ABSTRACT

Title of Document: **DEVELOPMENT OF TOOLS TO CHARACTERIZE PROTEIN-PROTEIN INTERACTIONS**

JIANGSONG JIANG, Doctor of Philosophy, 2010

Directed By: Assistant Professor, Shuwei Li,
Department of Chemistry and Biochemistry
Associate Professor, Richard Stewart
Department of Cell Biology and Molecular Genetics

Protein-protein interactions (PPIs) are crucial to most biological processes and activities. Large-scale PPI screening has been applied to model organisms as well as to human cells. Two approaches have been used extensively in high-throughput PPI studies: (i) the Yeast Two-Hybrid (Y2H) assay (a bottom-up method), and (ii) the tandem affinity purification (TAP) (a top-down method). However, a close examination of both techniques revealed issues that limit their effectiveness. Thus, it is important to develop new methods that can bridge the gap between the Y2H and the TAP. In this thesis, two approaches were developed to meet this need.

The first approach was a photoaffinity labeling tool, which was based on a photo-caged reactive intermediate para-quinone methide (pQM) to study protein-peptide associations. This system was developed and optimized by using the interaction

between catPTP1Bm and the EGFR peptide as a test case. Highly specific protein labeling was achieved, and mass spectrometry (MS) was used to identify the crosslinked site on the target protein. Interestingly, two peptides from catPTP1Bm detected by MS were found close to the enzyme-substrate binding interface in the three-dimensional structure of the complex, which demonstrated this method might be useful for the analysis of protein complex conformation.

The second approach, named “PCA plus”, took advantage of a technique referred to as “Protein-fragment Complementation Assay (PCA)”. A hydrolysis-deficient mutant β-lactamase (E166N) was used, which enabled interacting protein labeling in live cells. With this modification, the PCA plus method realized live cell imaging with subcellular resolution. Fluorescent microscopy and flow cytometry analysis demonstrated its potential applications. In addition, a new β-lactamase substrate was developed for the PCA plus method and was applied to enable purification, from living cells, of prey protein interacting with a bait protein. The observed enrichment of interacting partners suggested the system could be used for high-throughput PPI screening. Moreover, this method could also be useful for the characterization of low affinity and transient PPIs because of its capacity on labeling interacting protein inside cells.

DEVELOPMENT OF TOOLS TO CHARACTERIZE PROTEIN-PROTEIN INTERACTIONS

By

JIANGSONG JIANG

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2010

Advisory Committee:
Associate Professor, Richard Stewart, Chair
Assistant Professor, Shuwei Li, Co-chair
Professor, Marco Colombini
Professor, Steven Rokita
Associate Professor, Wenxia Song

Dedication

This dissertation is dedicated to my parents, my wife and my daughter, for bringing me the most beautiful things in my life.

Acknowledgements

I thank my advisor Dr. Shuwei Li for his guidance, support and help during my study. Shuwei's enthusiasm and passion on science has inspired me; His emphasis on research details has educated me; His open-mind and patience have established himself as a leading example in my future career. I'm grateful for the freedom he has kindly granted to me during my study. It is such a pleasure to work with him and share his beautiful mind and great ideals. Also, I thank my co-advisor Dr. Richard Stewart for his indispensable help and support on my study. In addition, I'm very thankful to my committee members Dr. Marco Colombini, Dr. Steven Rokita and Dr. Wenxia Song for their invaluable suggestion, constructive criticism, and encouraging remarks.

In addition, I would like to thank my lab mates Dr. Junxiang Zhang and Dr. Dexing Zeng for their corporation, and HHMI Undergraduate Fellow Puja Giri for her help. I also want to acknowledge Dr. Shunyuan Xiao, Dr. Illarion Turko and Dr. Yan Wang for their help. My special thanks go to Dr. Tom Stanton and Dr. Nancy Noben-Trauth for providing me teaching opportunities at the Shady Grove Campus. My gratitude extends to my colleagues and friends in the Institute for Bioscience and Biotechnology Research: Biao Ruan, Wei Yuan, Wei-Li Liao, Yiyuan Yin, Xinxiang Wang, thank you all for the help and friendship you have given to me.

Lastly, I want to thank my wife, my parents and family members. I don't know where I'd be without the love and support you have given to me throughout these years.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

3,4-dihydroxylphenylalanine	DOPA
Accessible surface area	ASA
Alkynyl-penicillin	AP
Calmodulin-binding peptide	CBP
Catalytic domain of protein tryrosine phosphotase 1B with C215 mutation	catPTP1Bm
Cell permeable alkynyl-penicillin	CP-AM
Chromatin Immunoprecipitation	ChIP
Co-Immunoprecipitation	Co-IP
Cytomegalovirus	CMV
Differential Interference Contrast	DIC
Dissociation constant	K_d
Dithiothreitol	DTT
DNA binding domain	BD
Dulbecco's Modified Eagle Medium	DMEM
Epidermal growth factor receptor	EGFR
Fetal Bovine Serum	FBS
Flash Laser Photolysis	FLP
Fluorescence in situ hybridization	FISH
Fluorescence resonance energy transfer	FRET
Fluorescence-Activated Cell Sorter	FACS
Glutathione	GSH
Glutathione disulfide	GSSG
Hank's Buffered Salt Solution	HBSS
High performance Liquid Chromatograph	HPLC
Horseradish peroxidase	HRP
IgG–binding unit of Protein A	ProtA
Isopropyl-β-D-thiogalactopyranoside	IPTG
Leucine zipper domain	Zip
Liquid Chromatography	LC
Mass spectrometry	MS
Mass-to-charge ratio	m/z
Non-specific esterase	NSE,
Nuclear Localization Signal	NLS
Nuclear magnetic resonance	NMR
Nuclear pore complex	NPC
Para-quinone methide	pQM
Phosphate buffered saline	PBS
Protein tyrosine phosphotase	PTP1B
Protein-fragment Complementation Assay	PCA

Protein-protein interaction	PPI
SRC Homology 3 Domain	SH3
SRC Homology 3 Domain binding peptide	PPLP
Tandem affinity purification	TAP
Transcription activation domain	AD
Tris (2-carboxyethyl) phosphine	TCEP
Tris [(1-benzyl-1H-1,2,3-triazol-4-yl)methyl] amine	TBTA
Yeast Two-Hybrid	Y2H
β-lactamase C-terminus	LacC
β-lactamase N-terminus	LacN
β-lactamase N-terminus with E166N mutation	NLacN
β-mercaptoethanol	βMe

Chapter One,

Introduction to Protein-Protein Interactions and Experimental Approaches Used to Study Them

Overview

Selective interactions between molecules play essential roles in biological system. Among these interactions are numerous protein-protein interactions (PPIs). For example, the structure and function of ribosome in *E. coli* is dependent on the interaction of 56 proteins (Nakao et al, 2004). Transcription in eukaryotes provides another example: it involves interactions among RNA polymerase II, five general transcription factors, a 20-protein complex called Mediator, and hundreds of transcription regulators (Kim et al, 1994). As another example, it was recently reported that the eukaryotic nuclear pore complex (NPC), one of the largest molecular machines, is assembled by about 30 nucleoporins (Elad et al, 2009). Besides these general functions, PPIs also play pivotal roles in specific biological processes. For example, in bacterial two-component signaling system, modular domains can adapt different conformational states to enable diverse regulatory mechanisms through stablizing or destablizing PPIs (Gao & Stock, 2009). In the innate immune system of mammals, adaptor proteins, such as Myd88, can activate different signaling pathways by selectively interacting with different effector proteins. On average, one protein interacts with five other proteins in a cell (Piehler, 2005). Therefore, characterizing PPIs is critical to understanding biological systems. In addition, because protein interactions can be highly dynamic as a result of spatial and temporal changes inside a cell, new characterization techniques that can detect diverse interactions could be important tools. In this chapter, several topics are reviewed: PPI diversity, principles

underlining PPI and experimental approaches used to study PPI. This review is then followed by an introduction to techniques that are involved in our method development.

Diversity of PPIs

PPIs are usually classified based on the affinity of their interaction, the lifetime of the complex, and the abundance of the interacting partners. Therefore, any given PPI might be referred to as: a strong or weak interaction, a stable or transient interaction, or a high or low abundance interaction. The affinity between two interacting proteins can be described quantitatively by the dissociation constant (K_d), which is associated to the free energy change of a protein interaction event. Usually, a PPI with a K_d higher than 10^{-6} M is considered to be a low affinity or weak interaction (Vaynberg et al, 2005). It is noteworthy that the affinity and the lifetime of a protein complex are the two sides of a coin, as K_d can also be determined by the ratio k_{off}/k_{on}. k_{on} is the association rate constant and k_{off} is the dissociation rate constant (Vaynberg & Qin, 2006). Accordingly, the affinity and the lifetime of a complex are often inversely related. High affinity complexes usually have a long lifetime, while low affinity complexes usually are short-lived. Typically, a low affinity complex (K_d 10^{-5}~10^{-6} M) will dissociate relatively quickly (half life of 0.1-1 s), and so it will be a transient PPI, continuously dissociating and reforming. By contrast, typical stable PPIs have binding affinities of 10^{-12}–10^{-9} M and half-life range of 12 min–19 h (Rudolph, 2007). PPI events are also related to the abundance of protein, defining low abundance and high abundance PPI. Protein concentrations vary dramatically inside cells. For example, in yeast, the absolute protein copy number ranges from 32 to 500,000 per cell (Gavin et al, 2006). Generally, cellular protein concentrations

are very low (less than 1nM), so some binary PPIs detected *in vitro* are considered nonspecific and therefore neglected if there is no further physiological evidence (Vaynberg et al, 2005).

The nature of PPIs

PPI modules

Protein domains and peptides are the functional units responsible for PPIs. Accordingly, PPIs have two forms: domain-domain association or domain-peptide association depending on the type of their interfaces. Although protein domains have highly diversified functions, many of them are exclusively dedicated to PPIs in a pathway related to specific cellular events. For example, the Death Domains (DDs) are more likely to mediate PPIs in an apoptosis signaling pathway (Hofmann & Tschopp, 1995). The Polyglutamine (PolyQ) domains are mostly involved in the interaction between transcriptional regulators (Palhan et al, 2005). The WD40 repeat domains are well known in cytoskeleton assembly and transcriptional activation by forming a platform for multi-protein assembly (Orlicky et al, 2003). The Leucine zipper domains can interact with each other to form two-, three-or four-stranded coiled coils (Harbury et al, 1993). All these are examples of domain-domain based PPIs involving either heterogeneous or homogeneous associations. Because of the importance of domain-domain interactions, some large-scale PPI maps have been constructed for them (Boxem et al, 2008; Prieto & De Las Rivas, 2009). Besides domain-domain interactions, domain-peptide interactions also govern critical cellular activities. There are also many examples in this category. The SH2 domains have high affinity for phosphotyrosine (pTyr) peptides (Filippakopoulos et

al, 2009). The PDZ domains are important interaction modules in cytoskeleton related signaling, which can bind the C-terminal of a peptide (Boxem et al, 2008). The SH3 domains recognize proline-rich peptides. Notably, there are over one hundred genes encoding SH3 domains in the human genome, making them attractive targets to study domain-peptide interactions (Kaneko et al, 2008). To summarize, PPI modules are building blocks in a PPI network, and understanding them is the key to unlock the PPI puzzle in cells.

PPI interface

The major difference between a domain-domain based PPI and a domain-peptide based PPI is the size of the interface, which is a key parameter to quantitatively describe a PPI. The PPI interface is defined by the change of the accessible surface area (ASA) of interacting partners before and after their binding. Specifically, if protein A and protein B form a complex, their interacting interface will be ASA(A) + ASA(B) - ASA(AB), that is, the difference between the sum of individual ASA of two proteins and the ASA of the protein complex (Janin et al, 2008). PPI interfaces range from 550 to 4900 $Å^2$, averaging 800 $Å^2$, and cover 6-30% of the monomer surface area (Archakov et al, 2003). The amino acid composition at different PPI interfaces can vary significantly, but arginine, histidine, asparagine, tryptophan, tyrosine, serine and hydrophobic amino acid residues are mostly observed (Veselovsky et al, 2002). Comparing large numbers of PPIs suggests that homocomplex interfaces are more hydrophobic than heterocomplex interfaces, and the hydrophobicity of a PPI interface is somewhere between that of protein interior (hydrophobic) and exterior (hydrophilic) (Jones & Thornton, 1996). In general, the

interfaces of short-lived protein complexes are similar to the active sites of enzymes, while the interfaces of stable protein complexes resemble protein cores (Archakov et al, 2003). An important way to describe the PPI interface is the "hotspot" theory. This theory states that at PPI binding interface most of binding energy just comes from a few key residues, which are surrounded by less important interactions from supportive residues (Clackson & Wells, 1995). Further analysis showes that these supportive residues may occlude solvent from hotspot residues (Bogan & Thorn, 1998).

Non-covalent factors in PPI events

The binding energy that drives the association of protein monomers is critical for PPIs. Many factors play roles in a biochemical reaction in terms of binding energy, but it is non-covalent factors that really matter for a PPI event, which, in a large way, may account for the dynamic and adaptive nature of life.

Hydrophobic interaction is one of the most important binding energy contributors for PPIs. It is formed through the concentration of non-polar groups in aqueous solution. The hydrophobic effect in a PPI interface tends to exist as a region called "patch", the number of which varies from 1 to 15 at an interface. Each hydrophobic patch may cover 200–400 $Å^2$ area and can reach up to 3000 $Å^2$ (Lijnzaad & Argos, 1997). The hydrophobic effect is more often observed in stable complexes than non-obligate interactions because of the more dynamic nature of the latter in water solution (Jones & Thornton, 1996).

The second important binding energy contributor for PPIs is hydrogen bond. It is formed

by the attraction of an electronegative atom (Acceptor) to the hydrogen atom covalently attached to another electronegative atom (Donor). On average, there is one hydrogen bond for each 100–200 $Å^2$ area at a PPI interface, totally about 10 hydrogen bonds for the whole interface(Veselovsky et al, 2002). Oxygen–nitrogen based hydrogen bonds are more prevalent than others given the polar group composition of protein surface. Amino acids side chains form about 76% of hydrogen bonds at the PPI interface (Xu et al, 1997). It is noteworthy that water molecules play a significant role at PPI interface in terms of hydrogen bond formation. Because a water molecule can have more than one hydrogen bond by interacting with an amino acid residue and another water molecule at the same time, they help form a network of hydrogen bonds at PPI interfaces (Xu et al, 1997).

The third important binding energy contributor for PPIs is the electrostatics. This is the attraction between two opposite charges. Consistent with the hydrophilic nature of protein surface, the average charge density at a protein surface is about 1.4 charged groups per 100 $Å^2$, totaling up to 12 charges per PPI interface (Barlow & Thornton, 1986). In contrast to the hydrogen bond formed only within the range of a dipole-dipole interaction, electrostatic attraction can mediate long-distance interaction between charged groups, which is very important for long-range protein associations (Janin et al, 2008).

Steric complementarity is another factor that can affect the association of two interacting proteins in terms of binding energy. Protein surface can be decomposed into "knobs", "holes" and "flats", which form the basis of the space complementarity at a PPI interface. Different types of PPIs have preference for interface shape complementarity. Based on a

statistic analysis, dimeric proteins have the most complementary shape, and antibody-protein interface shows the least complementarity, while protease-inhibitor complex displays moderate complementarity (Lawrence & Colman, 1993). Some PPI events may induce the conformation change of one or two interacting partners, which can significantly change the steric complementarity at the PPI interface.

The last binding energy contributor for PPIs is the van der Waals interaction, which is quite weak compared to other factors, but it may become important in some weak PPI events.

PPI kinetics and thermodynamics

Quantitative analysis of PPI events is very important, and kinetics and thermodynamics are two critical aspects to characterize PPIs in this way. As mentioned before, PPI kinetics and thermodynamics are connected by one important parameter: the disassociation constant K_d, which is used to define the PPI affinity and can be determined by the ratio of the disassociation rate constant k_{off} and the association rate constant k_{on} ($K_d = k_{off}/k_{on}$). In a simple one-step binding model between a receptor (R) and a ligand (L) to form RL, the K_d is equal to the concentration of R when half of L forms complex RL at equilibrium. Many methods have been developed to determine the $K_{d.}$, e.g. direct measurement of free ligand, indirect measurement of bound ligand, and competition methods with excessive free unlabeled ligand. Furthermore, some PPIs involve multiple proteins, which significantly complicate the experimental design. Therefore, it is very important to determine the binding stoichiometry for any protein binding experiment.

There are several critical points when measuring the K_d and stoichiometry of PPI events. Firstly, it is important to give enough time for the reaction to proceed for measuring K_d at equilibrium. Secondly, in order to accurately determine both binding stoichiometry and affinity in one experiment, protein concentration should be close to the binding constant, and an estimate range of 0.1 to 10 fold of K_d is ideal (Wilkinson, 2004). Thirdly, choosing a proper method to measure a specific PPI is critical. For different methods, equipment sensitivity at low concentration and protein aggregation at high concentration should be carefully weighed. For example, Nuclear magnetic resonance (NMR) may not be a good choice to measure a binding affinity tighter than micromolar because it needs protein at millimolar level to detect a signal(Wilkinson, 2004). However NMR is a perfect tool to measure low affinity PPIs. Another example is that gel filtration and electrophoresis cannot be used to measure PPIs with fast disassociation constants, since the equilibrium will be disturbed during the separation of proteins.

PPI thermodynamics is another critical part in understanding the nature of protein binding. The main thermodynamic parameters used to characterize PPIs include standard free energy change ($\triangle G^o$), Gibbs free energy ($\triangle G$), enthalpy change ($\triangle H$), and entropy change ($\triangle S$). The relationship between these parameters can be established as the following equations:

$$\triangle G^o = -RT\ln K_{eq}$$

$$\triangle G = \triangle G^{o} + RT\ln K_{eq}$$

$$\triangle G = \triangle H - T\triangle S$$

Where R=gas constant, T=temperature, and K_{eq} is the equilibrium constant.

Based on these interrelationships, we can determine the standard free energy change of a PPI by measuring the K_d ($1/K_{eq}$). For most PPIs, K_d falls into the range of 10^{-4}~10^{-14}M, corresponding to a binding energy range of 6~19kcal/mol (Janin, 2000).

The Gibbs free energy indicates the favorable direction of a process, and it is related to reaction temperature, changes of enthalpy and entropy. In PPI events, the enthalpy change is determined by hydrogen bond formation, electrostatic and van der Waals interactions, whereas the entropy change is decided by the freedom difference of elements in the system before and after binding, which can be further divided into conformation and association entropy change of protein partners and the entropy change of solvents (Archakov et al, 2003). Overall, if the system entropy change is the major contributor to the association of proteins, the PPI is entropy-driven. Otherwise, the enthalpy is the major contributor to the interaction (Veselovsky et al, 2002). However, in most cases the effects of enthalpy and entropy on Gibbs free energy are opposite to each other, leading to a smaller change to $\triangle G$ (Brady & Sharp, 1997).

PPI research tools

Protein interactome

Although PPIs are complex and diverse, significant progress has been achieved in high-throughput PPI screening. For some model organisms, there have been efforts to define

the “interactome”, which is a map of entire PPI network in an organism (Parrish et al, 2006). Two strategies have made great contribution to interactome studies: (i) detecting binary protein interaction with the Yeast Two-Hybrid (Y2H); and (ii) characterizing large-scale protein complex with Tandem Affinity Purification (TAP). Recently, several interactome maps based on binary interaction have been constructed, including maps from yeast (Fujimoto et al, 2001; Uetz et al, 2000), C. *elegans* (Burdine et al, 2004), Drosophila (Giot et al, 2003), and human (Rual et al, 2005; Stelzl et al, 2005). In addition to binary PPI screening, proteome-scale identification of protein complexes has also been carried out (Deane et al, 2002; Gavin et al, 2006; Gavin et al, 2002; Krogan et al, 2006), which can provide additional information about the PPI network organization. Interestingly, data obtained from these two strategies are largely complementary, which means neither of these strategies can cover the whole picture of an interactome.

Undoubtedly, the elucidation of these interactome maps from different organisms takes great advantage of genome projects. However, we should never forget the role of the technical advancement on the discovery and identification of PPIs. Especially, Two techniques are noteworthy: Y2H and TAP. It is very interesting to find that, current interactome maps mentioned above were all constructed on these two platforms: Y2H system for binary interaction mapping and TAP-MS for proteome scale complex identification.

Yeast Two-Hybrid (Y2H)

Why is the Y2H system used so extensively for interactome mapping? To reveal the answer, we have to take a close look at this technique. The Y2H system was first

introduced in 1989 (Fields & Song, 1989) and has been applied for high-throughput screening since the 1990s (Parrish et al, 2006). This method is based on the modular structure of transcriptional factors, initially represented by Gal4, which has a DNA binding domain (BD) and a transcription activation domain (AD). In principle, the two proteins (bait and prey) to be tested for interaction are fused to BD or AD and expressed in yeast cells (*eg*. BD-bait and AD-prey). BD-bait can bind the upstream of a reporter gene. If the bait and the prey interact, the AD-prey will be brought to a promoter, which controls the expression of the reporter gene. As a result, the interaction can be detected by measuring reporter activity (See Figure 1.1). The Y2H system has been improved in many aspects since it was introduced. For example, AD from VP16 of Herpes Virus, BD and AD from *E. coli* protein LexA and B42 can be utilized to replace Gal4 components (Brent & Finley, 1997). In addition, a variety of different reporter genes and promoters have been explored, such as Gal4-responsive promoters from *GAL2, GAL1, SPAL10, UASGAL1*, *MEL1* genes and reporter genes *ADE2*, *HIS3*, *LEU2* and *URA3*. More significantly, the basic design of the Y2H has been exploited to develop related methods, such as Yeast One-Hybrid system, Yeast Three-Hybrid system, Mammalian Two-Hybrid system, Bacterial Two-Hybrid system, and splitting ubiquitin system. In summary, the Y2H system inspired many technologies and has made great contribution to our understanding of biomolecular interactions.

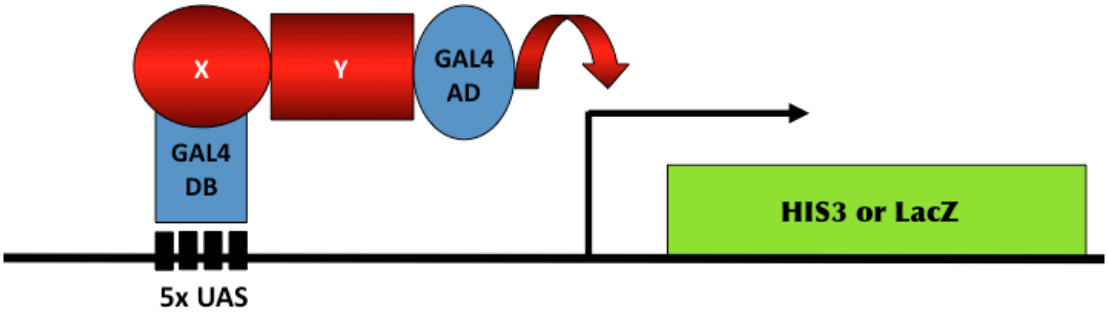


Figure1. 1 The principle of the Yeast Two-hybrid (Y2H).
The Y2H method is based on reporting yeast transcription activation to detect PPIs. The system includes two fusion proteins, bait (X) fused with Gal4 DB domain and prey (Y) fused with GAL4 AD domain, 5xUAS (upstream activating sequence in promoter), and a selection reporter (HIS3 or LacZ). Once X and Y interact, the system can activate survive or colorimetric selection system. See text for details. This figure is derived from reference (Fields & Song, 1989).

Although the Y2H system is a milestone in the development of biological research tools, it is not flawless. One limitation is that it can generate large numbers of false negative and false positive results. False negative results mean the Y2H system fails to detect some interactions. There are several situations that can cause this result. Firstly, the Y2H is based on transcription activation, but some proteins cannot fully function in the nucleus of the yeast cells. For example, membrane proteins and other proteins that require special conditions like low pH do not perform well in the Y2H assay. In addition, some proteins may contain a strong signal peptide, which will direct the protein to other organelles instead of the nucleus. Secondly, in yeast cells some posttranslational modifications, such as glycosylation and phosphorylation, cannot be recreated as those in the host cells (Van Criekinge & Beyaert, 1999). Thirdly, observable transcription activation in yeast cells requires more than 24 hours, making it difficult to detect some unstable and transient interactions. Fourthly, some proteins are toxic in yeast cells or become toxic when over-expressed at levels necessary for the Y2H detection. Lastly, some transcription repressors

can inhibit the reporter gene expression, so the Y2H system cannot be used to study them.

In addition to false negative results, the Y2H can also generate false positive results, meaning some interactions detected by the Y2H system never occur under physiological condition. Generally, the Y2H false positives can be divided into two groups: "biological" or "technical". A biological false positive is observed when an interaction is detected by the Y2H system, but the bait and the prey do not actually encounter one another in a normal host organism: they may be expressed at different times, under different conditions, in different organs or cell types, or in the same cell but at different compartments. There is no way to eliminate these false positives in the Y2H system. Technical false positives are also called "experimental" false positive results because they are generated from the procedure of the Y2H. There are several reasons accounting for this type of false positive results. Firstly, auto-activation of transcription caused by DB-bait chimerical protein can be a major issue. Some AD-prey fusion proteins can generate promoter specific false positive results. Therefore, it is important to use several reporter systems with different combination of reports and promoters (Brent & Finley, 1997; Vidalain et al, 2004). Secondly, over-expression of DB-bait and AD-prey fusion proteins may cause false positive results. Thirdly, several different AD-prey plasmids can be transformed into one yeast cell and are stably maintained during the process of Y2H, which increases the chance of misidentifying the real interaction partner (accounting for up to 42% of false positives).

False negatives and false positives have imposed significant limitations on the Y2H system. It has been estimated that about 50% or more of the PPIs detected by the Y2H high-throughput screening were not reliable (Deane et al, 2002; Fields, 2005; Sprinzak et al, 2003; von Mering et al, 2002). However, this is only part of the story. Because false negative results are “invisible”, we do not know to what degree the Y2H system fails to work. These limitations were highlighted by the first two yeast interactome papers (Fujimoto et al, 2001; Uetz et al, 2000): only ~20% of the interactions identified in both high-throughput Y2H studies were the same.

Tandem affinity purification and Mass spectrum (TAP-MS)

TAP-MS is another approach for identifying PPIs. It combines two methods: protein isolation (TAP, tandem affinity purification) and identification (MS, mass spectrometry). TAP is an advancement based on the protein detection, labeling, and purification techniques such as Co-Immunoprecipitation (Co-IP) and Pull-down Assay. Compared to other methods, the TAP can more efficiently remove non-specifically associated proteins. In principle, the TAP method introduces two "tandem" affinity tags joined by a protease cleavage site into the N or C-terminal of the target gene (bait). After the fusion protein is expressed in native cells, two sequential purification steps are applied to enrich the protein complex and remove nonspecifically bound proteins (See Figure 1.2)(Huber, 2003). Purified proteins are then separated by SDS-PAGE or Liquid Chromatography (LC) and identified by MS (Rigaut et al, 1999). It is noteworthy that during the purification steps, the protein complexes are maintained in mild wash condition, allowing isolation of active and functional protein complexes. In addition, nonspecifically associated proteins can be mostly removed after two rounds of affinity purification.

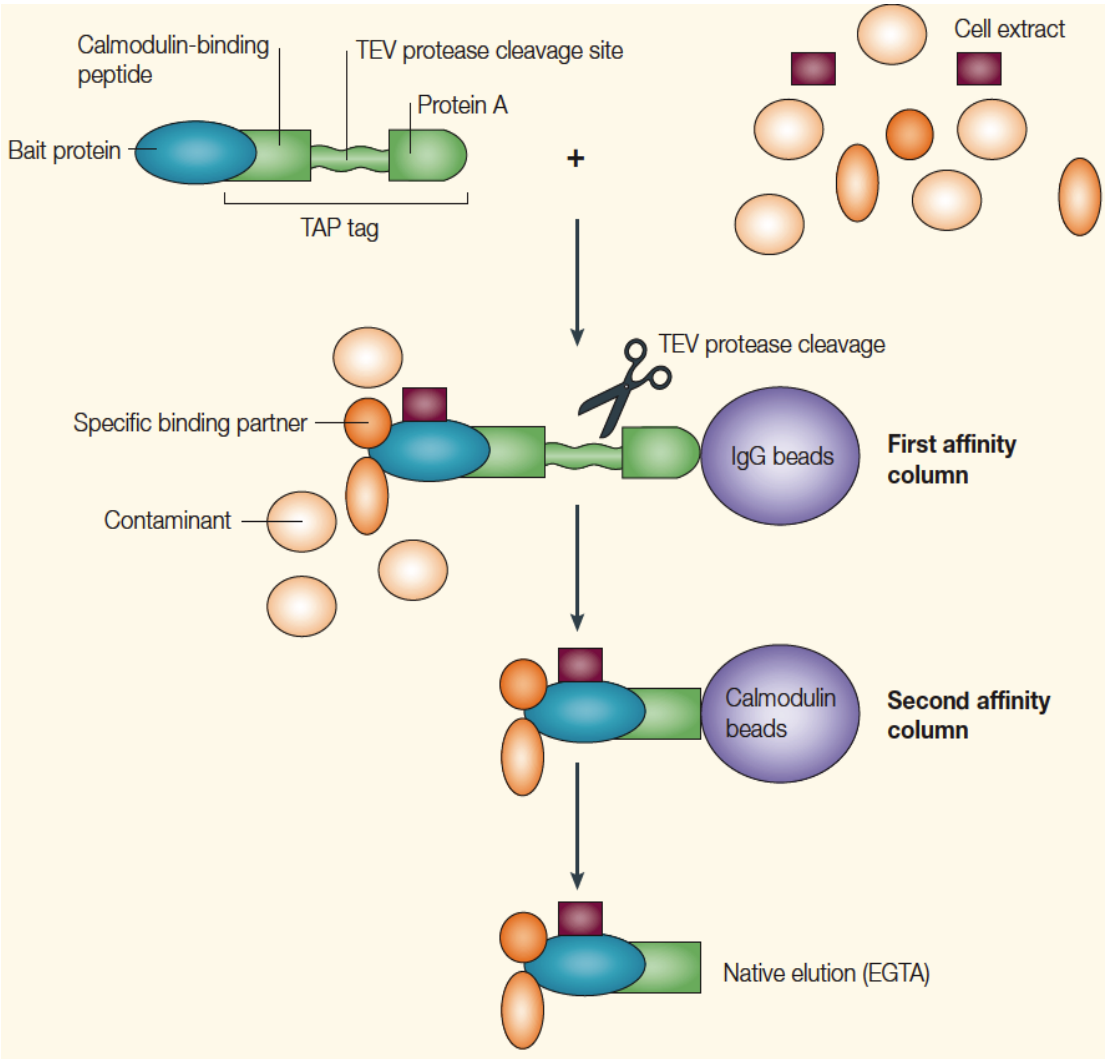


Figure 1. 2 Tandem affinity purification (TAP) for protein complex enrichment. This method uses two sequential steps to purify protein partners associated to the bait protein. The bait protein is expressed in host cells with two tags: ProtA (IgG–binding unit of Protein A) and CBP (calmodulin-binding peptide) joined by a TEV protease site. Two-step purification has two advantages: (i) they can remove much more contaminant than one step purification. (ii) the protein complex and the function can be preserved after two purifications under mild wash conditions. This figure is copied from reference (Huber, 2003)

By using tandem purification, this method reconciles the requirements for maintaining protein complex and eliminating as much impurity as possible (Kocher & Superti-Furga, 2007). Initially, Rigaut et al. tested several tags and found that a combination of ProtA (IgG–binding unit of Protein A) and CBP (calmodulin-binding peptide) tag jointed by a TEV protease recognition sequence could efficiently recover the protein complex at low concentration, so they coined the term "TAP (Tandem Affinity Purification)" (Rigaut et al, 1999). After a decade of development, this method has been successfully applied to many PPI studies (Danial et al, 2003; Khan et al, 2005; Ranish et al, 2004; Wang et al, 2006). A recent improvement for this method is to insert the whole TAP tag into the target gene locus in the host genome through site-specific DNA recombination, which allows the expression of the tagged gene under the control of its native promoter (Gavin et al, 2006; Gavin et al, 2002). The TAP-MS method has numerous advantages over other PPI identification methods: (i), only one component of the complex needs to be labeled, which minimizes the interference to the formation of protein complex. (ii), the interaction can happen at any location inside the cell, eliminating problems caused by differential protein localizations. (iii), this method can identify multiple protein partners in one experiment, which not only increases the efficiency but also provides information on the organization of protein interaction network and greatly facilitates the construction of PPI map. (iv), both purification steps are conducted under mild wash conditions, which is critical for preserving protein complexes and removing nonspecifically associated proteins. (v), cutting-edge LC and MS instruments can be exploited to increase the sensitivity, accuracy and throughput of the method.

