly synthesized proteome (after pull-down). (a) AHA and FC-Ak double incorporation; (b) HPG and Man-Az double incorporation; (c) HPG and Gal-Az double incorporation; (d) HPG and Glc-Az double incorporation; (e) HPG and FA-Az double incorporation; (f) HPG and GA-Az double incorporation; (g) HPG and MA-Az double incorporation; (h) AHA and PA-Ak double incorporation. The first and last lanes of each gel were control lanes: (1st) - DMSO was added in 0-12 hour period in place of a PTM probe; (last) - both CHX and a PTM probe were simultaneously added to the "chase" reaction in 0-12 hour period. Selected lanes (boxed in Red) from each fluorescence gel were reproduced as Figure 5.2a in the maintext.

proteomes and their changes which are normally undetectable in total proteomes (Kratchmarova, I. et al. 2005; Matsuoka, S. et al. 2007; Daub, H. et al. 2008; Pedersen, S. et al. 1978). Further quantitative fluorescence analysis of major fluorescent bands from each PTM event revealed different levels of temporal control on newly synthesized proteins (data not shown); most PTM probes, except GlcNAz, displayed the highest incorporation in the first 4-hour window upon protein synthesis, then gradually stopped in the subsequent two 4-hour windows, indicating most PTMs occurred quite rapidly as soon as the protein synthesis was complete. The most clear-cut temporal regulation was observed with the myristoylation profile, which showed little or no detectable Az-C12 incorporation after the first 4-h window, corroborating well with the co-translational nature of this type of PTM. Closer inspection of different bands in each gel also revealed dissimilar dynamic profiles amongst proteins undergoing the same type of PTM.

5.3.5 Monitoring the palmitoylation dynamics of newly synthesized proteomes

We further obtained a more detailed profile of palmitoylation dynamics by repeating the AHA/Alk-C18 double incorporation experiment with newly synthesized proteomes fed with 0.5 mM of AHA for 10 minutes, then with 0.5 mM of Alk-C18 for 20 minutes ending at 9 different time points (20 min, 40 min, 60 min, 2 h, 4 h, 8 h, 24 h, 48 h & 72 h, respectively, after protein synthesis). Prior to affinity purification, the total proteome (including both "old" and newly synthesized proteins) displayed highly similar palmitoylation profiles across the nine 20-min palmitoylation windows (Figure 5.8a; right gel). Upon affinity enrichment of only the newly synthesized proteome and in-gel fluorescence scanning, the resulting palmitoylation profiles revealed significant differences for different proteins across the different 20-min windows (Figure 5.8a; left gel). To further delineate the palmitoylation dynamics of each protein target, eight distinct fluorescent bands, one of which was unambiguously validated to be LCK (WB gels in Figure 5.9), were quantified and graphically plotted (Figure 5.8b); even by looking at only these proteins, highly diverse palmitoylation profiles were already evident. Rapid palmitoylation was observed for six of the eight protein bands, most of which peaked within 20 to 40 min following protein synthesis, and dropped to almost undetectable level after 2 hours. In contrast, enduring palmitoylation was observed for the two remaining protein bands. LCK is a well-known N-myristoylated and S-palmitoylated non-receptor tyrosine kinase, whose membrane localization has profound biological
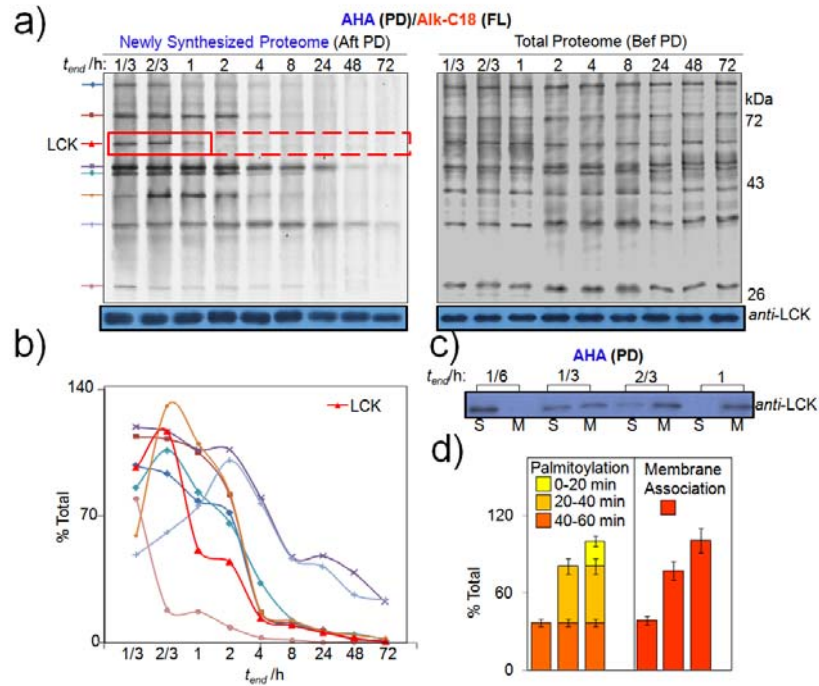
Figure 5.8 Monitoring the palmitoylation dynamics of newly synthesized proteomes. (a) Comparing the palmitoylation profiles (FL) generated from newly synthesized proteomes (left) and total proteomes (right), as obtained from double incorporation experiments (AHA: 10 min; Alk-C18: nine 20-min windows). The fluorescence intensity of eight major bands (arrowed) were analyzied and plotted in b). The band corresponding to LCK was unambigiously validated by WB with anti-LCK antibody. Each of the other protein bands were characterized and tentatively assigned a unique protein ID by comparing with the corresponding silver-stained gel and MSMS analysis (Figure 5.9). In the newly synthesized proteome (left gels), bands corresponding to LCK was boxed (in red). WB analysis (bottom gels) confirmed equal amount of LCK protein expression in each lane despite obvious differences in fluorescence intensities which represent different levels of LCK palmitoylation (top; boxed). (b) Quantitative analysis of 8 different palmitoylated targets identified from the newly synthesized proteome shown in (a), sampled at nine 20-min windows. Each fluorescent band from the newly synthesized proteome was quantified and divided by that from the total proteome, then plotted. This analysis produced a ratio of a newly synthesized protein's palmitoylation count to that of the total protein over the described period after the protein was synthesized. Line graphs with error bars calulcated from duplicated experiments revealing the palmitoylation change for each protein target are shown in Figure 5.9. (c) Membrane association dynamics of newly synthesized LCK. Lysates from AHA-incorporated cells, upon isolation into soluble (S) and membrane (M) fractions, were affinity-purified to isolate newly synthesized proteomes, then immunoblotted with anti-LCK antibody. Cells were collected at 10, 20, 40 or 60 min after addition of AHA, and analyzed in lanes 1/2, 3/4, 5/6 & 7/8 in the gel, respectively. (d) Comparing palmitoylation dynamics of newly synthesized

127

LCK with its membrane association within 60 min of protein synthesis. (Left) Relative % of accumulative palmitoylation level of LCK as obtained from (b). (Right) Relative % of membrane-associated LCK as obtained from (c). Each experiment was conducted in duplicate to ensure reproducibility. Error bars were generated from duplicated experiments. See Supporting Information for details.
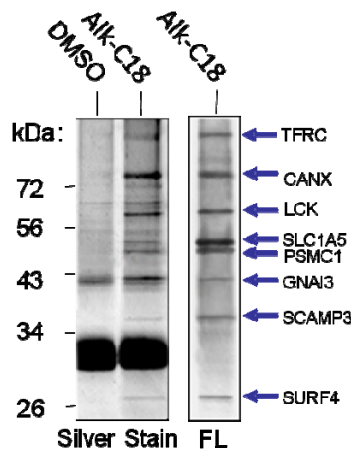
Figure 5.9 Silver-stained gel and western blotting validation of tandem pulled-down fractions from Alk-C18 labeled or DMSO-treated (negative control) Jurkat T cell lysates. (top) Only distinct bands in the pull-down fraction from Alk-C18-labeled sample matching those of the fluorescence gel were cut from the gel and analyzed by MSMS. (bottom) Western blotting analysis of pulled-down fractions with respective antibody (anti-LCK).

Implications (Zhang, M.M. et al. 2010). We were interested to know whether the observed palmitoylation dynamics of newly synthesized LCK from our double incorporation strategy is related to this protein's membrane association kinetics. To determine the subcellular localization changes of newly synthesized LCK, newly synthesized proteomes were isolated from the membrane and soluble fractions of Jurkat cells, analyzed (Figure 5.8c), and compared with the accumulated palmitoylation counts obtained from Figure 5.8b. As shown in Figure 5.8d, within one hour after protein synthesis, the increase of the membrane-associated LCK indeed coincided quite well with palmitoylation dynamics observed. It should be

highlighted that our findings herein would not have been possible, have we not successfully isolated and analyzed only the newly
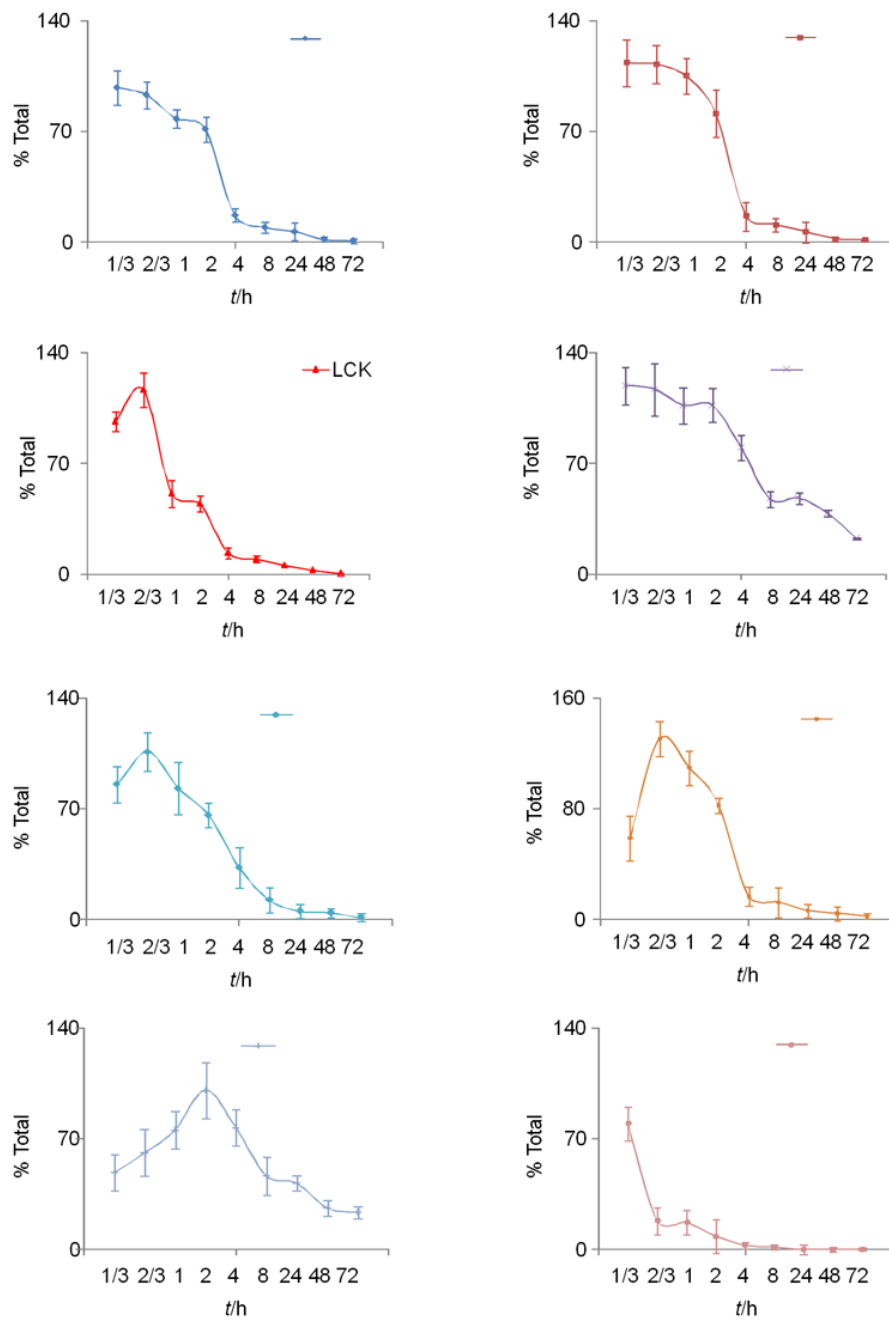
Figure 5.10 Quantitative analysis of the eight different palmitoylated proteins reproduced from Figure 5.8b. The intensity of each fluorescent band in Figure 5.8b from the newly synthesized sub-proteome in duplicates were quantified, averaged

and divided by that from the total proteome. Error bars were generated by calculation of the SE for the results of the duplicate experiments.

synthesized proteome. We concluded this double incorporation strategy enables proteome-wide dynamic profiling of PTMs on newly synthesized proteomes. By adjusting metabolic feeding windows, the strategy is applicable to both rapid and enduring PTM events.

5.3.6 Identification of up-regulated myristoylated PKA at the early stage (0-5 h after additon of BA) of BA-induced apoptosis.

Lastly, we explored the feasibility of this approach to discover novel PTM regulations on newly synthesized proteomes. It is known that butyric acid (BA)-induced protein synthesis in Jurkat T cells is critical for apoptosis to occur, but so far the molecular basis of this process is not well-understood (Medina, V. et al. 1997). We were interested to know whether the dynamic PTM regulation of any of the newly synthesized proteins at the early stage of apoptotic induction might play a role. By applying the double incorporation strategy (HPG/Az-C12) to BA-induced cells between the 0-5 hour window, followed by sequential Click chemistry (HPG labeled with Az-Biotin; Az-C12 labeled with Alk-Alx), enrichment of newly synthesized proteomes and in-gel fluorescence scanning, we obtained the corresponding myristoylation profiles (with and without myristoylation inhibitor HMA; lanes 2 & 3 respectively in the center gel in Figure 5.11a), as well as those of the corresponding total proteomes (i.e. before avidin pull down; left gel). A close examination of the newly synthesized proteomes revealed a significant fluorescence increase in a 43-

kDa band (labeled with an asterisk in Figure 5.11a) only in the BA-induced proteome.

This 43-kDa protein was subsequently identified to be myristoylated PKA catalytic

subunit alpha (PRKACA)
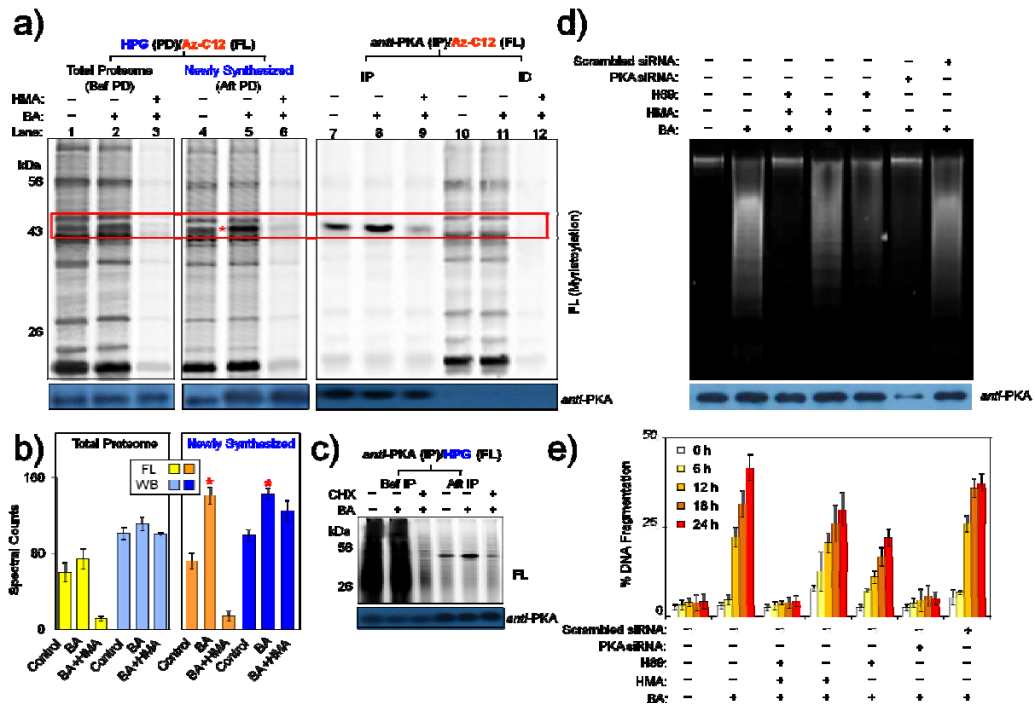
Figure 5.11 Identification of up-regulated myristoylated PKA at the early stage (0-5 h after additon of BA) of BA-induced apoptosis. (a) Comparing myristoylation profiles of the total proteome (left) and newly synthesized proteome (middle). In each gel, control lanes were obtained from cells treated with/without different BA/HMA (a myristoylation inhibitor) combinations. BA (5 mM) was added to growing Jurkat cells. Double incorporation was carried out by simultaneously feeding the cells, during 0-5 h, with HPG (50 M) and Az-C12 (0.2 mM). Newly synthesized proteomes were separated from total proteomes by affinity tagging and pull down (PD). Myristoylation profiles of the proteomes were visualized by fluorescence tagging. (Right) Validation of the 43-kDa band as myristoylated PKA catalytic subunit alpha by immunoprecipitation (IP) and immune-depletion (ID) with anti-PKA. To aid the view, the 43-kDa band corresponding to PKA in all three gels was boxed (in red). Western blotting results with anti-PKA were shown below each gel. (b) Quantitative analysis of the 43-kDa PKA bands from the first two gels in (a). Both the myristoylation counts (FL) and PKA expression count (WB) between the total proteome (left graph) and the newly synthesized proteome (right graph) were plotted. (*) highlighting the up-regulation of myristoylated PKA only became evident in the BA-induced, newly synthesized proteome. Means ± standard errors

were obtained from duplicates. (c) Validation of PKA's up-regulation in newly synthesized proteomes. BA-induced cells were fed with HPG, fluorescently labeled. After immunoprecipitation (IP) with anti-PKA (lanes 4-6), noticeable up-regulation of newly synthesized PKA in lane 5 was evident (compared to lane 4). (d) DNA fragmentation assay showing effects of H89+HMA combination and siRNA knockdown of PKA in blocking BA-induced apoptosis. Western blotting analysis of the corresponding protein samples with anti-PKA reveals the knock-down of PKA expression by PKA siRNA while not scrambled siRNA. (e) Time-dependent BA-induced apoptosis and its inhibition by H89+HMA combination or by siRNA knockdown of PKA. Each experiment was conducted in duplicate to ensure reproducibility. Means ± standard errors were obtained from duplicates.