However, the TAP-MS strategy still has some limitations when applied to high-throughput PPI screening, which has been indicated by the dramatically different datasets obtained from two similar studies (Goll & Uetz, 2006). There are several reasons underlying these differences. (i), protein complexes involve non-covalent association of multiple proteins. If subjected to extensive wash during purification, some complexes may dissociate. (ii), Although current protocol uses two mild washing steps to remove nonspecifically associated proteins, it is not sufficient to remove all of them (Gavin et al, 2006). (iii), this method can only be used to identify PPIs with an affinity higher than 50nM (Piehler, 2005). (iv), it can provide a snapshot of the interactions but cannot reveal the dynamic nature of biological events. (v), in most applications, purified proteins were separated, stained and digested in SDS-PAGE, which may become the limiting factors for detecting low abundant PPIs, because protein in-gel digestion efficiency is quite low (~20%) (Kocher & Superti-Furga, 2007),.

Additional methods for PPI research

Although the Y2H and TAP-MS are widely used for PPI detection, many other methods are also available, such as Fluorescence Resonance Energy Transfer (FRET), Bioluminescence Resonance Energy Transfer (BRET), Protein-fragment Complementation Assay (PCA), Phage Display, Far Western, *in vivo* cross-linking, Nuclear magnetic resonance (NMR), Surface Plasma Resonance (SPR), Atomic Force Microscopy (AFM), Analytical Ultracentrifugation, protein arrays, single-molecule imaging and bioinformatic tools. These methods can meet different criteria for diverse applications, each taking advantage of its unique feature (see table 1.1 for a summary).

For example, FRET is a widely used method to detect PPIs in cells. Its derivative, BRET, replaces the fluorescent donor with bioluminescent protein. So, the signal can be observed without using laser as an excitation light source, which significantly reduces the background compared with the FRET. Another modification of the FRET is the FLIM (Fluorescence Lifetime Imaging), which can measure the energy transfer efficiency, information that can be used to estimate the distance between two interacting proteins.

Table 1. 1 Features and applications of different techniques for protein interaction analysis*

Technique	Application	Identification	Localization	Affinity	Kinetics	Throughput	Membrane proteins	Kd range
Yeast two-hybrid	In vivo	++	–	–	–	++	–	<50 μM
Protein fragment complementation	In vivo	+	–	+/–	+/–	+	+	<10 μM
Tandem affinity chromatography	In vitro	++	+/–	–	–	++	+/–	<50 Nm
Protein arrays	In vitro	+	+/–	+	+/–	++	–	<1 μM
FRET/BRET	General	+/–	++	+	+/–	+/–	+/–	<10 μM
Solid-phase detection	In vitro	+/–	–	++	++	+	+	pM-μM
Single-molecule detection	General	–	++	++	++	+/–	+	nM-μM

*This table is copied from reference (Piehler, 2005)

PPI tools of interest

Crosslinking and Photoaffinity labeling

Chemical crosslinking, a process of chemically joining two molecules by a covalent bond, is a conventional tool for PPI analysis (Burdine et al, 2004). This method has long been used in biochemical studies to investigate the association of proteins, DNA/RNA and small molecules. Important chemical crosslinkers for PPI research include the general crosslinker formaldehyde and some bi- or tri- function crosslinkers.

Formaldehyde is known for fixing biological materials to keep their natural shape and position. For example, it has been used in Chromatin Immunoprecipitation (ChIP),

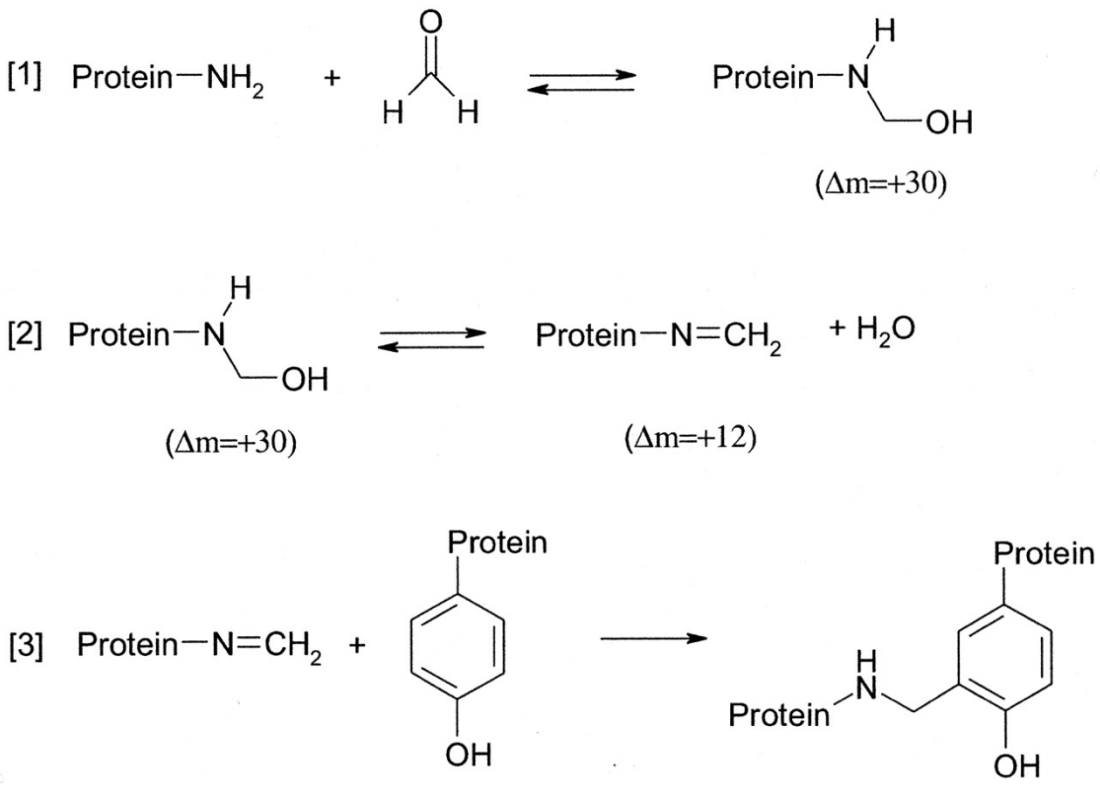


Figure1. 3 Mechanism of formaldehyde-based protein crosslinking.
The reaction is divided into three steps: [1], the formation of mythylol adduct; [2], dehydration; [3], meththylene bridge to the Schiff-base. A tyrosine residue is used in step [3] as an example. See text for detainsl. This figure is copied from reference (Metz et al, 2004).

Fluorescence in situ hybridization (FISH) and vaccine inactivation. Although formaldehyde's function is very clear, its chemical mechanism is very complicated. After decades of intensive investigation (Fraenkel-Conrat & Olcott, 1948; Metz et al, 2004), the mechanism has now started to be uncovered. When introduced into a biological system, formaldehyde works initially through the formation of methylol adducts on amine groups. Subsequently, these adducts can be partially dehydrated to give labile Schiff-base. Finally, a stable product is formed by crosslinking amino acid residues via methylene bridge to the Schiff-base (See Figure 1.3)(Metz et al, 2004). Before the introduction of the TAP method, formaldehyde crosslinking was a major tool for the

characterization of protein complexes. In recent years, however, its application on PPI detection has focused on analysis of transient and low affinity interactions, because pull-down or Co-Immunoprecipitation cannot purify loosely associated proteins without fixation (Guerrero et al, 2006; Schmitt-Ulms et al, 2004). Based on the mechanism of formaldehyde crosslinking, it is not surprising that highly heterogeneous crosslinking products can be generated if it is used for PPI studies. Actually, heterogeneous crosslinking has become the major hurdle for the detailed characterization of PPI with formaldehyde based crosslinking method, because large amounts of non-specific crosslink products can produce false positive results, making it difficult to identify true interactions. The situation can be further complicated for those proteins with heavy post-translational modifications(Borch et al, 2005; Vasilescu et al, 2004).

In addition to direct application of simple crosslinkers such as formaldehyde on PPI characterization, multiple functional crosslinkers have also been developed, of which the label transfer method is a very good example. Label transfer reagent was initially developed with heterobifunctional crosslinkers (e.g. a photo-activable moiety and an electrophile), that is, two crosslinkers joined by a cleavable linker like disulfide, azo or ester. When used for PPI detection, a heterobifunctional crosslinker is first tethered to a bait protein via the electrophile group at one end. Subsequently, the bait-crosslinker complex binds to prey protein, and ultraviolet light is used to trigger labeling the prey via the photo-activable moiety. Lastly, the cleavable linker is severed, realizing the transfer of the label to the prey protein(Fancy, 2000). In its early stage, label transfer strategy usually employed a radioactive group, such as in ^{125}I-derived label transfer compounds,

to report the signal (Koch et al, 1994; Schwartz, 1985). However, after decades of improvement many report groups have been used to enhance this method for PPI detection, such as fluorescent compounds and biotin. More than that, the overall design of the label transfer has also evolved, and many new techniques have been applied to revive this conventional method for PPI research. Recently, a brand new label transfer design using biarsenical FlAsH reagent to label bait protein has drawn a great deal of attention (Souza, 2007). In this design, the biarsenical FlAsH was combined with biotin and DOPA(Griffin et al, 1998), which could be oxidized by sodium periodate to generate a reactive ortho-quinone to crosslink a protein nearby. In this way, FlAsH moiety could be site specifically installed to the bait protein through an engineered FlAsH receptor peptide (FRP, CCPGCC). Based on this design, non-specific loading of label transfer reagent to unrelated proteins could be avoided and the label transfer efficiency could be increased significantly. In the report, Liu et al demonstrated the method by using a model system of 26S proteasome complex(Liu et al, 2007).

Although many crosslinking methods have been used for PPI study, non-specific and uncontrollable crosslinking makes it difficult to identify the targets of interest. To meet the sophisticated requirement for PPI characterization, photoaffinity labeling was developed, which features activating crosslinker by light as a remote control. Photoaffinity labeling was first introduced in 1962 (Singh et al, 1962), after so many years, the general principle still holds: upon light irradiation, the crosslink compound is activated and a covalent bond is formed to couple target molecules in the proximity (see Figure 1.4) (Jahn et al, 2004). Combined with cutting-edge analytical tools like LC/MS,

this method is becoming even more useful (Jahn et al, 2004). Although recent years have seen the applications of many photophores for photoaffinity labeling, most of them are derivates of three compounds: nitrene generated from acryl azides, carbene generated from diazirines, and photoreactive carbonyl group generated from benzophenone (Hatanaka & Sadakane, 2002). Among these leading photophores, acryl azides and diazirines are not very stable in ambient light. Because they can quickly react with a water molecule, they are not very efficient in photoaffinity labeling. On the other hand, benzophenone is preferred to acryl azides and diazirines because it is more stable to ambient light and protic solvents (Dorman & Prestwich, 2000; Jahn et al, 2004). However, the activated intermediate of benzophenone can randomly insert into proximate C-C or C-H bonds, leading to a huge population of heterogeneous crosslinked species that are difficult to characterize. Above all, current photoaffinity labeling systems are more or less defective. To address the challenges from PPI characterization, we need more powerful photoaffinity labeling tools that are more stable, controllable, and specific, generating more predictable result.

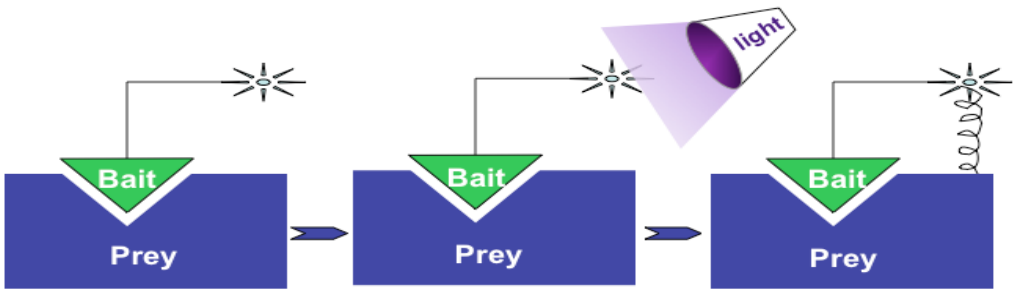


Figure 1. 4 General principle of the Photoaffinity labeling method.
The system contains the prey protein and the bait protein loaded with a photo-sensitive crosslinker. Upon light irradiation, the crosslinker can be activated and label the prey protein. Consequently, the prey protein can be enriched and identified. This figure is adapted from reference (Dorman, 2000).

β-lactamase based PCA

PCA is another interesting method for PPI characterization. It is based on the similar strategy to the Y2H. In the PCA system, a gene encoding an enzyme is firstly split into two fragments (for example, N and C terminal) and fused with a bait gene and a prey gene respectively. These fusion genes are then expressed in target cells to generate fusion proteins. If the bait and prey interact to form a complex, the two fragments of the enzyme will be brought into proximity. They will refold and become an active enzyme. Subsequently, by detecting the enzyme activity the interaction between the bait and the prey can be reported (See Figure 1.5) (Michnick, 2001). The difference between the PCA and the Y2H is in the reporter system. For the PCA, reporter itself is split; For the Y2H, a reporter system has to be introduced. Besides, the PCA allows the detection of PPI in its native compartment and original cells. However, in the Y2H, the interaction can only be detected in the nucleus of yeast cells. For the PCA system, two critical points must be noted. Firstly, PCA fragments should not refold spontaneously, otherwise, the system simply will not work (Michnick, 2001). Secondly, PCA is a switch-like system and it is an all-or-none process (See Figure 1.5c). In other words, the facilitated folding state of PCA fragments should be favored over any other intermediate forms if they are brought together (Michnick et al, 2007). Experimental results have shown that, in a PCA system, the fraction of refolded enzyme increases cooperatively as the ratio between the two fragments increases, resulting in a high dynamic range of reporter activity over a narrow range of protein ratio (See Figure 1.5c)(Galarneau et al, 2002; Michnick et al, 2000). Currently, there are several PCA systems available based on different reporters, such as DHFR (Dihydrofolate reductase), β-galactosidase (natural complementation between the

α and ω domain mutations), β-lactamase, Gaussia Luciferase and GFP (Michnick et al, 2007). Notably, PCA systems based on the complementation of fluorescent proteins are also called BiFC (Bimolecular Fluorescence Complementation). Because fluorescent protein based BiFC systems are not reversible, they have both advantages (for example, detecting low and transient interaction) and disadvantages (unable to show protein interaction dynamics)(Magliery et al, 2005; Remy & Michnick, 2006).

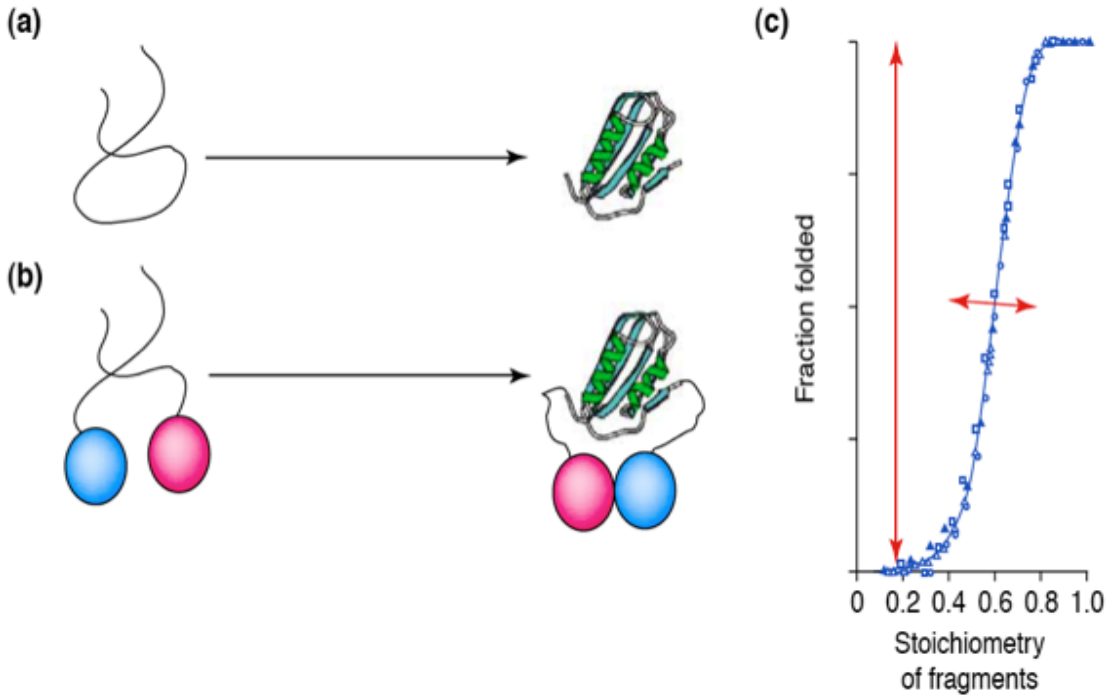


Figure 1. 5 Principle of the PCA.
This method is based on splitting protein reporter to detect PPIs. It was found that the refolding of reporter fragments (green protein) in a PCA facilitated by the interaction between the bait and prey (red and blue ball) in (b) has no difference from the natural process of protein refolding showed in (a). Moreover, it is demonstrated that PCA is an all-or-non process and determined only on the stoichiometry ratio of the two fusion proteins (c). This figure is copied from reference (Michnick, 2001 and 2003).

Among reporters used in the PCA system, β-lactamase (EC 3.5.2.6) is a very interesting one. β-Lactamase belongs to a group of enzymes that have evolved from penicillin-binding proteins (PBPs)(Davies et al, 2001). These enzymes can hydrolyze the β-lactam ring in penicillin antibiotics. Of all β-Lactamases, TEM-1 is very unique and has numerous merits, which make it a good choice for the PCA system. Firstly, TEM-1 β-lactamase is very efficient at hydrolyzing penicillin and cephalosporin, with rates close to the diffusion limit (10^8~$10^9 M^{-1}$ sec^{-1}) (Christensen et al, 1990; Minasov et al, 2002). Secondly, it is a small and monomeric enzyme, with a molecular weight of 29 KD. Thirdly, its structure has been well characterized. Fourthly, it is not toxic to mammalian cells, and there is no ortholog in eukaryotic organisms (Galarneau et al, 2002). Fifthly, many TEM-1 substrates have been developed for different applications. For example, Nitrocefin can be used for *in vitro* colorimetric assay to measure TEM-1 activity (Galarneau et al, 2002). Cell permeable fluorescent substrate CCF2/AM can be used for *in vivo* imaging (Zlokarnik et al, 1998). At the same time, Penicillin can be easily modified to make substrates with new properties, for example, biotin-penicillin and fluorescein-penicillin. Lastly, the catalytic mechanism of TEM-1 β-lactamase has long been unveiled (Figure 1.6). β-lactamase is a serine-based hydrolase and Ser70 is at the activity center. During the reaction, the enzyme and substrate form an acyl-enzyme intermediate through Ser70 (Dalbadie-McFarland et al, 1982; Fisher et al, 1980). Notably, β-lactamase E166N mutation can trap the substrate penicillin to the enzyme by stabilizing the acyl-enzyme intermediate (Adachi et al, 1991; Minasov et al, 2002). Above all, TEM-1 β-lactamase has many advantages, which are very crucial for its further development.

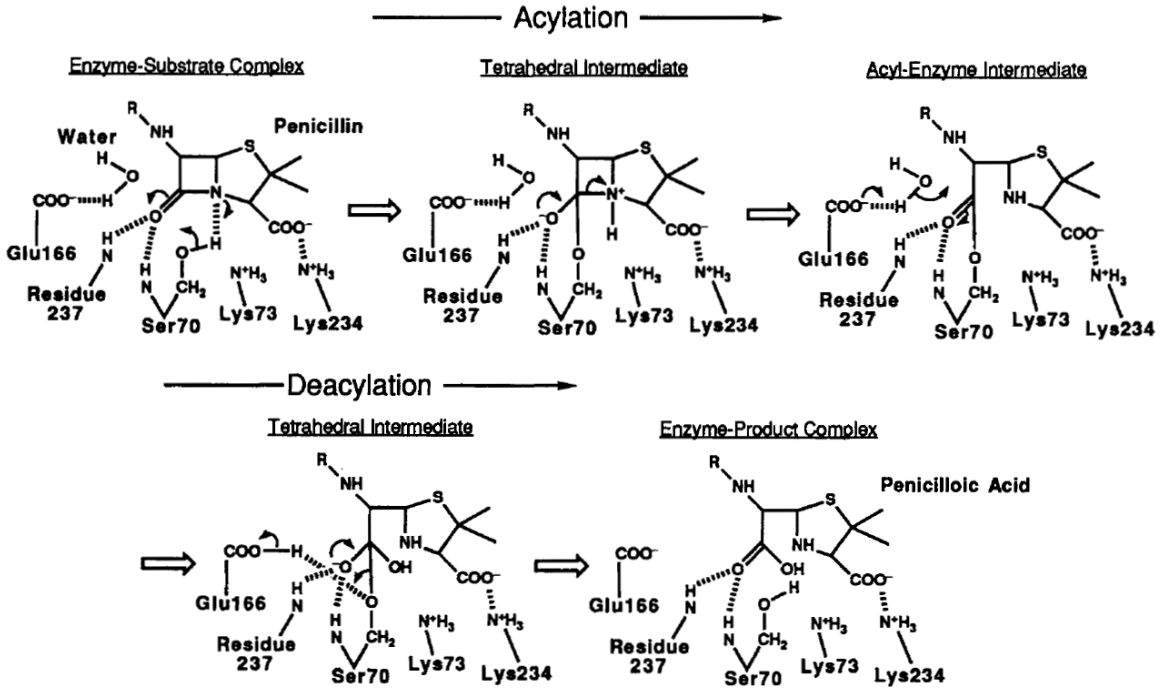


Figure 1. 6 Catalytic mechanism of TEM-1 β-lactamase.
TEM-1 β-lactamase is a serine-based hydrolase and works through two reaction steps: acylation and deacylation. Once the enzyme forms a complex with the substrate, they become a tetrahedral intermediate through Serine 70 residue, followed by the breakage of β-lactam ring and the formation of acyl-enzyme intermediate. In the presence of a water molecule activated by Glutamic acid residue at position 166, the acyl-enzyme intermediate is disrupted and the bond between the enzyme and substrate is broken. Subsequently, the reaction goes through deacylation and the enzyme is recycled for next round of catalytic reaction. This figure is copied from reference (Adachi et. al., 1991).

TEM-1 β-lactamase based PCA was developed by two groups at almost the same time (Galarneau et al, 2002; Wehrman et al, 2002). Each provided us a unique angle to understand this method. Wehrman et al. split the TEM-1 into two fragments: α197 (amino acid 25~197) and ω198 (amino acid 198~288). To test the idea of PCA, they fused these two fragments to several pairs of interacting proteins, including cJun and cFos, CD40ED and BW10-1, FKBP12 and FRB (with Rapamycin). They demonstrated that with different interacting pairs, TEM-1 β-lactamase based PCA could work in both *E. coli* and mammalian cells. They also showed that this method could be used to detect membrane and transient PPIs. Remarkably, they found that the signal-to-noise ratio of the method could increase by as many as 1,000,000 folds by introducing the tri-peptide NGR into the joint of α197-cJun, which probably works by stabilizing the α197 fragment and consequently the refolded β-lactamase (Wehrman et al, 2002). Similarly, Galarneau et al. dissected TEM-1 β-lactamase between Gly196 and Leu198 based on the structure. They also tested this splitting strategy with several known PPIs, including homodimerization of GCN4 leucine zipper domain (ZIP), heterodimerization of Bad and truncated Bcl2T, homodimerization of Smad3, and interaction between PKB and its substrate Bad. During their development, they introduced M182T mutation in TEM-1 β-lactamase to increase TEM-1 enzyme activity. Moreover, they made use of the interaction between FKBP and FRB triggered by rapamycin to titrate the refolding kinetics of the split fragments. Overall, these two pioneering studies introduced TEM-1 β-lactamase PCA to PPI research community. After that, several follow-up investigations were carried out(Schnee et al, 2006; Spotts et al, 2002). Now, TEM-1 β-lactamase PCA has been applied to many

PPI studies(Jun & Wickner, 2007; Lee et al, 2007; Lee et al, 2004; Park et al, 2007), making it an ideal candidate for further development.

Questions and aims

Even though technologies for PPI detection are fast evolving, the demand for more robust PPI study tools is also surging. Because current technologies cannot solve many problems, we need new tools for PPI study. However, many challenges need to be addressed for the development of a new PPI characterization tool. For example, Can it be used *in vivo*? Can it detect low affinity and transient interactions? Is it effective for low abundant proteins? Does it work for membrane proteins? Is it repeatable? Can it be easily scaled up for high-throughput application? To meet these practical requirements, we need to reexamine our current tools and integrate the power of different methods.

Based on previous discussion, we concluded that both PCA and crosslink methods could be adapted for many challenging applications. They were thus chosen for further development. The overall goal throughout this research was to label interacting target proteins with a strong covalent bond. A possible way to achieve this goal is to combine the advantages in both top-down and bottom-up strategies. Specifically, effort will be paid to integrate PCA, protein labeling and crosslinking, pull-down and MS to hit the target for better PPI study tools.

Chapter two,

Photo-generated Quinone Methide as an Affinity Crosslinker for Characterizing Protein-peptide Interaction

Abstract

Protein photoaffinity labeling is a useful method for PPI analysis. However, conventional photoaffinity labeling agents, such as carbene, nitrene and benzophenone are impaired by their limited photo-stability, low labeling efficiency and poor reaction specificity. In this study, a new photoaffinity labeling agent was developed through the photo-generation of a reactive intermediate, para-quinone methide (pQM). Photochemical analysis showed that two related photocaging groups, PC-1 and PC-2 could be effectively removed from the pQM precursor by UV light at 365nm. PC-2 could even be removed by visible light at 400nm. Reactivity characterization of the agent revealed that several amino acid residues could be labeled once they were in close proximity to the pQM. Protein labeling experiment, using a PPI pair catPTP1Bm and EGFR peptide carrying the pQM precursor (pep-1), found that the target protein could be specifically labeled at a concentration as low as 0.1nM in the presence of a thiol scavenger group. Moreover, MS mapping experiment indicated that two peptides from the target protein were labeled by the photoaffinity functional group, which further confirmed the crosslink specificity of this method. In addition, the MS experiment also showed that this method could greatly simplify the procedure of identifying crosslinked sites. In summary, it was demonstrated that the photocaged pQM could be a useful tool for the characterization of protein-peptide interactions.

Introduction

Photoaffinity labeling is a useful tool for the characterization of protein-ligand recognition or protein-protein association, especially for those with low affinity or short lifetime. This method features labeling the target protein with a covalent bond and can be divided into several stages. It starts with the interaction between the target protein and the bait probe. A light signal is then used to convert the crosslinker precursor to a short-lived but highly reactive intermediate, followed by the formation of a covalent bond bridging the bait probe and the target protein. In the end, the labeled target protein can be isolated and characterized by LC-MS/MS (PlaW, 1997). As we can see from this procedure, a photoaffinity labeling system contains these important components: a light harvesting structure, a cross-linker precursor, a bait probe, and a labeling marker.

Photocaging groups

The light harvesting structure plays a critical role in the photoaffinity labeling strategy. For traditional radical-based photo-crosslinkers, the light harvesting moiety and the crosslinker precursor are integrated into one group. However, it has been found that some specialized light harvesting groups, called photocaging groups, can be more efficiently removed by light at certain wavelengths. If these groups are applied to protect a chemical crosslinker, very likely, the photoaffinity labeling system will be more efficient. The desired properties for a photocaging group include chemical stability before photoactivation and rapid photolysis rate at a longer wavelength (Chowdhry & Westheimer, 1979). Among photocaging groups used for controlling biological

processes, nitrobenzyl derivates have been shown to be efficient, reliable and versatile in many biological applications (Young & Deiters, 2007). As an example, photocaged amino acid 4,5-dimethoxy-2-nitrobenzylserine (DMNB-Ser), an unnatural amino acid, has been genetically introduced into yeast cells to control protein phosphorylation (Lemke et al, 2007). In our study, nitrobenzyl derivates were chosen as our photocaging candidates (See Figure 2.1)

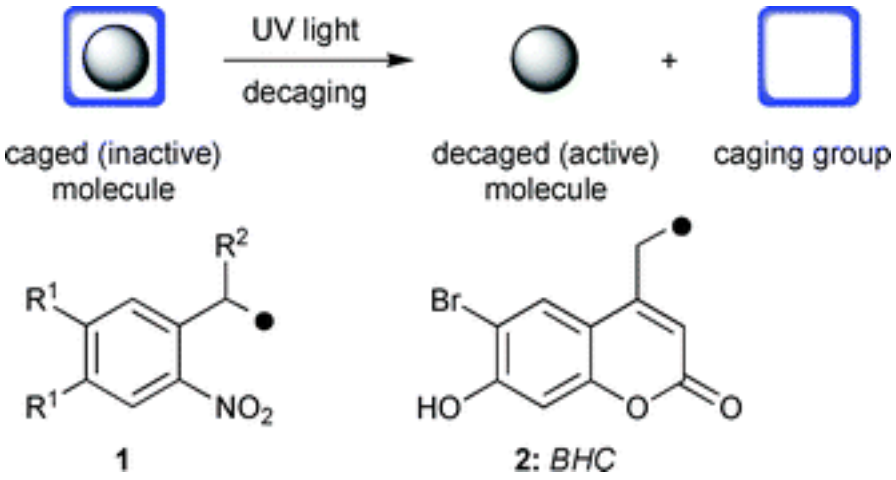


Figure 2. 1 Photocaging strategy and some caging groups.
The concept of photocaging (upper) and typical photocaging groups (lower). Substituent on 1: R1 = H, R2 = H: o-nitrobenzyl (ONB); R1 = OCH3, R2 = H: 4,5-dimethoxy-2-nitrobenzyl (DMNB); R1 =OCH3, R2 = CH3; (4,5-dimethoxy-2-nitrophenyl)ethyl (DMNPE). 2:6-bromo-7-hydroxycoumarin-4-ylmethyl (BHC). This figure is copied from reference (Young and Deiters, 2007).

Chemical intermediate pQM as a crosslinker

The most critical component in a photoaffinity labeling system is the crosslinker. As mentioned before, conventional photoaffinity crosslinkers have limitations in their reactivity, stability and efficiency. It might therefore be necessary to introduce new crosslinkers into the photoaffinity labeling toolkit to broaden the application of this method. Recently, chemical intermediate quinone methides have been used as

crosslinkers for the characterization of protein interactions. For example, they have been found to be efficient protein crosslinkers when activated by enzymatic reactions, such as dephosphorylation (Deane et al, 2002; Komatsu et al, 2006).

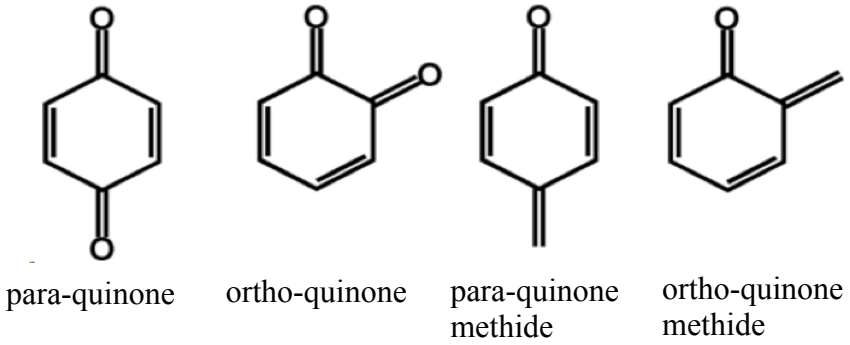


Figure 2. 2 Simple examples of quinone and quinone methide.
Quinones have two carbonyl oxygen groups in the six-member ring with two isoforms: ortho-quinone or para-quinone. Quinone methides have a carbonyl oxygen group and a methylene group also with ortho- and para-isoforms.

Quinone methides have a close relationship to quinones, which are a class of chemicals naturally found in organic world and widely used in chemical synthesis (see Figure 2.2). Quinone methides are different from quinones in that a methylene group and a carbonyl group occupy two positions on a cyclohexadiene ring. Based on the position of these two functional groups, the quinone methides have two isoforms: ortho- and para- quinone methides (Figure 2.2). Because quinone methides have two different functional groups, they are highly polarized and more active than their parental quinones. However, they are not as active as quinonedimethane, in which both carbonyl groups in quinone are replaced by methylene (Rokita, 2009a). At the same time, quinone methides have a dipolar nature and contain both cationic and anionic centers. They can react with both

nucleophiles and electrophiles to generate an aromatic phenol ring, which is a driving force for their reactivity (Chiang et al, 2000). The extensive exploration of quinone methide chemistry has established a solid foundation for the application of these compounds, and thus they were used as the crosslinker in our photoaffinity labeling method.

A new design of photoaffinity labeling method

In order to develop a new photoaffinity labeling method for characterizing protein-protein or protein-peptide recognition, we planned to use a new combination of different components in the photoaffinity labeling system, which featured following elements: Firstly, we adapted photocaging group nitrobenzyl derivates to protect the crosslinker. Secondly, we chose para-quinone methide as the crosslinker for a more specific labeling. Thirdly, we used a PPI pair between the PTP1B trapping mutant (catPTP1Bm, core domain with C215S mutation) and its binding peptide from EGFR (988–998) to demonstrate the advantage of the new system. Lastly, we added a biotin to the peptide and joined them to the crosslinker by an acid-labile linker. Over all, our aim was to develop a system that would follow the general principle of photoaffinity labeling strategy but be more sensitive, stable and specific.

In summary, a new photoaffinity labeling probe was designed, which had nitrobenzyl derivate as photocaging group, pQM as crosslinker, an acid cleavable biotin tag, and a bait peptide to target prey protein catPTP1Bm. Based on the design, UV light would be applied to remove the photocaging group and induce the formation of an active

crosslinker pQM. This reactive intermediate would then be attacked by the nucleophile group on catPTP1B surface to form a covalent bond. Next, the crosslinked target protein would be enriched and analyzed by MS.

Materials and methods

Chemicals and compounds

All chemicals, solvents and reagents used in this research are ACS Reagent Grade or Molecular Biology Grade, and they were purchased from Sigma, Fisher Scientific, EMD Merk or Acros Organic. Besides, some compounds were synthesized by Dr. Dexing Zeng in Li lab, and they were examined by MS and NMR to confirm their structure and purity. These compounds included: PC-1, PC-2, the template system to test amino acid activity to pQM and pep-1 used in protein crosslinking experiment. For detailed information about chemical synthesis of these compounds, please check reference (Jiang et al, 2009)

Determining the absorbance and photolysis efficiency of PC-1 and PC-2

Compound PC-1 and PC-2 were dissolved in 50mM HEPES buffer pH 7.2 with 30% acetonitrile. The absorbance was scanned with FluoroMax-2 Spectrofluorometer (ISA, Inc.). For photolysis experiment, 100μL solution was exposed to UV light (365 nm, Intensity = 4 mW/cm2). Repeat experiments were conducted for variable durations, and product from each experiment was then analyzed by HPLC. The percentage of remaining material in each fraction was plotted against exposure time to obtain photolysis efficiency dataset.

Measuring the reactivity between quinone methide and amino acids

A serial of nine compounds with nine selected amino acids in the one template were used for this experiment. See reference for synthesis and structure information (Jiang et al, 2009). Each compound was dissolved in a freshly prepared mixture of 50mM HEPES buffer (pH 7.2) and 50% acetonitrile to 100 μM, and then tested with two conditions. Under condition A, the compound was deprotected with UV light (365 nM, 4 mW/cm^2, the UV intensity was determined by Traceable^{TM} UV Light Meter, Fisher Scientific Inc.) for 5 min. 10 mM β-mercaptoethanol (βMe) was then added to stop the reaction. The final products were analyzed by HPLC with a detector at 473 nm. Under condition B, the compound was deprotected with UV light (365 nM, 4 mW/cm^2) for 5 min in the presence of 1 mM βMe. The final products were also analyzed by HPLC (gradient: from 10% acetonitrile in water to 90% acetonitrile in water in 30 min). All sample peaks were subsequently collected from HPLC and analyzed by MS.

Gene mutation and plasmid construction

CDS of PTP1B catalytic domain (amino acid 1~321) was amplified from a plasmid containing human PTP1B gene (Invitrogen, clone ID 4338750) using the high fidelity amplification kit (Bio-Rad, Cat. No.172-5301) with primers OSL038 and OSL041. This PCR product contained a SacII site at the N-terminal and a stop codon followed by a BamHI site at the C-terminal. Complementary primers OSL042 and OSL043 were then used to introduce the mutation C215S in the PTP1B catalytic domain. The mutation procedure included two rounds of PCR. OSL038 and OSL043 were used to amplify the N-terminal of the catalytic domain and generate fragment-1. OSL042 and OSL041 were used to amplify the C-terminal of the catalytic domain and generate fragment-2. After

fragment-1 and -2 were purified with agrose gel, they were combined as the template for another round of PCR with primers OSL038 and OSL041. The PCR product was purified by Qiaquick column (Qiagen), digested with SacII and BamHI (NEB). The expression vector was derived from pET11c(+) (Novagen). Briefly, a DNA fragment containing His-tag followed by a synthetic human ubiquitin gene and a SacII restriction site was amplified and cloned into the pET11c through NdeI and BamHI sites to replace the T7 tag in the vector. In the end, the PTP1B catalytic domain with C215S was cloned into the expression vector through SacII and BamHI sites.

Protein expression and purification

catPTP1Bm expression plasmid was transformed into BL21 (DE3) cells (Stratagene) and selected on a LB plate containing 50 μg/mL ampicillin. After growing overnight at 37°C, a single colony was picked and inoculated into 5 ml LB with 100 μg/ml ampicillin as seed culture. After growing overnight at 37°C, 1 ml seed culture was inoculated into 100 ml fresh LB with 100 μg/mL ampicillin, and grew at 37°C until OD600 reached ~0.6. IPTG was then added to induce the gene expression at a concentration of 1 mM. After growing 3 hours at 37°C, cells were cooled to 4°C and collected through a centrifugation at 10, 000 g.

Cells from 100 ml culture were lysed for 15 min at room temperature by adding 2 ml Bugbuster plus 2 μL of Benzonase (Novagen). Insoluble cell debris was removed by centrifugation at 45 000 g for 30 min at 4°C. The supernatant was mixed with 10 ml buffer A (50 mM Tris·HCl, pH 8.0; 500 mM NaCl; 10 mM imidazole), loaded onto 1 ml

(bed volume) pre-equilibrated Ni-column (Bio-Rad), and allowed to bind with the column at 4°C for 1 hr on a rotator/shaker. The column was washed extensively with 10 ml buffer A, and 50 ml buffer A containing 30 mM imidazole, then eluted with 10 ml buffer A with 400 mM imidazole. Ten fractions were collected and 15 μL of each fraction was loaded onto SDS-PAGE to determine which fraction contained target protein. Protein concentration was determined by the Bradford assay following the manufacturer's instruction (Sigma).

Photocrosslinking with different DTT concentrations

Purified catPTP1Bm and cell lysate prepared from strain BL21 DE3 were first dialyzed against deprotection buffer (50 mM HEPES, pH 7.2, NaCl 150 mM) overnight at 4 °C, and their concentrations were determined by the Bradford assay (Sigma). For crosslinking reaction, different concentration of DTT (0 nM, 100 nM, 1 μM, 10 μM, 0.1 mM, 0.5 mM, 1 mM, 5 mM) and 100 μg bacterial cell lysate were mixed in 100 μL solution with 10 nM catPTP1Bm and 1 μM pep-1. The deprotection was carried out with FluoroMax2 instrument for 5 min at 365nm (4 mW/cm2). Samples were then concentrated with SpeedVac and loaded onto a SDS-PAGE gel, followed by western blotting to detect crosslinked protein. After the membrane was blocked overnight in 5% milk in PBS with 0.05% Tween-20, strepavidin-HRP conjugate was added to the solution and incubated for 1 h at room temperature. The membrane was washed with TBS (10 mL, 20 mM Tris buffer, pH 8.0, 150 mM NaCl) for 3 times, 5 min each time. Subsequently, HRP substrate was added, and the chemiluminescence signal was recorded by a gel documentation system.

Photocrosslinking with pep-1 under different concentration of catPTP1Bm

The procedure was the same as that described above, but DTT concentration was kept at 100 μM and catPTP1Bm concentration changed (0.1 nM, 1nM, 5 nM, 10 nM, 25 nM, 100 nM, 1 μM) in each reaction.

Identification of crosslinked peptides with MS

After photocrosslinking reaction between catPTP1Bm and pep-1 peptide, 6 M urea and 10 mM DTT were added to denature and reduce the protein, then 55 mM bromoacetamide was added to the mixture and incubated at 37 °C for 2 hours in darkness. Excess DTT was added again to quench the unreacted bromoacetamide. Next, the sample was dialyzed overnight using 10KD MWCO dialysis cassette (Pierce). This step removed both small molecules and excess pep-1 in the sample. The sample was then mixed with 4 volumes of pre-chilled acetone, incubated at -20 °C for 1 h, and precipitated by a centrifuge at 45, 000 g for 20 min. The pellet was dissolved in 200 μl 6 M urea, 100 mM NH_4HCO_3, and the solution was diluted with 1.8 ml 100 mM NH_4HCO_3. 10 μg of trypsin (Sigma, proteome grade) was added to digest the protein for 20 h at 37 °C. After digestion, trypsin was removed by incubating with immobilized trypsin inhibitor (Sigma). Then, the biotinylated peptide was enriched with monomeric avidin agrose by following manufacture's instruction (Pierce). The purified peptides were treated with acid cocktail (88% TFA, 5% H_2O, 5% phenol, 2% Triisopropylsilane), dried with lyophilizer, mixed with MS matrix (10mg/ml α-cyano-4-hydroxycinnamic acid in 50% acetonitrile, 0.1% TFA) and examined by MS instrument (ABI 4700 Proteomics Analyzer).

Result

Characterizing two photocaging groups derived from nitrobenzyl moieties

Photocaging group in our photoaffinity labeling design is critical because it can affect photolysis efficiency, crosslinking activity and probe stability. We decided to use nitrobenzyl derivates as our photocaging groups, and two pQM precursors PC-1 and PC-2 with two slightly different nitrobenzyl species were synthesized (See Figure 2.3).

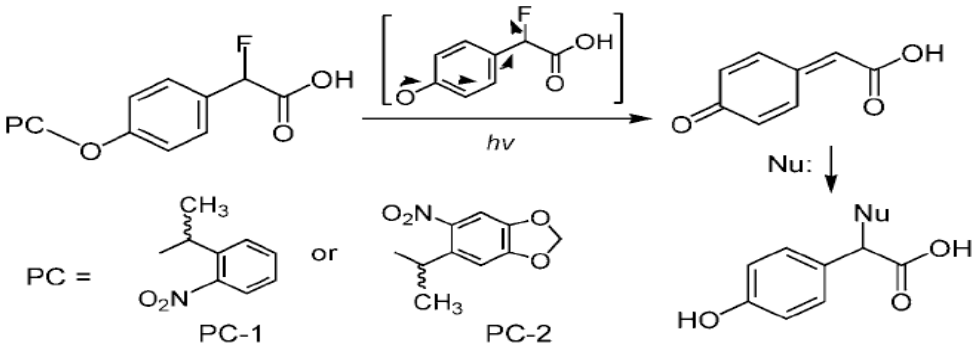


Figure 2. 3 Photocaging groups PC-1 and PC-2.
Two photocaging groups PC-1 and PC-2 were derived from nitrobenzyl group. They were installed to protect pQM. During application, they could be deprotected to label the nucleophile groups in a protein.

In order to carry out photoaffinity labeling, a light source is used to deliver energy to the probe. Depending on their different photochemical properties, photocaging groups can be photolysed by light with different wavelength. Currently, most photoaffinity labeling studies use high energy UV or Flash Laser Photolysis (FLP) to activate crosslinking. The major problem associated with these methods is that high-level energy during photolysis may cause significant damage to biological molecules. As an initial step of our study, we characterized our photocaging groups and compared their photochemical properties, so that we could find an optimal photocaging group from PC-1 and PC-2. An absorbance

scan for both PC-1 and PC-2 was firstly conducted. Our data showed that compound PC-2 had an absorbance peak at 348nm, and the absorbance extended to over 400nm. In the mean time, compound PC-1 did not show significant absorbance at longer wavelength. Instead, it showed an absorbance peak at 264nm (Figure 2.4A). Next, we deprotected PC-1 and PC-2 with 365nm UV light (4mW/cm^2), PC-2 with 400nm UV light (4mW/cm^2) for variable durations. The products from these reactions were then analyzed by HPLC to measure the amount of materials left after deprotection. After the experiment, the ratio of PC-1 and PC-2 leftovers to starting materials against different time points was plotted (Figure 2.4B). In order to compare the photo-deprotection efficiency of these compounds, the time required to deprotect 50% of PC-1 or PC-2 was defined as $T_{50\%.}$ Our result showed that with 365nm, 4mW/cm^2 UV light, the $T_{50\%}$ of PC-1 and PC-2 was 1.75min and 0.80min respectively, and PC-2 had a $T_{50\%}$ of 4min with 400nm light. These data demonstrated that both PC-1 and PC-2 could be deprotected by UV light and our design of photocaging groups PC-1 and PC-2 was successful. Moreover, it was also shown that PC-2 was better than PC-1 in terms of photolysis efficiency at longer wavelength. Based on these data, it was concluded that PC-2 was an ideal photocaging group, and it would be used in our photoaffinity labeling experiment.

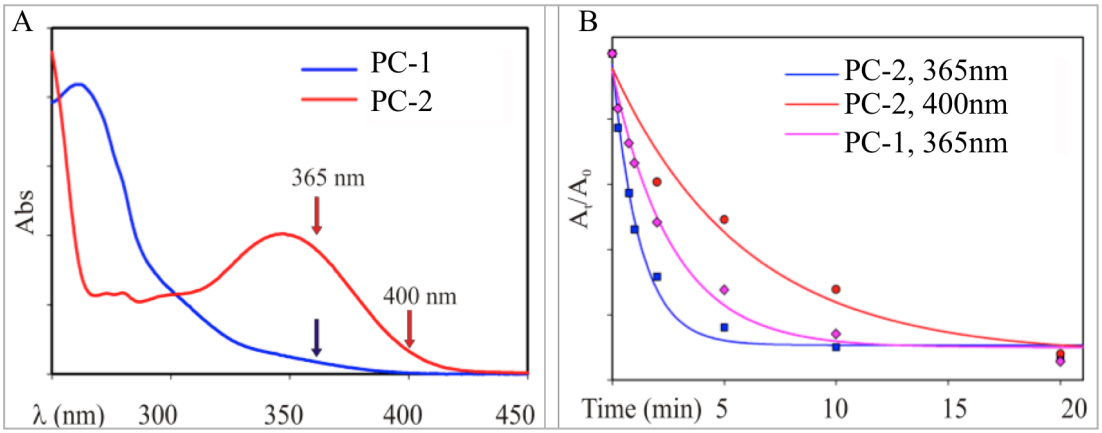


Figure 2. 4 Photochemical characterization of PC-1 and PC-2 caged pQM.
A, absorbance scan for pQM precursor caged by PC-1 and PC-2. B, photolsyis experiment to test deprotection efficiency for both compounds. For photolysis, the compounds were reconstituted to 500μM in HEPES buffer, pH 7.2 with 30% acetonitrile. UV light (4 mW/cm$^{2)}$ was used to deprotect the material followed by HPLC analysis. The ratio of residue compound to starting material was plotted against time and got the graph. This result showed that the photolysis of PC-2 was more efficient than that of PC-1, because it could even be deprotected by visible light with the same intensity.

Investigating the reactivity between pQM and amino acids

To apply pQM-based photoaffinity labeling strategy on the characterization of protein-peptide or protein-protein interaction, we had to firstly know which amino acid could be crosslinked by the pQM. Previously, the reactivity of QM species to biomolecules, such as DNA, peptide and glutathione, has been intensively investigated. It was found that amines and sulfide in amino acids and glutathione could be alkylated by QM through Michael addition. However, because most of the investigations about amino acids focused on the reactivity between QM and free amino acids, it was necessary to characterize the reaction between pQM and amino acid side chains in a protein. In a recently study, complementary PNA molecules (peptide nuclear acid) were used to bring a DOPA derived ortho-quinone and amino acids together to test their reactivity. It was

found that α- and γ-amino of Lys, imidazole of His, and thiol of Cys could react with DOPA derived ortho-quinone (Liu et al, 2006). To advance our understanding on the pQM crosslinking mechanism, a template molecule was designed to bring the pQM precursor and an amino acid together for deprotection. We hoped this experiment would reveal the reactivity between pQM and various amino acids under a condition that mimicked protein crosslinking reaction (Figure 2.5A).

The template system used in our study was composed of a crosslinker precursor with the photocaging group, an amino acid side chain, and a chromogenic dye Dabcyl, which had strong absorbance at 473 nm and would help us on HPLC analysis of reaction products (Figure 2.5A). For this template system, nine amino acids were selected to represent 20 natural amino acids with different side chains: lysine, arginine, tyrosine, aspartic acid, asparagine, histidine, serine, methionine and tryptophan. Accordingly, nine compounds were synthesized, purified by HPLC and analyzed by MS. A final concentration of 100 μM of each compound was used to run two experiments: procedure A and B. In procedure A, the sample was first irradiated with UV light, and βMe was then added to 10mM to quench unreacted intermediate. In procedure B, the sample was irradiated with UV in the presence of 1mM βMe. After reaction, both samples were analyzed by HPLC (acetonitrile gradient: 10% to 90% in 30 min) and all product peaks were collected and analyzed by MS. Based on the design of procedure A and B (Figure 2.5B), the compound I could be deprotected by UV light to form compound II, which could generate an intramolecular crosslink product compound III if the amino acid side chain could react with pQM. Otherwise, it would form a thiol adduction product compound IV, or a water

adduction product compound V if free thiol group was absent in the system. There were several reasons to add βMe in the procedure A and B. Firstly, compound II and III had the same molecular weight and were difficult to tell from each other simply by MS analysis. However, compound II was an active intermediate, in the presence of free thiol group, it could be quickly quenched and form thiol adduction compound IV. Besides, by adding βMe to both procedures, we could tell which amino acid side chain could out-compete free thiol group for pQM to form intramolecular crosslink product, which is especially important when this method is used for the characterization of protein and peptide recognition in live cells, where free thiol concentration can be up to 10mM (Bremer et al, 1981).

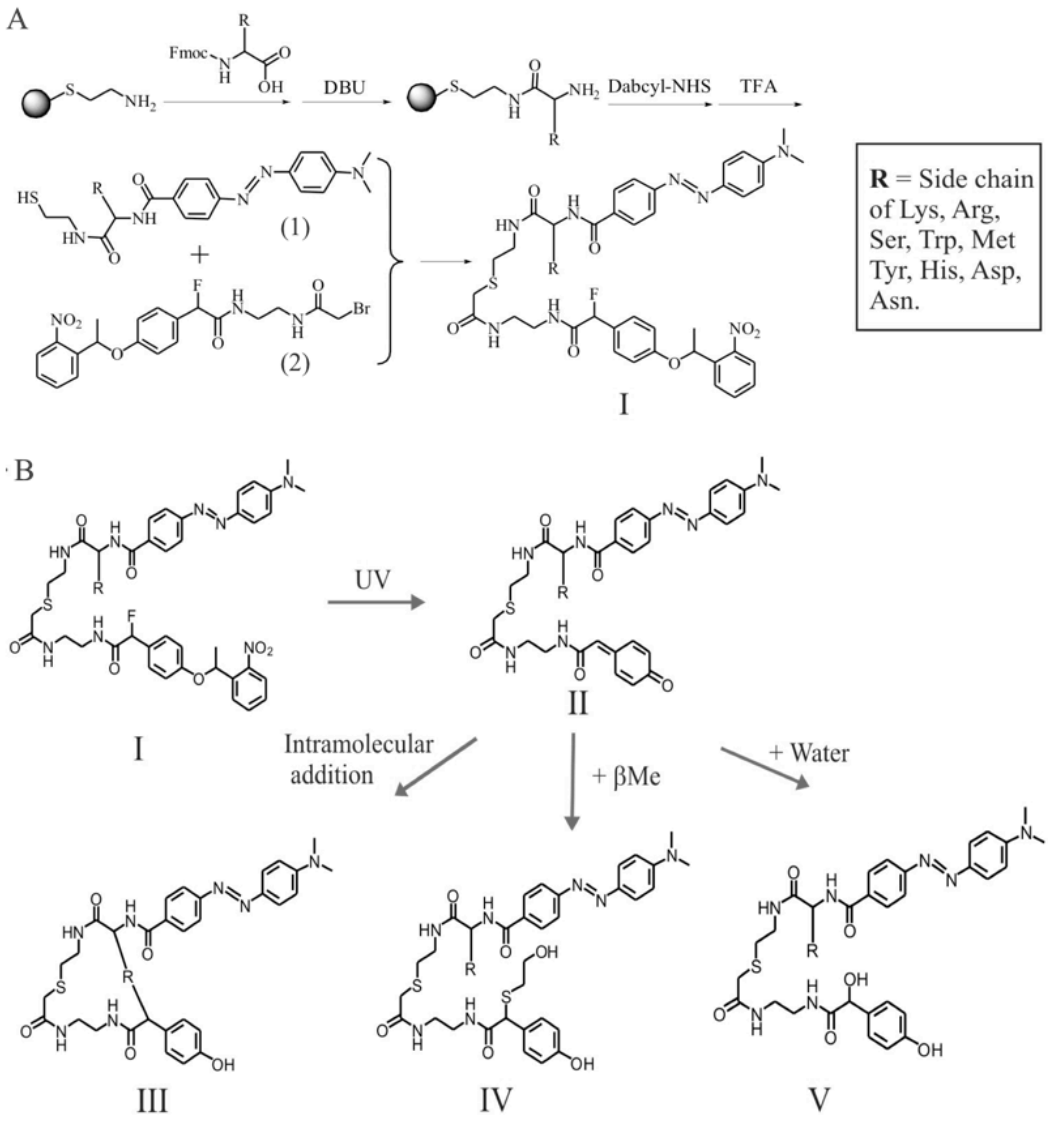


Figure 2. 5 Amino acid deprotection and crosslinking assay.
A, Generation of nine compound Is with different amino acid residues. For detailed synthesis procedure, please see reference (Jiang et al, 2009). Each compound I contains a pink dye Dabcyl, a PC-2 caged pQM precursor and an amino acid residue. The dye has an absorbance at 473nm and was used to facilitate HPLC analysis because there would be no appropriate chromophore in the compound otherwise. B, different reaction possibilities after the deprotection. Compound I was deprotected by UV at 365nm, and an unstable intermediate was firstly formed. Consequently, active pQM was generated through molecular rearrangement (II), followed by reaction with either the amino acid side chain or thiol scavenger or water to generate intromolecular crosslink (III) or thiol adduct (IV) or water adduct (V) products respectively.

After reaction, HPLC and MS data were analyzed. It was found that all nine amino acids under investigation could be divided into three groups based on their reactivity toward pQM (see Figure 2.6 and Table 2.1). Group I amino acids were those actively crosslinked by pQM, including cysteine, histidine, lysine, aspartic acid and glutamic acid. Cysteine contains a thiol group, which is very reactive to pQM and was used as a quench group in this study. Histidine, aspartic acid and glutamic acid were also very reactive and could be crosslinked, even in the presence of thiol group (see Figure 2.6). Lysine is also reactive, but there was only a small fraction of intramolecular crosslinking in the presence of thiol group. Group II amino acids included tyrosine and tryptophan. These two amino acids could moderately react with pQM without thiol (see Figure 2.6 and Table 2.1). With thiol in the system, however, we didn't see intramolecular crosslinked product (See Figure 2.6 Tyr-B and Trp-B). Instead, thiol adduction product dominated the reaction. These observations meant although tyrosine and tryptophan could react with pQM, they could not out-compete free thiol group for the reaction. Group III amino acids included serine, methionine, arginine, asparine, glutamine and all other amino acids without a good nucleophile side chain that could not crosslink pQM (see Figure 2.6 and Table 2.1).

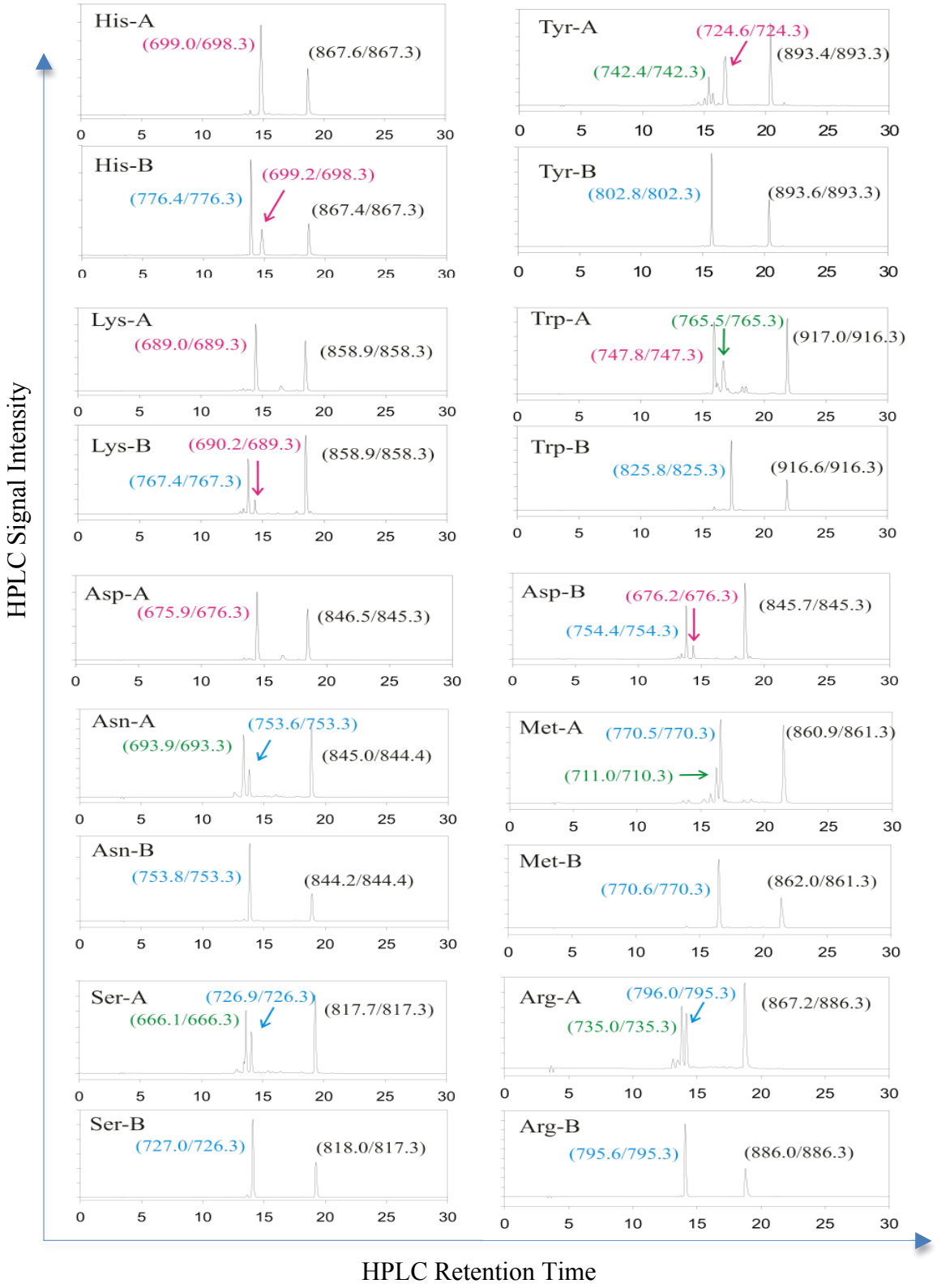


Figure 2.6, to be continued in next page.