by immunoprecipitation (IP) and immune-depletion experiments (right gel in Figure 5.11a). Further quantitative analysis was carried out on this fluorescent band (indicating PKA myrisoylation counts) and its Western-blotted counterpart (indicating PKA expression counts) for both the total proteomes and the newly synthesized proteomes (Figure 5.11b); up-regulation of both the myristoylated PKA and overall PKA expression in the newly synthesized proteomes was clearly evident (right graph), but not so in the total proteomes (left graph). Under our assay conditions, the newly synthesized proteome made up only a small fraction of the total proteome. This might have obscured the detection of PKA up-regulation in the total proteome of BA-induced apoptotic cells. Protein myristoylation is normally a co-translational event, and PKA myristoylation has been postulated to be involved in modulation of its translocation and membrane association (Breitenlechner, R. et al. 2004). As a key signaling enzyme involved in energy metabolism, PKA might play a vital role in controlling BA-induced apoptosis of Jurkat cells by up-regulation of its expression level, which leads to myristoylation and subsequent modulation of its enzymatic activity (Bijlmakers, M.J. et al. 2003; Franklin, R. et al. 2000). To confirm this, we determined the effect of H89 (a PKA inhibitor), HMA (a myristoylation

inhibitor) and siRNA of PKA in blocking BA-induced apoptosis (Figure 5.11d & e). DNA fragmentation assay revealed that only a combined H89+HMA treatment, or siRNA knockdown of PKA, was able to sufficiently block BA-induced apoptosis, but not H89 or HMA alone or scrambled siRNA (Figure 5.11d). A time-dependent assay lends further support to our observations (Figure 5.11e), thus unequivocally proving up-regulation of myristoylated PKA is needed for the occurrence of BA-induced apoptosis in Jurkat cells.

5.4 Conclusion

In summary, our double metabolic incorporation approach is capable of proteome-wide profiling of PTM dynamics on newly synthesized proteins. With this strategy, we had for the first time identified myristoylated PKA as the key enzyme in regulating butyric acid-induced apoptosis in Jurkat cells. Work is underway to further expand the utility of this new chemical proteomic tool with other powerful mass spectrometric techniques such as ICAT and SILAC (Gygi, S. P. *et al.* 1999；Ong, S. E. *et al.* 2003).

Chapter 6.

Experimental Procedures

6.1 General

All chemicals used in this dissertation were purchased from commercial vendors and used without further purification unless otherwise stated. All chemicals probes were synthesised by my collaborators in Chemistry lab (Dr. Kalesh for kinase cross-linkers; Shi Haibin for the hydroxylethyl-based probes and inhibitor library for aspartic proteases; Yang Pengyu for THL-based probes and Rhodamine/Biotin alkynes and azides).

6.2 Chapter 2

6.2.1 General

All chemicals were purchased from commercial vendors and used without further purification, unless indicated otherwise. The ^{1}H-NMR spectra were taken on a Bruker 300 MHz NMR spectrometer. Chemical shifts are reported in parts per million referenced with respect to residual solvent (CHCl$_3$ = 7.26 ppm). All enzymes were expressed in *E. coli* strain BL21-DE3 and purified as described previously (Seeliger, M.A., 2007; Slice, L. W., 1989; Khokhlatchev, A., 1997; Bougeret, C., 1993).They include ABL (human c-Abl, residues 229-512), CSK (human c-src

tyrosine kinase, residue 1-450), ERK1 (human mitogen activated protein kinase 3), ERK2 (human mitogen activated protein kinase 1), SRC (chicken c-src residue 251-533), PKA (protein kinase A catalytic subunit). After purification, all kinases were prepared as 2-5 mg/ml solution in 100 mM sodium phosphate buffer (pH 7.5), 500 mM NaCl, 10 % glycerol and stored at -20 °C as stock solution until use. Staurosporine was obtained from i-DNA Biotechnology Pte Ltd (Singapore). Fluorescence scanning of the SDS-PAGE gels was carried out with Typhoon 9200 fluorescence gel scanner (Amersham), and where applicable, the bands were quantified with the ImageQuant software installed on the scanner. For inhibition experiments, the IC_{50} curves were generated using the Graphpad Prism software v.4.03 (GraphPad, San Diego, USA).

6.2.2 Chemical Synthesis

The synthesis of o-phthalaldehyde adenosine (OPA-AD) and naphthalene-2,3-dicarboxaldehyde (NDA-AD) was carried out by my collaborator Dr. Kalesh and the synthesis procedure was based on previously reported literature (Maly, D. J., 2004; Liu, K., 2008). Five pseudosubstrates targeting 5 different kinases were made. based on known peptide substrate sequences in the literature (Casnellie, J.E., 1991 Cheng, H.C., 1992; Gonzalez, F.A., 1991; Kemp, B.E., 1977; Sekimoto, H., 2003) and a cysteine residue was introduced in place of the Ser/Thr/Tyr phosphorylation site present in the original sequences. Where applicable, the original peptide substrate sequences were also synthesized.

135

Panel of Pseudosubstrates

    CSKtide (Fluorescein-GG-KKKKEEICFFF).   Pseudosubstrate for Csk kinase. Calculated MW: 1960; ESI-MS: m/z $[M+H]^{2+}$ = 980.9.

    ERKtide (Fluorescein-GG-ELVEPLCPSGEAPNQ). Pseudosubstrate for Erk1 and Erk2 kinases. Calculated MW: 2096; ESI-MS: m/z $[M+H]^{2+}$ = 1048.9.

    SCRtide (Fluorescein-GG-KVEKIGEGTCGVVYK). Pseudosubstrate for Src kinase. Calculated MW: 2122; ESI-MS: m/z $[M+H]^{2+}$ = 1062.0.

    ABLtide (Fluorescein-GG-EAICAAPFAKKK). Pseudosubstrate for Abl kinase. Calculated MW:1790; ESI-MS: m/z $[M+H]^{2+}$ = 895.9.

    PKAtide (Fluorescein-GG-LRRACLG). Pseudosubstrate for Pka kinase. Calculated MW: 1300; ESI-MS: m/z $[M+H]^{2+}$ = 651.3.

Original Peptide Substrates

Csk:	KKKKEEIYFFF,
Erk (1 & 2):	ELVEPLSPSGEAPNQ,
Src:	KVEKIGEGTYGVVYK,
Abl:	EAIYAAPFAKKK and
Pka:	LRRASLG

6.2.3 Cross-Linking Experiments with Purified Kinases

Unless otherwise indicated, all cross-linking experiments were carried out with the following optimized conditions: kinase (100-140 nM), pseudosubstrate (1 μM), NDA-adenosine or OPA-adenosine (20 μM) in the reaction buffer (25 mM HEPES at pH 7.5, 150 mM NaCl, 2 mM MgCl$_2$) were incubated for 20 min at RT before SDS-PAGE analysis and fluorescence scanning.

6.2.4 Dose-dependent cross-linking studies

To optimize the concentration of the pseudosubstrate in the cross-linking reaction, cross-linking reaction mixtures made up of varied concentrations of the pseudosubstrate peptide (0 to 1 μM), NDA-adenosine (20 μM), kinase (125 nM for Pka and 140 nM for Csk, respectively) in the reaction buffer (25 mM HEPES at pH

7.5, 150 mM NaCl, 2 mM MgCl$_2$).The reactions were incubated at RT for 20 min before SDS-PAGE analysis. Results showed that the cross-link reaction using peptide pseudosubstrate (1 μM) gave the strongest labeling signal with minimum background. To optimize the concentration of the NDA-adenosine in the cross-linking reaction, cross-linking reaction mixtures was made up of the pseudosubstrate (1 μM), kinase (125 nM for Pka and 140 nM for Csk respectively), and NDA-adenosine (0 to 20 μM) in the reaction buffer (25 mM HEPES at pH 7.5, 150 mM NaCl, 2 mM MgCl$_2$). The reactions were incubated at RT for 20 min before SDS-PAGE analysis. Results showed that the cross-link reaction using NDA-adenosine (20 μM) gave the strongest labeling signal with minimum background. To determine whether the cross-linking experiments could be used to quantitatively measure the amount of kinase present in the reaction, and the lowest kinase detection limit, cross-linking reaction mixtures made up of the pseudosubstrate (1 μM), kinase (0 to 125 nM for Pka and 0 to 140 nM for Csk, respectively), and NDA-adenosine (20 μM) in the reaction buffer (25 mM HEPES at pH 7.5, 150 mM NaCl, 2 mM MgCl$_2$). The reactions were incubated at RT for 20 min. After SDS-PAGE, the fluorescence bands of the labeled proteins were quantified with the ImageQuant software (Amersham).

6.2.5 Effect of exogenous thiols and amines on cross-linking efficiency.

To determine cross-linking reaction's tolerance by the presence of an exogenous competing thiol (b-mercaptoethanol) or amine (lysine). Cross-linking experiments with Pka and PKAtide were set up. Briefly, PKAtide (1 μM), Pka (125 nM), NDA-adenosine (20 μM) in the reaction buffer (25 mM HEPES at pH 7.5, 150

mM NaCl, 2 mM MgCl$_2$), in the presence of varied amounts of the exogenous competitor (0 to 200 μM) for 20 min at RT before SDS-PAGE analysis.

6.2.6 Competition effects of ATP/peptide substrate on cross-linking.

To determine whether there is competition effect of ATP/peptide substrate on crosslinking reaction, ATP and LRRASLG-OH, a known Pka peptide substrate (same as PKAtide except with Ser instead of Cys) were used as competitors in the cross-linking reactions. Briefly, cross-linking reactions with PKAtide (1 μM), Pka (125 nM), NDA-adenosine (20 μM) in the reaction buffer (25 mM HEPES at pH 7.5, 150 mM NaCl, 2 mM MgCl$_2$), in the presence of the competitor (0-20 mM for ATP; 0 to 5 mM for LRRASLG-OH) were incubated for 20 min at RT before SDS-PAGE analysis.

6.2.7 Inhibition experiments with Staurosporine.

Pka and Csk were taken as the model kinases and cross-linking experiments were carried out with the corresponding pseudosubstrates, PKAtide and CSKtide, respectively, with or without Staurosporine. Briefly, pseudosubstrate (1 μM), kinase (125 nM for Pka and 140 nM for Csk), NDA-adenosine (20 μM), and Stauosporine (0 to 10^4 nM) in the reaction buffer (25 mM HEPES at pH 7.5, 150 mM NaCl, 2 mM MgCl$_2$) were incubated for 20 min at RT before SDS-PAGE analysis and fluorescence scanning, followed by determination of IC$_{50}$

6.2.8 Cross-linking specificity of experiments using various kinase pseudosubstrates.

In order to examine whether the cross-linking reaction of a given kinase was

specific only towards its cognate peptide substrate (in our case the pseudosubstrate), we took Pka and Csk as model proteins and label them against the various pseudosubstrates. In addition, we carried out *in vitro* phosphorylation assay using the cognate peptide substrates (that is, the peptide substrates with a Ser/Thr/Tyr instead of a Cys mutation).

9.2.8.1 Crosslinking specificity experiments

For the cross-linking experiment with different pseudosubstrates, Pka or Csk (125 nM and 140 nM respectively), a pseudosubstrate (1 μM), NDA-adenosine (20 μM) in the reaction buffer (25 mM HEPES at pH 7.5, 150 mM NaCl, 2 mM $MgCl_2$) were incubated for 20 min at RT before SDS-PAGE analysis. % activity of the kinase against a peptide was calculated by comparing the relative cross-linking efficiency with that of the target pseudosubstrate (taken as 100 %). It appears that for Csk, in addition to its pseudosubstrate CSKtide, the kinase was able to cross-link SRCtide (~ 40 % activity). For Pka, in addition to its pseudosubstrate PKAtide, the kinase was able to cross-link SRCtide (~ 15 % activity).

9.2.8.2 enzymatic phosphorylation specificity experiments

In order to further examine whether the cross-linking profiles of the kinase against different pseudosubstrates were indeed due to the enzyme's intrinsic promiscuous activity to phosphorylate different peptides, we carried out *in situ* enzymatic assay by measuring the phosphorylation ability of the kinase against the original peptide substrates. A Kinase-Glo™ Plus Kit (Promega) was used.[5] Briefly, the kinase (12.5 nM for Pka and 14 nM for Csk), an original peptide substrate (100 μM) and ATP (100 μM) in kinase reaction buffer (20 μl; 25 mM HEPES at pH 7.5,

10 mM MgCl, 0.1 % b-mercaptoethanol, 100 μM Na$_3$VO$_4$) were incubated for 20 min at RT. Subsequently, 20 μL of the Kinase-Glo^{TM} Plus reagent was added. The resulting luminescence generated was detected by a Tecan microplate reader, and the amount of phosphorylation was calculated for each kinase/substrate pair following the vendor's instructions. % activity of the kinase against a peptide was calculated by comparing the relative phosphorylation efficiency with that of the target peptide substrate (taken as 100 %).

6.2.9 Cross-linking of Pka/Csk with their pseudosubstrates in crude lysates.

In a typical reaction, lysate (6 μg), kinase (125 nM for Pka and 140 nM for Csk, corresponding to 3.3 % and 6 % by weight, respectively, of the total proteome), pseudosubstrate (1 μM), NDA-adenosine (20 μM) in the reaction buffer (25 mM HEPES at pH 7.5, 150 mM NaCl, 2 mM MgCl$_2$; final volume of reaction = 20 μL) were incubated for 20 min at RT before SDS-PAGE analysis and fluorescence scanning.

6.2.10 Detection limits of cross-linking in crude lysates.

In order to determine the lowest amount of detectable kinase present in the crude lysate, different amounts of the kinase was spiked. Briefly, lysate (6 μg), kinase (0 to 800 ng for Csk and 0 to 400 ng for Pka, corresponding to 0-12 % and 0-6 % of the total proteome), pseudosubstrate (1 μM), NDA-adenosine (20 μM) in the reaction buffer (final volume = 20 μL) were incubated for 20 min at RT before SDS-PAGE analysis and fluorescence scanning.

6.2.11 Inhibition experiments with Staurosporine

Pka and Csk spiked in the crude lysate were taken as the model kinases and cross-linking experiments were carried out with the corresponding pseudosubstrates, PKAtide and CSKtide, respectively, without or with Staurosporine. Briefly, pseudosubstrate (1 μM), kinase (125 nM for Pka and 140 nM for Csk), NDA-adenosine (20 μM), and stauosporine (0 to 160 μM) in the reaction buffer (25 mM HEPES at pH 7.5, 150 mM NaCl, 2 mM MgCl$_2$) were incubated for 20 min at RT before SDS-PAGE analysis and fluorescence scanning, followed by determination of IC$_{50}$

6.2.12 Multiplexed Kinase Detection and Inhibition in the Crdue Proteome.

In order to demonstrate our approach could be used for simultaneous detection of multiple kinases in the crude proteome, and screening of their potential inhibitors, such that both potency and selectivity of a putative inhibitor could be concurrently evaluated, we carried out multiplexed kinase detection/inhibition experiments. Briefly, both Csk (400 ng) and Pka (0 to 200 ng) were spiked into the same lysate (6 μg), together with NDA-adenosine (20 μM) in the reaction buffer (final volume = 20 μL) and 1 μM of either (a) CSKtide, (b) PKAtide, or (c) a CSKtide/PKAtide mixture (1 μM each). The reactions were incubated for 20 min at RT before SDS-PAGE analysis and fluorescence scanning. For inhibition experiments, similar reactions were set up, in the presence of Staurosporine (0 to 25 μM final concentration), and cross-linked with a mixture of CSKtide and PKAtide (1 μM each).