Figure 2. 6 Reactivity of different amino acids toward the pQM.
Each compound I with a unique amino acid residue shown in Figure 2.5 was dissolved to get a 100 μM solution and tested with two conditions. Under condition A, deprotection was carried out with UV light (365 nm, 4 mW/cm^2) for 5 min, then 10 mM βMe was added to stop the reaction. Under condition B, the same compound was deprotected with UV light (365 nm, 4 mW/cm^2) for 5 min in the presence of 1 mM βMe. The final products from both reactions were analyzed with HPLC and major peaks were collected and analyzed by MS. This figure shows the HPLC result, X-axis representing retention time, Y-axis representing signal intensity. For each reaction, the intramolecular adducts, βMe adducts, and water adducts are colored in red, blue, and green respectively. The unreacted starting material is shown in black. Numbers in parentheses are the m/z (mass to charge ratio) of the peak, and the measured m/z over the calculated molecule weight is separated by a slash.

Table 2. 1 Summary of amino acid deprotection assay

Group	Amino acid	Starting material	Intramolecular crosslinking	Thiol addition	Water addition
I	**Lys A**	YES	YES	NO	NO
	Lys B	YES	YES	YES	NO
	His A	YES	YES	NO	NO
	His B	YES	YES	YES	NO
	Asp A	YES	YES	YES	NO
	Asp B	YES	YES	YES	NO
II	**Trp A**	YES	YES	NO	YES
	Trp B	YES	NO	YES	YES
	Tyr A	YES	YES	NO	YES
	Tyr B	YES	NO	YES	NO
III	**Arg A**	YES	NO	YES	YES
	Arg B	YES	NO	YES	NO
	Ser A	YES	NO	YES	YES
	Ser B	YES	NO	YES	NO
	Met A	YES	NO	YES	YES
	Met A	YES	NO	YES	NO

Note: Letter A or B behind each amino acid represent different reaction conditions, and A refers to the deprotection in the absence of thiol, but thiol was added to 10mM to quench the rest pQM after the deprotection; B refers to the deprotection in the presence of 1mM thiol. In this summary, amino acids are categorized into three groups based on their reactivity to the pQM.

Our investigation on amino acids and pQM reactivity showed that many amino acids on a protein surface could be crosslinked by the pQM compound if they were brought into the proximity. Moreover, these data provided some guidelines for the characterization of pQM crosslinking site on a protein, which would be our next interest.

Labeling catPTP1Bm with pQM based photoaffinity method

To develop pQM based photoaffinity system for PPI study, we used a PPI model to demonstrate the method. In this interaction model, prey protein was the catalytic domain of protein tyrosine phosphotase 1B C215S trapping mutant (catPTP1Bm) (Flint et al, 1997), and the bait was its binding peptide DADEpYLIPQQG from epidermal growth factor (EGF) receptor (988-998) (Milarski et al, 1993). The affinity between these two proteins is 395nM (Jia et al, 1995), which is moderate and makes this pair a good representative for different PPIs.

For this experiment, the bait probe (pep-1) was synthesized (Jiang et al, 2009), which contained the PC-2 photocaging group, pQM precursor, an acid-labile linker, a biotin tag, and EGFR peptide with phosphorylated tyrosine at position 992 (See Figure 2.7A). At the same time, the prey protein catPTP1Bm was expressed in a modified pET11(c)+ vector (Figure 2.7B) and purified with Ni-column (Figure 2.7C).

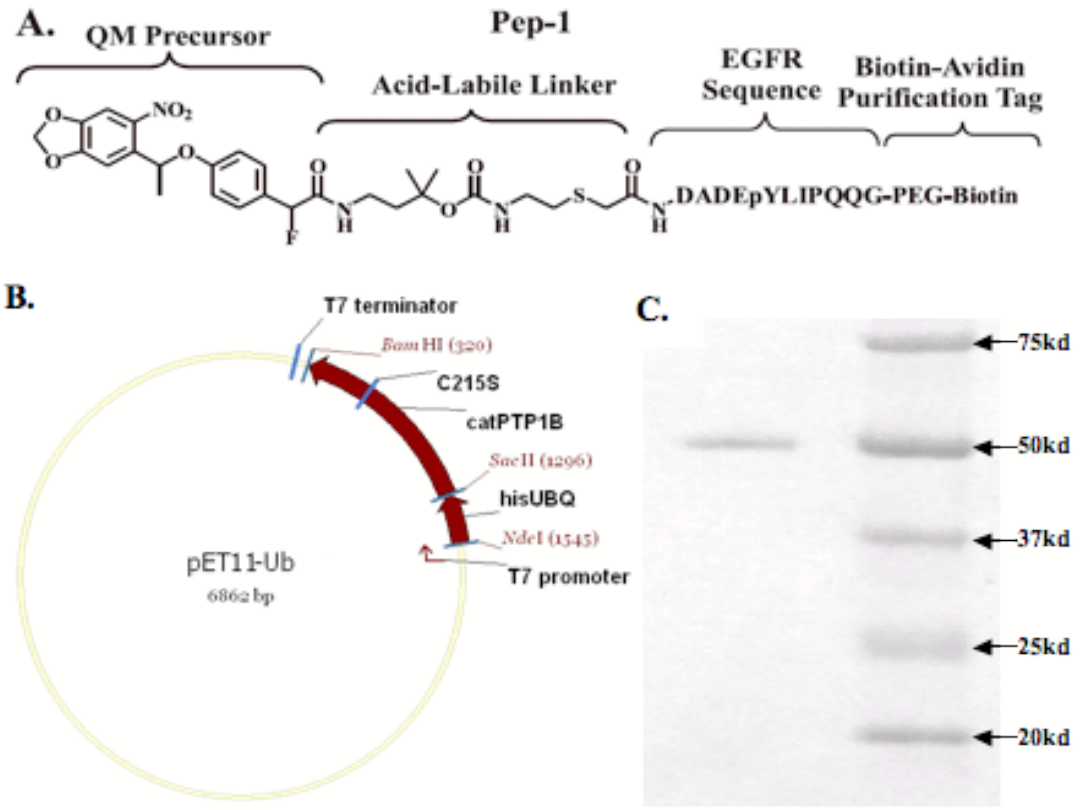


Figure 2. 7 Preparation of pep-1 and catPTP1Bm for photoaffinity labeling.
A, pep-1 containing EGFR peptide (988-998), affinity tag, acid labile linker and PC-2 caged pQM precursor; B, the map of the plasmid constructed for expressing catPTP1Bm; C, purified protein catPTP1Bm.

Next, 10nM catPTP1Bm and 1μM peptide probe (pep-1) were mixed in the presence of 0.5mM DTT as scavenger at room temperature. The sample was irradiated by UV light for 5 min at 365nm (4 mW/cm^2) and then separated with SDS-PAGE. After that, we did a western blotting experiment with streptavidin conjugated HRP to detect the protein product generated from the photoaffinity labeling. Our result showed that the target protein catPTP1Bm was successfully labeled (see Figure 2.8, left). Although this experiment was carried out with a purified protein, it demonstrated that the overall design of our photoaffinity labeling reagent was working.

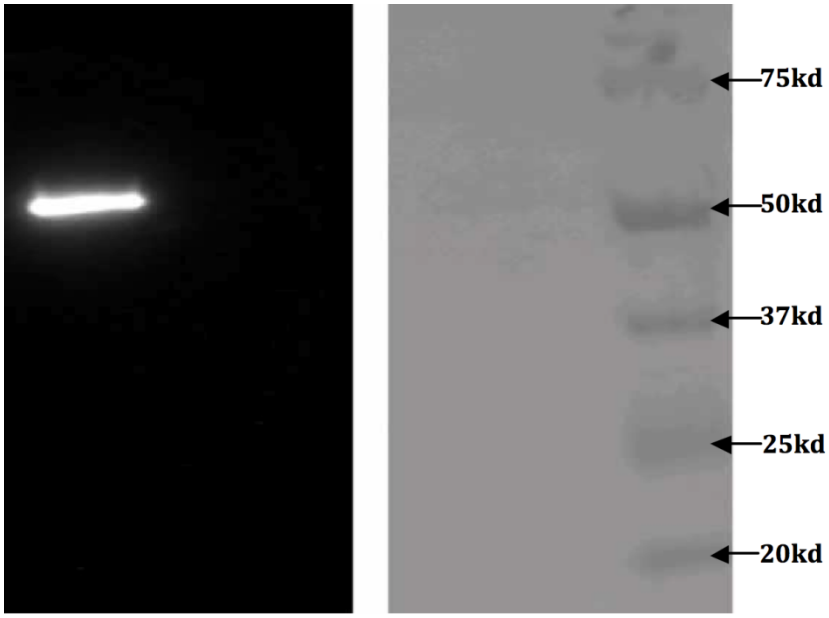


Figure 2. 8 Photoaffinity labeling assay with purified protein.
Photoaffinity labeling of target protein catPTP1Bm was done with pep-1 containing photocaged pQM precursor. Purified catPTP1Bm was mixed with the pep-1, and the deprotection was carried out in the presence of 0.5mM DTT. After deprotection, the sample was then resolved by SDS-PAGE followed by western blotting for the detection of biotin tag. Left, western blotting result; Right, the same membrane was aligned with the western image.

pQM based photoaffinity labeling was specific and sensitive

When we were developing this method, an interesting question was how it would perform in physiological condition. It has long been revealed that the physiological condition in a living cell is a reducing environment, which is maintained by the exchange between glutathione (GSH) and glutathione disulfide (GSSG). It has also been found that the physiological glutathione concentration varies from 0.1mM to 10mM (Bremer et al, 1981). In our previous experiment, we found that thiol group had high reactivity toward our photoaffinity labeling reagent. However, some amino acids could compete with free thiol if they were close to the photoaffinity labeling moiety. In the mean time, free thiol had been used as a scavenger in our protein labeling experiment. Given the wide

concentration range of free thiol in physiological condition and its high reactivity toward our pQM based photoaffinity labeling reagent, it was necessary to investigate how our method would perform in physiological concentrations, that is, could free thiol in physiological condition be used as a scavenger for our photoaffinity labeling method?

To address this issue, a DTT titration experiment was conducted. Specifically, a series of samples containing 10nM protein catPTP1Bm, 1μM peptide probe, 1μg/μl bacterial cell lysate and different concentration of DTT (from 0 to 5mM) were prepared. The purpose to add bacterial cell lysate in these samples was to test the photoaffinity labeling specificity in a complex environment. All samples except a control were irradiated by UV light under the same conditions as used before. After reaction, samples were separated with SDS-PAGE and blotted with strepavidin-HRP conjugate to detect labeled products. As a comparison, another SDS-PAGE loaded with the same set of samples was stained with Coomassie blue to show the total amount of proteins in each lane. In the end, it was found when DTT concentration was low, multiple protein bands were observed, which might be caused by non-specific labeling between the peptide probe and some proteins. Once DTT concentration was higher than 0.1mM, however, only a strong catPTP1Bm band and a faint unidentified protein band were still visible, indicating thiol at these levels could dramatically improve the specificity of our photoaffinity labeling reaction (see Figure 2.9). In order to quantify this result, we determined the specificity of our crosslinking result by measuring the main band intensity and the background signal level with ImageJ software. Our result showed that when DTT concentration was higher than 0.1mM, the signal/background ratio was increased significantly (See table 2.2). Given the

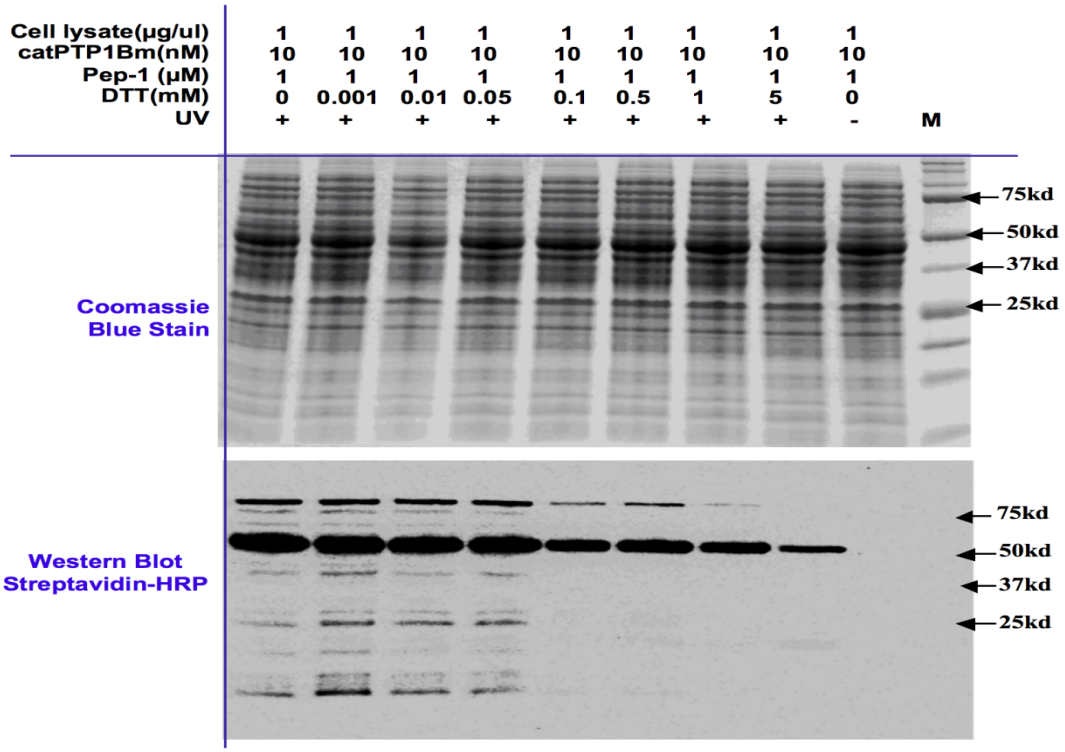


Figure 2. 9 DTT titration assay for pQM based photoaffinity labeling.
The purpose of this experiment was to test the effect of different thiol concentrations on the photoaffinity labeling efficiency and specificity. Purified catPTP1Bm with pep-1 was spiked in bacterial cell lysate and deprotected with UV light at different concentrations of DTT except the last sample lane, which was a control and not deprotected. After deprotection, the samples were resolved in two SDS-PAGE gels: one for coomassie blue stain (upper) and one for western blotting using HRP conjugated streptavidin (lower). In upper picture, we could see the cell lysate concentrations were the same for each lane, while in lower picture, western signal was getting more specific to the catPTP1Bm with the increase of scavenger concentration. The measurement of the mainband intensity, background signal intensity and their ratio is provided in Table 2.2.

Table 2. 2 Quantifying protein crosslinking specificity for the DTT titration assay*

Lane	TA	MB	BG	MB/BG	NTA	NMB	NBG	NMB/NBG
1	374708	156676	218032	0.72	243650	145013	98637	1.47
2	387871	144065	243806	0.59	256813	132402	124411	1.06
3	360958	144942	216016	0.67	229900	133279	96621	1.38
4	352715	141819	210896	0.67	221657	130156	91501	1.42
5	250545	103883	146662	0.71	119487	92220	27267	3.38
6	277783	126207	151576	0.83	146725	114544	32181	3.56
7	236558	104574	131984	0.79	105500	92911	12589	7.38
8	170211	45214	124997	0.36	39153	33551	5602	5.99
9	131058	11663	119395	0.10	0	0	0	0

* The measurement was done by ImageJ, and the unit is "Gray value". The red part of the form is normalized by lane 9. TA=Total; NTA = Normalized Total; MB= Main Band; NMB = Normalized Main Band; BG = Background; NBG = Normalized Background.

concentration of glutathione varies from 0.1mM to 10 mM in living cells (Bremer et al, 1981), which was similar to the range of thiol group in our DTT titration assay, we concluded that pQM based photoaffinity labeling would be highly specific even inside cells.

Protein concentration is another critical issue that has to be taken into consideration when developing a PPI characterization tool. It has been documented that in yeast the copy number of different proteins could vary from 32 to 500,000 copies per cell, equivalent to about 30pM to 500nM considering the average size of yeast cells (Gavin et al, 2006). So our next experiment was to evaluate the effectiveness of pQM based photoaffinity labeling method with various concentrations of the target protein. The experiment was carried out by using conditions similar to the DTT titrations assay. But this time different concentrations of catPTP1Bm (0.1nM to 1000nM) were added to each reaction mixture, while DTT concentration was fixed to 0.1mM DTT. After the reaction, samples were also analyzed by Western blotting and Coomassive blue staining. In the end, it was found that even at 0.1nM, the catPTP1Bm band was still clearly visible (see Figure 2.10), indicating that minute amount of catPTP1Bm could be labeled by the peptide probe through this method in the presence of 10^6-fold excess of the thiol scavenger (0.1mM vs. 0.1nM).

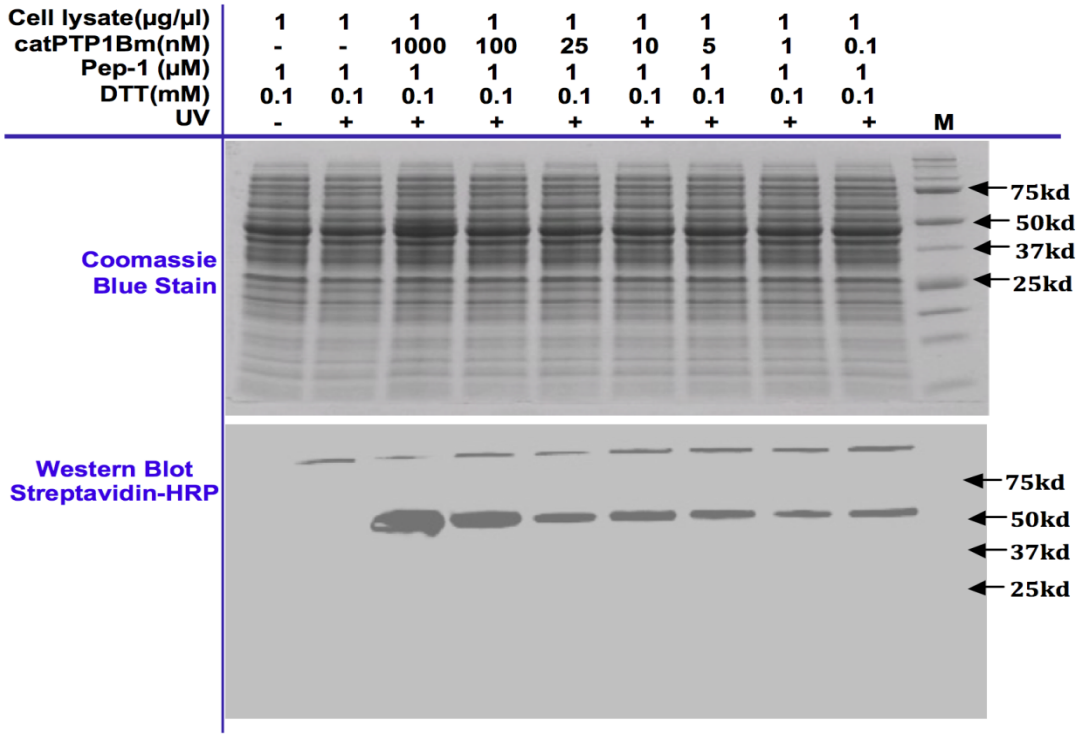


Figure 2. 10 Protein concentration effect on pQM based photoaffinity labeling. This experiment used the same deprotection condition as the DTT titration assay to test the performance of the method with different concentrations of target protein catPTP1Bm. Protein concentrations were changed from 1μM to 0.1nM, which covered most of the physiological protein concentrations. With protein concentration reducing in the system, the crosslink signal decreased accordingly. Even though the protein concentration was as low as 0.1nM, there was still a clear crosslinked band.

Mapping crosslink sites for pQM based photoaffinity labeling

Chemical crosslinking has long been used as a tool to explore the protein structure and PPI binding interface. Although the resolution may not be as high as X-ray crystallography or NMR (Jin Lee, 2008; Sinz, 2006), its ability to characterize protein structure and function in physiological condition makes it a useful tool. For example, chemical crosslinking combined with MS analysis has been used to study conformational change of kinase Akt during its activation (Huang & Kim, 2006) and calmodulin-melittin complexs in response to calcium binding (Nadeau et al, 2007). With these preceding

examples, it would be very interesting to analyze the crosslinking sites on the target protein after photoaffinity labeling with our method. We wanted to see whether pQM based photoaffinity labeling strategy could provide useful structure information about protein-peptide recognition. In the mean time, if we could identify the crosslinking site, specifically, the crosslinked amino acid, we would have a better understanding to the photoaffinity labeling procedure and mechanism.

To map the protein crosslink site, MS analysis was carried out after the photoaffinity labeling of the target protein with the same procedure as before. Briefly, a mixture of 1μM catPTP1Bm protein, 1 μM peptide, 1μg/μl cell lysate and 100μM DTT was used for the photoaffinity labeling reaction. The sample was then denatured by urea, reduced by TCEP, alkylated by bromoacetamide and digested by trypsin. Next, the biotin containing peptides were enriched by monomeric avidin resin and cleaved by acid, which significantly reduced the size of the crosslinking peptide. In the end, the peptide sample was analyzed by MS. According to the experiment design (see Figure 2.7), the peptide generated from the photoaffinity labeling should contain a 217.13D marker,

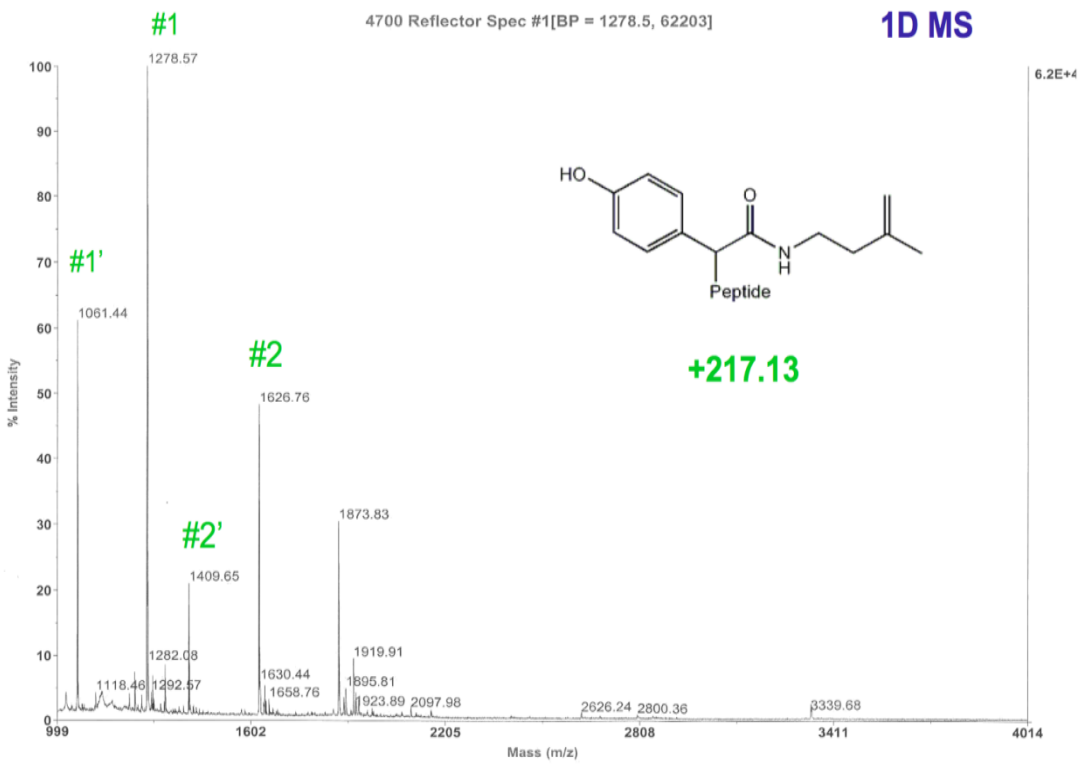
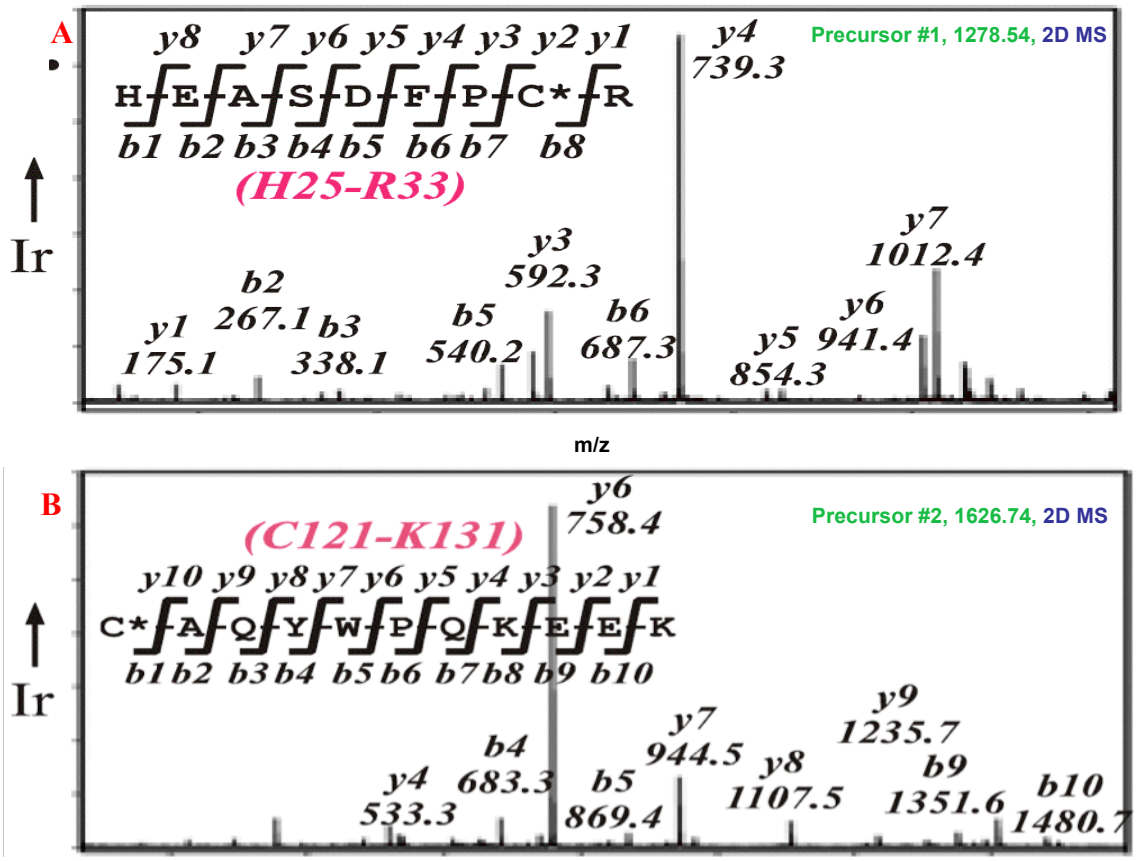


Figure 2. 11 Enriched peptides from the photoaffinity labeling experiment.
Protein catPTP1Bm was crosslinked with the same as those for DTT titration assay except that DTT concentration was fixed to 100μM. After deprotection, samples were pretreated, digested, enriched, cleaved and analyzed. MS result showed two peptides (#1 and #2) carried the crosslink footprint (the structure was shown inside), which was 217.13D and generated after acid cleavage. Besides, two peptides without the marker (#1' and #2') were also identified, which were generated during MS experiment from the crosslinked peptides.

which was derived from the cleaved probe. To our interest, two peptides from catPTP1Bm were found bearing this marker (Figure 2.11). Tandem MS analysis revealed their amino acid sequence and the crosslinking sites Cys32 and Cys 121 (see Figure 2.12): peptide#1 HEASDFPC*R (25-33, *=marker, predicted m/z=1278.57, measured m/z=1278.57) and peptide#2 C*AQYWPQKEEK (121-131, *=marker, predicted m/z=1626.78, measured m/z=1626.76, one miss cut). Surprisingly, two satellite peaks, peptide#1' and peptide#2', were found accompanying two crosslinked peptide peaks. Tandem MS analysis showed that they had the same amino acid sequence as peptide#1 and peptide#2 except there was no crosslinking marker (peptide#1' predicted m/z=1061.44, measured m/z=1061.44; peptide#2' predicted m/z=1409.65, measured m/z=1409.65). Our explanation to this phenomenon was that the crosslinked products were not stable under MS condition and a fraction of peptides could be disrupted from the final crosslinked product by the high energy in laser beam, generating the peptides without 217.13Dmarker. This speculation was consistent with the reversible nature of QM based crosslinking, which has long been recognized and applied to DNA alkylation research (Rokita, 2009c).

In the end, we checked the complex structure of catPTP1Bm with its substrate EGFR peptide 988-998 (DADEpYLIPQQG). Interestingly, it was found that two cysteine residues crosslinked by our peptide probe were flanking the N-terminus of the ligand peptide in the complex structure.



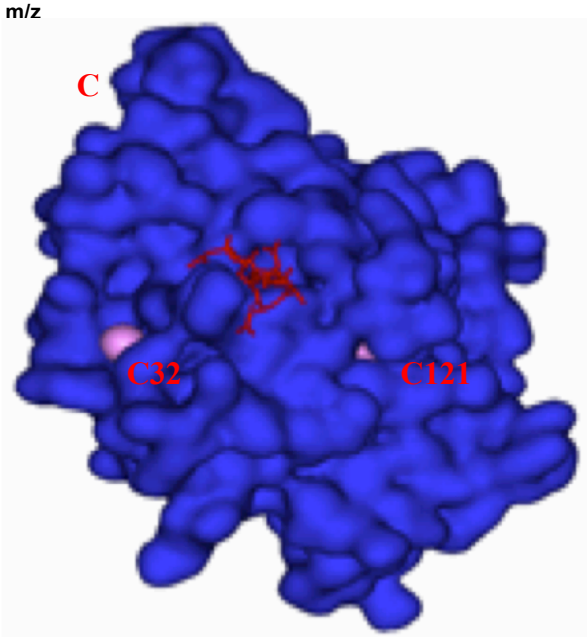


Figure 2. 12 mapping crosslinked sites with MS analysis.
To identify the crosslinked amino acid residues, we used the tandem MS to analyze peptide precursors #1 and #2 (in Figure 2.11). Several ions derived from each peptide precursor were found. Based on these ions, we unambiguously tracked the crosslinked sites to Cysteine 32 in peptide precursor #1 (A) and Cysteine 121 in peptide precursor #2 (B). Interestingly, the PTP1B-EGFR peptide complex structure (PDB ID: 1PTU) revealed that two cysteine residues were flanking to the protein-peptide binding interface (C).

Discussion

In this project, a photoaffinity labeling method based on pQM was developed to characterize protein-peptide interaction. At the beginning, the light harvest functional group was optimized and the amino acid residues that could be crosslinked by the reactive intermediate pQM were investigated. Then, the performance of this method was tested under different conditions. Lastly, the detailed crosslinking result was characterized through MS based peptide analysis. In general, this method incorporated cutting-edge technologies from the field of photochemistry, protein chemistry and reactive intermediate chemistry. It was a combination of label transfer strategy, activity based protein assay and traditional photoaffinity labeling, as some essential elements in these diverse methods were integrated in this one method.

This new method significantly improved protein crosslinking specificity compared to other methods. Crosslinking specificity is a double-edge sword for a crosslinker. If it is too specific, for example, only targeting one functional group, the crosslinking efficiency will be very low and the method may not work in the absence of such a functional group, resulting false negatives. On the other hand, if it is not specific enough, everything in the proximity will be crosslinked, generating false positives. In our method, moderate reactive intermediate pQM and a thiol scavenger were combined to increase crosslinking specificity. It was showed that in the presence of scavengers at physiological level the target protein spiked in cell lysate could be specifically crosslinked (Figure 2.9). This result was consistent with previous reports about the crosslink capacity of the quinone-derived crosslinkers on the detection of protein-protein recognition(Liu et al, 2007; Liu et al, 2006). Interestingly, an earlier work using an activity based crosslinking strategy with

pQM identified several crosslinked peptides from the target protein PTP1B, which were randomly distributed on the protein surface. In contrast to that report, our new method identified two peptides flanking the protein-peptide binding interface. More importantly, based on the complex structure of catPTP1Bm and EGFR peptide, we found that the distances between the N-terminal of the peptide and Cys32 and Cys121 on catPTP1Bm were 20.7 Å and 21.1 Å respectively, while our compound with the acid cleavable linker could extend about 21.05 Å from the N-terminal of the peptide. The consistency of these distances strongly suggested that our crosslinker could specifically label the target residue within the proximity. We thus argued that this method might be useful for the characterization of protein-peptide binding interface, representing a significant improvement on analyzing protein complex structure. After a detailed comparison between our method and previously reported method, it was reasoned that the different specificity between these two studies was caused by the application of PTP1B binding peptide and scavenger, because they could dramatically increase the specificity of crosslinking.

There are other interesting aspects in this research. For example, the photolysis analysis of photo-caging groups at different conditions revealed flexibility for different biological applications. Furthermore, our new phototaffinity labeling method was compatible with different concentrations of thiol group, which was significant given the critical role of GSH (thiol) in physiological conditions. Lastly, the successful application of acid cleavable linker in our peptide probe was also a notable feature.

It was particularly exciting to find that our photocaged crosslinker could completely react with some functional groups, such as side chains in cysteine, histidine, lysine and aspartic acid (see Figure 2.6). Moreover, it was found that the crosslinked peptides on cysteine residue could be fragmented during MS analysis (see Figure 2.11), which might become a useful tool for chemical biology research.

Chapter Three,

Detecting Protein Interactions in Live Cells via Complementation of a Hydrolysis-deficient β-Lactamase

Abstract

β-Lactamase PCA, combined with a FRET based substrate CCF2/AM, is a useful tool for the detection of PPI in live cells. However, the current method cannot be used in PPI subcellular localization and high-throughput screening. Previous work demonstrated that β-lactamase mutant E166N has an altered catalytic cycle that causes the protein to permanently attach a substrate molecule to itself via a covalent bond (Adachi et al, 1991). Based on this intriguing finding, it was reasoned that if the E166N β-lactamase mutant was used in the PCA system, the prey fusion protein could be covalently labeled. Furthermore, if CCF2/AM was used as the substrate, this method could be used for PPI subcellular localization in live cells. In addition, if the substrate was switched to alkynyl-penicillin, a β-lactamase substrate containing an affinity tag, this method could be used for high-throughput screening to identify PPIs. Because this new method featured covalently labeling prey protein, it was named "PCA plus". To test the effectiveness of PCA plus, a series of three proof-of-concept experiments were carried out. First, an *in vitro* experiment utilizing an interacting pair of purified protein (SH3 and PPLP peptide) showed that N-terminus of β-lactamase with the E166N mutation could be labeled by substrates fluorescein-penicillin and alkynyl-penicillin using the PCA procedure. Second, PCA plus was used in combination with fluorescence microscopy to visualize a PPI (Leucine zipper domain dimerization) in living mammalian cells. Additional results

indicated that this application of PCA plus could be useful for subcellular localization of PPIs. Third, PCA plus was utilized as an approach for covalent tagging of prey protein with a small molecule that enabled easy and quick enrichment strategies in preparation for LC-MS/MS. In conclusion, the results described here demonstrate that PCA plus is a useful method, in particular because it enables high resolution imaging of PPIs in live cells and because it can be applied to high-throughput PPI screening. These advances could make PCA plus a valuable tool for future PPI studies.

Introduction

The concept of PCA plus

β-Lactamase based PCA has been well established, and its advantages in characterizing PPIs have been demonstrated in many applications. For example, this method was used to screen for inhibitors that disrupt the interaction between TLR4 and Myd88 (Lee et al, 2007). In another application, β-lactamase based PCA mapped the binding domain between herpesvirus protein UL34 and UL31 family members (Schnee et al, 2006). Yet another example of applying this method is differentiating two types of antibodies against herpes simplex virus (HSV) (Fry et al, 2008). These examples clearly showed that β-lactamase based PCA is a reliable and versatile method for the study of protein interactions. However, because complemented β-lactamase can quickly hydrolyze many substrate molecules, which can quickly diffuse throughout a cell, this method is useless in PPI subcellular localization. Another limitation of the current version of PCA is that it is very difficult for high-throughput PPI screening. We devised a modified version of PCA to circumvent these limitations. In our new version of PCA (PCA plus), PPI results in

covalent labeling of the prey protein with a tag. This approach is made possible by using a mutant version of β-lactamase in which the active site is affected by a E166N mutation(Adachi et al, 1991).

β-Lactamase is a serine-based hydrolase. It can hydrolyze the β-lactam ring in penicillin through a two-step reaction: acylation and deacylation (See Figure 1.6). During the acylation reaction, the enzyme and the substrate firstly associate with each other. The Ser70 in the enzyme catalytic center attacks the β-lactam carbonyl carbon to form a tetrahedral intermediate. Consequently, the four-member β-lactam ring opens and an acyl-enzyme complex is generated through the covalent bond between the Oγ atom of Ser70 and carbonyl carbon in the open β-lactam ring. During the subsequent deacylation reaction, a water molecule, activated by the side chain of Glu166, breaks the covalent bond bridging the enzyme and the product. The enzyme is thus recycled, product released and deacylation finished (See Figure 1.6) (Adachi et al, 1991).

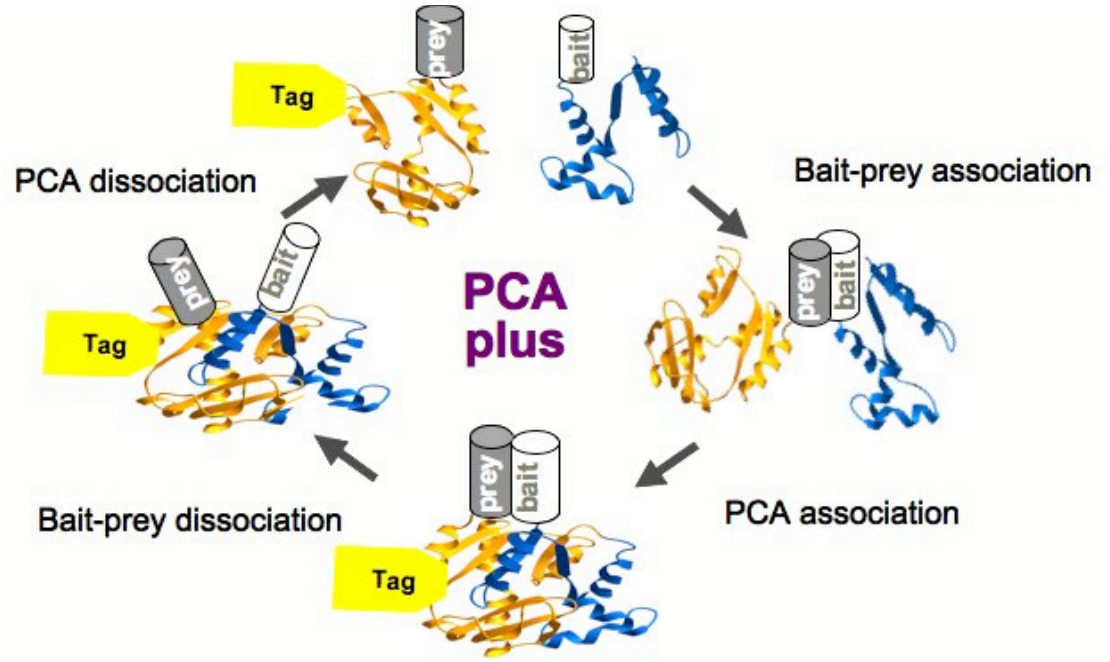


Figure 3. 1 Overview of PCA plus method.
Prey and bait protein are represented by the gray and white cylinder respectively. They are fused with the N- and C-terminal of the β-lactamase (yellow and blue structure in this Figure). The tag is shown as a bright yellow box. See text for the detailed description about the principle.

As we can see from the β-lactamase enzymatic reaction, there is an enzyme-substrate intermediate bridged by a covalent bond. Interestingly, in a site-directed mutation study, it was found that β-lactamase with the E166N mutation could stabilize the enzyme-substrate intermediate, that is, the mutated enzyme could be conjugated by its substrate (Adachi et al, 1991). If we use β-lactamase E166N mutant in the PCA system, a penicillin molecule will be conjugated to the prey protein through the N-terminal of β-lactamase (E166N) after the PCA reaction. Furthermore, if we add a tag to penicillin and use it as the substrate in the PCA reaction, we can covalently label the prey protein through the N-terminal of β-lactamase (E166N) with this tag. Subsequently, we can use

the tag to image PPI or enrich the prey and identify it. Because this method features covalently adding a tag to the target protein through the PCA reaction, it was named "PCA plus" (Figure 3.1).

To summarize the PCA plus method design: just like many other PPI detecting tools, PCA plus is built on reporter protein splitting and protein fusion technology. Here, the reporter protein, the enzyme β-lactamase, is split into N and C terminal segments fused to the bait and prey proteins respectively. As shown Figure 3.1, the bait and the prey proteins interact and bring the N and C terminal of the β-lactamase into proximity, which in turn fold to form an active enzyme. After the enzymatic reaction, the bait and the prey protein may dissociate and separate the β-lactamase N- and C- terminal fragments. PCA plus uses the hydrolysis deficient β-lactamase E166N, which can be labeled by a substrate molecule through a stable covalent bond. So after the bait-prey interaction, a tag derived from β-lactamase substrate will be attached to the fusion prey protein, leaving a permanent marker on the prey to record the PPI event. In PPI live cell imaging mode, the tag will be a fluorescent dye, which can be visualized to track the target protein in live cells and reveal protein dynamics. In PPI high-throughput screening mode, the tag will be an affinity label moiety, which can facilitate enrichment/purification of the target protein to enable identification (eg. by MS). Notably, although in this research β-lactamase is used as a PCA plus prototype model due to some exceptional properties with this enzyme and the great flexibility of its substrates, many other reporters can be used for the same purpose, and among them, Halo tag protein is a very good candidate (Los et al, 2008).

PCA plus for PPI live cell imaging

β-Lactamase PCA is a PPI imaging method and the application of β-lactamase substrate CCF2/AM has made it a great success. For us, it was a natural choice to use the PCA plus for PPI imaging because this new method had two remarkable features compared to conventional PCA. Firstly, PCA plus could be used for PPI subcellular localization. Because a hydrolysis-deficient β-lactamase would be used in the PCA plus system, only one fluorescent substrate molecule could be hydrolyzed and consequently conjugated to the prey protein, generating a footprint of PPI on the prey protein. Secondly, this method could track the movement of the prey target protein after PPI event in live cells, which may be even more interesting than PPI detection, as revealing PPI function is our ultimate goal for a PPI study.

PCA plus for interacting protein enrichment and "Click Chemistry"

The PCA plus method featured the covalent label of interacting protein with a tag, which could be used for target protein enrichment. There were several requirements for this affinity tag. Firstly, the tag should have a minimal effect on the β-lactamase enzymatic reaction. Secondly, the tag should be small and cell permeable in order to be used as an *in vivo* labeling reagent. Thirdly, the tag should greatly facilitate down-stream protein enrichment. More specifically, the tag must be compatible with harsh wash conditions during affinity purification, including wash with detergent, organic solvent and other denaturing reagents. Lastly, the tag should be stable and easy to synthesize, purify and manipulate. After a careful search, we focused on an alkyne tag based on the "Click Chemistry", because we especially favored the Huisgen's 1,3-dipolar cycloaddition between alkynes and azides.

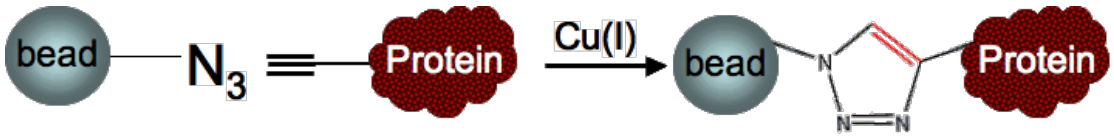


Figure 3. 2 "Click Chemistry" enrichment of labeled protein.
"Click Chemistry" is a class of chemical reactions that are powerful, highly reliable, selective and easy to do. The most widely used "Click" reaction is Huisgen 1,3-dipolar cycloaddition between alkynes and azides to form 1,4-disubsituted-1,2,3-triazoles using Cu(I) as catalyst. Shown here is the scheme of its application on protein enrichment.

There are several advantages to use alkyne and azide based “Click Chemistry” in our system. Firstly, azide and alkyne are stable and easy to be installed to other compounds. Secondly, the “Click” reaction is quick, specific and can proceed to completion in aqueous solution when supplemented with a catalyst of Cu(I). Actually, water molecule plays an important role in this reaction and is indispensable (Rostovtsev et al, 2002). Thirdly, the reaction product 1,2,3-triazole is very stable because it forms a rigid linking unit like a peptide bond (Bock et al, 2006; Kolb & Sharpless, 2003). Based on this information, we thought that alkyne and azide based “Click” reaction was a good system for protein enrichment and it would be our choice of the tag for *in vivo* interacting protein labeling and *in vitro* protein enrichment.

In summary, we have laid out a plan for the development of an *in vivo* labeling strategy for PPI characterization. We would take advantage of β-lactamase based PCA and introduce the mutation E166N to the N-terminal fragment of β-lactamase in the system. Subsequently, we would use cell permeable substrates, CCF2/AM or alkynyl-penicillin, to test the PCA plus in live cells, through which either a fluorescent compound or an alkyne group could be attached to the target protein and used for PPI subcellular imaging

or protein enrichment and identification.

Materials and methods

Critical compounds

β-lactamase substrate fluorescein-ampicillin, alkynyl-penicillin, alkynyl-penicillin cell permeable derivate (CP-AM), and azide beads for protein enrichment were synthesized by Dr. Junxiang Zhang in Li lab, and their molecular weight, structure and chemical properties were confirmed by appropriate methods accordingly. For detailed synthesis information, please check reference(Jiang et al, 2010)

Gene mutation and plasmid construction

Plasmid for the expression of β-lactamase with E166N mutation: Full-length β-lactamase gene was amplified with primer Nde-lac-5end and Xho-lac-3end from plasmid pDEST14. E166N mutation was introduced by three PCR reactions: PCR1 with primer pair OSL109 and Xho-lac-3end, PCR2 with primer pair OSL110 and Nde-lac-5end. Subsequently, PCR products from both reactions were gel purified (Qiagen). PCR3 was conducted by combining purified PCR1 and PCR2 product as the templates with primer pair Nde-lac-5end and Xho-lac-3end. Subsequently, PCR3 product was digested by restriction enzyme NdeI and XhoI and gel purified. All PCR reactions for plasmid construction in this research used high fidelity amplification system iproof (Bio-Rad). Expression vector pET26(b)+ was also digested by the same enzymes and purified. Generated DNA fragments of mutated β-lactamase gene and the vector were then ligated and transformed into *E. coli* DH5α. Positive clones were identified by colony PCR and plasmid digestion, followed by DNA sequencing to further confirm the construct.

Plasmid for the expression of NLacN-SH3 protein: SH3 domain coding sequence from yeast gene *Sho1p* was codon-optimized for *E. coli.* expression and a serial of 5 primers OSL082, OSL083, OSL084, OSL085 and OSL086 were synthesized to cover the modified gene. PCR reactions were used to join these primers to obtain whole SH3 domain DNA fragment. NLacN fragment was generated by PCR reaction with primer OSL020 and OSL039 from β-lactamase (E166N) gene. Generated DNA fragment was further extended by two rounds of PCR reaction with primer pair OSL039-OSL111 and OSL039-OSL082 to join NLacN-Linker with SH3 coding DNA to obtain NLacN-SH3 fusion gene. The fusion gene was digested by SacII and BamHI and cloned into pET11(c)+ derived vector with 6xHis tag and Ubiquitin gene (HisUBQ vector).

Plasmid for the expression of PPLP-LacC protein: two primers OSL087 and OSL088 were designed to cover Sho1p SH3 domain binding peptide QQIVNKPLPPLPVAGSS of Pbs2p protein (amino acid 88-104) (Marles et al, 2004), and codons was also optimized for *E. coli.* expression. β-lactamase C-terminal (amino acid 198-286) DNA fragment was amplified with primer OSL035-OSL021 and extended with primers OSL088 and OSL087 to obtain fusion gene for SH3 binding peptide, linker and β-lactamase C-terminal, named PPLP-LacC.

Plasmid for mammalian cell expression of Zip-NLacN and Zip-LacC: two plasmids for mammalian cell expression of Zip-Bla(1) and Zip-Bla(2) in vector pcDNA3.1/zeo were requested and received from Dr. Michnick's lab based on their report (Galarneau et al, 2002). In this thesis, they are renamed Zip-LacN pcDNA3.1/zeo and Zip-LacC

pcDNA3.1/zeo accordingly. E166N mutation was introduced into Zip-LacN plasmid by the same method as that for β-lactamase E166N mutation, and primers were OSL100, OSL109, OSL110 and OSL105. Mutated gene was cloned back into the same vector and obtained plasmid Zip-NLacN pcDNA3.1/zeo. In order to increase Zip-NLacN and Zip-LacC co-expression efficiency in mammalian cells, Zip-LacC furion gene with its regulatory elements was transferred to the Zip-NLacN pcDNA3.1/zeo. Primers OSL136 and OSL137 with a BglII site in each of them were used to amplify the Zip-LacC fusion gene with its regulatory elements. Generated DNA fragment was digested by BglII and inserted into the only BglII site in plasmid Zip-NLacN pcDNA3.1/zeo. In the end, we obtained the plasmid with two sets of regulatory elements to control the expression of both Zip-NLacN and Zip-LacC in one pcDNA3.1/zeo vector. After construction, *E. coli* strain DH5α was used to amplify and produce plasmids, and QIAGEN Plasmid Midi Kit was used to prepare highly pure plasmids for mammalian cell transfection experiment.

Plasmid for mammalian cell expression of NLS-Zip-NLacN and NLS-Zip-LacC: The amino acid sequence of the Nuclear Localization Signal (NLS) in GCN4 protein is DPAALKRARNTEAARRSRARKLQRMK, which was predicted by software predictNLS (Cokol et al, 2000) and cNLS Mapper (Kosugi et al, 2009). Primer OSL167 and OSL168 were used to generate the fusion gene with this NLS. Specifically, primer OSL168 and OSL105 (BGH reverse primer) was firstly used to amplify the Zip-NLacN and Zip-LacC fusion genes. OSL167 and OSL105 were used as the second round of PCR to obtain the fusion gene NLS-Zip-NLacN and NLS-Zip-LacC and add restriction site BamHI. Generated DNA fragments were cloned back into the vector pcDNA3.1/zeo.

These two plasmids were further combined to make one plasmid that expressed both NLS-Zip-NLacN and NLS-Zip-LacC in mammalian cells. The plasmid construction procedure was exactly the same as mentioned above.

Protein expression and purification

Proteins were expressed in *E. coli.* Strain BL21 (DE3). Plasmid was transformed into competent cells, and bacterial cells were grown overnight on plates with appropriate antibiotics. Next day, bacterial cells were inoculated into LB media with antibiotics. Cell density was determined by measuring OD600. When it reached 0.6, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a finial concentration of 1mM. Cells were allowed to grow for additional 3 hours and then collected and stored at -70^{o}C for future use.

For protein purification, bacterial pellet was thawed at room temperature, and Bugbuster (Novagen) was added to break down cells according to manufacturer's instruction. Cell lysate was centrifuged at 40,000g and supernatant was kept for purification. Ni-resin (Bio-Rad) was used to purify these proteins according to the instruction. Briefly, cell lysate was diluted with 3 volumes of binding solution (50mM Tris buffer, pH7.5, 500mM NaCl, 5mM imidazole) and loaded to Ni column equilibrated with large amount of binding solution. Column was rotated at 4^{o}C to allow the binding to occur in batch mode. The column was then washed intensively with wash buffer (50mM Tris buffer, pH7.5, 500mM NaCl, 10 to 20mM imidazole). Sequentially, target protein was eluted with elution buffer (50mM Tris buffer, pH7.5, 500mM NaCl, 250mM imidazole), and fractions were collected and resolved by SDS-PAGE. Protein was further dialyzed

against appropriate reaction buffer and concentrated by ultra-filtration tubes (Millipore). Protein concentration was estimated by either SDS-PAGE or Bradford assay (Sigma).

Labeling β-lactamase E166N mutant with penicillin derivative substrates

β-lactamase (E166N) was prepared as above, and the concentration was adjusted to 10μM. Substrate fluorescein-ampicillin and alkynyl-penicillin were dissolved in DMSO and added to the protein solution to a final concentration of 100μM. Reactions were carried out at 37^{o}C and allowed to proceed for a maximum of 3 hours. After reaction, protein samples in both reactions were precipitated by adding 4 volumes of cold acetone. For fluorescent labeling experiment, the sample was dissolved in 6M Urea and resolved by SDS-PAGE. Fluorescent gel imaging was conducted with a gel documentation system (Alpha Innotech) using UV as excitation light, after that, the gel was stained with Coommassive blue. For alkynyl-penicillin labeling experiment, the sample was resuspended in 50% acetonitrile, and 1.5μl protein sample was then mix with equal volume of matrix Sinapinic acid (10mg/ml) and examined with MS instrument (Applied Biosystem).

***in vitro* PCA experiment with purified proteins using fluorescein-ampicillin**

For *in vitro* PCA experiment, purified NLacN-SH3 and PPLP-LacC proteins were mixed with equal molar ratio and β-lactamase substrate fluorescein-ampicillin was added to 100μM. Two control experiments with the same concentration of NLacN-SH3 and PPLP-LacC were also set up individually at the same time. All reactions were maintained at 37^{o}C for 3 hours. After reaction, all samples were precipitated with cold acetone,

dissolved in Urea and resolved by SDS-PAGE. The same gel imaging system as mentioned above was used to document the signal from the PCA result, and again the gel was stained with Coommassive blue thereafter.

Target protein enrichment by "Click Chemistry" method

in vitro PCA experiment was conducted by following the same experimental procedure as stated above except that the substrate was change to 100μM alkynyl-penicillin. After that, the sample was dialyzed against 100 mM HEPES buffer, pH7.0 to remove excessive free substrate and change the buffer. The "Click Chemistry" reaction was carried out by using azide conjugated glass beads (binding capacity 136 μmol/g). In the reaction, 5mg glass beads were mixed with the dialyzed protein sample, and 1mM copper sulphate ($CuSO_4$), 1mM ascorbic acid and 0.1mM TBTA were used as a compound catalytic system. The reactants were mixed and rotated at room temperature for 1 hour. After the "Click" reaction, samples were transfered into a column and washed intensively with 2% CHAPS, 8M urea, and acetonitrile. In the end, the glass beads were rinsed by water and equilibrated with ammonium bicarbonate (40mM) and digested by trypsin (Sigma). Peptides generated from this experiment were analyzed by MS.

Mammalian cell culture and transfection

Human cell line Hela and HEK293T were grown in complete DMEM, supplemented with 10 mM HEPES, 10% fetal calf serum, 1% L-glutamine, nonessential amino acids (Invitrogen). Cells were grown in a 5% CO_2 incubator at 37°C. One day before the transfection, cells were seeded to a proper density. For live cell imaging, Hela cells were used and seeded at a low density. For protein enrichment experiment, HEK293T

cells were used and seeded to a relatively higher density so that the next day the confluence could reach 90%. At the beginning of the experiment, purified plasmid DNA was transfected into mammalian cells by Effectene reagent (Qiagen). Briefly, DNA was firstly mixed with DNA-condensation buffer EC and Enhancer was added at a mass to volume ratio of 1:8. The sample was incubated at room temperature for 5min. Effectene reagent was then added to the mixture, followed by vortex for 10 seconds. The sample was kept at room temperature for 10min. Cells were prepared during this 10min and fresh medium was added. After 10min incubation, transfection complex was first mixed with fresh media and then added to cells in a drop-wise way. After transfection, cells were maintained in 5% CO_2 incubator at 37°C until further use.

Live cell imaging with CCF2/AM substrate

Live cell imaging was carried out with Hela cells because of their exceptional adhesive property and clear morphological feature. Poly-lysine coated cell culture container (BD) were used to grow cells. Similar transfection protocol was used as mentioned above but at a smaller scale proportional to the size of the cell culture chamber. 48 hours after transfection, cells were washed with 1X HBSS (Hyclone) twice. CCF2/AM dissolved in DMSO was prepared by following manufacture's instruction, and added to the cells at a final concentration of 1μM. Cells were kept at room temperature for 2 hours and then examined with fluorescent microscope (Zeiss) with filter setting specified by the document with the substrate CCF2/AM.

Flow cytometry experiment with live cells

Cells transfected with PCA plus plasmid were maintained in DMEM medium supplemented with 10% FBS. 48 hours after transfection, cells were washed twice with 1X PBS and digested with trypsin. After digestion, DMEM medium with 10% FBS was added to stop the tryptic reaction. Cells were pipetted to make single cell suspension and then collected by centrifuging at 200g for 5min at room temperature and washed with 1X HBSS twice to remove trace amount of FBS. Cell density was adjusted to 10^6 cells/ml and stained by CCF2/AM at 1 μM at room temperature for 2 hours. After staining, cells were analyzed with FACSdiva flow cytometry instrument (BD) and Flowjo software was used to further process the data (Becton Dickinson, Mountain View, CA).

Labeling and enriching the target protein in live cells

For mammalian cell enrichment experiment, HEK293T cells were used and transfected with the PCA plus plasmid. When the confluence reached about 90% 24 hours after transfection, cells were washed with PBS, trypsinized, resuspended and seeded into 3 volumes of the original flask. After another 24~36 hours, cells were ready for next experiment. Cell permeable substrate CP-AM was dissolved in DMSO and loaded to cells transfected with PCA plus plasmid in the same way as CCF2/AM. After incubating at room temperature for two hours, cells were washed once with 1X HBSS to remove extra substrate and then lysed in 0.5% CHAPS, 1X PBS solution on ice for 30min. Next, cell lystae was separated from insoluble fraction by centrifuging at

40,000g. Target protein enrichment, digestion and identification were then carried out using the same protocol as *in vitro* PCA plus experiment stated above.

Result

Labeling β-lactamase E166N mutant with two substrates

In order to carry out the PCA plus project, a new β-lactamase substrate alkynyl-penicillin based on the experiment design was synthesized (Jiang et al, 2010). This substrate was formed by conjugating alkyne group to the amine of (+)-6-aminopenicillanic acid with a short linker. The purpose to introduce the alkyne group was to take advantage of highly specific and efficient "Click Chemistry" to enrich the target protein once it was coupled by the substrate after the PCA plus procedure (see Figure 3.3 C). At the same time, a cell permeable version of this substrate (CP-AM) was synthesized by protecting the only carboxylic acid group through the formation of an ester bond (see Figure 3.3B), which made this compound permeable to cell membrane through diffusion. Once inside live cells, substrate precursor CP-AM could be cleaved by cytoplasmic non-specific esterase to form the active substrate alkynyl-penicillin. In addition, ampicillin was conjugated with fluorescein and got substrate fluorescein-ampicillin (Figure 3.3A), which would be used to visualize *in vitro* protein labeling.

Although the alkyne group is small and β-lactamase can accommodate many different functional groups at this position, such as ampicillin, it was still necessary to test whether the E166N mutant of β-lactamase could be labeled by this substrate. So we introduced the E166N mutation to the full-length β-lactamase, expressed and purified the mutated

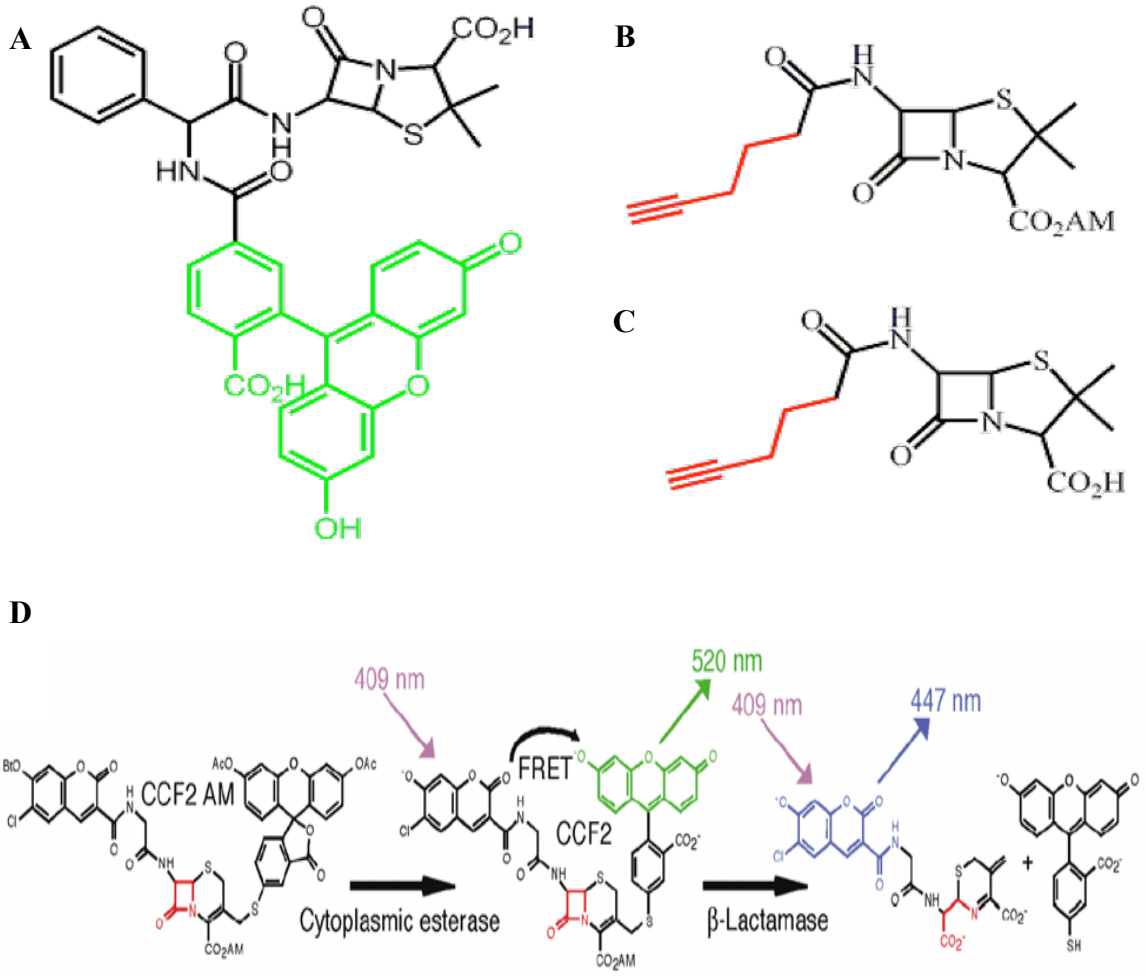


Figure 3. 3 Substrates for the β-lactamase PCA plus system.
A, fluorescein-ampicillin, which contained a fluorecein molecule and β-lactam ring. This compound was generated by coupling fluorescein-NHS ester to the amine of Ampcillin, which was used to visualize the *in vitro* PCA plus result. B, the cell permeable version of alkynyl-penicillin (CP-AM). C, alkynyl-penicillin, in which an alkyne functional group was attached to the amine of (+)-6-Aminopenicillanic acid. D, FRET based substrate CCF2/AM, which is also cell permeable. This substrate contained a FRET donor and acceptor joined by a cephalosporin ring. Once the cephalosporin ring was cleaved by β-lactamase, FRET acceptor would be released and quenched, allowing the emission light of the donor to be observed.