6.3 Chapter 3

6.3.1 General

All chemicals were purchased from commercial vendors and used without further purification, unless indicated otherwise. The ^{1}H-NMR spectra were taken on a Bruker 300 or 500 MHz NMR spectrometer. Chemical shifts are reported in parts per million referenced with respect to residual solvent (CDCl$_3$ = 7.26 ppm). Pepsin and Renin (from Mucor miehei) were from Fluka. Enzyme stocks were made fresh using 10 mM HCl (for pepsin) or water (for renin). Plasmepsin I (PM-I), Plasmepsin II (PM-II), histoaspartic protease (HAP) and HAP mutant H34A, S37A, K78A E278A were expressed and purified as described previously (Helm, K.V.D., 1994; Parr, C. L., 2008)Anti-PM-I mouse serum mAb 1C6-24 , anti-PM-II rabbit serum 737, anti-PM-IV mouse mAb 13.9.2 were requested from MR4 (USA). The anti-HAP mouse mAb was kindly provided by Dr. Daniel E. Goldberg. In-gel fluorescence scanning of the SDS-PAGE gels was carried out with Typhoon 9200 fluorescence gel scanner (GE), and where applicable, the bands were quantified with the ImageQuant™.[2] For inhibition experiments, the IC$_{50}$ curves were generated using the Graphpad Prism software v.4.03 (GraphPad, San Diego, USA).

6.3.2 Chemical synthesis

The synthesis of all chemicals used in Chapter 3was carried out by my collaborator Dr.Shi Haibin and the synthesis procedure was reported elsewhere (Liu, K., 2009).

6.3.3 Labeling of recombinantly purified aspartic proteases

The 7 A*f*BPs (**A-G**) were characterized with various recombinant aspartic proteases, respectively pepsin, renin, PM-I, PM-II and HAP. Each of the 7 probes was tested with each aspartic protease individually. Details of the labeling experiment are described below. Pepsin stocks were made fresh using 10 mM HCl. Renin stocks were made fresh by using water. Stocks of PM-I (350 ng/µl in 100 mM NaOAc, 10% glycerol, pH 4.5), PM-II (250 ng/µl in 100 mM NaOAc, 10% glycerol, pH 4.5) and HAP (1.5 mg/ml in 100 mM MES, 10% glycerol, pH 6.5) were diluted in appropriate buffers before use (100 mM NaOAc, 10% glycerol, pH 4.5 for PM-I and PM-II, 100 mM NaOAc, 10% glycerol, pH 6.5 for HAP, respectively). Labeling experiments were carried out with the following optimized conditions: desired enzyme amounts (100 ng for pepsin, 100 ng for renin, 50 ng for PM-I, PM-II and HAP, respectively) were individually incubated with each probe (5 µM final concentration) in appropriate buffer for 20 min at RT. For pepsin and renin, reaction buffer conditions were as follows: 50 mM Tris Cl at pH 2.0. For PM-I, PM-II and HAP, reaction buffer condition was 100 mM NaOAc, 10% glycerol, pH 4.5. After 20 min incubation at RT, samples were irradiated on ice for 25 min using a B100A lamp (UVP) at a distance of 5 cm. After irradiation, samples were boiled for 10 min with 4 µl of 6 x loading buffer. The sample mixture (24 µl) was resolved by 12% SDS-PAGE followed by in-gel scanning for fluorescence with a Typhoon 9200 gel scanner. Labeled bands were quantified by ImageQuant™ software. Results were analyzed/represented by cluster analysis software Treeview™ as previously described.(Wang, J., 2006; Uttamchandani, M; 2008; Kalesh, K.A., 2007),

6.3.4 Labeling of Plasmepsins in malaria parasite lysates

Briefly, *P. falciparum* 3C7 strain was used in this study. Parasites were cultured in RPMI medium 1640 (Invitrogen, USA) supplemented with 0.29225 g of L-glutamine, and 0.05 g of hypoxanthine dissolved in 1 ml of 1 M NaOH. Parasites were synchronized twice 16 hours apart at ring stage using 2.5% sorbitol. Cultures were stored at 37 $^{\circ}$C after gassing with a 5% CO_2, 3% O_2 and 92% N_2 gas mixture and their hematocrit maintained at 2.5%. Parasitized red blood cells were collected by centrifugation and treated with 0.1% Saponin in PBS for 15 min at room temperature with shaking. The parasites were centrifuged and the pellet was resuspended (in 50 mM sodium Acetate at pH 5.5, 1 mM DTT, 0.1% NP-40) and homogenized. Proteome Labeling experiments were performed with desired amounts (20 μg) of parasite lysates from each blood stage in 200 μl reactions containing Reaction Buffer (100 mM sodium acetate, 10 % glycerol, pH 5.0) and 5 μM of each of the A*f*BP (**A-G**). Inhibition experiments were carried out in the presence of 1 μM Pepstatin. After 30 min incubation at RT, samples were irradiated on ice for 25 min using a B100A lamp (UVP) at a distance of 5 cm. After irradiation, samples were concentrated to 20 l using a Microcon column (GE healthcare). Then samples were boiled for 10 min with 4 μl of 6 x SDS loading buffer, resolved on a 12% SDS-PAGE followed by in-gel fluorescence scanning. The relative fluorescence intensity of the 37-KDa bands labeled by each of the 7 AfBPs were quantified using ImageQuant™ software.

6.3.5 2D-PAGE analysis of AfBP labeled parasite proteome

After labeling as earlier described, the sample was desalted and concentrated by acetone precipitation overnight at -20 $^{\circ}$C. Precipitates were collected by

centrifugation at 13000 rpm. Samples were resuspended in 1 x re-swelling buffer (8 M urea, 2% CHAPS and 2% IPG buffer and 0.002% bromphenol blue), and briefly sonicated. Just prior to use, DTT was added (final concentration: 3 mg/ml). Rehydration buffer and stock DTT (62.5 mg/ml) were stored at 20 °C. Rehydrated IEF strips were isoelectrically focused at RT under low viscosity oil with a gradient voltage of 0-200 V for 1 min, 200-3,500 V for 1.5 h, and a constant voltage of 3,500 V for 1.5 h. After IEF separation, the gel strip was reduced and alkylated. The reduction step was performed for 15 min in 10 ml of equilibration buffer 1 (0.5% w/v DTT in 50 mM Tris-HCl pH 8.8, 6 M urea, 30% v/v glycerol, 2% w/v SDS, and 0.002% w/v bromphenol blue). The alkylation step was performed for 15 min in 10 ml of equilibration buffer 2 (4.5% w/v iodoacetamide in 50 mM Tris-HCl pH 8.8, 6 M urea, 30% v/v glycerol, 2% w/v SDS, and 0.002% w/v bromphenol blue). The IEF strip was equilibrated in equilibration buffer stock solution (50 mM Tris-HCl pH 8.8, 6 M urea, 30% v/v glycerol, 2% w/v SDS, and 0.002% w/v bromphenol blue). Next, the equilibrated IEF strips were each placed on a preparative well and sealed using 1% agarose plus 0.002% (w/v) bromphenol.

6.3.6 MSMS analysis of AfBP labeled protein

After electrophoresis the two-dimensional gels were scanned with Typhoon fluorescence scanner. Fluorescence spots corresponding to each target gel spot were excised directly from the gel. Trypsin digestion was performed with In-Gel Trypsin Digestion Kit from Pierce. After digestion, digested peptides were then extracted from the gel with 50% ACN and 1% formic acid. Tryptic peptide extracts were evaporated by speedvac and reconstituted with 10 μl of 0.1% TFA, a volume of 2 μl

of the peptide extracts were manually spotted onto a Prespotted AnchorChip MALDI target plate for MALDI-TOF Mass Spectrometry (Bruker Daltonics) and incubated for 3 min, followed with washing step with 10 µl of 10 mM ammonium phosphate in 0.1% TFA, then dried. MALDI TOF mass spectra were recorded using Ultraflex III TOF/TOF mass spectrometer (Bruker Daltonics) with the compass 1.2 software package including flexControl 3.0 and flexAnalysis 3.0, calibrated with PAC peptide calibration standards. MS/MS analysis for the major peaks in PMF spectra were carried out by autoLIFT on the MALDI-TOF/TOF instrument. MS and MS/MS Peak lists with intensity value were submitted to Matrix Science Mascot server (http://www.matrixscience.com/) through BioTools 3.0 (Bruker Daltonics) using database NCBInr with species of Plasmodium falciparrum (malaria parasite), variable modifications of carbamidometyl on cysteine (C) and oxidation on methione (M), allowing maxium of 1 trypsin missed cleavage, peptide mass tolerance at 200 ppm; MS/MS mass tolerance of 0.7 Da.

6.3.7 Pulldown of AfBP labeled protein targets

Briefly, 1 mg of the lysate was labeled by probe **G** (5 µM) as earlier described, and acetone precipitated followed by resolubilization in 0.1% SDS in PBS with brief sonication. This resuspended sample was then precleared using protein G sepharose beads for 1 hr. The protein supernatant was further incubated with anti-TMR antibody (100 µg/1.0 mg of protein) at 4 $^{\circ}$C for 4 h. The samples were then added to pre-washed protein G Sepharose beads (300 µl/1 mg of protein) at 4 $^{\circ}$C for 2 h. Following centrifugation, the beads were washed with 1% SDS in PBS for 4 times. After washing, the beads were boiled in elution buffer (200 mM Tris pH 6.8,

400 mM DTT, 8% SDS). The supernatant was collected after centrifugation and precipitated by acetone precipitation for 4 h at -20 ^{o}C. This pull-down sample was then separated on a 12% SDS PAGE gel (un-labeled lysate was run aloneside as a negative control). After in-gel fluorescence scanning, the gel was then transferred to a PVDF membrane and subsequently blocked with 2.5% (w/v) BSA/PBST. Membranes were incubated for 1 h at RT with antibodies that recognize each of the four plasmepsins (anti-PM-I mouse serum mAb 1C6-24, (1:5000), anti-PM-II rabbit serum 737 (1:5000), anti-HAP mouse mAb (1:5000), anti-PM-IV mouse mAb 13.9.2 (1:5000)) in 2.5% (w/v) BSA/PBST, followed by appropriate secondary antibody incubation. After wash with PBST for three times the SuperSignal West Pico kit (Pierce) were used to develop the blot.

6.3.8 Membrane/Soluble sub-proteome analysis

Highly synchronized parasite total lysates were prepared as earlier described. Total lysates were further centrifuged at 13000g for 15 min and split into a pellet fraction and a soluble fraction. Protein concentrations of each fraction were quantified. Total lysates, Soluble and insoluble fraction lysates (20 μg each) from each intraerythrocytic stage were labeled by probe **G** under condition described above. Same samples were resolved on SDS-PAGE and transferred to PVDF membrane for analysis with specific antibodies.

6.3.9 In-situ inhibitor screening

In situ inhibitor screening and identification was carried out by incubating parasite lysates with each inhibitor, followed by incubation with probe **G** and UV irradiation. In detail, parasite lysates (20 μg) was incubated with each inhibitor at a

final concentration of 10 μM individually for 0.5 h in 20 μl Reaction buffer used for proteome labeling experiment in section 6. After incubation with inhibitors, **G** was added to a final concentration of 5 μM and incubated for another 0.5 h followed by UV irradiation on ice for 25 min using a B100A lamp (UVP) at a distance of 5 cm. Then samples were boiled for 10 min with 4 μl of 6 x SDS loading buffer, resolved on a 12% SDS-PAGE followed by in-gel fluorescence scanning. The relative fluorescence intensity of the 37-Kda band labeled by **G** were quantified using ImageQuant™ software. The potency of specific inhibitor scaffolds was measured as a ratio of the percent residual labeled proteases after inhibitor treatment relative to an untreated control. For analysis, the inhibition data were displayed in a colorimetric format and clustered on the basis of similarities in inhibitor profiles using treeview software.

6.3.10 Inhibition assay of live parasite-infected RBC cultures

Parasites were grown to 5% parasitemia after synchronization. Cultures were treated with various concentrations of individual inhibitors at 40 h post invasion, and were harvested after 12 h of treatment. The parasites were then viewed after giemsa staining, and the percentage of ring and schizont stages present was calculated after counting 1500 cells. Synchronous 0.5% DMSO-treated cultures were used as controls.

6.3.11 Cytotoxicity Analysis of Selected Compounds

The cytotoxicity of the selected compounds were analyzed using a simple cytotoxicity assay in which Hela cells were treated with different concentrations of

each compound individually for 24 h followed by measurement of the mitochondrial dehydrogenase activity using XTT reagent. The XTT assay was performed as described below: 25 μl of XTT/PMS solution was added per 100 μl of culture giving a final concentration of 0.2 mg/ml XTT and 25 μM PMS. After incubation at 37 °C in a humidified atmosphere with 5% CO_2 for 4 h the OD was determined on a Tecan scanner using a sample wavelength of 450 nm and a reference wavelength of 620 nm which was expressed in the data as corrected absorbance 450 nm. Cytotoxicity was measured by the inhibition of growth rate by toxic compounds expressed as growth rate of compound treated well over the growth rate of non-treated well. Cytotoxicity of compounds in various concentrations from 2 μM to 100 μM was evaluated, plotted.

6.3.12 Molecule Modeling Prediction of the G16 binding mode in FV plasmepsins

Molecular modelling and graphic manipulations were performed using Sybyl7.2 docking system. Published crystal structures of plasmepsins was retrieved from PDB database and was used as the structure model after deleting water molecule for docking experiment, respectively 1W6I for PM-II, 3FNT for HAP and 1LS5 for PM-IV. The catalytic aspartic acid 214 in each structure was chosen as the centre of the docking site.

6.4 Chapter 4

6.4.1 general

Orlistat (98%), Tris(2-carboxyethyl) phosphine (TCEP), and the click chemistry ligand, tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine ("ligand"), were purchased from Sigma-Aldrich. Antibody against FAS (Cat No. 610963) was from BD Transduction Labs (San Diego, CA). Antibodies against eIF2 (#9722), phospho-eIF2 (#9721), and cleaved caspase-8 (#9746) were from Cell Signaling Technologies (Beverly, MA). Antibodies against HSP90b (sc-1057), Annexin A2 (sc-1492) and RPL14 (KQ-16; sc-100826) were from Santa Cruz Biotechnology, Inc.. Antibodies against RPL7a (ab70753) and RPS9 (ab74711) were from Abcam.

6.4.2 Chemicals synthesis

The synthesis of all chemicals used in Chapter 4 was carried out by my collaborator Dr.Yang Pengyu and the synthesis procedure was reported elsewhere (Yang, P-Y., 2010).

6.4.3 Cell lines and culture conditions

Cell lines were obtained from the National Cancer Institute Developmental Therapeutics Program (NCI60 cell line panel). HepG2 and HEK293 were grown in DMEM (Invitrogen, Carlsbad, CA) containing 10% heat-inactivated fetal bovine serum (FBS; Gibco Invitrogen), 100 U/mL penicillin and 100 μg/mL streptomycin (Thermo Scientific, Rockford, IL) and maintained in a humidified 37 ^{o}C incubator with 5% CO_2. MCF-7 and PC-3 were maintained in RPMI 1640 medium supplemented with 10% FBS and 100 U/mL penicillin and 100 μg/mL streptomycin. To generate protein lysates, cells were washed twice with cold phosphate-buffered saline (PBS), and harvested with a cell scraper, and collected by centrifugation. Cell pellets were resuspended in PBS and lysed by sonication. Protein concentration was

determined by the Bradford assay. Cell lysates were diluted with PBS to achieve final concentration of ~1 mg/mL for labeling reactions.

6.4.4 Cell proliferation assay

Cell viability was determined using the XTT colorimetric cell proliferation kit (Roche) following manufacturer's guidelines. Briefly, cells were grown to 20-30% confluence (since they will reach ~90% confluence within 48 to 72 h in the absence of drugs) in 96-well plates under the conditions described above. The medium was aspirated, and then washed with PBS, and then treated, in duplicate, with 0.1 mL of the medium containing different concentrations of THL analogs (1-50 μM) or Orlistat (1-50 μM; as a positive control). Probes were applied from DMSO stocks whereby DMSO never exceeded 1% in the final solution. The same volume of DMSO was used as a negative control. Fresh medium, along with THL analogs or Orlistat, were added every 24 h. After a total treatment time of 72 h, proliferation was assayed using the XTT colorimetric cell proliferation kit (Roche) following manufacturer's guidelines (read at 450 nm). Data represent the average ± s.d. for two trials.

6.4.5 Western Blotting for eIF2 and caspase-8

To monitor the effects of THL analogs on inducing phosphorylation of eIF2 , PC-3 cells were treated with indicated concentrations of Orlistat and THL analogs for 16 h. Samples from treated cells were then separated on 12% SDS/PAGE gel and further transferred to PVDF membranes. Membranes were blocked with 5% BSA in TBS. After blocking, membranes were incubated with anti-eIF2 (#9722 from

Cell Signaling, 1/5000) or anti-phospho-eIF2 (#9721 from Cell Signaling, 1/2000). After incubation, membranes were washed with TBST for three times and then incubated with an appropriate secondary antibody [anti-mouse conjugated HRP (1/5000) or anti-rabbit conjugated HRP (1/5000)]. After secondary incubation, blots were washed again with TBST before the development with SuperSignal West Pico kit (Pierce). To monitor the effects of THL analogues on inducing activation of caspase-8 pathway, MCF-7 cells were treated with indicated concentrations of Orlistat and THL analogs for 36 h. Samples from treated cells were then separated on a 12% SDS/PAGE gel and further transferred to PVDF membranes. Membranes were blocked with 5% BSA in TBS. After blocking, membranes were incubated with anti-caspase 8 (#9746 from Cell Signaling, 1/2000). After incubation, membranes were washed with TBST for three times, and then incubated with anti-mouse conjugated HRP (1/5000). After secondary incubation, blots were washed again with TBST before the development with SuperSignal West Pico kit (Pierce).

6.4.6 Measurement of Protein Synthesis.

Live HepG2 cells were treated with the indicated concentrations of Orlistat/THL-R or CHX (Cycloheximide, an inhibitor of protein biosynthesis) for 12 h, washed twice with PBS, and then pulsed with AHA (L-Azidohomoalanine; 20 μM) for 4 h. Cells were collected, washed, and cell lysates were prepared and subjected to click chemistry with rhodamine-azide 33, SDS-PAGE analysis, and in-gel fluorescence scanning.

6.4.7 *In vitro* and *in situ* proteome labeling and analysis

For *in vitro* proteome labeling, probes were added to cell lysates (50 μg) in 50 μL PBS at a final concentration of 1-20 μM in the presence or absence of excess Orlistat or Maleimide competitor (a final concentration of 100 μM). Unless indicated otherwise, samples were incubated for 2 h with varying concentrations of probe at room temperature. After incubation, 10 μL freshly premixed click chemistry reaction cocktail in PBS [rhodamine-azide (100 μM, 10 mM stock solution in DMSO), tris(2-carboxyethyl)phosphine hydrochloride (TCEP) (1 mM, 50 mM freshly prepared stock solution in deionized water), tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl] amine (TBTA) (100 μM, 10 mM stock solution in DMSO) and $CuSO_4$ (1 mM, 50 mM freshly prepared stock solution in deionized water)] was added and vortexed, then incubated for 2 h at room temperature with gentle mixing. The reactions were terminated by the addition of pre-chilled acetone (0.5 mL), placed at -20 °C for 30 min and centrifuged at 13000 rpm for 10 min at 4 °C to precipitate proteins. The supernatant was discarded and the pellet washed two times with 200 μL of pre-chilled methanol. The protein pellets were allowed to air-dry for 10 min, resuspended in 25 μL 1×standard reducing SDS-loading buffer and heated for 10 min at 95 °C; ~ 20 μg of protein was loaded per gel lane for separation by SDS-PAGE (12% or 8-16% gradient precast gel), then visualized by in-gel fluorescent scanning using a Typhoon 9410 Variable Mode Imager scanner.

For *in situ* labeling, Cells were grown to 80-90% confluence in 24-well plate under the conditions described above. The medium was removed, and then cells were washed twice with cold PBS, and treated with 0.5 mL of DMEM-containing probe (1-20 μM), with or without Orlistat or Maleimide (100 μM). Probes were applied

from DMSO stocks whereby DMSO never exceeded 1% in the final solution. The same volume of DMSO was used as a negative control. After 2 h of incubation at 37 ^{o}C/5% CO_2, the growth medium was aspirated, and cells were washed twice with PBS to remove the excessive probe, trypsined, and pelleted by centrifugation. The cell pellet was resuspended in PBS (50 μL), homogenized by sonication, and diluted to ~1 mg/mL with PBS. Probe targets were detected by click chemistry with rhodamine-azide, SDS/PAGE analysis, and in-gel fluorescence scanning.

6.4.8 Hydroxylamine Treatment of Gels.

After the proteins were separated by SDS/PAGESDS-PAGE gel, the gel was soaked in 40% MeOH, 10% acetic acid, shaking overnight at room temperature, washed with deionized water (2 × 5 min) and scanned for the prehydroxylamine treatment fluorescence. The gel was then soaked in PBS, shaking 1 h at room temperature, followed by boiling in neutralized hydroxylamine (Alfa Aesar) (2.5% final concentration) for 5 min, washing with deionized water (2 × 5 min), and soaking in 40% MeOH, 10% acetic acid, shaking overnight at room temperature. The gel was washed with deionized water (2 × 5 min) and scanned for the post-hydroxylamine treatment fluorescence.

6.4.9 Pull-Down and Mass spectrometry Identification

To identify the *in situ* targets of THL-R in HepG2 cell line, pull-down followed by immunoblot or MS/MS identification experiments were carried out as described below. *In situ* labeling sample of HepG2 by THL-R were prepared as described previously. After *in situ* labeling, cells were detached from T75 culturing flask and pelleted by centrifuge for 1000 rpm for 15 min. 5 mg of the lysates were

154

reacted by click chemistry with biotin-azide under the conditions described above, acetone precipitated, and resolubilized in 0.1% SDS in PBS with brief sonication. This resuspended sample was then incubated with avidin-agarose beads (100 μL/1mg protein) at room temperature for 30 min. After centrifugation, supernatant were removed and the beads were washed with 1% SDS in PBS for 4 times. After washing, the beads were boiled in elution buffer (200 mM Tris pH 6.8, 400 mM DTT, 8% SDS). For immunoblot analysis, this pull down sample was then separated on 8-16% gradient precast gel (Biorad), transferred to PVDF membrane and probed with purified mouse anti-FAS (Cat No. 610963, BD Transduction Laboratories). 265 kDa FAS band was observed in the pull-down elute fraction while not in the negative control pull-down fraction. For MSMS identification, proteins from pull-down fraction were separated on 12% SDS/PAGE gel, followed by coomassie staining. Trypsin digestion was performed with In-Gel Trypsin Digestion Kit from Pierce for respective visible protein bands. After digestion, digested peptides were then extracted from the gel with 50% acetonitrile and 1% formic acid. Tryptic peptide extracts were evaporated by speedvac and reconstituted with 10 μL 0.1% TFA, a volume of 2 μL of the peptide extracts were manually spotted onto a Prespotted AnchorChip MALDI target plate for MALDI-TOF Mass step with 10 μL of 10 mM ammonium phosphate in 0.1% TFA, and allowed to dry at ambient temperature. MALDI TOF mass spectra were recorded using Ultraflex III TOF/TOF mass spectrometer (Bruker Daltonics) with the compass 1.2 software package including flexControl 3.0 and flexAnalysis 3.0, calibrated with PAC peptide calibration standards. MS/MS analysis for the major peaks in PMF spectra were carried out by

autoLIFT on the MALDI-TOF/TOF instrument. MS and MS/MS Peak lists with intensity value were submitted to Matrix Science Mascot server (http://www.matrixscience.com/) through BioTools 3.0 (Bruker Daltonics) using IPI database, variable modifications of carbamidometyl on cysteine (C) and oxidation on methione (M), allowing maxium of trypsin missed cleavage, peptide mass tolerance at 200 ppm; MS/MS mass tolerance of 0.7 Da.

6.4.10 Target Validation by Western Blotting.

Pull-down sample from labeled lysates was then separated on 12% SDS-PAGE gel together with pull-down sample from DMSO-treated, un-labeled lysates (negative controls). After SDS-PAGE gel separation, proteins were then transferred to a PVDF membrane and subsequently blocked with 2.5% (w/v) BSA/PBST. Membranes were incubated for 1 h at room temperature with the respective antibodies (i.e. anti-RPL7a, anti-Annexin A2, anti-HSP90b, anti-RPS9, anti-GAPDH and anti-b-Tubulin). After 3 X washes with PBST, blots were further incubated with appropriate secondary antibody for 1 h at room temperature. After incubation, the blot was washed again with PBST for 3 times and the SuperSignal West Pico kit (Pierce) was used to develop the blot.

6.4.11 Target Validation of 4 targets by Rrecombinant Protein Eexpression in HEK-293T Ccells.

Mammalian expression vectors overexpressing each of the four targets were purchased from Origene. Vectors were transfected with Lipofectamine reagent at 80% confluence. After 24 hr of transfection, either non-transfected and transfected cell were incubated with THL-R. After THL-R incubation, small fractions of the samples

156

were analyzed by western blot with 1/2000 anti-c-Myc antibody from SantaCruz. The rest of the cells were lysed and immunopurified with c-Myc agarose beads (from Santa Cruz). Pull-down fractions were clicked with rhodamine-azide as previously described. After click chemistry, samples were separated on SDS-/PGAGE gel and fluorescence scanned by in-gel fluorescence with Typhoon Scanner.

6.4.12 Identification of the labeling Ssite of GAPDH.

The mMammalian expression vector overexpressing GAPDH wasere used as template for generating the GAPDH active- site mutant. Cys151 were mutated to Ala chosen as the mutation site. Primers used for site-directed mutagenesis design was as shown bellow:

5' -ATCAGCAATGCCTCCGCCACCACCAACTGC-3'

5'-GCAGTTGGTGGTGGCGGAGGCATTGCTGAT-3'

Vectors overexpressing both wild-type and mutant GAPDH were transfected to HEK293T cell line. Recombinant protein was purified, labeled and clicked as previously described. Further, samples were separated on SDS gel and fluorescence scanned with Typhoon Scanner. After fluorescence scan, gel was fixed and silver stained to visualize the protein bands.

6.4.13. Cellular Imagingolocalizing in situ targets of THL-R analogs with FAS.

HepG2 cells were seeded onto 24-well plates containing sterile glass coverslips and grown until 70-80% confluence under the conditions described above. After 24 h, the growth medium was removed, and cells were washed twice with PBS. Further, cells were treated with 0.5 mL of DMEM with 20 μM THL-R or DMSO.

After fixation in 3.7% paraformaldehyde in PBS for 15 min at room temperature, cells were washed twice with cold PBS again. Cells were permeabilized with 0.1% Triton X-100 in PBS 10 min at room temperature, and blocked with 2% BSA in PBS for 30 min at room temperature, and then washed twice with PBS. Cells were then treated with a freshly premixed click chemistry reaction solution in a 250 μL volume at final concentrations of the following reagents: 1 mM CuSO4, 1 mM TCEP, 100 μM TBTA and 10 μM rhodamine-azide in PBS for 2 h at room temperature with vigorous shaking. Cells were washed with PBS three times, and then washed with 20 mM HEPES, pH 7.5, 500 mM NaCl, 2% triton for overnight at room temperature, and then washed again by PBS for three times. If neccessary, indirect immunofluorescence and Hoechst staining were carried out thereafter. For colocalizing in situ targets of THL analogs with FAS, cells were further incubated with anti-FASN primary antibodies (1:200) for 1 h at room temperature (or overnight at 4 ^{o}C), and washed twice with PBS. The cells were incubated with FITC-conjugated anti-mouse IgG (1:500) for 1 h, washed again. Cells were stained with 1 μg/mL Hoechst for 10 min at room temperature, washed again before mounting. Colocalizing in situ targets of THL analogs with ER. For colocalizing in situ targets of THL-R with ER, cells were further incubated with ER-TrackerTM Green (glibenclamide BODIPY® FL) for 1 h at room temperature (or overnight at 4 ^{o}C), and washed twice with PBS. Cells were stained with 1 μg/mL Hoechst for 10 min at room temperature, washed again before mounting.

6.5 Chapter 5

6.5.1 General

All chemicals were purchased from commercial vendors and used without further purification, unless indicated otherwise. Click-iT® AHA (L-azidohomoalanine) C10102 azide, Click-iT® HPG (L-homopropargylglycine) alkyne (C10186); Click-iT® farnesyl alcohol, azide (C10248) Click-iT® geranylgeranyl alcohol, azide (C10249); Click-iT® fucose alkyne (tetraacetylfucose alkyne) (C10264); Click-iT® myristic acid, azide C10268 Click-iT® myristic acid, azide (C10268); Click-iT® Gal-Az (tetraacetylated N-azidoacetylgalactosamine) azide (C33365); Click-iT® Man-Az (tetraacetylated N-azidoacetyl-D-mannosamine) azide (C33366); Click-iT® Glc-Az (tetraacetylated N-azidoacetylglucosamine) azide (C33367) were commercially available from Invitrogen. In-gel fluorescence scanning was carried out with Typhoon 9410 fluorescence gel scanner (GE Amersham). All other reagents were purchased from Invitrogen. PA-Ak (17-Octadecynoic acid) were purchased from Cayman Inc.

6.5.2 chemical synthesis

The synthesis of azide/alkyne-containing fluorescent and affinity tags used in Chapter 5 was carried out by my collaborator Dr.Yang Pengyu and the synthesis procedure was reported elsewhere (Yang, P-Y., 2010).

6.5.3 Metabolic Labeling with AHA and HPG

Jurkat cells were grown in Growth Medium (RPMI medium containing 10% FBS and 2 mM glutamine). After cells reached confluence and before labeling, cells were centrifuged (1500 rpm x 10 min) to remove the medium. Cell pellets were

resuspended in an appropriate amount of Met-free Growth Medium (Met-free RPMI containing 10% dialyzed FBS and 2 mM glutamine). The cell density was adjusted to 1×10^6 cells/ml. Subsequently, cells were starved for 30 min before addition of 50 µM of AHA. Upon further incubation, cells were harvested at the desired time points. Briefly, the medium was removed by centrifugation (1500 rpm x 10 min), resuspended and washed with PBS for 3 times. To tag AHA-labeled, newly synthesized proteins with a fluorophore or biotin, cell pellets were re-suspended in 1% SDS in PBS. When the pellets were completely dissolved, the volume of the solution was diluted 10-folds with PBS (to give a final 0.1% SDS solution). The lysate was reacted with Alexa-647 Alkyne in 1 ml of PBS for 10 hour at 10 °C with gentle mixing. Standard click chemistry conditions involve the addition of 1-2 mg/ml of the proteome, 100 µM of Alexa-647 Alkyne (dissolved in DMSO), 1 mM of Tris(2-carboxyethyl)phosphine (TCEP, Sigma-Aldrich; dissolved in water), 100 µM of Tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA) (Sigma-Aldrich; dissolved in DMSO), and 1 mM $CuSO_4$ in PBS. The reactions were terminated by the addition of a 4-times volume of pre-chilled acetone, placed at -20 °C for 30 min and centrifuged (13000 rpm x 10 min) at 4 °C to precipitate proteins. The supernatant was discarded and the pellet washed with pre-chilled methanol (2 x). The protein pellets were allowed to air-dry and re-suspended in $1 \times$ standard reducing SDS-loading buffer and heated for 10 min at 95 °C. An appropriate amount of protein was loaded per gel lane for separation by 12% SDS-PAGE, then visualized by in-gel fluorescence scanning using a Typhoon 9410 fluorescence scanner as described previously (Liu, K., 2009). To confirm the labeling of newly synthesized protein by

AHA, we added protein synthesis inhibitor Cycloheximide (CHX) to Jurkat T cell cultures alone with AHA. Fluorescence profile was quantified by ImageQuant[TM] software as described previously (Liu, K., 2009). Above procedures were similarly followed to determine the time-dependent metabolic labeling of HPG. Briefly, Jurkat T cells were cultured to confluence, labeled with 50 μM of HPG, tagged with Rhodamine-Azide, washed, air-dried and re-suspended in $1 \times$ standard reducing SDS-loading buffer and heated for 10 min at 95 °C; appropriate amount of protein was loaded per gel lane for separation by 12% SDS-PAGE, then visualized by in-gel fluorescence scanning and quantified

6.5.4 Metabolic Labeling with PTM probe

To perform metabolic incorporation for each PTM probe, Jurkat T cells were first grown to confluence in Growth Medium. Upon centrifugation, cell pellets were collected, then resuspended in Maintainance Medium to minimize the effect of any trace amount of lipids or carbohydrates that might be present in the serum. The tested PTM probe was added (to a final concentration of 50 μM). The cells were further grown and collected at desired time points. Preparation of protein samples and fluorescence tagging by Click Chemistry were carried out as previously described.

6.5.5 Affinity pulldown of Post-Translationally Modified Proteins in Jurkat T cells

To identify the targets of all 8 types of PTMs in Jurkat T cells, Jurkat T cells were incubated with each of the 8 PTM probes (50 μM x 3 days). After incubation, biotin tagging with Click Chemistry (i.e. 100 μM of Biotin-Azide or Biotin-Alkyne was used in place of Rhodamine-Azide or Alexa 647-Alkyne), washing and air-

drying of PTM labeled protein samples were carried out as described in previous. Next, the protein samples were dissolved in 1% SDS in PBS by vortexing, then diluted 10-folds with PBS, and centrifuged (13000 rpm x 30 min) to remove any insoluble proteins. Subsequently, avidin-agarose beads (50 μL/mg protein) were added and the mixture was further incubated at room temperature for 30 min. After centrifugation, the supernatant was removed and the beads were washed with 0.1% SDS in PBS (3 times), 0.4% SDS in PBS (3 times), 0.1% SDS (3 times), then boiled with 50 μl of 1 x SDS loading dye at 95°C for 10 min. The samples were then separated on a 12% SDS-PAGE gel.