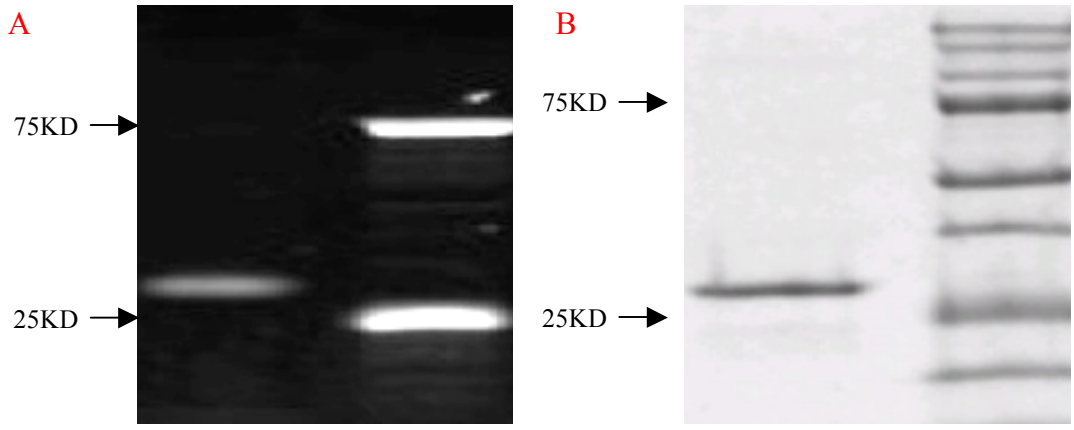


Figure 3. 4 Labeling β-lactamase E166N by fluorescein-ampicillin.
Mutated TEM-1 was expressed in *E. coli.* and purified by Ni-resin. Purified protein was incubated with fluorescein-ampicillin at 37°C for 3 hours and then precipitated. The protein sample was dissolved in urea and resolved by SDS-PAGE. The fluorescein signal from the protein band was captured with a gel imaging system (A) and the gel was stained with coomassie Blue (B).

enzyme (see Figure 3.4). As an initial step, we tested this mutated enzyme with substrate fluorescein-ampicillin. The protein was incubated with the substrate at 37°C for 3 hours, precipitated by pre-chilled acetone, and dissolved by 8M Urea. Subsequently, the protein sample was resolved by SDS-PAGE and visualized by gel documentation system. The result confirmed that the mutated β-lactamase could be labeled by the substrate fluorescein-ampicillin (See Figure 3.4).

Next, we carried out the protein labeling experiment with substrate alkynyl-penicillin. The same β-lactamase with E166N mutation was used to incubate with alkynyl-penicillin. After reaction, protein sample was precipitated by cold acetone at -20 °C and resuspended with 50% acetonitrile. At the same time, a protein sample without adding the substrate was prepared with the same protocol. In order to show that the β-lactamase with

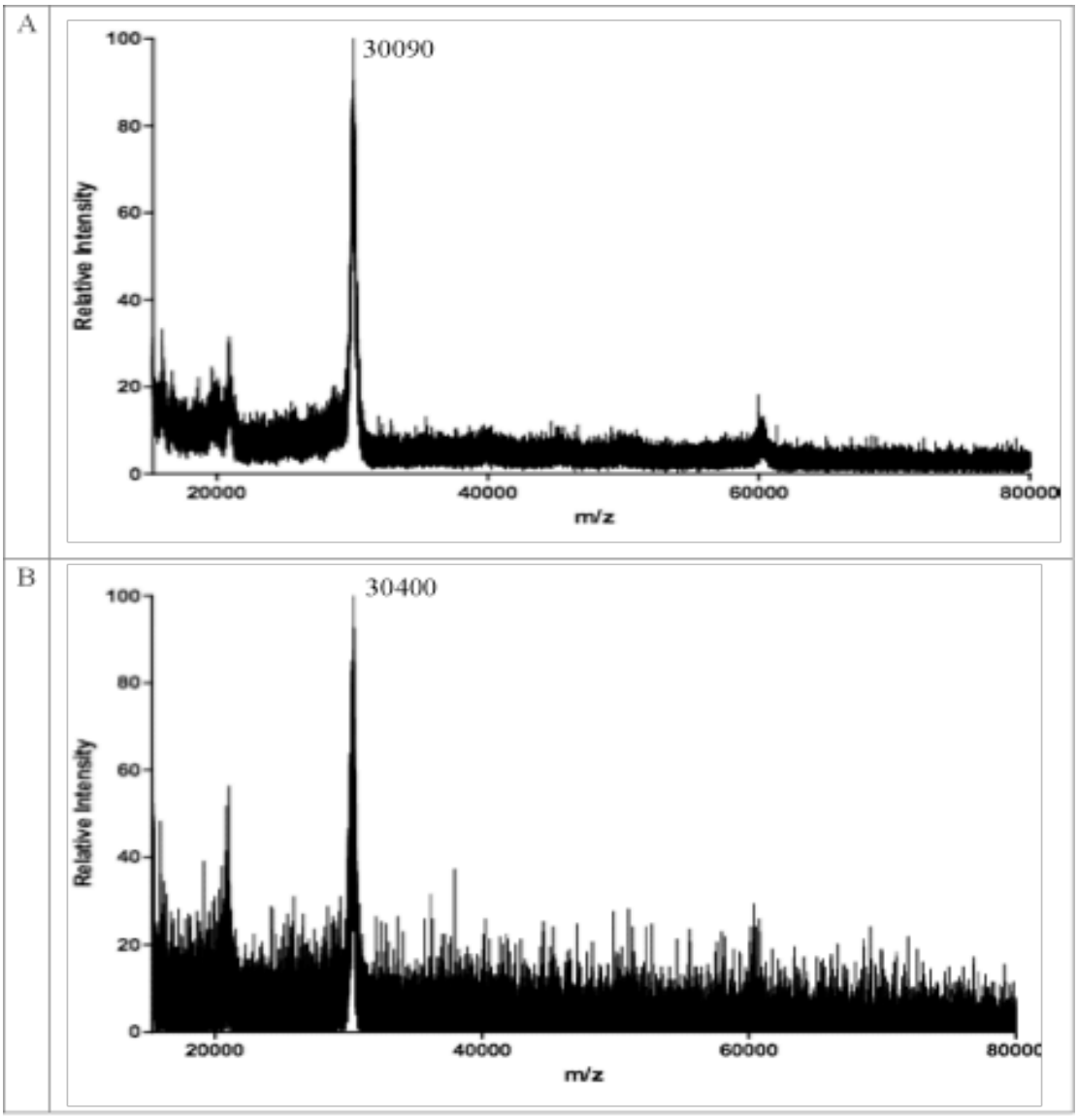


Figure 3. 5 Full-length β-lactamase labeling experiment.
MS result of full-length β-lactamase E166N alone (A) or labeled by alkynyl-penicillin substrate (B). Purified β-lactamase E166N was incubated with the substrate alkynyl-penicillin at 37^{o}C for 3hours. The sample was then precipitated by acetone, resuspended in 50% acetonitrile and examined by MS instrument. A control sample was also precipitated, resuspended and examined. For the control sample, a peak of 30090 was indicated on MS (A), compared to the predicted molecular mass 30087 for full-length β-lactamase E166N as shown in. For labeled sample, MS shows a peak of 30400, compared to the predicted molecular mass 30397 of β-lactamase E166N conjugatged with substrate alkynyl-penicillin (B).

E166N mutation could be labeled by alkynyl-penicillin, we used MS to analyze both samples (with and without adding the substrate). Our result showed that without the substrate, the mutated β-lactamase had a peak of 30090, compared to the predicted molecular weight 30087. After the labeling experiment with alkynyl-penicillin, a MS peak of 30400 was detected, compared to the predicted molecular weight 30397 (see Figure 3.5). These data demonstrated that the mutated β-lactamase with E166N could be labeled with modified substrate fluorescein-ampicillin and alkynyl-penicillin, which cleared the way for us to further study PCA plus with these substrates.

***in vitro* PCA plus using a PPI model between SH3 domain and its binding peptide**

in vitro experiment could provide critical information about a biochemical reaction, so we carried out *in vitro* PCA plus experiment with a well characterized PPI model: the SH3 domain from yeast protein Sho1p (amino acid 302-367) and its binding peptide QQIVNKPLPPLPVAGSS from protein Pbs2p (amino acid 88-104)(Marles et al, 2004). This SH3 domain was quite small with just 66 amino acids. A serial of primers were synthesized to cover the SH3 domain and used for PCR to join these primers to obtain the SH3 domain coding sequence. We also made the mutation E166N and M184T in the N-terminal fragment (corresponding to amino acid 23-196) of TEM-1 β-lactamase as E166N is a hydrolysis-deficient mutation and M184T can increase the enzyme activity and stability(Galarneau et al, 2002). This mutated N-terminal fragment of TEM-1 β-lactamase was named NLacN. We then joined the coding sequence of SH3 domain and the NLacN by a flexible linker $(G_4S)_3$ to obtain NLacN-SH3. At the same time, another serial of primers were synthesized to obtain the coding sequence for peptide

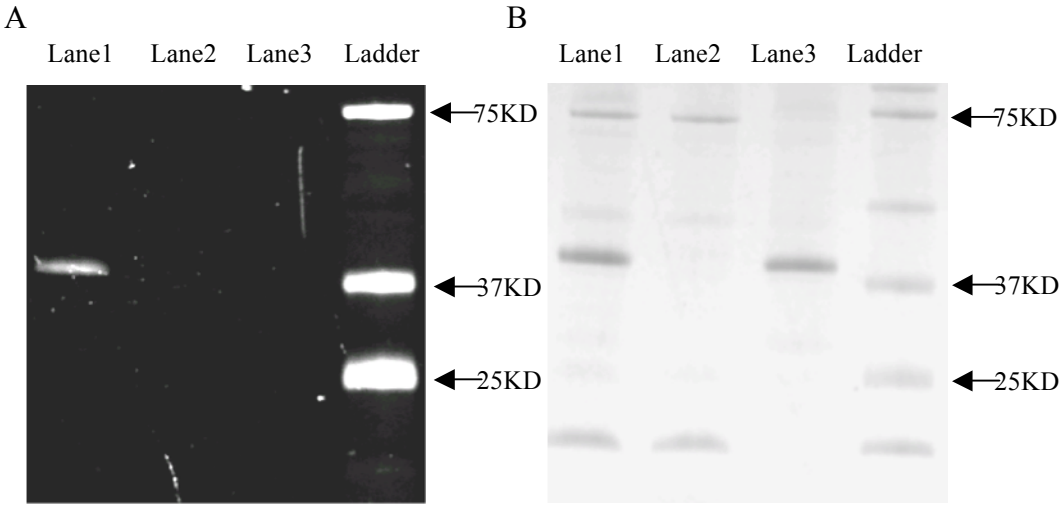


Figure 3. 6 *in vitro* PCA plus with substrate fluorescein-ampicillin.
This experiment was carried out with an interacting model SH3 domain and its binding peptide. Fluorescein-ampicillin was used as the substrate. Fusion proteins NLacN-SH3 and PPLP-LacC were expressed and purified. A serial of three reactions were set up for the *in vitro* PCA plus experiment: NLacN-SH3 and PPLP-LacC (lane 1), PPLP-LacC alone (lane 2) and NLacN-SH3 alone (lane3). All samples were incubated with fluorescein-ampicillin at 37^{o}C for three hours. After reaction, samples were precipitated by acetone, dissolved in Urea and resolved by SDS-PAGE. Fluorescein imaging was firstly carried out with a UV gel documentation system (A) and the gel was then stained with Coomassie blue (B).

(QQIVNKPLPPLPVAGSS), which was then added to the C-terminal fragment (corresponding to amino acid 198-286) of TEM-1 β-lactamase. Also, we introduced a flexible linker between these two coding sequences and got PPLP-LacC. We cloned NLacN-SH3 and PPLP-LacC into expression vectors, expressed both proteins in *E. coli* and purified them with Ni column.

Next, we used the substrate fluorescein-ampicillin to test PCA plus method *in vitro* (see Figure 3.6). Individual protein NLacN-SH3 or PPLP-LacC or both of them were incubated with the substrate at 37^{o}C for 3 hours. They were then precipitated by cold acetone, dissolved and separated by SDS-PAGE. The gel was firstly irradiated by UV

light to visualize the fluorescein labeled protein band and then stained with Coomassive blue. Our result showed that the target protein NLacN-SH3 could be labeled by fluorescein-ampicillin when mixed with PPLP-LacC protein, while either NLacN-SH3 or PPLP-LacC could not be labeled (see Figure 3.6). This result demonstrated that the PCA plus could be used to label the target protein in an *in vitro* system.

At the same time, we tested the substrate alkynyl-penicillin with the *in vitro* PCA plus using the same experiment procedure as fluorescein-ampicillin. Instead of visualizing the protein labeling with fluorescent gel imaging system, we used MS to analyze these samples. Our data showed that in the presence of both NLacN-SH3 and PPLP-LacC protein, the substrate alkynyl-penicillin could be successfully conjugated to the target protein NLacN-SH3 and generated a MS peak of 37623, compared to a predicted molecular weight of 37624 (see Figure 3.7A), which was the molecular weight of NLacN-SH3 (MW=37313) plus alkynyl-penicillin (MW=310). On the contrary, individual protein NLacN-SH3 or PPLP-LacC alone could not be labeled, and their MS peaks were 37313 and 22014 respectively, compared to predicted molecular weight 37313 and 22018 (see Figure 3.7 B and C).

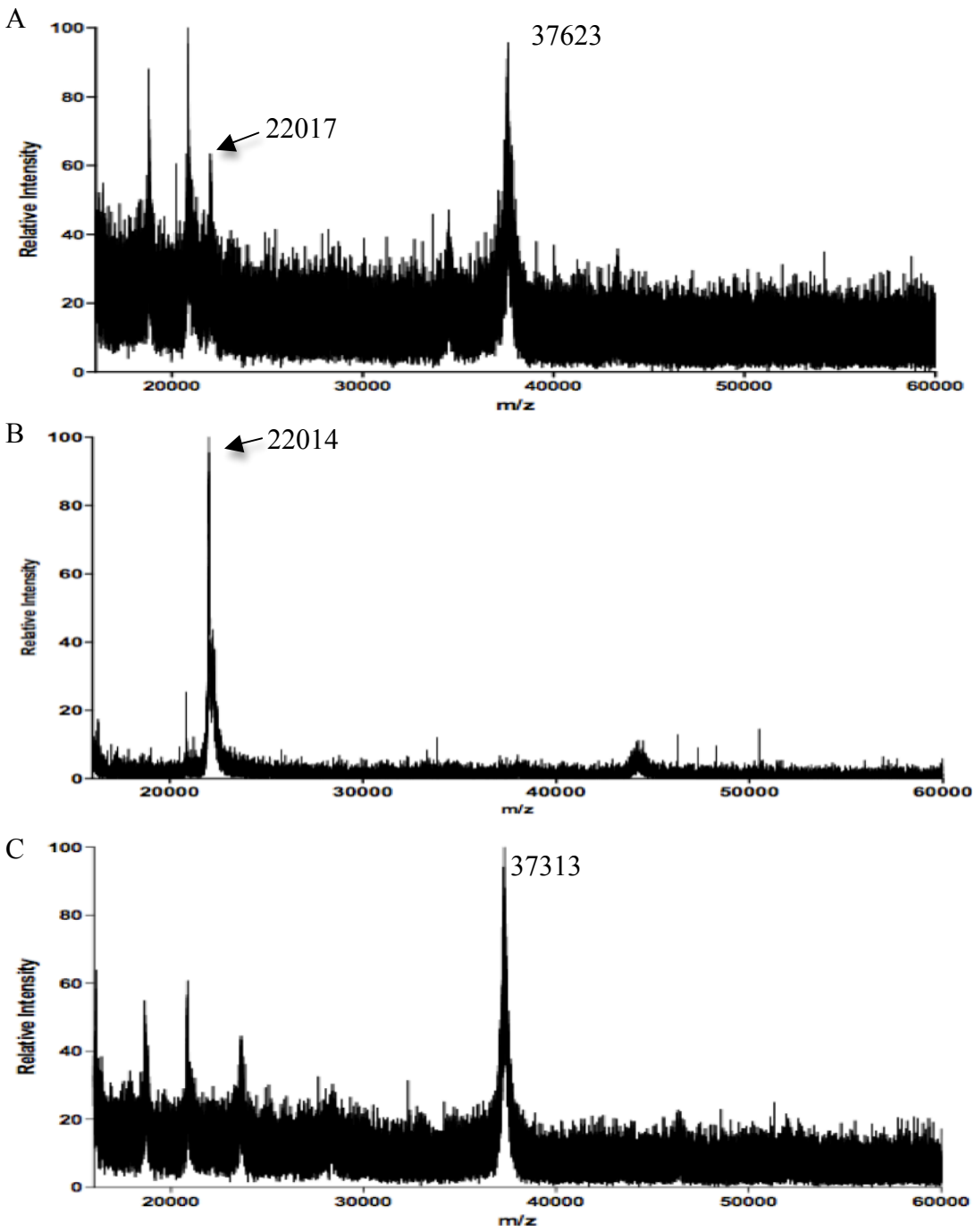


Figure 3. 7 *in vitro* PCA plus with alkynyl-penicillin to label the target protein. Purified protein NLacN-SH3 and PPLP-LacC mixture (A), PPLP-LacC alone (B), and NLacN-SH3 alone (C) were used to incubate with substrate alkynyl-penicillin in 3 parallel reactions. After labeling, protein samples were concentrated and examined by MS. In this result, only in the presence of both NLacN-SH3 and PPLP-LacC proteins, the NLacN-SH3 could be labeled by the substrate, generating a peak of 37623 (A), and we could see PPLP-LacC and labeled NLacN-SH3 peaks in (A); while NLacN-SH3 alone and PPLP-LacC alone could not be labeled by the substrate (B and C).

The purpose to use alkynyl-penicillin substrate was to facilitate target protein enrichment after it was labeled through the PCA plus method. To test this enrichment method, we prepared glass beads with azide functional groups on the surface. With this tool in hand, we used "Click Chemistry" to enrich protein labeled through *in vitro* PCA plus experiment. In the "Click" reaction, 1mM copper sulphate ($CuSO_4$) was reduced by 1mM ascorbic acid to generate Cu (I), which was further stabilized by TBTA and used as catalyst for protein enrichment. After the Huisgen cyclization "Click" reaction, the glass beads were washed with non-denaturing detergent CHAPS (2%), denaturing reagent urea (8M) and acetonitrile. In the end, the glass beads were equilibrated with ammonium bicarbonate (40mM) and digested by Trypsin, and peptides were analyzed by MS instrument. Notably, it was benefited from the specific, efficient and robust "Click Chemistry" that we were able to immobilize the target protein with the glass beads and carried out on-column digestion. The first dimension of MS analysis showed that there were several peaks that matched predicted peptides from the target protein, and further MS/MS analysis confirmed their sequences (see Figure 3. 8A and B).

Here we tested the *in vitro* PCA plus method with purified protein by using substrates fluorescent-ampicillin and alkynyl-penicillin. Our data showed that the fusion protein could be labeled by both substrates after going through the PCA plus procedure, and our protein gel imaging, protein enrichment, digestion and peptide MS analysis provided detailed information about this result.

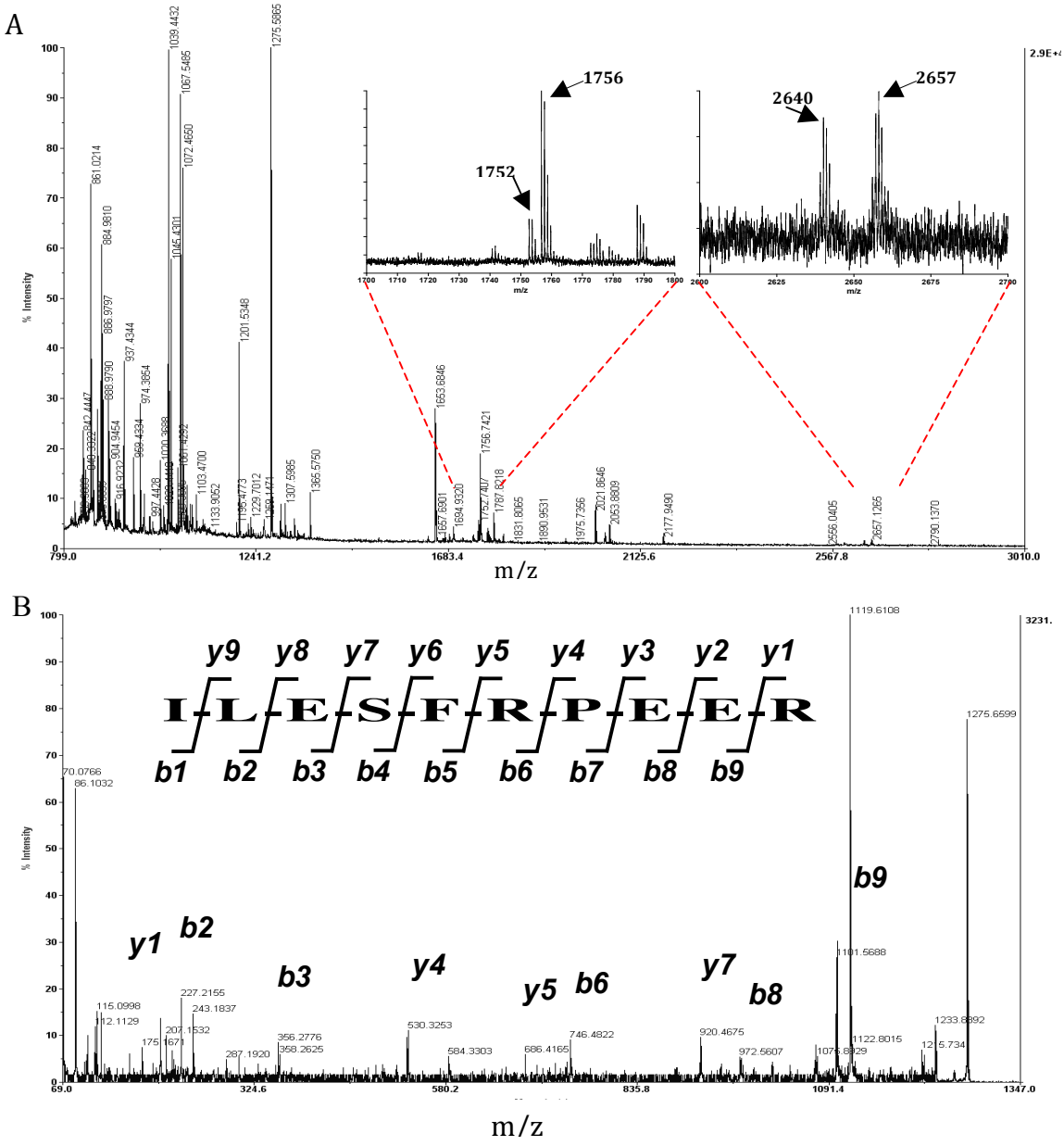


Figure 3. 8 MS analysis of enriched peptides from "Click Chemistry".
Target protein was labeled with substrate alkynyl-penicillin in an *in vitro* PCA plus experiment. The reaction condition was the same as protein labeling with substrate fluorescein-ampicillin. Glass beads coated with azide were then used to enrich labeled protein. In the end, the purified protein was digested and examined by MS instrument. A, MS of peptides derived from the enriched protein NLacN-SH3. Inlets show two peptide peaks (1752 and 2640) from SH3 domain. B, MS/MS spectrum of the precursor 1275 ion in (A).

PPI live cell imaging with PCA plus

TEM-1 β-lactamase PCA was first developed as a PPI imaging tool using FRET-based substrate CCF2/AM. However, once refolded during PCA, one active β-lactamase molecule can hydrolyze many substrate molecules. As a result, these hydrolyzed substrate molecules can diffuse to all over the cell, making it impossible to localize PPI with subcellular resolution, which actually is a very important goal for many PPI studies. With the new design of the PCA plus, it would be a totally different story. Because the N-terminal fragment of TEM-1 β-lactamase with E166N mutation can site-specifically couple the substrate molecule during the PCA plus reaction, thus label itself with a fluorescent dye when CCF2/AM is used as the substrate, enabling PPI live cell imaging with subcellular resolution.

We conducted live cell imaging with the same PPI pair used by those who firstly developed TEM-1 β-lactamase PCA. Two plasmids for the expression of fusion proteins Zip-NLacN and Zip-LacC in mammalian cell expression vector pcDNA3.1/Zeo were provided by Dr. Michnick's lab. "Zip" was a leucine zipper homodimerization domain coming from yeast transcription factor GCN4. We introduced the E166N mutation to the Zip-LacN fusion gene and obtained plasmid Zip-NLacN pcDNA3.1/zeo. Firstly, we transiently transfected plasmid Zip-NLacN pcDNA3.1/zeo and plasmid Zip-LacC pcDNA3.1/zeo into Hela cells, after 48 hours, cells were washed and incubated with substrate CCF2/AM for 2 hours at room temperature. Fluorescent microscopy examination found there were positive cells that showed blue fluorescent signal but the efficiency appeared low (Less than 10%). We reasoned that the co-transfection of two

plasmids might result in a low chance for both fusion proteins Zip-NLacN and Zip-LacC to be co-expressed in the same cell and thus lower PCA plus efficiency. To overcome this problem, we transferred the Zip-LacC fusion gene and its expression control cassette to the plasmid Zip-NLacN pcDNA3.1/Zeo. Theoretically, this single plasmid would express two fusion proteins at the same time and same level when introduced into mammalian cells. We transfected this single plasmid into mammalian cells. After following the same protocol for imaging, we found the live cell PPI imaging result was significantly improved (See actual result in Figure 3.9C), while our control cells transfected with only Zip-NLacN pcDNA3.1/Zeo plasmid or empty vector plasmid showed no significant signal (see Figure 3.9A and B).

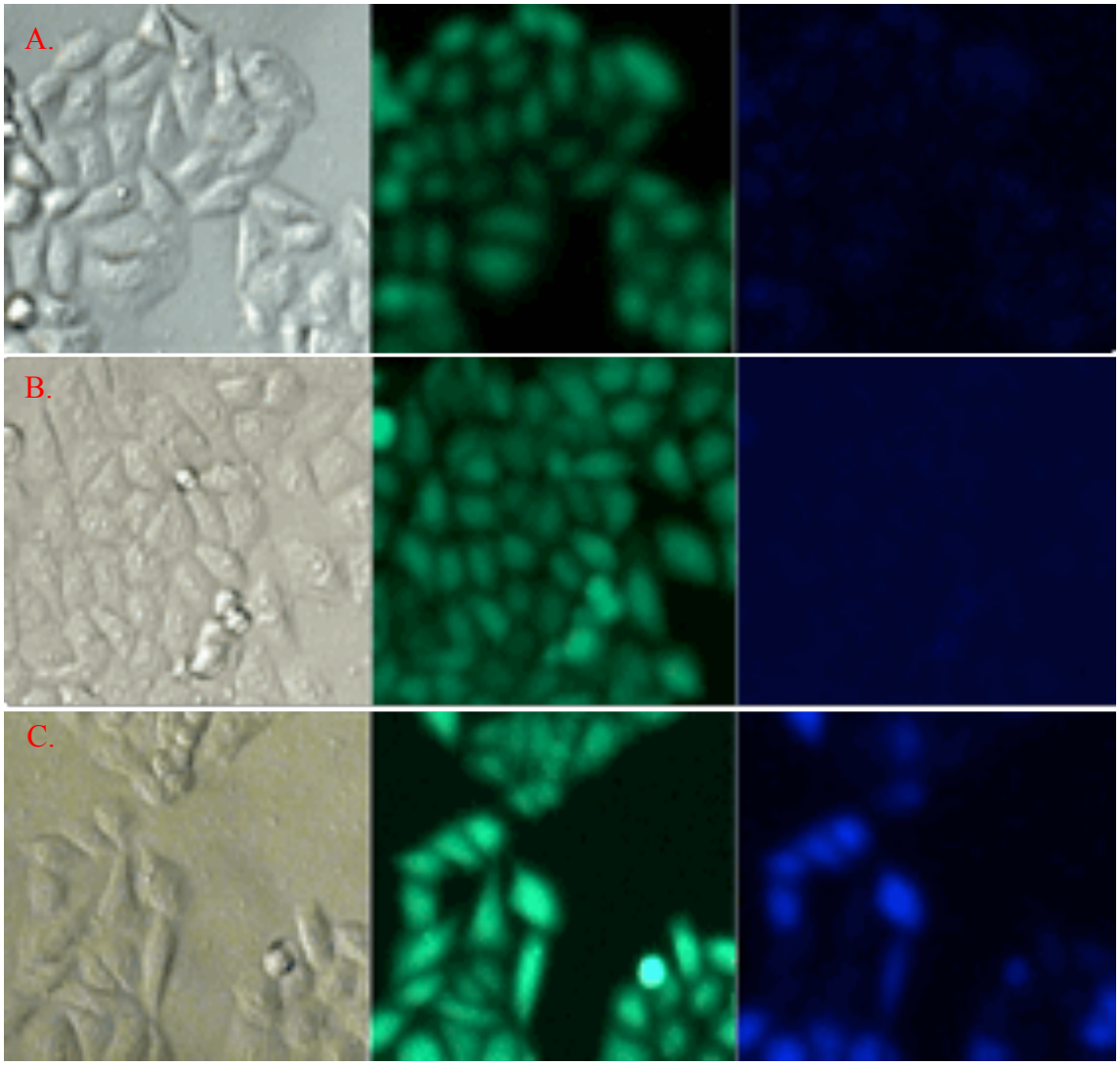


Figure 3. 9 PCA plus live cell imaging with a PPI model.
Hela cell line was used for imaging. 48 hours after transfection, cells were stained with CCF2/AM and imaged with fluorescent microscope. **A**, negative control group transfected with plasmid vector pcDNA3.1/zeo; **B,** negative control group transfected with plasmid expressing Zip-NLacN alone; **C**, experiment group transfected with plasmid expressing ZIP-NLacN and ZIP-LacC. In each group, three pictures were taken using the settings for DIC, green fluorescent signal and blue fluorescent signal. The microscope objective is set at 40X. For control group A and B, no significant blue fluorescent signal was observed, while in experiment group transfected with the plasmid expressing Zip-NLacN and Zip-LacC proteins, clear blue signal was observed. This result demonstrated that PCA plus could be used for live cell PPI imaging.