6.5.6 LC-MS/MS identification of pull-down proteins.

Each lane was cut into 1 mm slices. Proteins in the gel slices were reduced with DTT, alkylated by incubation with IAA and digested with modified porcine trypsin (Promega Corp., Madison, WI), as previously described. The resultant peptide mixture was extracted from the gel slices using a 60% ACN solution (containing 5% formic acid). The extract peptide was dried *in vacuo* and reconstituted to 40 μL with 0.1% formic acid before LCMS analysis. The LTQ-FT ultra (Thermo Electron, Bremen, Germany) was coupled with an online Shimadzu UFLC systems utilizing nanospray ionization. Peptides were first enriched with a Zorbax 300SB C_{18} column (5 mm × 0.3 mm, Agilent Technologies, Santa Clara, CA) followed by elution into an integrated nanobore column (75 μm × 100 mm, New Objective, Woburn, MA) packed with C_{18} material (5 μm particle size, 300 Å pore size, Michrom BioResources Inc.). Mobile phase A (0.1% formic acid in H_2O) and mobile phase B (0.1% formic acid in acetonitrile) were used to establish the 90-min

162

gradient, consisting of 0-5% B (3 min), then 5-30% B (52 min), 30-60% B (12 min), 80% B (8 min), then 5% B (15 min) for re-equilibration. The MS was operated in the data-dependent mode. Samples were injected into the MS with an electrospray potential of 1.8 kV without sheath and auxiliary gas flow, with an ion transfer tube temperature of 180 °C and a collision gas pressure of 0.85 mTorr. A full survey MS scan (350-2000 m/z range) was acquired in the 7-T FT-ICR cell at a resolution of 100 000 and a maximum ion accumulation time of 1000 ms. Precursor ion charge state screening was activated. The linear ion trap was used to collect peptides where 10 most intense ions were selected for collision-induced dissociation (CID) in MS2, which were performed concurrently with a maximum ion accumulation time of 200 ms. Dynamic exclusion was activated for this process, with a repeat count of 1 and exclusion duration of 30 s. For CID, the activation Q was set at 0.25, isolation width (m/z) 2.0, activation time 30 ms, and normalized collision energy of 35%. The extract_msn (version 4.0) program found in Bioworks Browser 3.3 (Thermo Electron, Bremen, Germany) was used to extract tandem MS spectra in the dta format from the raw data of LTQ-FT ultra. These dta files were then converted into MASCOT generic file format using an in-house program. Intensity values and fragment ion m/z ratios were not manipulated. This data was used to obtain protein identities by searching against the IPI human protein database (version 3.34; 67758 sequences) by means of an in-house MASCOT server (version 2.2.03) (Matrix Science, Boston, MA). The search was limited to maximum 2 missed trypsin cleavages; #13C of 2; mass tolerances of 10 ppm for peptide precursors; and 0.8 Da mass tolerance for fragment ions. Fixed modification was carbamidomethyl at Cys residue, whereas

163

variable modifications were oxidation at Met residue, and phosphorylation at Ser, Thr or Tyr residues. Only proteins with a MOWSE score higher than 39, corresponding to $p < 0.05$ were considered significant.

6.5.7 Dynamic Profiling of Palmitoylation on Newly Synthesized Proteins

9 identical newly synthesized proteomes ("pulsed" with 0.5 mM AHA x 10 min) from Jurkat T cells, were prepared, as previously described. They were each "chased" with PA-Ak (0.5 mM) for a duration of 20 min, at each of the 9 different time windows (assuming the start of AHA pulsing is time 0, PA-Ak was chased, each for 20-min duration, ending at 20 min, 40 min, 60 min, 2 hrs, 4 hrs, 8 hrs, 24 hrs, 48 hrs and 72 hrs past, respectively), and resulting palmitoylation activities of these subproteomes were tracked by affinity tagging, isolation of newly synthesized proteins, fluorescence tagging, as previously described.

6.5.8 DNA Fragmentation Assay

To set up a butyric acid (BA)-induced apoptosis model system, we treated Jurkat T cells with 5 mM of butyric acid in Growth Medium and the induction of Jurkat cell apoptosis was determined by DNA fragment gel electrophoresis. In detail, after reaching confluence, Jurkat cells were resuspended in Growth Medium. Cells were cultured in 24-well tissue culture plates in the presence or absence of butyric acid. After 20 hours of BA treatment, cells were harvested and centrifuged (1500 rpm x 15 min) and washed twice with ice-cold PBS. Cells were resuspended in 400 l of hypotonic lysis buffer (0.2% Triton X-100, 10 mM Tris, 1 mM EDTA; pH 8.0) and centrifuged (15 min x 13,000 rpm). The supernatant was treated with an equal

volume of absolute isopropyl alcohol and 0.5 M NaCl to precipitate the DNA and stored at -20°C overnight. After centrifugation (13,000 rpm x 15 min), the pellet was washed with 70% ethanol and allowed to dry at room temperature. The DNA was resuspended in 20 l of TE solution (10 mM Tris-HCl, 1 mM EDTA; pH 7.4) and 3 ml of loading buffer (50% glycerol, 13% TAE, 10% saturated bromphenol blue, 1% xylene cyanol), incubated at 37°C for 20 min, and then electrophoresed on 1.7% agarose gel containing 0.71 mg of ethidium bromide per ml for 1 h. Gels were photographed by using UV transillumination.

Chapter 7.

Concluding Remarks

7.1 Conclusion
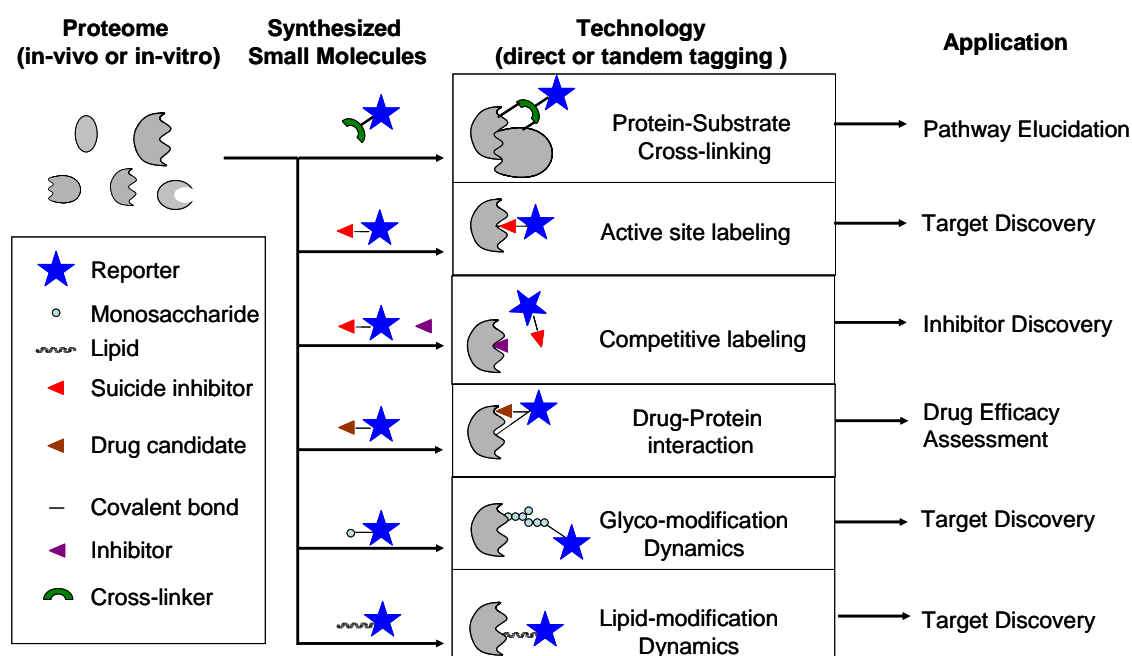
Figure 7.1 Strategies for proteomic labeling with synthesized small molecules in vitro and in vivo. (From top to bottom) Enzyme-substrate interactions can be captured with mechanism based cross-linkers (Chapter 2). Enzymes that are catalytically active can be specifically labeled with activity/affinity based probes bearing chemical reporters (Chapter 3). Inhibitor library can be screened with competitive probe labeling (Chapter 3). Drug efficacy can be assessed by profiling small molecule-protein interaction (Chapter 4). Glyco- or lipid-modification dynamics can be measured by assaying modification on newly synthesized proteins (Chapter 5).

Herein I have desribed a series of chemical proteomics strategies (Figure 7.1). The strategy describe in Chapter 2 potentially facilitate the identification of upstream excution kinases of phosphorylation substrates. For the first time we have demonstrated a general chemical approach for the identification of multiple kinase activities (not only derine/threonine kinases but also tyrosine kinases) directly from the whole proteome. This approach was useful for multiplexed detection and kinase activities present in a proteome, and is amenable for potential screenings of potent and selective inhibitors of kinases in their native environments. The establishment of the highly specific and sensitive NDA-adenosine guided crosslinking reactions with desired kinase-substrate pairs in their native states represents a step forward towards the creation of novel chemical tools in cell signaling and drug discovery. This approach was further improved by Shokat and colleagues to detect endogenous excutioner kinases of known phosphorylation substrate in mamalian proteome (Statsuk, A.V., *et Al.* 2008).

Strategies described in Chapter 3 realized the proteomic profiling of activities and inhibition of FV family aspartic proteases (FV plasmepsins) in *P. falciparum.* For the first time we have reported AfBPs that targeting FV plasmepsins. Subsequent in situ screening of parasites with these probes has led to the identification of a compound, G16, which show good inhibition against all 4 PMs and parasite growth in infected red blood cells (RBCs). Our finding indicates that feasibility of using ABP approaches for identification of inhibitors against less-characterized enzymes (i.e. Histoaspartic Protease, or HAP). Our finding indicates that feasibility of using AfBP approaches for identification of inhibitors against less-characterized enzymes

167

(i.e. Histoaspartic Protease, or HAP).

Chapter 4 describes an activity-based profiling of small molecule-protein interactions approach for unbiased, empirical identification of off-targets of development-stage or even marketed drugs. We were able to demonstrate, for the first time, this approach is suitable for the identification of previously unknown cellular targets of Orlistat. This method determines inhibitor's selectivity in a proteome-wide scale, unlike selectivity determined by standard ABP/AfBP approach only representing the selectivity within certain class of enzyme targeted by the probe. . Our findings have important implications in the consideration of Orlistat as a potential anticancer drug at its early stages of development for cancer therapy. Our strategy should be broadly useful for off-target identification against quite a number of existing drugs and/or candidates, which are also covalent modifiers of their biological targets.

In Chapter 5, we presented a methodology that enables the proteome-wide profiling of PTMs on proteins synthesized in defined time window, as well as their dynamics, by using a double metabolic incorporation strategy. We validated the feasibility of this approach with a number of proteins covering a total of eight different types of PTMs which occur on different time scales (rapid and enduring). We further applied the strategy to monitor the myristoylation of newly synthesized proteins in apoptotic Jurkat cells, and successfully identified Protein Kinase A (PKA), a key signaling enzyme, whose myristoylation appeared to be up-regulated in response to butyric acid (BA)-induced apoptosis. We anticipate this new chemical proteomic tool may facilitate the discovery of primary PTM changes associated with

different extracellular and intracellular cues.

These tools have been tested successfully and the results indicate that such chemical proteomic tools would provide comprehensive insight into the understanding of the proteome in various levels other than protein expression. These proteomic profiling methods contributes towards our understanding of proteins in their native environment and accelerate the process of drug target identification (by potentially identifying upstream executioner kinases, profiling FV family aspartic proteases, identifying regulated PTMs in cellular events); drug lead discovery (by facilitating inhibitor's potency and selectivity screening in complex proteome) and drug efficacy assessement (by identifying drug's potential off-targets in living cell). The probes developed may be applied for the annotation and discovery of novel enzymes. The inhibitors discovered may also be developed as next generation chemical probes for proteomic profiling or as potential drugs. To conclude, the technological advances here offer frontier capabilities towards the rapid evaluation and characterization of proteins in various levels other than protein expression.

Chapter 8.

References

Adam G.C., Cravatt B.F., Sorensen E.J., Profiling the specific reactivity of the proteome with non-directed activity-based probes. *Chem. Biol.* **8,** 81-95 (2001)

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

Aebersold, R. & Goodlett, D.R. Mass spectrometry in proteomics. *Chem. Rev.* **101**, 269-295 (2001).

Aebersold, R.H., Leavitt, J., Saavedra, R.A., Hood, L.E., Kent, S.B. Internal amino acid sequence analysis of proteins separated by one- or two-dimensional gel electrophoresis after in situ protease digestion on nitrocellulose. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6970-6974 (1987)

Agard, N.J., Prescher, J.A. & Bertozzi, C.R. A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J. Am. Chem. Soc.* **126**, 15046-15047 (2004)

Agard, N.J., Baskin, J.M., Prescher, J.A., Lo, A. & Bertozzi, C.R. A comparative study of bioorthogonal reactions with azides. *ACS Chem. Biol.* **1**, 644-648 (2006).

Agard, N. J., Bertozzi, C. R. Chemical Approaches to Perturb, Profile, and Perceive Glycans.*Acc. Chem. Res.* **42,** 788-797 (2009)

Anderson, N.L., and NG Anderson, Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis* **19,** 1853-1861 (1998)

Barglow, K. T., Cravatt, B. F. Discovering disease-associated enzymes by proteome reactivity profiling. *Chem. Biol.* **11**, 1523-31. (2004)

Baselga, J., Targeting tyrosine kinases in cancer: the second wave. Science **312**, 1175-1178 (2006)

Baskin, J. M., Prescher, J. A., Laughlin, S. T., Agard, N. J., Chang, P. V., Miller, I. A., Lo, A., Codelli, J. A., Bertozzi, C. R. Copper-free click chemistry for dynamic in vivo imaging *Proc. Natl. Acad. Sci. U.S.A.* **104**, 16793- 16797 (2007)

Beatty, K. E., Liu, J. C., Xie, F., Dieterich, D. C., Schuman, E. M., Wang, Q., Tirrell, D. A., Fluorescence visualization of newly synthesized proteins in mammalian cells. *Angew. Chem. Int. Ed. Engl.* **45**, 7364-7367 (2006)

Berkers, C.R., Verdoes, M., Lichtman, E., Fiebiger, E., Kessler, B.M., Anderson, K.C., Ploegh, H.L., Ovaa H. ,and Galardy, P.J., Activity probe for in vivo profiling of the specificity of proteasome inhibitor bortezomib, *Nat Methods* **2,** 357-362. (2005)

Bijlmakers, M. J., Marsh, M., The on-off story of protein palmitoylation. *Trends Cell Biol*. **13**, 32-42 (2003)

Blair, J. A., Rauh, D., Kung, C., Yun, C.H., Fan, Q.W., Rode, H., Zhang, C., Eck, M.J., Weiss, W.A., Shokat, K.M., Structure-guided development of affinity probes for tyrosine kinases using chemical genetics. *Nat. Chem. Biol.* **3,** 229-238 (2007)

Bogyo M ,McMaster J S ,Gaczynska M ,Tortorella D ,Goldberg A L ,Ploegh H Covalent modification of the active site threonine of proteasomal beta subunits and the Escherichia coli homolog HslV by a new class of inhibitors. Proc Natl Acad Sci USA **94**, 6629-6634 (1997)

Bolden J.E., Peart M.J., Johnstone R.W., Anticancer activities of histone deacetylase inhibitors. *Nat. Rev. Drug Discov.* **5,** 769-784 (2006)

Borodovsky, A., Kessler, B.M., Casagrande, R., Overkleeft, H.S., Wilkinson K.D., and Ploegh, H.L., A novel active site-directed probe specific for deubiquitylating enzymes reveals proteasome association of USP14, EMBO J. **20**, 5187-5196 (2001)

Boss, C., Corminboeuf, O., Grisostomi, C., Meyer, S., Jones, A. F., Prade, L., Binkert, C., Fischli, W., Weller, T., Bur, D., Achiral, cheap, and potent inhibitors of Plasmepsins I, II, and IV *ChemMedChem* **1**, 1341 - 1345 (2006)

Bo¨ttcher, T., Sieber, S. A. β -Lactones as Privileged Structures for the Active-Site Labeling of Versatile Bacterial Enzyme Classes *Angew. Chem., Int. Ed.* **47**, 4600- 4603 (2008)

Bo¨ttcher, T., Sieber, S. A. Beta-lactones as specific inhibitors of Cl attenuate the production of extracellular virulence factors of Staphylococcus aureus. *J. Am. Chem. Soc.* **130**, 14400-14401 (2008)

Bo¨ttcher, T., Sieber, S. A. Structurally Refined β-Lactones as Potent Inhibitors of Devastating Bacterial Virulence Factors *ChemBioChem* **10**, 663-666 (2009)

Bougeret, C., Rothhut, B., Jullien, P., Fischer, S., Benarous, R., Recombinant Csk expressed in Escherichia coli is autophosphorylated on tyrosine residue(s). *Oncogene* **8,** 1241-1247 (1993)

Breitenlechner, C., Engh, R. A., Huber, R., Kinzel, V., Bossemeyer, D., Gassel, M., The Typically Disordered N-Terminus of PKA Can Fold as a Helix and Project the Myristoylation Site into Solution. *Biochemistry* **43**, 7743-7749 (2004)

Brik, A., J. Muldoon, Y.-C. Lin, J.H. Elder, D.S. Goodsell, A.J. Olson, V.V. Fokin, K.B. Sharpless, C.-H. Wong, Rapid diversity-oriented synthesis in microtiter plates for in situ screening of HIV protease inhibitors *ChemBioChem* **4,** 1246-1248 (2003)

Buglino, J.A., and Resh, M.D., Hhat is a palmitoylacyltransferase with specificity for N-palmitoylation of Sonic Hedgehog, *J Biol Chem* **283,** 22076-22088 (2008),

Campillos, M., Kuhn, M., Gavin, A. -C., Jensen, L. J., Bork, P. Drug target identification using side-effect similarity. *Science* **321,** 263-266 (2008)

Casnellie,J.E., Assay of protein kinases using peptides with basic residues for phosphocellulose binding. *Methods Enzymol.* **200,** 115-120 (1991,)

Chan E.W., Chattopadhaya S, Panicker R.C., Huang X., Yao S.Q. Developing photoactive affinity probes for proteomic profiling: hydroxamate-based probes for metalloproteases. *J. Am. Chem. Soc.* **126,** 14435-46 (2004)