To further characterize the PCA plus imaging result, we used flow cytometry to measure the fluorescent color and intensity change. In FACS assay, we set up a serial of control groups including plasmid vector pcDNA3.1/Zeo transfection and Zip-NLacN pcDNA3.1/Zeo plasmid transfection. In the experimental group we used the single plasmid construct to express both Zip-NLacN and Zip-LacC proteins. Moreover, we transfected different amount of plasmid DNA (0.25, 0.5 and 1 μg) to test the influence of protein expression level on the efficiency of PCA plus result. For all control groups, we did not see fluorescent signal change on both intensity and color (see Figure 3.10). For experimental groups transfected with plasmid co-expressing fusion protein Zip-NLacN and Zip-LacC, significant amount of cells shift toward higher blue signal, while green signal level was the same. More interestingly, with the increase of DNA amount transfected (0.25, 0.5 and 1 μg), cells showing blue fluorescent signal increased accordingly (15.7%, 28.1%, 38.7%) (Figure 3.10). Taken together, our experiment result showed that PCA plus assay was working with current mammalian cell model, and it was compatible with fluorescent microscopy and flow cytometry analysis.

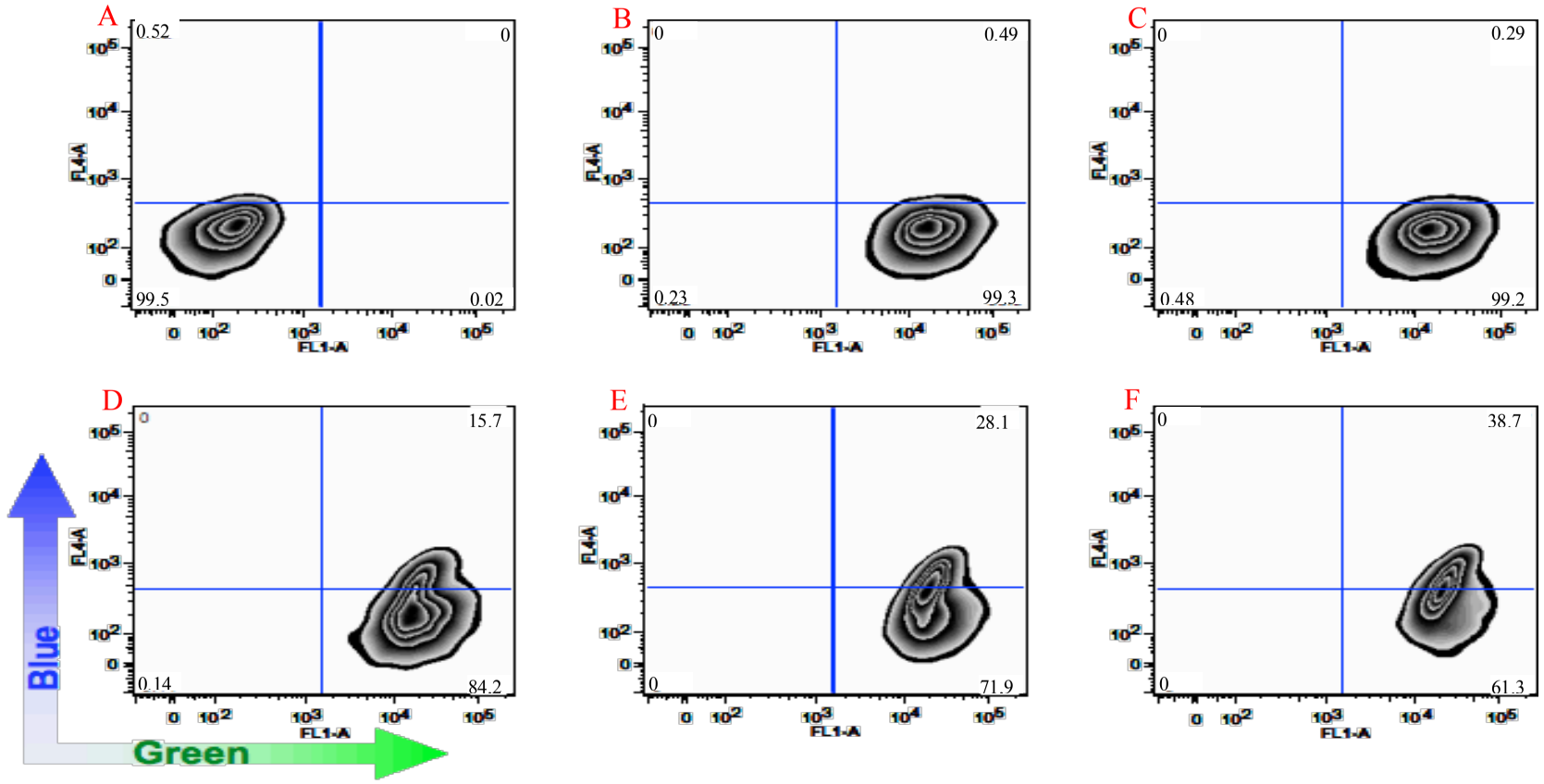


Figure 3.10 Flow cytometry analysis of PPI imaging with PCA plus method.
In order to quantify the PCA plus live cell imaging result, we conducted flow cytometry analysis. Cells were grown in 6-well plate and treated with different conditions. A, negative control and not stained; B, negative control transfected with vector pcDNA3.1/zeo and stained with CCF2/AM; C, negative control transfected with plasmid expressing Zip-NLacN protein alone and stained with CCF2/AM; D, E, F are cells transfected with 0.25, 0.5 and 1 μg of plasmid DNA expressing both Zip-NLacN and Zip-LacC, and they were all stained with CCF2/AM. In all plots, X-axis shows the intensity of Green fluorescent signal and Y-axis shows that of blue signal. In this result, we can see A, B and C show background blue signal, while D, E, F show cell population shift on Y-axis. Notably, the blue signal in D, E and F increases with the increasing amount of plasmid DNA transfected, suggesting the blue signal is related to Zip-NLacN and Zip-LacC expression level. Shown here is a representative result from three indipendent experiments.

PCA plus can image PPI with subcellular resolution

Protein spatial localization is an important regulatory mechanism in many biological processes, and PPI localization is even more critical for studying protein cellular functions. On the other hand, PPIs are dynamic events inside a cell. So it will be very useful if a tool can readily reveal these critical characters of PPIs without using costly instrument and highly specialized technologies like single molecule imaging or NMR. Although the PCA plus method hold the potential for both PPI live cell imaging with subcellular resolution and PPI dynamic imaging, here we wanted to first demonstrate that we could image PPI in a specific organelle, that is, PPI subcellular localization.

In order to test whether the PCA plus could localize PPI in an organelle, we added a Nuclear Localization Signal (NLS) to our PCA plus model Zip-NLacN and Zip-LacC. A sequence analysis revealed that yeast GCN4 protein has a NLS element, which is DPAALKRARNTEAARRSRARKLQRMK. We synthesized primers to add this NLS to the fusion protein Zip-NLacN and Zip-LacC to obtain NLS-Zip-NLacN and NLS-Zip-LacC. As always, we constructed one plasmid for the expression of NLS-Zip-NLacN and NLS-Zip-LacC. In this plasmid, each fusion gene was controlled by an individual CMV promoter. We transfected this plasmid into Hela cells and followed the same experiment procedure for CCF2/AM imaging as mentioned above. Interestingly, we indeed observed that cells showed blue signal in a subcellular organelle when transfected with NLS-Zip-NLacN and NLS-Zip-LacC expression plasmid (See Figure 3.11C), while control cells showed no significant signal in any specific subcellular position (See Figure 3.11A and B). These data demonstrated that the PCA plus strategy could indeed be used for PPI subcellular imaging.

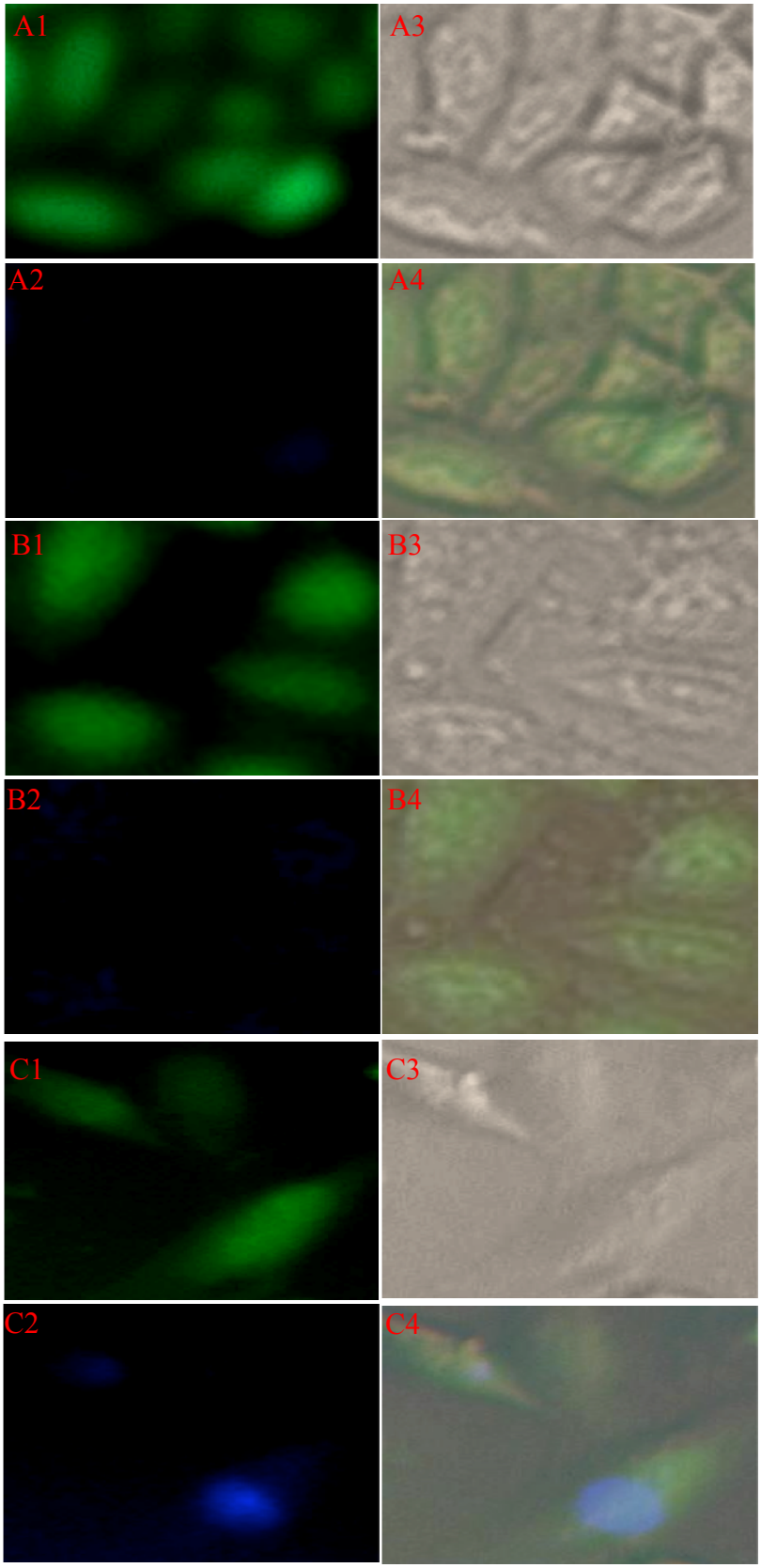


Figure 3.11 Nuclear PPI imaged by PCA plus method in live cells with CCF2/AM.

Live cell imaging was carried out with the proteins with NLS expressed in Hela cells. A, cells transfected with only plasmid vector; B, cells transfected with plasmid for the expression of NLS-Zip-NLacN only; C, cells transfected with plasmid expressing both NLS-Zip-NLacN and NLS-Zip-LacC. For each experiment group, three pictures were taken for green fluorescent (1), blue fluorescent signal (2), and DIC (3), followed by a merge of all three channels (4). As a result, we can only see blue fluorescent signal from the group C. Most importantly, the blue signal is limited to a small area of the cell, showing subcellular localization of leucine zipper domain (Zip) dimerization inside the live cells.

The PCA plus method can enrich interacting protein from live cells

High-throughput and high-fidelity characterization of PPI events is an ultimate goal for the development of a PPI study tool. However, conventional PCA systems have seldom been used for high-throughput PPI screening in live mammalian cells despite their advantage of detecting PPI under physiological conditions. That might suggest there are some limitations that hindered this method from the high-throughput applications. On the contrary, the PCA plus method should label the interacting protein with an affinity tag, which would make it feasible to enrich the target protein by affinity purification. When combined with powerful LC-MS/MS system, hopefully, this method could realize high-throughput PPI screening in mammalian cells.

In order to demonstrate that the PCA plus was potential for high-throughput PPI screening, we carried out *in vivo* labeling, *in vitro* pull-down and MS identification experiments with this method. We transfected the plasmid to express both Zip-NLacC and Zip-LacC proteins in mammalian cells. After 48 hours, cells were loaded with cell permeable substrate CP-AM (alkynyl-penicillin), in which a protection Group (Acetoxymethyl) was used to block the carboxylic acid group in penicillin (See Figure 3.3). Two hours later, substrate was removed by an extensive wash and cells were lysed on ice by adding lysis buffer (PBS, 0.5% CHAPS). Next, target proteins were enriched by following "Click Chemistry" protocol: incubating the sample with azide beads in the presence of 1mM ascorbic acid, 1mM Copper (II) sulfate and 0.5mM TBTA at room temperature for 1 hour. The beads were then thoroughly washed and proteins were digested with trypsin. Subsequently, LC-MS/MS system was used to separate the

peptides generated from the enrichment and the target peptides were identified by searching the database with Mascot.

Interestingly, our LC-MS/MS result showed that several peptides from the target protein Zip-NLacN were identified (See Figure 3.12C). The tandem MS spectrum of one of these peptides, IDAGQEQLGR was illustrated (Figure 3.12 B). More importantly, a peptide "NYHLENEVAR" from the leucine zipper domain was also identified (See Figure 3.12C). In this experiment, we demonstrated that PCA plus method could be used to label and enrich interacting proteins from live cells, and it was compatible with proteomics method for protein identification. Although this was a proof-of-concept experiment and there is a long way to go for high-throughput library screening, we were encouraged by the potential of this method.

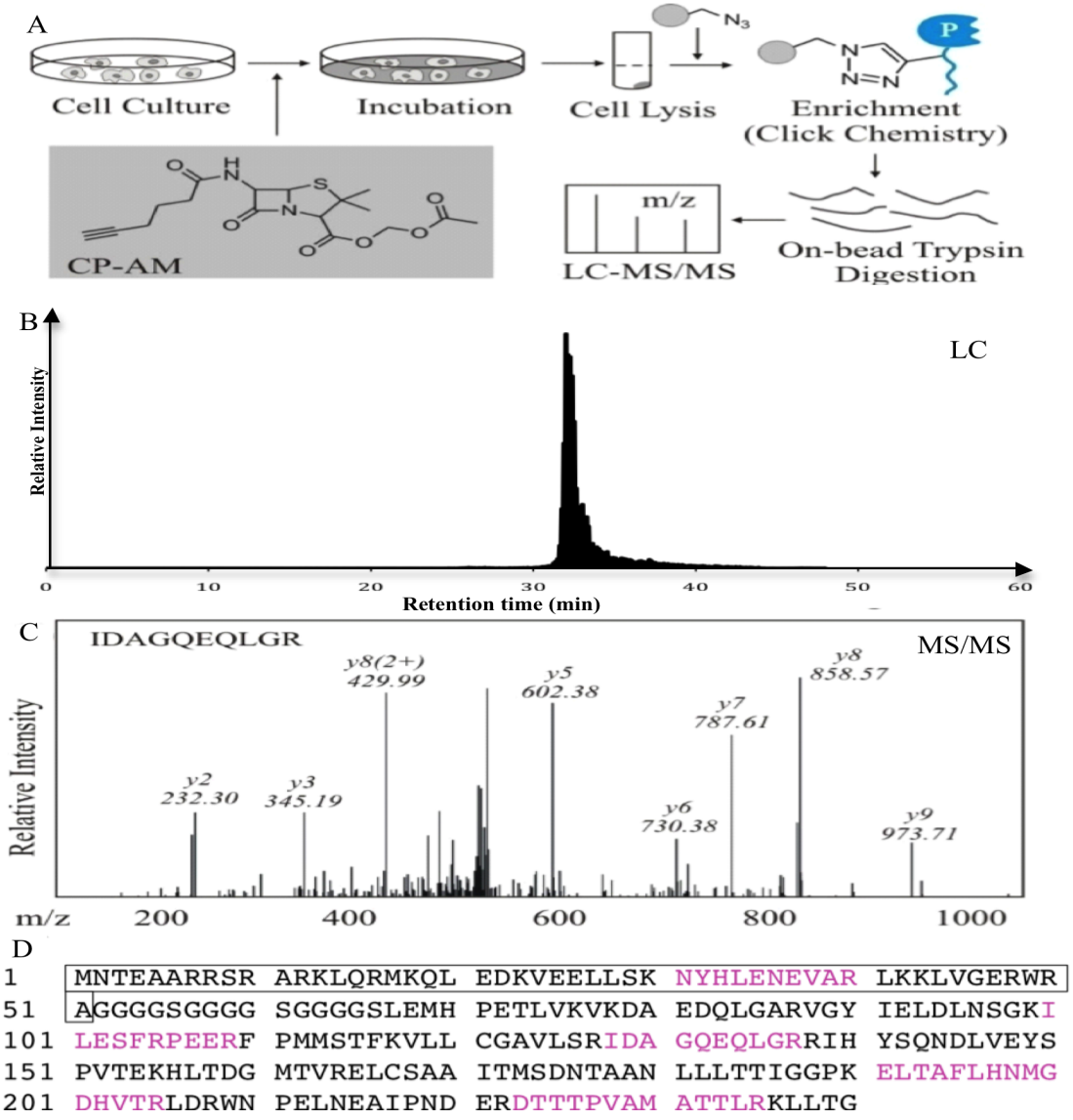


Figure 3. 12 Interacting protein enrichment and identification with PCA plus.
A, a scheme of the experimental procedure. Cells were first transfected with plasmid to express protein Zip-NLacN and Zip-LacC. Subsequently, substrate CP-AM was loaded followed by cell lysis and protein enrichment through "Click Chemistry". In the end, target protein was digested and peptides were analyzed by LC-MS/MS. B, LC graph for the peptides obtained from enriched protein sample. C, Tandem MS spectrum generated from one of the peptides. The spectrum was consistent with manual fragmentation of peptide "IDAGQEQLGR", and b- and y- ions were labeled. D, Amino acid sequence of Zip-NLacN. Tryptic peptides identified by LC-MS/MS and Mascot were colored in red. The sequence in the box was from leucine zipper domain, and a peptide from this domain (m/z 1243.6) was also identified by Mascot.

Discussion

In this project, a new method "the PCA plus" was developed based on β-lactamase PCA strategy to address challenges for PPI characterization. By introducing a point mutation E166N to the N-terminus of β-lactamase, the PCA method was modified to enable subcellular PPI localization and MS based PPI screening. In addition, the new method has the potential to detect low affinity and transient PPIs.

The concept of the PCA plus was firstly demonstrated by an *in vitro* experiment. Purified fusion proteins NLacN-SH3 and PPLP-LacC were used to react with two penicillin-derived substrates, fluorescein-ampicillin and alkynyl-penicillin. After the PCA plus procedure, the prey protein NLacN-SH3 could be labeled by both substrates. This *in vitro* system was important because it provided the most direct information about this method compared to cell-based assays. Also, it could serve as a platform for the future optimization and analysis of the PCA plus strategy.

After *in vitro* exploration, we used PCA plus to determine whether we could visualize homodimerization of a leucine zipper domain in live cell imaging experiments. These experiments took great advantage of a remarkable substrate of β-lactamase: CCF2/AM (See Figure3.3D). This substrate is cell permeable and can be trapped inside cells for a couple of days without significantly affecting cellular functions (Zlokarnik et al, 1998). Also, this substrate undergoes large color change to fluorescently green upon entry into the cell and a distinct change to fluorescently blue upon hydrolysis by β-lactamase (Zlokarnik et al, 1998). These fluorescent changes allowed us to monitor uptake by

mammalian cells and to visualize its covalent attachment to the prey using fluorescent microscopy. Flow cytometry experiments further confirmed the conclusion that bait-prey interaction resulted in an easy detected fluorescence signal. Subsequently, we asked a question: Could the PCA plus detect a PPI with specific subcellular localization. To answer this question, a NLS was added to the same set of fusion proteins used above and the PCA plus system was tested with these nuclear proteins. Interestingly, after going through the PCA plus imaging procedure, PPI signal was detected from a specific location inside a live cell. Although we did not directly demonstrate that this location was, in fact, the nucleus, our results are certainly consistent with this possibility and they suggested that PCA plus could be useful for PPI subcellular localization analysis.

Lastly, the PCA plus method was tested as an approach for pulling interacting protein out of whole cell extracts of mammalian cells. In order to facilitate this protein affinity enrichment, a cell permeable substrate alkynyl-penicillin (CP-AM) was applied. Also, a plasmid that expressed both bait and prey fusion proteins in the same cell and at the same level was used. This could significantly improve the efficiency of PCA plus (Fujimoto et al, 2001; Lee et al, 2009). In the end, the target protein was labeled inside live cells through the PCA plus method, enriched by "Click Chemistry" and identified by LC-MS/MS analysis (Figure 3.12).

An underlying concern with the PCA plus method was whether sufficient protein could be enriched for MS analysis. In order to answer this question, a comparison between conventional TAP method and PCA plus was required. Based on a review

(Burckstummer et al, 2006), a typical TAP experiment requires 5 X 10^7 mammalian cells, equivalent to 35mg total protein with 70μg recombinant bait protein. After two rounds of affinity purification, about 2μg bait protein and 7μg prey proteins (0.2~4μg each) can be enriched. Subsequently, SDS-PAGE is used to resolve protein complex followed by protein band recovery and in gel digestion. There are two issues related to this procedure. (i), proteins are diluted during cell lysis, column wash and TEV cleavage, and are maintained for a long time under this condition, which can cause the protein complex to dissociate according to PPI kinetics. (ii), in-gel digestion can further decrease the overall efficiency of the method, because the yield of this step is about 20% (Kocher & Superti-Furga, 2007). Compared to the TAP method, our PCA plus only need one step of purification, which is based on covalent label of target protein and highly efficient "Click Chemistry". Therefore, if low efficient TAP method can enrich enough protein for MS analysis from mammalian cells, a more robust PCA plus method should be able to achieve the same goal, which has actually been demonstrated by our LC-MS/MS experiment (Figure 3. 12).

Chapter Four,

Summary and Perspective

PPI studies provide important information for many fields in life science research, including proteomics, cell biology, molecular genetics, biophysics and biochemistry. An ideal approach for PPI analysis would combine high fidelity and high-throughput PPI detection. High fidelity means detecting a PPI with physiological implications, and validating the physiological role of any PPI can require a lot of follow up experiments, that are usually the focus of hypothesis-driven projects. On the other hand, technique-driven projects, such as genome wide PPI mapping with the Y2H system, generally focus on exploring the likelihood of protein association based on their physical structures but not on their physiological functions. High-throughput PPI detection is another important issue in this field. Although hypothesis-driven PPI studies are more reliable, the most cost-effective and timesaving manner for PPI research is still high-throughput PPI screening, given the huge workload for PPI network mapping. The dilemma between these two aspects of current PPI research must be addressed before comprehensive PPI network mapping is achieved.

The two projects carried out in this thesis research represent our attempts to develop PPI tools that can provide both high fidelity and high throughput. Although both projects are still at their primary stages, they appear to have great potential. For the photoaffinity labeling project, in the future we can make use of the unnatural amino acid system to incorporate the photoactivable moiety into a bait protein and label its interacting proteins inside cells. Previously, through a pair of amber tRNA and tRNA synthetase (tRNA/ RS),

which can load an unnatural amino acid to an amber tRNA in a cell, the unnatural amino acid o-nitrobenzyl-tyrosine (oNBTyr) has been incorporated into a protein at the position corresponding to the amber nonsense codon. This approach has been used successfully to incorporate ^{15}N-labeled oNBTyr into a protein to facilitate its structural analysis by NMR (Cellitti et al, 2008; Deiters et al, 2006). Because oNBTyr is very similar to our compound PC-1 (See Figure 2.3), it is very likely that we can incorporate PC-1 into a bait protein using the same approach. In this way, we can generate the PC-1 labeled bait protein in cells and use UV light to trigger its crosslinking to interacting proteins. Alternatively, we can make use of site-specific protein labeling technology (Chen & Ting, 2005) to attach PC-1 or PC-2 to a bait protein prior to crosslinking it to interacting partner inside a cell. Currently, there are many efficient systems available to incorporate small molecules into a target protein through site-specific protein labeling, such as FlAsH (Souza, 2007), Halo tag (Los et al, 2008), Sortase (Popp et al, 2009) and PRIME (Uttamapinant et al, 2010). In addition, directly adding PC-1 or PC-2 to a peptide ligand during synthesis and then using this ligand to characterize cell surface receptors through UV triggered protein crosslinking is another option for applying our photo-caging crosslinker.

There are also many possible future applications of the PCA plus method. Firstly, we can apply this method for high-throughput PPI screening. Our result indicates that it is best to express bait-LacC and prey-NLacN in one plasmid construct to ensure equal expression levels of bait and prey. Thus, our starting vector would carry the bait-LacC fusion gene, and we would insert into this plasmid a library of prey-NLacN fusions.

Some practical issues for the PCA plus method are: how many cells to use and how much plasmid to transfect into cells. To address these issues, it might be necessary to systematically test different parameters before carrying out a large-scale PPI screening.

Another potential future direction for our PCA plus method would be to characterize transient and low affinity PPIs. These types of PPIs are not easily detected by Y2H or TAP methods. Because PCA plus method may covalently attach a tag to even transiently or weakly interacting proteins, it will be very exciting to see how it can be used to study these challenging PPIs. Currently, we are working on using PCA plus to identify protein substrates for a protein tyrosine phosphotase SHP2 (Mohi & Neel, 2007). These SHP2-substrate protein interactions are expected to be transient. Through this effort, we hope to discover some new targets for SHP2, a critical protein involved in growth factor and integrin signaling. More importantly, in this ongoing study we set the goal to further develop the PCA plus and add new features to it by applying some newly designed substrates for β-lactamase.

Although we are confident of the future application of both methods, we recognize they have limitations. In order to better use them in different situations, we want to clarify some conditions that may affect their effectiveness. For the photocrosslinking method, one of the concerns is self-crosslinking to the bait protein when it is applied to identify the unknown interacting proteins. For example, if we incorporate the PC-1 compound to a bait as an unnatural amino acid and apply it in a cell, there is a chance that the

deprotected crosslinker will conjugate the amino acid residues, especially cysteine residue, on the bait protein, because these amino acid residues may be closer to the crosslinker than those on the surface of prey proteins. To circumvent this problem, the location of the PC-1 in the bait protein 3-dimensional structure must be carefully designed to avoid positioning it next to a nucleophile group on the bait protein surface. Another concern for our photocrosslinking approach is the method used to load the photocaged crosslinker onto the bait. Although we have discussed that there are many available methods for achieving this goal, they all involve some chemical synthesis, for example loading the compound to a ligand peptide, which may hinder their use by laboratories lacking expertise in chemical synthesis. Our last concern about the crosslinking method is the stability and reversibility of the crosslinked product under MS conditions. Reversibility of quinone methide based crosslinking is an interesting phenomenon that has been widely investigated (Rokita, 2009b). From our crosslinking site mapping experiment, we found that a portion of crosslinked peptides from both crosslinked sites could shed off the crosslinking footprint during MS analysis (See Figure 2. 11). This feature may facilitate the identification of crosslinked peptides with MS. On the other hand, it reminds us that the crosslinked peptide may not be stable or undergo reverse crosslinking under special conditions, and either possibility will cause the decrease of MS signal intensity and will affect the efficiency of characterizing the crosslinking site on the target protein.

For the PCA plus method, there are also several critical issues that need special attention. Firstly, the sensitivity of the PCA plus is not as good as conventional PCA. Because the

PCA plus uses a hydrolysis deficient enzyme to label the target protein, the PPI signal will not be amplified in this system. On the contrary, conventional PCA systems use very efficient enzymes to amplify PPI signal, they are the best choice if detecting sensitivity is the most important requirement for a PPI study, such as the study of low abundant PPIs. Secondly, protein overexpression may cause false positive for the PCA plus method. This method is built on protein overexpression strategy for the detection of PPI inside live cells. This overexpression may give rise to artifacts (false positives), and this possibility can be even more significant when one realizes that the local concentration of recombinant protein in a cellular compartment may be much higher than the average protein concentration in a cell. Thus, if the bait and prey protein happen to be concentrated in the same organelle, although they are physiologically irrelevant to each other, the chance to generate false positive result will be increased significantly, which should be taken into consideration when we apply this method to PPI detections. Using a weaker promoter for the PCA plus method or a validation procedure after high-throughput PPI screening may be necessary in order to solve this problem. Thirdly, fusion protein orientation may affect the PCA plus result and generate false negative results. In the PCA plus method, bait and prey will be fused to the C and N terminal of β-lactamase respectively. It is possible that bait and prey interaction will make the two fragments of β-lactamase far from each other if the orientation of fusion protein is not appropriate. This type of false negative result can be corrected by switching the orientation of one of the fusion proteins. Another potential concern is that fusing a fragment of β-lactamase protein to one end of the target protein may interfere with its function. In this case, we may also have to change fusion protein orientation. Lastly,

unhealthy cells may affect CCF2/AM de-esteration and generate strong background signal for live cell imaging, which is a general concern for using this substrate for β-lactamase (Remy et al, 2007). Thus, we recommend using very healthy cells for this experiment (for example, low passage cells and cells with good morphological characteristics) and avoiding exposing cells under harsh conditions for too long.