Charron, G., Zhang, M. M., Yount, J. S., Wilson, J., Raghavan, A. S., Shamir, E., Hang, H. C. Robust fluorescent detection of protein fatty-acylation with chemical reporters. *J. Am. Chem. Soc.* **131,** 4967-4975 (2009)

Cheng, H.C., Nishio, H., Hatase, O., Ralph, S., Wang, J. H., A synthetic peptide derived from p34cdc2 is a specific and efficient substrate of src-family tyrosine kinases. *J. Biol. Chem.* **267**, 9248-9256 (1992)

Chino, M., Wakao, M., Ellman, J.A., Efficient method to prepare hydroxyethylamine-based aspartyl protease inhibitors with diverse P1 side chains *Tetrahedron,* **58**, 6305-6310 (2002)

Codelli, J. A., Baskin, J. M., Agard, N. J., Bertozzi, C. R. Second-generation difluorinated cyclooctynes for copper-free click chemistry *J. Am. Chem. Soc.* **130**, 11486- 11493 (2008)

Cole, R.B. Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation and Applications. *Annu. Rev. Biomed. Eng.* **11**, 49-79 (2009)

Corthals, G.L., Wasinger, V.C., Hochstrasser, D.F., and Sanchez J.C., The dynamic range of protein expression: a challenge for proteomic research. *Electrophoresis* **21,** 1104-1115 (2000)

Crunkhorn, S. Metabolic disease: New oortunity for serotonin receptor agonists *Nat. Rev. Drug Discov.* **7**, 729-729 (2008)

Daub, H., Olsen, J. V., Bairlein, M., Gnad, F., Oermann, F. S., Korner, R., Greff, Z., Keri, G., Stemmann, O., Mann, M. Kinase-selective enrichment enables quantitative phosphoproteomics of the kinome across the cell cycle *Mol. Cell* **31,** 438-448 (2008)

Davies, B. S., Yang, S. H., Farber, E., Lee, R., Buck, S. B., Andres, D. A., Spielman, H. P., Agnew, B. J., Tamanoi, F., Fong, L. G., Young, S. G. Increasing the length of progerin's isoprenyl anchor does not worsen bone disease or survival in mice with Hutchinson-Gilford progeria syndrome. *J. Lipid Res.* **50**, 126-134 (2008)

Deal, R. B., Henikoff, J. G., Henikoff, S. Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science* **328**, 1161-1164 (2010)

Dejkriengkraikhul, P., Wilairat, P., Requirement of malarial protease in the invasion of human red cells by merozoites ofPlasmodium falciparum *Parasitenkd.* **69**, 313-317 (1983)

Dieterich, D.C., Link, A.J., Graumann, J., Tirrell D.A., and Schuman, E.M. Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT), *Proc Natl Acad Sci USA* **103,** 9482-9487 (2006)

Dieterich, D.C., Hodas, J.J.L., Gouzer, G., Shadrin, I.Y., Ngo, J.T., Triller, A., Tirrell, D.A., and Schuman, E.M., In situ visualization and dynamics of newly synthesized proteins, *Nature Neuroscience* **13,** 897-905 (2010)

Drew, M. E., Banerjee, R., Uffman, E.W., Gilbertson, S., Rosenthal, P. J., Goldberg, D. E., Plasmodium food vacuole plasmepsins are activated by falcipains.*J. Biol. Chem.* **283,** 12870-12876 (2008)

Dube, D.H. Prescher ,J.A. Quang C.N. and Bertozzi, C.R. Probing mucin-type O-linked glycosylation in living animals, *Proc Natl Acad Sci USA* **103,** 4819-4824 (2006)

Eisenberg, D., Marcotte, E.M., Xenarios, I. & Yeates, T.O. Protein function in the post-genomic era. *Nature* **405**, 823-826 (2000).

Ersmark, K., Samuelsson, B., Hallberg, A., Plasmepsins as potential targets for new antimalarial therapy. Med. Res. Rev. 26, 626 - 666. (2006) *Med. Res. Rev.* **26**, 626-666 (2006)

Evans, M.J.，Cravatt, B.F. Mechanism-Based Profiling of Enzyme Families *Chem. Rev.* **106**, 3279-3301 (2006)

Evans, M. J., Saghatelian, A., Sorensen, E. J., Cravatt, B. F. Target discovery in small-molecule cell-based screens by in situ proteome reactivity profiling. Nat. Biotechnol. **23**, 1303-1307 (2005)

Everley, P.A., Krijgsveld, J., Zetter, B.R., Gygi, S.P. Quantitative Cancer Proteomics: Stable Isotope Labeling with Amino Acids in Cell Culture (SILAC) as a Tool for Prostate Cancer Research. *Molecular & Cellular Proteomics.* **3,** 729-735 (2004)

Hsieh, Y.L.F, Wang, H.Q, Elicone, C., Mark, J., Martin S.A., and Regnier, F., Automated Analytical System for the Examination of Protein Primary Structure *Anal. Chem.* **68** 455-462 (1996)

Hannoush, R. N., Sun, J. L. The chemical toolbox for monitoring protein fatty acylation and prenylation. Nat. Chem. Biol. **6** 498-506 (2010)

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

Fidock, D. A., Eastman, R. T., Ward, S. A., Meshnick, S. R. Trends Parasitol. 2008, 24, 537 - 544; Rosenthal, P. J., Cysteine proteases of malaria parasites *Int. J. Parasitol.* **34**, 1489 - 1499 (2004)

Filiatos, T. D., Derdemezis, C. S., Gazi, I. F., Nakou, E. S., Mikhailidis, D. P., Elisaf, M. S. Orlistat-associated adverse effects and drug interactions: a critical review *Drug Saf.* **31**, 53-65 (2008)

Franklin R.A., McCubrey J.A. Kinases: positive and negative regulators of apoptosis . Leukemia **14**, 2019-2034 (2000)

Freilech, S., Spriggs, R.V., George, R.A., Al-Lazikani, B., Swindells, M., and Thornton, J.M. The complement of enzymatic sets in different species. *J. Mol. Biol* **349**, 745-763 (2005)

Garrels, J.I. McLaughlin, C.S., Warner, J.R., Futcher, B., Latter, G.I., Kobayashi, R., Schwender, B., Volpe, Andersen, D.S., Mesquita-Fuentes R and Payne WE, Proteome studies of Saccharomyces cerevisiae: identification and characterization of abundant proteins. *Electrophoresis* **18**, 1347-1360 (1997)

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

Giacomini, K. M., Krauss, R. M., Roden, D. M., Eichelbaum, M., Hayden, M. R., Nakamura, Y. When good drugs go bad. *Nature* **446**, 975-977 (2007)

Gorg, A., Weiss, W. & Dunn, M.J. Current two-dimensional electrophoresis technology for proteomics. *Proteomics* **4**, 3665-3685 (2004).

Goldberg, D. E., Discovery of new targets for antimalarial chemotherapy *Curr. Top. Microbiol. Immunol.* **295**, 275-291 (2005)

Gonzalez,F.A., Raden, D. L., Davis, R. J., Identification of substrate recognition determinants for human ERK1 and ERK2 protein kinases. *J. Biol. Chem.* **266**, 22159-22163 (1991,)

Greenbaum, D.C., A. Baruch, L. Hayrapetian, Z. Darula, A. Burlingame, K.F. Medzihradszky and M. Bogyo, Chemical aroaches for functionally probing the proteome, *Mol. Cell. Proteomics* **1,** 60-68 (2002)

Greenbaum D.C., Arnold WD, Lu F, Hayrapetian L, Baruch A, et al. Small molecule affinity fingerprinting. A tool for enzyme family subclassification, target identification, and inhibitor design. *Chem. Biol.* **9,** 1085-1094 (2002)

Greenbaum, D.C., A. Baruch, M. Grainger, Z. Bozdech, K.F. Medzihradszky, J. Engel, J. DeRisi, A.A. Holder and M. Bogyo, A role for the protease Falcipain 1 in host cell invasion by the human malaria parasite. *Science* **298,** 2002-2006 (2002)

Guerciolini, R., Mode of action of orlistat. *Int. J. Obes. Relat. Metab. Disord.* **21**, S12-23. (1997)

Guengerich F.P., Wu ZL, Bartleson CJ.Function of human cytochrome P450s: characterization of the orphans. *Biochem Biophys Res Commun.* **338,** 465-469 (2005)

Gurcel, C., & Anne-Sophie Vercoutter-Edouart & Catherine Fonbonne & Marlène Mortuaire & Arnaud Salvador & Jean-Claude Michalski & Jérôme Lemoine Identification of new O-GlcNAc modified proteins using a click-chemistry-based tagging *Anal Bioanal Chem* **390,** 2089-2097 (2008)

Gygi, S.P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., Aebersold, R. Direct analysis of protein complexes using mass spectrometry *Nat. Biotechnol.* **17**, 994-999. (1999)

Gygi, S.P., GL Corthals, Y Zhang, Y Rochon and R Aebersold, Evaluation of two dimensional gel electrophoresis-based proteome analysis technology. *Proc Natl Acad Sci USA* **97**, 9390-9395 (2000)

Gygi S.P. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology* **17**, 994-999 (1999)

Hanash, S. Disease proteomics. Disease proteomics. *Nature* **422,** 226-232 (2003).

Hang, H.C., C. Yu, D.L. Kato and C.R. Bertozzi, A metabolic labeling aroach toward proteomic analysis of mucin-type O-linked glycosylation, *Proc Natl Acad Sci USA* **100,** 14846-14851 (2003)

Hang, H.C., E.J. Geutjes, G. Grotenbreg, A.M. Pollington, M.J. Bijlmakers and H.L. Ploegh, Chemical probes for the rapid detection of fatty-acylated proteins in mammalian cells, J Am Chem Soc **129,** 2744-2745 (2007)

Heal, W.P., Tate, E.W., Getting a chemical handle on protein post-translational modification. *Org. Biomol. Chem.* **8,** 731-738 (2010)

Helm, K.V.D., Seelmeier, S., Kisselev, A., Nitschko, H., Identification, purification, and cell culture assays of retroviral proteases *Methods Enzymol* **241**, 89-104 (1994)

Hewick, R. M., Hunkapiller, M. W., Hood, L. E., Dreyer, W. J. A gas-liquid solid phase peptide and protein sequenator. *J. Biol. Chem.* **256**, 7990-7997 (1981)

Hieter, P. & Boguski, M. Functional genomics: it's all how you read it. *Science* **278**, 601-602 (1997).

Hoving, S., Voshol, H., and van Oostrum, J., Towards high performance two-dimensional gel electrophoresis using ultrazoom gels. *Electrophoresis* **21,** 2617-2621 (2000)

Hof, F., Schutz, A. Fah, C. Meyer, S. Bur, D. Liu, J. Goldberg, D. E. Diederich, F. Angew. Chem. 2006, 118, 2193 - 2196; *Angew. Chem. Int. Ed.* 45, 2138-2141 (2006)

Ho, Y.,Gruhler, A., Heilbut, A., Bader, G.D., Moore, L., Adams, S.L., Millar, A., Taylor, P., Bennett, K., Boutilier K., Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature* **415**, 180-183 (2002)

Hsu, T. L., Hanson, S. R., Kishikawa, K., Wang, S. K., Sawa, M., Wong, C. H. Alkynyl sugar analogs for the labeling and visualization of glycoconjugates in cells *Proc. Natl. Acad. Sci. U.S.A.* **104**, 2614- 2619 (2007)

Hagenstein, M.C.，Mussgnug, J.H., Lotte, K. Plessow, R., Brockhinke, A., Kruse, O., Sewald, N., Affinity‐Based Tagging of Protein Families with Reversible Inhibitors: A Concept for Functional Proteomics *Angew. Chem. Int. Ed.* **42**, 5635-5638 (2003)

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y., A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4569-4574 (2001)

James, P., Quadroni, M., Carafoli, E., Gonnet, G. Protein identification by mass profile fingerprinting. *Biochem. Biophys. Res. Commun.* **195,** 58-64 (1993)

Jessani, N., Liu, Y.S., Humphrey, M., Cravatt, B.F., Enzyme activity profiles of the secreted and membrane proteome that depict cancer cell invasiveness *Proc Natl Acad Sci USA* **99,** 10335-10340 (2002)

J. A. Johnson, Y. Y. Lu, J. A. Van Deventer, D. A. Tirrell, Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications *Curr. Opin. Chem. Biol.* **14,** 774-780 (2010)

Joubert，R. Two-dimensional gel analysis of the proteome of lager brewing yeasts *Yeast* **16,** 511-522 (2000)

Kalesh, K. A., Yang, P. -Y., Srinivasan, R., Yao, S. Q. Click Chemistry as a High-Throughput Amenable Platform in Catalomics *QSAR Comb. Sci.* **26**, 1135-1144 (2007)

Kidd D, Liu Y, Cravatt BF. Profiling serine hydrolase activities in complex proteomes. *Biochemistry* **40,** 4005-4015 (2001)

Kiick, K.L., Saxon, E., Tirrell D.A., and Bertozzi, C.R., Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation, *Proc Natl Acad Sci USA* **99**, 19-24 (2002)

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

Karas, M., Hillenkamp, F. Fast Screening of Low Molecular Weight Compounds by Thin-Layer Chromatography and "On-Spot" MALDI-TOF Mass Spectrometry *Anal. Chem.* **60**, 2299 (1995)

Knight,Z.A., K. M. Shokat, Protein Kinase Inhibitors for the Treatment of Inflammation - An Overview *Chem. Biol.* **12**, 621-637 (2005)

Knight, Z. A., B. Schilling, R. H. Row, D. M. Kenski, B. W. Gibson, K. M. Shokat, Phosphospecific proteolysis for mapping sites of protein phosphorylation. *Nat. Biotechnol.* **21**, 1047-1054 (2003)

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

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

Khokhlatchev, A., Xu, S. , English, J. , Wu, P. , Schaefer, E., Cobb, M. H. , Reconstitution of mitogen-activated protein kinase phosphorylation cascades in bacteria *J Biol. Chem.* **272**, 11057-11062 (**1997**)

Klose, J., and U Kobalz, Two-dimensional electrophoresis of proteins: an update protocol and implications for a functional analysis of the genome, *Electrophoresis* **16**, 1034-1059 (1995)

Kostiuk, M. A., Corvi, M. M., Keller, B. O., Plummer, G., Prescher, J. A., Hangauer, M. J., Bertozzi, C. R., Rajaiah, G., Falck, J. R., Berthiaume, L. G. Identification of palmitoylated mitochondrial proteins using a bio-orthogonal azido-palmitate analogue *FASEB J.* **22**, 721-732 (2008)

Kramer, G., Sprenger, R. R., Back, J., Identification and quantitation of newly synthesized proteins in Escherichia coli by enrichment of azidohomoalanine-labeled peptides with diagonal chromatography. *Mol. Cell. Proteomics* **8**, 1599-1611 (2009)

Kridel, S. J., Axelrod, F., Rozenkrantz, N., Smith, J. W. Orlistat is a novel inhibitor of fatty acid synthase with antitumor activity. *Cancer Res.* **64**, 2070-2075 (2004)

Knowel, L. M., Yang, C., Osterman, A., Smith, J. W. Inhibition of fatty-acid synthase induces caspase-8-mediated tumor cell apoptosis by up-regulating DDIT4. *J. Biol. Chem.* **283,** 31378-31384 (2008)

Kemp,B.E., D. J. Graves, E. Benjamini, E. G. Krebs, Role of multiple basic residues in determining the substrate specificity of cyclic AMP-dependent protein kinase. *J. Biol. Chem.* **252**, 4888-4894 (1977)

Kingston, D. G., Newman, D. J. Combinatorial solid-phase natural product chemistry. *Curr. Opin. Drug Discov. Dev*. **10**, 130-144 (2007)

Kratchmarova, I., Blagoev, B., Haack-Sorensen, M., Kassem, M., Mann, M. Mechanism of divergent growth factor effects in mesenchymal stem cell differentiation. *Science* **308**, 1472-1477 (2005)

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).

Laughlin, S. T., Baskin, J. M., Amacher, S. L., Bertozzi, C. R. In vivo imaging of membrane-associated glycans in developing zebrafish *Science* **320**, 664- 667 (2008)

Larive, C. K., Lunte, S. M., Zhong, M., Perkins, M. D., Wilson, G. S., Gokulrangan, G., Williams, T., Afroz, F., Schoneich, C., Derrick, T. S., Middaugh, C. R., Bogdanowich-Kni, S. Mass Spectrometry in Proteomics *Anal. Chem.* 7**1**, 398 **(**1999**)**

Le Bonniec, S., C. Deregnaucourt, V. Redeker, R. Banerjeei, P. Grellier, D. E. Goldbergi, J. Schrevel, Plasmepsin II, an Acidic Hemoglobinase from thePlasmodium falciparum Food Vacuole, Is Active at Neutral pH on the Host Erythrocyte Membrane Skeleton *J. Biol. Chem.* **274**, 14218- 14223 (1999)

Lemieux, G. A., De Graffenried, C. L., Bertozzi, C. R. A fluorogenic dye activated by the staudinger ligation *J. Am. Chem. Soc.* **125**, 4708-4709 (2003)

Leung, D., C. Hardouin, D.L. Boger and B.F. Cravatt, Discovering potent and selective inhibitors of enzymes in complex proteomes. *Nat. Biotechnol.* 21, 687-691 (2003)

Liu, J., E. S. Istvan, I. Y. Gluzman, J. Gross, D. E. Goldberg, Plasmodium falciparum ensures its amino acid supply with multiple acquisition pathways and redundant proteolytic enzyme systems *Proc. Natl. Acad. Sci. USA* **103**, 8840 - 8845 (2006)

Lindsay, M.A., Target discovery Nature Reviews Drug Discovery **2**, 831-838 (2003)

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

Lockhart, D.J. & Winzeler, E.A. Genomics, gene expression and DNA arrays. *Nature* **405**, 827-836 (2000).

Lopez, M.F. K. Berggren, E. Chernokalskaya, A. Lazarev, M. Robinson and W.F. Patton, A comparison of silver stain and SYPRO Ruby protein gel stain with respect to protein detection in two-dimensional gels and identification by peptide mass profiling. *Electrophoresis* **21**, 3673-3683 (2000)

Liu, K., Kalesh, K.A,; Ong L.B., Yao, S.Q., An Improved Mechanism-Based Cross-Linker for Multiplexed Kinase Detection and Inhibition in A Complex Proteome. *ChemBioChem*, **9**, 1883-1888 (2008)

Liu, K., Shi, H., Xiao, H., Chong, A.G.L., Bi, X., Chang, Y.T., Tan, K., Yada, R.Y., Yao, S.Q. Functional Profiling, Identification and Inhibition of Plasmepsins in Intraerythrocytic Malaria Parasites *Angew. Chem. Intl. Ed.*, **48**, 8293-8297 (2009)

Liu, Y., Zhu, X., Liao, S., Tang, Q., Liu, K., Guan, X., Zhang, J., Feng, Z. Identification of differential expression of genes in hepatocellular carcinoma by suppression subtractive hybridization combined cDNA microarray. *Oncol. Rep.* **18**, 943-951 (2007)

Little, J. L., Wheeler, F. B., Fels, D. R., Koumenis, C., Kridel, S. J. Disruption of Crosstalk Between the Fatty Acid Synthesis and Proteasome Pathways Enhances Unfolded Protein Response Signaling and Cell Death *Cancer Res.* **67,** 1262-1269 (2007)

Malone, J.P., M.R. Radabaugh, R.M. Leimgruber and G.S. Gerstenecker, Practical aspects of fluorescent staining for proteomics alications. *Electrophoresis* 22, 919-932 (2001)

Mann, M., Hendrickson, R.C. & Pandey, A. Analysis of proteins and proteomes by mass spectrometry. *Annu. Rev. Biochem.***70**, 437-473 (2001).

Mann, M., Jensen, O. N., Proteomic analysis of post-translational modifications. *Nat. Biotechnol.* **21**, 255-61 (2003)

Matsuoka, S., Ballif, B. A., Smogorzewska, A., McDonald III, E. R., Hurov, K. E., Luo, J., Bakalarski, C. E., Zhao, Z., Solimini, N., Lerenthal, Y., Shiloh, Y., Gygi, S. P., Elledge, S. J. ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* **316**, 1160-1166 (2007)

Manning, G., Whyte, R. M. D., Hunter, T., Sudarsanam, S., The protein kinase complement of the human genome. *Science* **298**, 1912-1934 (2002)

Marks PA, Breslow R. Dimethyl sulfoxide to vorinostat: development of this histone deacetylase inhibitor as an anticancer drug. *Nat. Biotechnol.* **25,** 84-90 (2007)

Maly, D. J., Allen, J. A. K. M. Shokat, A mechanism-based cross-linker for the identification of kinase-substrate pairs. *J. Am. Chem. Soc.* **126,** 9160-9161 (2004)

Macbeath G. and S.L. Schreiber, Printing proteins as microarrays for high-throughput function determination. *Science* **289,** 1760-1763 (2000)

Martin, B.R., and Cravatt, B.F., Large-scale profiling of protein palmitoylation in mammalian cells, *Nat Methods* **6** 135-138 (2009)

Martin, D.O., Gonzalo L. Vilas, Jennifer A. Prescher†, Gurram Rajaiah‡, John R. Falck‡, Carolyn R. Bertozzi† and Luc G. Berthiaume Rapid detection, discovery, and identification of post-translationally myristoylated proteins during apoptosis using a bio-orthogonal azidomyristate analog. *The FASEB Journal*. **22**, 797-806 (2008)

Moche, M., Schneider, G., Edwards, P., Dehesh, K., Lindqvist, Y. Structure of the complex between the antibiotic cerulenin and its target, beta-ketoacyl-acyl carrier protein synthase. *J. Biol. Chem.* **274**, 6031-6034 (1999)

Ma, G., Zancanella, M., Oyola, Y., Richardson, R. D., Smith, J. W., Romo, D. Total synthesis and comparative analysis of orlistat, valilactone, and a transposed orlistat derivative: Inhibitors of fatty acid synthase. *Org. Lett.* **8**, 4497-4500 (2006)

Mckusick, V.A. Genomics: structural and functional studies of genomes. *Genomics* **45**, 244-249 (1997)

McLachlin, D. T., B. T. Chait, Analysis of phosphorylated proteins and peptides by mass spectrometry. *Curr. Opin. Chem. Biol.* **5**, 591-602 (2001)

Medina, V., Edmonds, B., Young, G. P., James, R., Aleton, S., Zalewski, P. D. Induction of caspase-3 protease activity and apoptosis by butyrate and trichostatin A (inhibitors of histone deacetylase): dependence on protein synthesis and synergy with a mitochondrial/cytochrome c-dependent pathway. *Cancer Res.* **57**, 3697-3707 (1997)

Meggio, F. , A. Donella Deana, M. Ruzzene, A. M. Brunati, L. Cesaro, B. Guerra, T. Meyer, H. Mett, D. Fabbro, P. Furet, Different susceptibility of protein kinases to staurosporine inhibition. Kinetic studies and molecular bases for the resistance of protein kinase CK2. *Eur. J. Biochem.* **234**, 317-322 (1995)

Menendez, J. A., Lupu, R. Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. *Nat. Rev. Cancer* 7, 763-777 (2007)

Michael S. Finnin1, Jill R. Donigian2, Alona Cohen2, Victoria M. Richon3, Richard A. Rifkind3, Paul A. Marks3, Ronald Breslow4 & Nikola P. Pavletich1 Structures of a histone deacetylase homologue bound to the TSA and SAHA inhibitors *Nature* **401**, 188-193 (1999)

Mori, R., Wang, Q., Danenberg, K. D., Pinski, J. K., Danenberg, P. V. Both beta-actin and GAPDH are useful reference genes for normalization of quantitative RT-

PCR in human FFPE tissue samples of prostate cancer. *Prostate* **68**, 1555-1560. (2008)

Ngo, J.T., J.A. Champion, A. Mahdavi, I.C. Tanrikulu, K.E. Beatty, R.E. Connor, T.H. Yoo, D.C. Dieterich, E.M. Schuman and D.A. Tirrell, Cell-selective metabolic labeling of proteins, *Nat Chem Biol* **5**, 715-717 (2009)

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

Nezami, A., T. Kimura, K. Hidaka, A. Kiso, J. Liu, Y. Kiso, D. E. Goldberg, E. Freire, High-Affinity Inhibition of a Family of Plasmodium falciparum Proteases by a Designed Adaptive Inhibitor *Biochemistry* **42**, 8459 - 8464 (2003)

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

Ong, S. E., Kratchmarova, I., Mann, M. *J. Proteome Res.* **2**, 173-181 (2003)

Opiteck, G. J., Lewis, K. C., Jorgenson, J. W., Anderegg, R. Comprehensive on-line LC/LC/MS of proteins *J. Anal. Chem.* **69**, 1518-1524 (1997)

Opiteck, G. J., Jorgenson, J. W., Anderegg, R. Two-Dimensional SEC/RPLC Coupled to Mass Spectrometry for the Analysis of Peptides *J. Anal. Chem.* **69**, 2283-2291 (1997)

Pamela V. Chang, Jennifer A. Prescher, Matthew J. Hangauer, and Carolyn R. Bertozzi Imaging Cell Surface Glycans with Bioorthogonal Chemical Reporters *J. Am. Chem. Soc.,* **129,** 8400-8401 (2007)

Pandey, A., and M Mann, Proteomics to study genes and genomes. *Nature* **405**, 837-846 (2000)

Parr, C. L., Tanaka, T., Xiao, H. , Yada, R. Y., The catalytic significance of the proposed active site residues in Plasmodium falciparum histoaspartic protease *Febs J* **275**, 1698-1707 (2008)

Pasa-Tolic, L., Jensen, P.K., Anderson, G.A., Lipton, M.S., Peden, K.K., S. Martinovic, S., Tolic, N., Bruce J.E., and Smith, R.D., High throughput proteome-wide precision measurements of protein expression using mass spectrometry. *J Am Chem Soc* **121,** 7949-7950 (1999)

Patterson, S.D. From electrophoretically separated protein to identification: strategies for sequence and mass analysis. *Anal. Biochem.* **221**, 1-15 (1994)

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

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

Paulsen, I.T. Sliwinski, M.K. Nelissen, B. Goffeau, A. and Saier, Jr, M.H. Unified inventory of established and putative transporters encoded within the complete genome of Saccharomyces cerevisiae. *FEBS Lett* **430,** 116-125 (1998)

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

Prescher J.A., and Bertozzi, C.R. Chemistry in living systems, *Nat. Chem. Biol.* **1**, 13-21 (2005)

Pemble, C. W., Johnson, L. C., Kridel, S. J., Lowther, W. T. Crystal structure of the thioesterase domain of human fatty acid synthase inhibited by Orlistat. *Nat. Struct. Mol. Biol.* **14**, 704-709. (2007)

Parr, C. L., Tanaka, T. Xiao, H. Yada, R. Y., The catalytic significance of the proposed active site residues in Plasmodium falciparum histoaspartic protease. *Febs J* **275**, 1698-1707; (2008)

Phadke, M. S., Krynetskaia, N. F., Mishra, A. K., Krynetskiy, E. J. Inhibition of glycolysis modulates prednisolone resistance in acute lymphoblastic leukemia cells Pharmacol. Exp. Therapeut. **331**, 77-86 (2009)

Pedersen, S., Bloch, P. L., Reeh, S., Neidhardt, F. C. Patterns of protein synthesis in E. coli: a catalog of the amount of 140 individual proteins at different growth rates. *Cell* **14**, 179-190 (1978)

Rabuka, D., Hubbard, S.C. S.T. Laughlin, S.P. Argade and C.R. Bertozzi, A chemical reporter strategy to probe glycoprotein fucosylation, *J Am Chem Soc* **128** 12078-12079 (2006)

Resh, M.D., Use of analogs and inhibitors to study the functional significance of protein palmitoylation, *Methods* **40,** 191-197 (2006)

Rostovtsev V.V. et al., A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes, *Angew Chem. Int. Ed. Engl.* **41,** 2596-2599 (2002)

Roche, F.K., B.M. Marsick and P.C. Letourneau, Protein synthesis in distal axons is not required for growth cone responses to guidance cues, *J Neurosci* **29,** 638-652 (2009)

Richardson, R. D., Ma, G., Oyola, Y., Zancanella, M., Knowel, L. M., Cieplak, P., Romo, D., Smith, J. W. Synthesis of novel beta-lactone inhibitors of fatty acid synthase. *J. Med. Chem.* **51**, 5285-5296 (2008)

Saghatelian, A. & Cravatt. B.F. Assignment of protein function in the postgenomic era *Nature Chemical Biology* **1**, 130-142 (2005)

Saghatelian A, Jessani N, Joseph A, Humphrey M, Cravatt BF. Activity-based probes for the proteomic profiling of metalloproteases. *Proc. Natl. Acad. Sci. USA* **101,** 10000-10005 (2004)

Santoni, V., T. Rabilloud, P. Doumas, D. Rouquié, M. Mansion, S. Kieffer, J. Garin and M. Rossignol, Towards the recovery of hydrophobic proteins on two-dimensional electrophoresis gels. *Electrophoresis* **20,** 705-711 (1999)

Sawa, M., Hsu, T. L., Itoh, T., Sugiyama, M., Hanson, S. R., Vogt, P. K., Wong, C. H. Glycoproteomic probes for fluorescent imaging of fucosylated glycans in vivo *Proc. Natl. Acad. Sci. U.S.A.* **103**, 12371-12376 (2006)

Seeliger, M.A., Nagar, B., Frank, F., Cao, X., Henderson, M. N., Kuriyan, J., c-Src binds to the cancer drug imatinib with an inactive Abl/c-Kit conformation and a distributed thermodynamic penalty *Structure* **15**, 299-311 (2007)

Sieber S.A., Niessen S, Hoover H.S., Cravatt B.F. Proteomic profiling of metalloprotease activities with cocktails of active-site probes. *Nat. Chem. Biol.* **2,** 274-281 (2006)

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

Saxon, E., Luchansky, S.J., Hang, H.C., Yu, C., Lee, S.C., and Bertozzi, C.R. Investigating cellular metabolism of synthetic azidosugars with the Staudinger ligation. *J Am Chem Soc* **124,** 14893-14902 (2002)

Salisbury C.M., Cravatt B.F. Activity-based probes for proteomic profiling of histone deacetylase complexes. *Proc. Natl. Acad. Sci. USA* **104,** 1171-1176 (2007)

Salisbury C.M., and Cravatt, B.F. Optimization of activity-based probes for proteomic profiling of histone deacetylase complexes. *J. Am. Chem. Soc.*, **130,** 2184-2194 (2008)

Sekimoto, H., Boney, C. M. C-terminal Src kinase (CSK) modulates insulin-like growth factor-I signaling through Src in 3T3-L1 differentiation. *Endocrinology* **144,** 2546-2552 (2003)

Schmidinger,H., Birner-Gruenberger,R., Riesenhuber, G., Saf, R., Susani-Etzerodt H,. and Hermetter,A. Novel fluorescent phosphonic acid esters for discrimination of lipases and esterases, ChemBioChem **6**, 1776-1781 (2005)

Shults, M. D., Janes, K. A., Lauffenburger, D. A., Imperiali, B. A multiplexed homogeneous fluorescence-based assay for protein kinase activity in cell lysates. *Nat. Methods* **2**, 277-283 (2005)

Sletten, E. M., Bertozzi, C. R. Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality. Angew. Chem., Int. Ed. **48**, 6974-6998 (2009)

Slice, L. W., Taylor, S. S., Expression of the Catalytic Subunit of CAMP-dependent Protein Kinase in Escherichia coli *J Biol. Chem.* **264**, 20940-20946 (1989)

Sondhi, D., Xu, W. Songyang, Z. Eck, M. J. Cole, P. A. Peptide and protein phosphorylation by protein tyrosine kinase Csk: insights into specificity and mechanism. *Biochemistry* **37**, 165-172 (1998)

Snow, R.W., Guerra, C. A., Noor, A. M., Hay, S. I., The global distribution of clinical episodes of Plasmodium falciparum malaria. *Nature* **434**, 214 - 217 (2005)

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

Speers, A.E., Adam, G.C. & Cravatt, B.F. Activity-based protein profiling in vivo using a copper(i)-catalyzed azide-alkyne [3 + 2] cycloaddition. *J. Am. Chem. Soc.* **125**, 4686-4687 (2003)

Srinivasan; R., Tan, L.P., Wu, H., Yang, P.-Y., Kalesh, K.A,; Yao, S.Q. High-Throughput Synthesis of Azide Libraries Suitable for Direct "Click" Chemistry and in situ Screening, *Org. Biol. Chem.,* 7, 1821-1828 (2009)

Srinivasan; R., Tan, L.P., Wu, H., Yao, S.Q., Solid-phase assembly and in situ screening of protein tyrosine phosphatase inhibitors. *Org. Lett.,* **10**, 2295-2298 (2008)

Srinivasan, R., Uttamchandani, M., Yao, S.Q., Rapid assembly and in situ screening of bidentate inhibitors of protein tyrosine phosphatases. *Org. Lett.,* **8**, 713-716 (2006)

Steven H. L., Bogyo, M. A Mild Chemically Cleavable Linker System for Functional Proteomic Applications Angewandte Chemie, **119**, 1306-1308 (2007)

Stasche, R. et al. A bifunctional enzyme catalyzes the first two steps in N-acetylneuraminic acid biosynthesis of rat liver—Molecular cloning and functional expression of UDP-N-acetyl-glucosamine 2-epimerase/N-acetylmannosamine kinase. *J. Biol. Chem.* **272**, 24319−24324 (1997)

Tornoe C.W. et al., Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(i)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides, *J. Org. Chem.* **67**, 3057-3064 (2002)

Uetz, P., Giot, L. Cagney, G. Mansfield, T.A. Judson, R.S. Knight, J.R. Lockshon, D. Narayan, V. Srinivasan, M., Pochart P., et al., A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* **403,** 623-627 (2000)

Ubersax, J. A., Ferrell Jr. J. E., Mechanisms of specificity in protein phosphorylation. *Nat. Rev. Mol. Cell Biol.* **8**, 530-541 (2007)

Uttamchandani, M; Li. J., Sun, H., Yao, S.Q. Activity-Based Profiling: New Developments and Directions in Protein Fingerprinting ChemBioChem, **9**, 667-675 (2008)

Uttamchandani, M., Lu, C. H. S., Yao, S. Q. Next Generation Chemical Proteomic Tools for Rapid Enzyme Profiling *Acc. Chem. Res.* **42**, 1183-1192 (2009)

Vocadlo, D.J. ; H.C. Hang, E.J. Kim, J.A. Hanover and C.R. Bertozzi, A chemical aroach for identifying O-GlcNAc-modified proteins in cells. *Proc Natl Acad Sci USA* **100** 9116-9121 (2003)