Appendix

1. Primers used in both studies

ID	NAME	SEQUENCE
OSL109	E166NF	5’-GATCGTTGGAACCCGGAGCTG-3’
OSL110	E166NR	5’-CAGCTCCGGGTTCCAACGATC-3’
OSL020	LAC196N	5’-AACTCGAGGCAGAATTCGCTACC ACCACCACCGCCAGTTAATAG-3’
OSL039	3MLACNN-SacII	5’–AAGGCTCCGCGGTGGTCACCCAGAAAC-3’
OSL082	SH3#1	5’-GTGCTTTTGCCTTGTAAATGCTACCACCAC CACCGCTACCACCACCACCGCTACCACCAC-3’
OSL083	SH3#2	5’- CAAAAGCACTGTACCCATACGACGCTGAT GACGATGATGCTTACGAAATCTCTTTTGAAC -3’
OSL084	SH3#3	5’- GCCTTCCACCAGCGGCCTTCAATGTCAGA GACTTGCAGGATTTCATTTTGTTCAAAAGAG -3’
OSL085	SH3#4	5’- GTGGAAGGCACGCCGTGCAAACGGTGAAA CGGGTATTATTCCAAGCAATTATGTTCAACT-3’
OSL086	SH3#5	5’- GCAGCCGGATCCTTAACGATGCATTTCTTC TGGACCATCGATCAGTTGAACAT-3’
OSL112	LINKER REVERSE	5’- TGAACCTCCACCTCCGGACCCACCACCT CCAGAGCCACCGCCACC-3’
OSL111	LINKER FORWARD	5’- GGTGGCGGTGGCTCTGGAGGTGGT GGGTCCGGAGGTGGAGGTTCA-3’
OSL087	PLAC1	5’- AAGGCTCCGCGGTGGTCAGCAGAT TGTTAATAAGCCGCTGCCGCCGCTGCCGGTAGC-3’
OSL088	PLAC2	5’-CCGCTGCCGGTAGCAGGCAGCTCTG GTGGTGGTGGTAGCGGTGGTGGTGGTAGCGGT-3’
OSL035	LACCC BamHI	5’- GCAGCCGGATCCTTACCAATGCTTAATCAG -3'
OSL021	LAC198C	5’-GTGGTAGCGGTGGTGGTGGTAGCACTA GTCTACTTACTCTAGC -3
OSL100	T7PROBamHI	5’-TATCGGGGATCCGAAATTAATACGACTCACTATAG-3’
OSL105	BGH REVERSE	5’-CTAGAAGGCACAGTCGAGGC-3’
OSL136	2IN1F	5'-ATCGGGAGATCTCCCGATCCCCTATGG-3'
OSL137	2INR	5’-CAGCTGAGATCTCCAGCTGGTTCTTTCCGCC-3’
OSL038	CATPTP1BF	5'-AAGGCTCCGCGGTGGTGAGATGGAAAAG-3'
OSL041	CATPTP1BR	5’- CATGTGGATCCTTAATTGTGTGGCTCCAGG-3’
OSL042	C215S PTP1BF	5’-GTGGTGCACTCCAGTGCAG-3’
OSL043	C215S PTP1BR	5’- CTGCACTGGAGTGCACCAC - 3’
	NDE-LAC-5’	5’-GCGGGAATTCCATATGCACCCAGAAACGCTGGTG-3’
	XHO-LAC-3’	5’- CAATCCTCCGCTCGAGCCAATGCTTAATCAGTGAG-3’

2. Publications

Jiangsong Jiang, Junxiang Zhang, Shuwei Li (2010) Detecting Protein Interactions in Live Cells via Complementation of a Hydrolysis-Deficient ß-Lactamase. ***Chem. Commun***. DOI:10.1039/C0CC01998D,

Jiangsong Jiang, Dexing Zeng, Shuwei Li (2009) Photogenerated quinone methides as protein affinity labeling reagents. ***ChemBioChem*** 10(4): 635-638

Reference

Adachi H, Ohta T, Matsuzawa H (1991) Site-directed mutants, at position 166, of RTEM-1 beta-lactamase that form a stable acyl-enzyme intermediate with penicillin. *J Biol Chem* 266(5): 3186-3191

Archakov AI, Govorun VM, Dubanov AV, Ivanov YD, Veselovsky AV, Lewi P, Janssen P (2003) Protein-protein interactions as a target for drugs in proteomics. *Proteomics* 3(4): 380-391

Barlow DJ, Thornton JM (1986) The distribution of charged groups in proteins. *Biopolymers* 25(9): 1717-1733

Bock VD, Hiemstra H, Maarseveen JHv (2006) CuI-Catalyzed Alkyne–Azide "Click" Cycloadditions from a Mechanistic and Synthetic Perspective. *European Journal of Organic Chemistry* 2006(1): 51-68

Bogan AA, Thorn KS (1998) Anatomy of hot spots in protein interfaces. *J Mol Biol* 280(1): 1-9

Borch J, Jorgensen TJ, Roepstorff P (2005) Mass spectrometric analysis of protein interactions. *Curr Opin Chem Biol* 9(5): 509-516

Boxem M, Maliga Z, Klitgord N, Li N, Lemmens I, Mana M, de Lichtervelde L, Mul JD, van de Peut D, Devos M, Simonis N, Yildirim MA, Cokol M, Kao HL, de Smet AS, Wang H, Schlaitz AL, Hao T, Milstein S, Fan C, Tipsword M, Drew K, Galli M, Rhrissorrakrai K, Drechsel D, Koller D, Roth FP, Iakoucheva LM, Dunker AK, Bonneau R, Gunsalus KC, Hill DE, Piano F, Tavernier J, van den Heuvel S, Hyman AA, Vidal M (2008) A protein domain-based interactome network for C. elegans early embryogenesis. *Cell* 134(3): 534-545

Brady GP, Sharp KA (1997) Entropy in protein folding and in protein-protein interactions. *Curr Opin Struct Biol* 7(2): 215-221

Bremer HJ, Duran M, Kamerling H, Rozyrenbek H, Wadman SK (1981) *Disturbances of Amino Acid Metabolism: Clinical Chemistry and Diagnosis*: Baltimore : Urban & Schwarzenberg.

Brent R, Finley RL, Jr. (1997) Understanding gene and allele function with two-hybrid methods. *Annu Rev Genet* 31: 663-704

Burckstummer T, Bennett KL, Preradovic A, Schutze G, Hantschel O, Superti-Furga G, Bauch A (2006) An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells. *Nat Methods* 3(12): 1013-1019

Burdine L, Gillette TG, Lin HJ, Kodadek T (2004) Periodate-triggered cross-linking of DOPA-containing peptide-protein complexes. *J Am Chem Soc* 126(37): 11442-11443

Cellitti SE, Jones DH, Lagpacan L, Hao X, Zhang Q, Hu H, Brittain SM, Brinker A, Caldwell J, Bursulaya B, Spraggon G, Brock A, Ryu Y, Uno T, Schultz PG, Geierstanger BH (2008) In vivo incorporation of unnatural amino acids to probe structure, dynamics, and ligand binding in a large protein by nuclear magnetic resonance spectroscopy. *J Am Chem Soc* 130(29): 9268-9281

Chen I, Ting AY (2005) Site-specific labeling of proteins with small molecules in live cells. *Curr Opin Biotechnol* 16(1): 35-40

Chiang Y, Kresge† AJ, Zhu Y (2000) Reactive intermediates. Some chemistry of quinone methides. *Pure applied chemistry* 72(12): 2299-2308

Chowdhry V, Westheimer FH (1979) Photoaffinity labeling of biological systems. *Annu Rev Biochem* 48: 293-325

Christensen H, Martin MT, Waley SG (1990) Beta-lactamases as fully efficient enzymes. Determination of all the rate constants in the acyl-enzyme mechanism. *Biochem J* 266(3): 853-861

Clackson T, Wells JA (1995) A hot spot of binding energy in a hormone-receptor interface. *Science* 267(5196): 383-386

Cokol M, Nair R, Rost B (2000) Finding nuclear localization signals. *EMBO Rep* 1(5): 411-415

Dalbadie-McFarland G, Cohen LW, Riggs AD, Morin C, Itakura K, Richards JH (1982) Oligonucleotide-directed mutagenesis as a general and powerful method for studies of protein function. *Proc Natl Acad Sci U S A* 79(21): 6409-6413

Danial NN, Gramm CF, Scorrano L, Zhang CY, Krauss S, Ranger AM, Datta SR, Greenberg ME, Licklider LJ, Lowell BB, Gygi SP, Korsmeyer SJ (2003) BAD and glucokinase reside in a mitochondrial complex that integrates glycolysis and apoptosis. *Nature* 424(6951): 952-956

Davies C, White SW, Nicholas RA (2001) Crystal structure of a deacylation-defective mutant of penicillin-binding protein 5 at 2.3-A resolution. *J Biol Chem* 276(1): 616-623

Deane CM, Salwinski L, Xenarios I, Eisenberg D (2002) Protein interactions: two methods for assessment of the reliability of high throughput observations. *Mol Cell Proteomics* 1(5): 349-356

Deiters A, Groff D, Ryu Y, Xie J, Schultz PG (2006) A genetically encoded photocaged tyrosine. *Angew Chem Int Ed Engl* 45(17): 2728-2731

Dorman G, Prestwich GD (2000) Using photolabile ligands in drug discovery and development. *Trends Biotechnol* 18(2): 64-77

Elad N, Maimon T, Frenkiel-Krispin D, Lim RY, Medalia O (2009) Structural analysis of the nuclear pore complex by integrated approaches. *Curr Opin Struct Biol* 19(2): 226-232

Fancy DA (2000) Elucidation of protein-protein interactions using chemical cross-linking or label transfer techniques. *Curr Opin Chem Biol* 4(1): 28-33

Fields S (2005) High-throughput two-hybrid analysis. The promise and the peril. *Febs J* 272(21): 5391-5399

Fields S, Song O (1989) A novel genetic system to detect protein-protein interactions. *Nature* 340(6230): 245-246

Filippakopoulos P, Muller S, Knapp S (2009) SH2 domains: modulators of nonreceptor tyrosine kinase activity. *Curr Opin Struct Biol* 19(6): 643-649

Fisher J, Belasco JG, Khosla S, Knowles JR (1980) beta-Lactamase proceeds via an acyl-enzyme intermediate. Interaction of the Escherichia coli RTEM enzyme with cefoxitin. *Biochemistry* 19(13): 2895-2901

Flint AJ, Tiganis T, Barford D, Tonks NK (1997) Development of "substrate-trapping" mutants to identify physiological substrates of protein tyrosine phosphatases. *Proc Natl Acad Sci U S A* 94(5): 1680-1685

Fraenkel-Conrat H, Olcott HS (1948) Reaction of formaldehyde with proteins; cross-linking of amino groups with phenol, imidazole, or indole groups. *J Biol Chem* 174(3): 827-843

Fry SR, Li J, de las Heras R, McCourt JA, Arel E, Kachab EH, Hazell SL, Huang CY (2008) Detection of HSV type-1 and type-2 IgG using an in vitro PCA based homogeneous immunoassay. *Biochem Biophys Res Commun* 372(4): 542-546

Fujimoto I, Bruses JL, Rutishauser U (2001) Regulation of cell adhesion by polysialic acid. Effects on cadherin, immunoglobulin cell adhesion molecule, and integrin function and independence from neural cell adhesion molecule binding or signaling activity. *J Biol Chem* 276(34): 31745-31751

Galarneau A, Primeau M, Trudeau LE, Michnick SW (2002) Beta-lactamase protein fragment complementation assays as in vivo and in vitro sensors of protein protein interactions. *Nat Biotechnol* 20(6): 619-622

Gao R, Stock AM (2009) Biological insights from structures of two-component proteins. *Annu Rev Microbiol* 63: 133-154

Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dumpelfeld B, Edelmann A, Heurtier MA, Hoffman V, Hoefert C, Klein K, Hudak M, Michon AM, Schelder M, Schirle M, Remor M, Rudi T, Hooper S, Bauer A, Bouwmeester T, Casari G, Drewes G, Neubauer G, Rick JM, Kuster B, Bork P, Russell RB, Superti-Furga G (2006) Proteome survey reveals modularity of the yeast cell machinery. *Nature* 440(7084): 631-636

Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM, Remor M, Hofert C, Schelder M, Brajenovic M, Ruffner H, Merino A, Klein K, Hudak M, Dickson D, Rudi T, Gnau V, Bauch A, Bastuck S, Huhse B, Leutwein C, Heurtier MA, Copley RR, Edelmann A, Querfurth E, Rybin V, Drewes G, Raida M, Bouwmeester T, Bork P, Seraphin B, Kuster B, Neubauer G, Superti-Furga G (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415(6868): 141-147

Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, Hao YL, Ooi CE, Godwin B, Vitols E, Vijayadamodar G, Pochart P, Machineni H, Welsh M, Kong Y, Zerhusen B, Malcolm R, Varrone Z, Collis A, Minto M, Burgess S, McDaniel L, Stimpson E, Spriggs F, Williams J, Neurath K, Ioime N, Agee M, Voss E, Furtak K, Renzulli R, Aanensen N, Carrolla S, Bickelhaupt E, Lazovatsky Y, DaSilva A, Zhong J, Stanyon CA, Finley RL, Jr., White KP, Braverman M, Jarvie T, Gold S, Leach M, Knight J, Shimkets RA, McKenna MP, Chant J, Rothberg JM (2003) A protein interaction map of Drosophila melanogaster. *Science* 302(5651): 1727-1736

Goll J, Uetz P (2006) The elusive yeast interactome. *Genome Biol* 7(6): 223

Griffin BA, Adams SR, Tsien RY (1998) Specific covalent labeling of recombinant protein molecules inside live cells. *Science* 281(5374): 269-272

Guerrero C, Tagwerker C, Kaiser P, Huang L (2006) An integrated mass spectrometry-based proteomic approach: quantitative analysis of tandem affinity-purified in vivo cross-linked protein complexes (QTAX) to decipher the 26 S proteasome-interacting network. *Mol Cell Proteomics* 5(2): 366-378

Harbury PB, Zhang T, Kim PS, Alber T (1993) A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science* 262(5138): 1401-1407

Hatanaka Y, Sadakane Y (2002) Photoaffinity labeling in drug discovery and developments: chemical gateway for entering proteomic frontier. *Curr Top Med Chem* 2(3): 271-288

Hofmann K, Tschopp J (1995) The death domain motif found in Fas (Apo-1) and TNF receptor is present in proteins involved in apoptosis and axonal guidance. *FEBS Lett* 371(3): 321-323

Huang BX, Kim HY (2006) Interdomain conformational changes in Akt activation revealed by chemical cross-linking and tandem mass spectrometry. *Mol Cell Proteomics* 5(6): 1045-1053

Huber LA (2003) Is proteomics heading in the wrong direction? *Nat Rev Mol Cell Biol* 4(1): 74-80

Jahn O, Eckart K, Tezval H, Spiess J (2004) Characterization of peptide-protein interactions using photoaffinity labeling and LC/MS. *Anal Bioanal Chem* 378(4): 1031-1036

Janin J (2000) Kinetics and thermodynamics of protein-protein interactions In *Protein-protein Recognition*, Kleanthous C (ed), pp 1-32. New York: Oxford University Press

Janin J, Bahadur RP, Chakrabarti P (2008) Protein-protein interaction and quaternary structure. *Q Rev Biophys* 41(2): 133-180

Jia Z, Barford D, Flint AJ, Tonks NK (1995) Structural basis for phosphotyrosine peptide recognition by protein tyrosine phosphatase 1B. *Science* 268(5218): 1754-1758

Jiang J, Zeng D, Li S (2009) Photogenerated quinone methides as protein affinity labeling reagents. *Chembiochem* 10(4): 635-638

Jiang J, Zhang J, Li S (2010) Detecting protein interactions in live cells via complementation of a hydrolysis-deficient beta-lactamase. *Chem Commun (Camb)*

Jin Lee Y (2008) Mass spectrometric analysis of cross-linking sites for the structure of proteins and protein complexes. *Mol Biosyst* 4(8): 816-823

Jones S, Thornton JM (1996) Principles of protein-protein interactions. *Proc Natl Acad Sci U S A* 93(1): 13-20

Jun Y, Wickner W (2007) Assays of vacuole fusion resolve the stages of docking, lipid mixing, and content mixing. *Proc Natl Acad Sci U S A* 104(32): 13010-13015

Kaneko T, Li L, Li SS (2008) The SH3 domain--a family of versatile peptide- and protein-recognition module. *Front Biosci* 13: 4938-4952

Khan SM, Franke-Fayard B, Mair GR, Lasonder E, Janse CJ, Mann M, Waters AP (2005) Proteome analysis of separated male and female gametocytes reveals novel sex-specific Plasmodium biology. *Cell* 121(5): 675-687

Kim YJ, Bjorklund S, Li Y, Sayre MH, Kornberg RD (1994) A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* 77(4): 599-608

Koch T, Suenson E, Korsholm B, Henriksen U, Buchardt O (1994) Pitfalls in characterization of protein interactions using radioiodinated crosslinking reagents. Preparation and testing of a novel photochemical 125I-label transfer reagent. *Bioconjug Chem* 5(3): 205-212

Kocher T, Superti-Furga G (2007) Mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nat Methods* 4(10): 807-815

Kolb HC, Sharpless KB (2003) The growing impact of click chemistry on drug discovery. *Drug Discov Today* 8(24): 1128-1137

Komatsu T, Kikuchi K, Takakusa H, Hanaoka K, Ueno T, Kamiya M, Urano Y, Nagano T (2006) Design and synthesis of an enzyme activity-based labeling molecule with fluorescence spectral change. *J Am Chem Soc* 128(50): 15946-15947

Kosugi S, Hasebe M, Tomita M, Yanagawa H (2009) Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc Natl Acad Sci U S A* 106(25): 10171-10176

Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, Li J, Pu S, Datta N, Tikuisis AP, Punna T, Peregrin-Alvarez JM, Shales M, Zhang X, Davey M, Robinson MD, Paccanaro A, Bray JE, Sheung A, Beattie B, Richards DP, Canadien V, Lalev A, Mena F, Wong P, Starostine A, Canete MM, Vlasblom J, Wu S, Orsi C, Collins SR, Chandran S, Haw R, Rilstone JJ, Gandi K, Thompson NJ, Musso G, St Onge P, Ghanny S, Lam MH, Butland G, Altaf-Ul AM, Kanaya S, Shilatifard A, O'Shea E, Weissman JS, Ingles CJ, Hughes TR, Parkinson J, Gerstein M, Wodak SJ, Emili A, Greenblatt JF (2006) Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. *Nature* 440(7084): 637-643

Lawrence MC, Colman PM (1993) Shape complementarity at protein/protein interfaces. *J Mol Biol* 234(4): 946-950

Lee H, Chen R, Lee Y, Yoo S, Lee C (2009) Essential roles of CKIdelta and CKIepsilon in the mammalian circadian clock. *Proc Natl Acad Sci U S A* 106(50): 21359-21364

Lee HK, Brown SJ, Rosen H, Tobias PS (2007) Application of beta-lactamase enzyme complementation to the high-throughput screening of toll-like receptor signaling inhibitors. *Mol Pharmacol* 72(4): 868-875

Lee HK, Dunzendorfer S, Tobias PS (2004) Cytoplasmic domain-mediated dimerizations of toll-like receptor 4 observed by beta-lactamase enzyme fragment complementation. *J Biol Chem* 279(11): 10564-10574

Lemke EA, Summerer D, Geierstanger BH, Brittain SM, Schultz PG (2007) Control of protein phosphorylation with a genetically encoded photocaged amino acid. *Nat Chem Biol* 3(12): 769-772

Lijnzaad P, Argos P (1997) Hydrophobic patches on protein subunit interfaces: characteristics and prediction. *Proteins* 28(3): 333-343

Liu B, Archer CT, Burdine L, Gillette TG, Kodadek T (2007) Label transfer chemistry for the characterization of protein-protein interactions. *J Am Chem Soc* 129(41): 12348-12349

Liu B, Burdine L, Kodadek T (2006) Chemistry of periodate-mediated cross-linking of 3,4-dihydroxylphenylalanine-containing molecules to proteins. *J Am Chem Soc* 128(47): 15228-15235

Los GV, Encell LP, McDougall MG, Hartzell DD, Karassina N, Zimprich C, Wood MG, Learish R, Ohana RF, Urh M, Simpson D, Mendez J, Zimmerman K, Otto P, Vidugiris G, Zhu J, Darzins A, Klaubert DH, Bulleit RF, Wood KV (2008) HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem Biol* 3(6): 373-382

Magliery TJ, Wilson CG, Pan W, Mishler D, Ghosh I, Hamilton AD, Regan L (2005) Detecting protein-protein interactions with a green fluorescent protein fragment reassembly trap: scope and mechanism. *J Am Chem Soc* 127(1): 146-157

Marles JA, Dahesh S, Haynes J, Andrews BJ, Davidson AR (2004) Protein-protein interaction affinity plays a crucial role in controlling the Sho1p-mediated signal transduction pathway in yeast. *Mol Cell* 14(6): 813-823

Metz B, Kersten GF, Hoogerhout P, Brugghe HF, Timmermans HA, de Jong A, Meiring H, ten Hove J, Hennink WE, Crommelin DJ, Jiskoot W (2004) Identification of formaldehyde-induced modifications in proteins: reactions with model peptides. *J Biol Chem* 279(8): 6235-6243

Michnick SW (2001) Exploring protein interactions by interaction-induced folding of proteins from complementary peptide fragments. *Curr Opin Struct Biol* 11(4): 472-477

Michnick SW, Ear PH, Manderson EN, Remy I, Stefan E (2007) Universal strategies in research and drug discovery based on protein-fragment complementation assays. *Nat Rev Drug Discov* 6(7): 569-582

Michnick SW, Remy I, Campbell-Valois FX, Vallee-Belisle A, Pelletier JN (2000) Detection of protein-protein interactions by protein fragment complementation strategies. *Methods Enzymol* 328: 208-230

Milarski KL, Zhu G, Pearl CG, McNamara DJ, Dobrusin EM, MacLean D, Thieme-Sefler A, Zhang ZY, Sawyer T, Decker SJ, et al. (1993) Sequence specificity in recognition of the epidermal growth factor receptor by protein tyrosine phosphatase 1B. *J Biol Chem* 268(31): 23634-23639

Minasov G, Wang X, Shoichet BK (2002) An ultrahigh resolution structure of TEM-1 beta-lactamase suggests a role for Glu166 as the general base in acylation. *J Am Chem Soc* 124(19): 5333-5340

Mohi MG, Neel BG (2007) The role of Shp2 (PTPN11) in cancer. *Curr Opin Genet Dev* 17(1): 23-30

Nadeau OW, Anderson DW, Yang Q, Artigues A, Paschall JE, Wyckoff GJ, McClintock JL, Carlson GM (2007) Evidence for the location of the allosteric activation switch in the

multisubunit phosphorylase kinase complex from mass spectrometric identification of chemically crosslinked peptides. *J Mol Biol* 365(5): 1429-1445

Nakao A, Yoshihama M, Kenmochi N (2004) RPG: the Ribosomal Protein Gene database. *Nucleic Acids Res* 32(Database issue): D168-170

Orlicky S, Tang X, Willems A, Tyers M, Sicheri F (2003) Structural basis for phosphodependent substrate selection and orientation by the SCFCdc4 ubiquitin ligase. *Cell* 112(2): 243-256

Palhan VB, Chen S, Peng GH, Tjernberg A, Gamper AM, Fan Y, Chait BT, La Spada AR, Roeder RG (2005) Polyglutamine-expanded ataxin-7 inhibits STAGA histone acetyltransferase activity to produce retinal degeneration. *Proc Natl Acad Sci U S A* 102(24): 8472-8477

Park JH, Back JH, Hahm SH, Shim HY, Park MJ, Ko SI, Han YS (2007) Bacterial beta-lactamase fragmentation complementation strategy can be used as a method for identifying interacting protein pairs. *J Microbiol Biotechnol* 17(10): 1607-1615

Parrish JR, Gulyas KD, Finley RL, Jr. (2006) Yeast two-hybrid contributions to interactome mapping. *Curr Opin Biotechnol* 17(4): 387-393

Piehler J (2005) New methodologies for measuring protein interactions in vivo and in vitro. *Curr Opin Struct Biol* 15(1): 4-14

PlaW M (1997) Photoaffinity Labeling Introduction. *Photochemistry and Photobiology* 65(2): 193-194

Popp MW, Antos JM, Ploegh HL (2009) Site-specific protein labeling via sortase-mediated transpeptidation. *Curr Protoc Protein Sci* Chapter 15: Unit 15 13

Prieto C, De Las Rivas J (2009) Structural domain-domain interactions: assessment and comparison with protein-protein interaction data to improve the interactome. *Proteins* 78(1): 109-117

Ranish JA, Hahn S, Lu Y, Yi EC, Li XJ, Eng J, Aebersold R (2004) Identification of TFB5, a new component of general transcription and DNA repair factor IIH. *Nat Genet* 36(7): 707-713

Remy I, Ghaddar G, Michnick SW (2007) Using the beta-lactamase protein-fragment complementation assay to probe dynamic protein-protein interactions. *Nat Protoc* 2(9): 2302-2306

Remy I, Michnick SW (2006) A highly sensitive protein-protein interaction assay based on Gaussia luciferase. *Nat Methods* 3(12): 977-979

Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* 17(10): 1030-1032

Rokita SE (2009a) Introduction. In *Quinone Methides*, Rokita SE (ed), Introduction, p xv. Hobken, New Jersey: Jonh Willey & sons, Inc.

Rokita SE (2009b) Reversible Alkylation of DNA by Quinone Methides. In *Quinone Methides*, Rokita SE (ed), 9. Reversible Alkylation of DNA by Quinone Methides, pp 297-327. Hobken, New Jersey: Jonh Willey & sons, Inc.

Rokita SE (2009c) Reversible alkylation of DNA by quinone methides. In *Quinone Methides*, Rokita SE (ed), Introduction, pp 297-323. Hobken, New Jersey: Jonh Willey & sons, Inc.

Rostovtsev VV, Green LG, Fokin VV, Sharpless KB (2002) A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew Chem Int Ed Engl* 41(14): 2596-2599

Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N, Klitgord N, Simon C, Boxem M, Milstein S, Rosenberg J, Goldberg DS, Zhang LV, Wong SL, Franklin G, Li S, Albala

JS, Lim J, Fraughton C, Llamosas E, Cevik S, Bex C, Lamesch P, Sikorski RS, Vandenhaute J, Zoghbi HY, Smolyar A, Bosak S, Sequerra R, Doucette-Stamm L, Cusick ME, Hill DE, Roth FP, Vidal M (2005) Towards a proteome-scale map of the human protein-protein interaction network. *Nature* 437(7062): 1173-1178

Rudolph J (2007) Inhibiting transient protein-protein interactions: lessons from the Cdc25 protein tyrosine phosphatases. *Nat Rev Cancer* 7(3): 202-211

Schmitt-Ulms G, Hansen K, Liu J, Cowdrey C, Yang J, DeArmond SJ, Cohen FE, Prusiner SB, Baldwin MA (2004) Time-controlled transcardiac perfusion cross-linking for the study of protein interactions in complex tissues. *Nat Biotechnol* 22(6): 724-731

Schnee M, Ruzsics Z, Bubeck A, Koszinowski UH (2006) Common and specific properties of herpesvirus UL34/UL31 protein family members revealed by protein complementation assay. *J Virol* 80(23): 11658-11666

Schwartz MA (1985) A 125I-radiolabel transfer crosslinking reagent with a novel cleavable group. *Anal Biochem* 149(1): 142-152

Singh A, Thornton ER, Westheimer FH (1962) The photolysis of diazoacetylchymotrypsin. *J Biol Chem* 237: 3006-3008

Sinz A (2006) Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrom Rev* 25(4): 663-682

Souza Nd (2007) FlAsHing the neighbors. *Nature Methods* 4(12): 990

Spotts JM, Dolmetsch RE, Greenberg ME (2002) Time-lapse imaging of a dynamic phosphorylation-dependent protein-protein interaction in mammalian cells. *Proc Natl Acad Sci U S A* 99(23): 15142-15147

Sprinzak E, Sattath S, Margalit H (2003) How reliable are experimental protein-protein interaction data? *J Mol Biol* 327(5): 919-923

Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S, Timm J, Mintzlaff S, Abraham C, Bock N, Kietzmann S, Goedde A, Toksoz E, Droege A, Krobitsch S, Korn B, Birchmeier W, Lehrach H, Wanker EE (2005) A human protein-protein interaction network: a resource for annotating the proteome. *Cell* 122(6): 957-968

Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, Qureshi-Emili A, Li Y, Godwin B, Conover D, Kalbfleisch T, Vijayadamodar G, Yang M, Johnston M, Fields S, Rothberg JM (2000) A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* 403(6770): 623-627

Uttamapinant C, White KA, Baruah H, Thompson S, Fernandez-Suarez M, Puthenveetil S, Ting AY (2010) A fluorophore ligase for site-specific protein labeling inside living cells. *Proc Natl Acad Sci U S A* 107(24): 10914-10919

Van Criekinge W, Beyaert R (1999) Yeast Two-Hybrid: State of the Art. *Biol Proced Online* 2(1): 1-38

Vasilescu J, Guo X, Kast J (2004) Identification of protein-protein interactions using in vivo cross-linking and mass spectrometry. *Proteomics* 4(12): 3845-3854

Vaynberg J, Fukuda T, Chen K, Vinogradova O, Velyvis A, Tu Y, Ng L, Wu C, Qin J (2005) Structure of an ultraweak protein-protein complex and its crucial role in regulation of cell morphology and motility. *Mol Cell* 17(4): 513-523

Vaynberg J, Qin J (2006) Weak protein-protein interactions as probed by NMR spectroscopy. *Trends Biotechnol* 24(1): 22-27

Veselovsky AV, Ivanov YD, Ivanov AS, Archakov AI, Lewi P, Janssen P (2002) Protein-protein interactions: mechanisms and modification by drugs. *J Mol Recognit* 15(6): 405-422

Vidalain PO, Boxem M, Ge H, Li S, Vidal M (2004) Increasing specificity in high-throughput yeast two-hybrid experiments. *Methods* 32(4): 363-370

von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, Bork P (2002) Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* 417(6887): 399-403

Wang X, Venable J, LaPointe P, Hutt DM, Koulov AV, Coppinger J, Gurkan C, Kellner W, Matteson J, Plutner H, Riordan JR, Kelly JW, Yates JR, 3rd, Balch WE (2006) Hsp90 cochaperone Aha1 downregulation rescues misfolding of CFTR in cystic fibrosis. *Cell* 127(4): 803-815

Wehrman T, Kleaveland B, Her JH, Balint RF, Blau HM (2002) Protein-protein interactions monitored in mammalian cells via complementation of beta -lactamase enzyme fragments. *Proc Natl Acad Sci U S A* 99(6): 3469-3474

Wilkinson KD (2004) Quantitative analysis of protein-protein interactions. *Methods Mol Biol* 261: 15-32

Xu D, Tsai CJ, Nussinov R (1997) Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein Eng* 10(9): 999-1012

Young DD, Deiters A (2007) Photochemical control of biological processes. *Org Biomol Chem* 5(7): 999-1005

Zlokarnik G, Negulescu PA, Knapp TE, Mere L, Burres N, Feng L, Whitney M, Roemer K, Tsien RY (1998) Quantitation of transcription and clonal selection of single living cells with beta-lactamase as reporter. *Science* 279(5347): 84-88