Vaarala, M. H., Porvari, K. S., Kyllonen, A. P., Mustonen, M. V. J., Lukkarinen, O., Vihko, P. T. Several genes encoding ribosomal proteins are over-expressed in prostate-cancer cell lines: confirmation of L7a and L37 over-expression in prostate-cancer tissue samples. *Int. J. Cancer* **78**, 27-32 (1998)
Van den Bergh, G. & Arckens, L. Fluorescent two-dimensional difference gel electrophoresis unveils the potential of gel-based proteomics. *Curr. Opin. Biotechnol.* **15**, 38-43 (2004)

Watson, J.D. & Cook-Deegan, R.M. Origins of the Human Genome Project. FASEB J. 5, 8-11 (1991)

Wang Q. et al., Bioconjugation by copper(I)-catalyzed azide-alkyne [3 + 2] cycloaddition, J. Am. Chem. Soc. **125**, 3192-3193 (2003)

Wang, J., Uttamchandani, M., Li, J., Hu, M., Yao, S.Q. "Click" Synthesis of Small Molecule Probes for Activity-Based Fingerprinting of Matrix Metalloproteases *Chem. Commun.,* 3783-3785. (2006)

Wang, Y., Cheong, D., Chan, S., Hooi, S. C. Ribosomal protein L7a gene is up-regulated but not fused to the tyrosine kinase receptor as chimeric trk oncogene in human colorectal carcinoma. *Int. J. Oncol.* **16,** 757-762 (2000)

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

Xiao, H., Sinkovits, A. F., Bryksa, B. C., M. Ogawa, R. Y. Yada, Recombinant expression and partial characterization of an active soluble histo-aspartic protease from Plasmodium falciparum. *Protein Expression Purif.* **49**, 88 - 94 (2006)

Xiao, H., T. Tanaka, M. Ogawa, R. Y. Yada, Expression and enzymatic characterization of the soluble recombinant plasmepsin I from Plasmodium falciparum *Protein Eng. Des. Sel.* **20**, 625 - 633 (2007)

Yates, J. R., III; Speicher, S., Griffin, P. R., Hunkapiller, T. Peptide mass maps: a highly informative approach to protein identification. *Anal. Biochem.* **214**, 397-408 (1993)

Yan, J.X. R.A. Harry, C. Spibey and M.J. Dunn, Post-electrophoretic staining of proteins separated by two-dimensional gel electrophoresis using SYPRO dyes. *Electrophoresis* **21**, 3657-3665 (2000),

Yang, P-Y., Liu, K., Ngai, M.H., Lear, M.J.., Wenk, M., Yao, S.Q. Activity-Based Proteome Profiling of Potential Cellular Targets of Orlistat - An FDA-Approved Drug with Anti-Tumor Activities. *J. Am. Chem. Soc.* **132**, 656-666 (2010)

Yee M.C., Fas S.C., Stohlmeyer M.M., Wandless T.J., Cimprich K.A. A cell-permeable, activity-based probe for protein and lipid kinases. *J. Biol. Chem.* **280,** 29053-29059 (2005)

Zhang, J., R. E. Campbell, A. Y. Ting, R. Y. Tsien, Creating new fluorescent probes for cell biology. *Nat. Rev. Mol. Cell Biol.* **3**, 906-918 (2002)

Zhu, Y., Lin, H., Li, Z., Wang, M., Luo, J. Modulation of expression of ribosomal protein L7a (rpL7a) by ethanol in human breast cancer cells. *Breast Cancer Res. Treat.* **69**, 29-38 (2001)

Zhang, M. M., Tsou, L. K., Charron, G., Raghavan, A. S., Hang, H. C. Tandem fluorescence imaging of dynamic S-acylation and protein turnover. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 8627-8632 (2010)

Chapter 9.

Appendix

9.1 Supplemental Tables

Table 9.1 List of 152 compound in the inhibitor library targeting plasmepsins in *P. Falciparum*. Identity of each compound as well as its quality confirmed by LC/MS

#	Product ID	Alkyne Warhead	Azide	LCMS Results			NMR & Scale up
				Est % Purity	Cal. MW	Obs. MW	
1	A1	A		25	571.17	572.168	-
2	B1	B		30	547.25	548.253	-
3	C1	C		80	537.19	538.183	-
4	D1	D		50	513.27	514.266	-
5	E1	E	1	80	587.17	588.187	-
6	F1	F		30	563.25	564.268	-
7	G1	G		30	523.17	524.191	-
8	H1	H		70	499.25	500.296	-
9	A2	A		>90	585.19	586.222	-
10	B2	B		>90	561.27	562.275	-
11	C2	C		>90	551.21	552.222	-
12	D2	D		-	527.29	-	-
13	E2	E	2	-	601.19	-	-
14	F2	F		-	577.27	-	-
15	G2	G		-	537.19	-	-
16	H2	H		-	513.27	-	-
17	A3	A		>95	551.17	552.187	-
18	B3	B		>90	527.25	528.261	-
19	C3	C		>95	517.19	518.202	-
20	D3	D	3	-	493.27	-	-
21	E3	E		-	567.17	-	-
22	F3	F		63	543.25	544.258	-
23	G3	G		-	503.17	-	-

24	H3	H		-	479.25	-	v
25	A4	A		60	551.17	552.179	-
26	B4	B		80	527.25	528.261	-
27	C4	C		>90	517.19	518.198	-
28	D4	D	4	70	493.27	494.277	-
29	E4	E		-	567.17	-	-
30	F4	F		40	543.25	544.258	-
31	G4	G		-	503.17	-	-
32	H4	H		-	479.25	-	-
33	A5	A		60	579.16	580.173	-
34	B5	B		50	555.24	556.258	-
35	C5	C		>95	545.18	546.172	-
36	D5	D	5	-	521.26	-	-
37	E5	E		40	595.16	596.15	-
38	F5	F		50	571.24	572.236	-
39	G5	G		70	531.16	532.192	-
40	H5	H		60	507.24	508.236	-
41	A6	A		70	555.12	556.111	-
42	B6	B		70	531.2	532.169	-
43	C6	C		50	521.14	522.126	-
44	D6	D	6	-	497.22	-	-
45	E6	E		-	571.12	-	-
46	F6	F		60	547.2	548.191	-
47	G6	G		>90	507.12	508.116	-
48	H6	H		-	483.2	-	-
49	A7	A		>90	539.15	540.144	-
50	B7	B		>90	515.23	516.226	-
51	C7	C		>90	505.17	506.163	v
52	D7	D	7	>90	481.25	504.221	-
53	E7	E		>85	555.15	556.136	-
54	F7	F		>90	531.23	532.215	-
55	G7	G		-	491.15	-	-
56	H7	H		>90	467.23	468.226	-
57	A8	A		-	557.14	-	-
58	B8	B		-	533.22	-	-
59	C8	C	8	-	523.16	-	-
60	D8	D		-	499.24	-	-
61	E8	E		-	573.14	-	-
62	F8	F		-	549.22	-	-

63	G8	G		50	509.14	510.134	-
64	H8	H		-	485.22	-	-
65	A9	A		>95	553.16	554.16	-
66	B9	B		-	529.24	-	-
67	C9	C		>90	519.18	542.177	-
68	D9	D	9	-	495.26	-	-
69	E9	E		60	569.16	570.153	-
70	F9	F		>95	545.24	-	-
71	G9	G		60	505.16	506.16	-
72	H9	H		-	481.24	-	-
73	A10	A		70	546.15	547.147	-
74	B10	B		-	522.23	-	-
75	C10	C		>95	512.17	513.169	-
76	D10	D		>95	488.25	489.242	-
77	E10	E	10	40	562.15	563.158	-
78	F10	F		40	538.23	539.221	-
79	G10	G		>85	498.15	499.167	-
80	H10	H		-	474.23	-	-
81	A11	A		40	607.14	608.1371	-
82	B11	B		70	583.22	584.2211	-
83	C11	C		60	573.16	574.1565	-
84	D11	D		>95	549.24	550.2378	-
85	E11	E	11	>95	623.14	624.132	-
86	F11	F		>95	599.22	600.213	-
87	G11	G		80	559.14	560.136	-
88	H11	H		80	535.22	536.222	-
89	A12	A		50	607.14	608.137	-
90	B12	B		50	583.22	584.219	-
91	C12	C		70	573.16	574.158	v
92	D12	D		70	549.24	550.24	-
93	E12	E	12	>95	623.14	624.147	-
94	F12	F		>95	599.22	600.21	-
95	G12	G		60	559.14	560.149	-
96	H12	H		>90	535.22	536.222	-
97	A13	A		>95	593.11	594.11	-
98	B13	B		>95	569.19	570.191	-
99	C13	C	13	50	559.13	560.129	-
100	D13	D		>95	535.21	536.26	-
101	E13	E		40	609.11	610.105	-

102	F13	F		70	585.19	586.189	-
103	G13	G		>90	545.11	546.112	-
104	H13	H		60	521.19	522.191	-
105	A14	A		>90	599.14	600.136	-
106	B14	B		70	575.22	576.223	-
107	C14	C		80	565.16	566.157	-
108	D14	D		70	541.24	542.232	-
109	E14	E	14	-	615.14	-	-
110	F14	F		>85	591.22	592.217	-
111	G14	G		-	551.14	552.138	-
112	H14	H		>90	527.22	528.218	v
113	A15	A		70	585.16	586.159	-
114	B15	B		50	561.24	562.236	-
115	C15	C		50	551.18	552.177	-
116	D15	D		50	527.26	528.254	-
117	E15	E	15	>95	601.16	602.161	-
118	F15	F		>90	577.24	578.235	-
119	G15	G		>90	537.16	538.1544	v
120	H15	H		70	513.24	514.248	-
121	A16	A		60	591.09	592.078	-
122	B16	B		-	567.17	568.161	-
123	C16	C		>90	557.11	558.109	v
124	D16	D		>85	533.19	534.184	-
125	E16	E	16	90	607.09	608.081	-
126	F16	F		>90	583.17	584.168	-
127	G16	G		70	543.09	544.084	v
128	H16	H		70	519.17	520.161	-
129	A17	A		50	633.16	634.157	-
130	B17	B		60	609.24	610.239	-
131	C17	C		90	599.18	600.148	-
132	D17	D		70	575.26	576.239	-
133	E17	E	17	>95	649.16	650.134	-
134	F17	F		>95	625.24	626.209	-
135	G17	G		>95	585.16	586.139	-
136	H17	H		60	561.24	562.221	-
137	A18	A		70	602.11	603.094	-
138	B18	B	18	70	578.19	579.173	-
139	C18	C		70	568.13	569.11	-
140	D18	D		80	544.21	545.191	-

141	E18	E		30	618.11	619.088	-
142	F18	F		-	594.19	-	-
143	G18	G		60	554.11	555.097	-
144	H18	H		-	530.19	-	-
145	A19	A	19	70	599.17	600.149	-
146	B19	B		70	575.25	576.224	-
147	C19	C		80	565.19	566.174	-
148	D19	D		70	541.27	542.245	-
149	E19	E		80	615.17	616.136	-
150	F19	F		>95	591.25	592.226	-
151	G19	G		-	551.17	-	-
152	H19	H		-	527.25	-	v

9.2 Supplemental Figures

Eight hydroxyethyl-based WH (A-H).
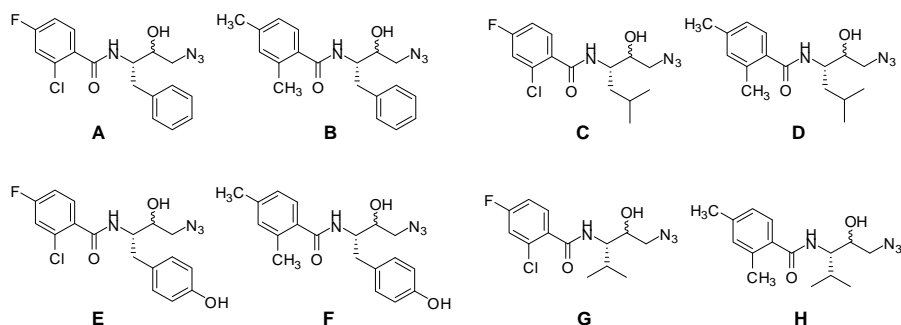
Figure 9.1. Chemical structures of the eight hydroxyethyl-based warheads **WH** (**A-H**). The eight Warheads (**A-H**; Figure 9.1) were synthesized and purified as mixtures of diasteromers by modifications of published procedures,[3] and characterized as the followings.
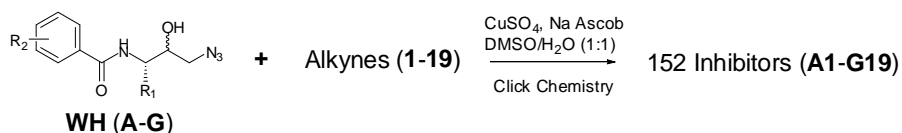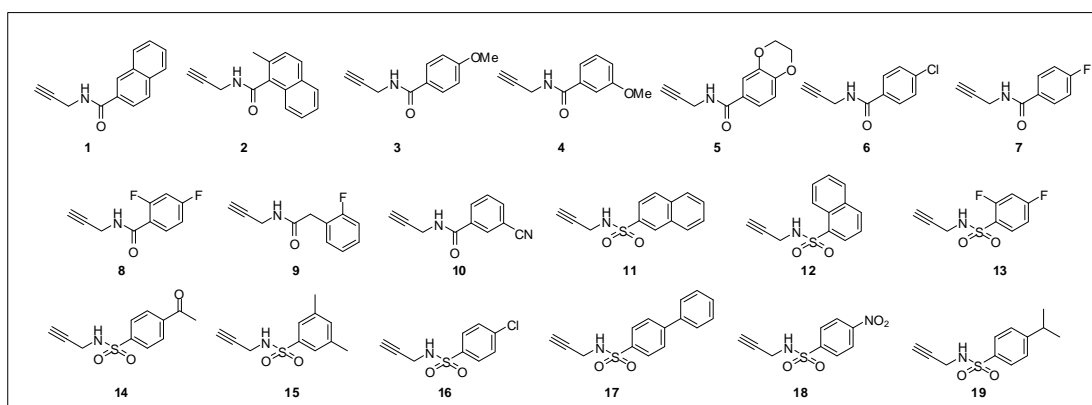
Figure 9.2. Structures of the 19 alkynes used, and the "click" synthesis of 152-member plasmepsin inhibitors.
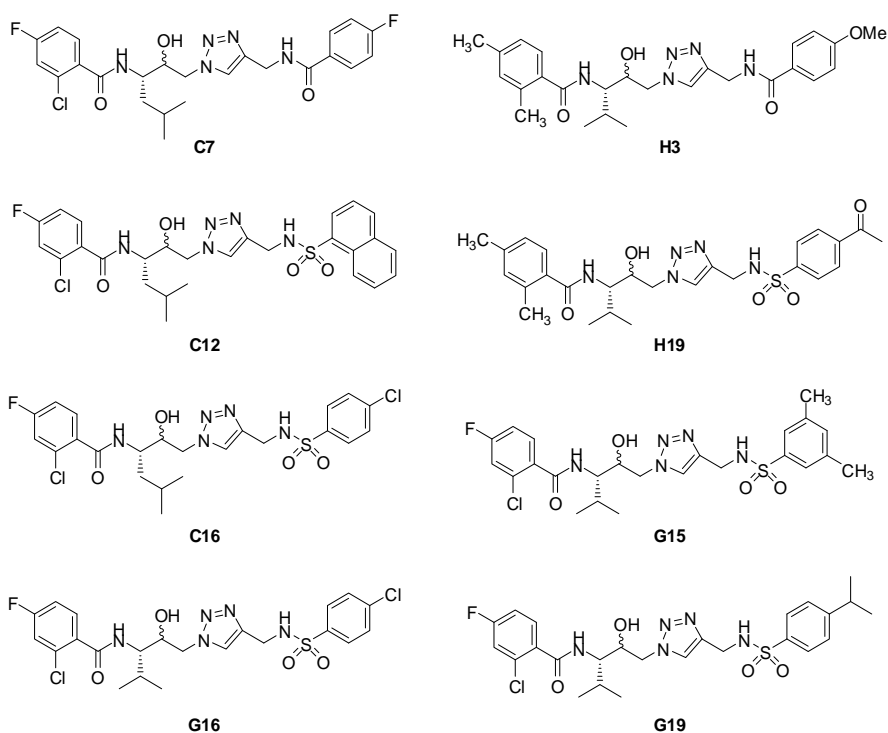
Alkyne-containing linker (≡-BP-TER).
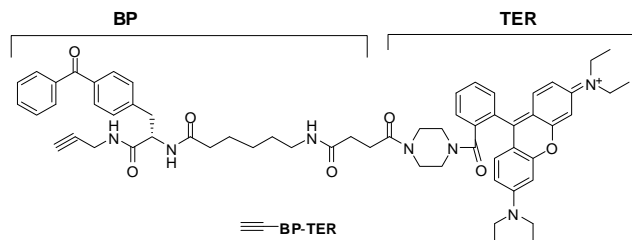
Figure 9.3. Chemical structure of the alkyne-containing **BP-TER** linker. (≡−**BP**−**TER**). The alkyne-containing linker (Figure 9.3) was synthesized based on previously published procedures.[4]

Charaterization of 8 putative "hits" against plasmepsins

Figure 9.4 Chemical Structures of the 8 selected "hits" from in-situ screening with parasite extracts and screening with purified enzymes.