DEVELOPMENT OF CHEMICAL STRATEGIES FOR SPECIFICALLY PROBING AND IDENTIFYING SULFUR CARRIER PROTEINS AND VITAMIN B6-DEPENDENT ENZYMES IN BACTERIA

A Dissertation

by

YIQUAN LIU

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Tadhg P. Begley
Committee Members,	David H. Russell
	Frank M. Raushel
	Wenshe Liu
	James C. Hu
Head of Department,	Francois P. Gabbai

December 2015

Major Subject: Chemistry

ABSTRACT

Activity based protein profiling (ABPP) is a functional proteomic technology that uses chemical probes to detect mechanistically related classes of enzymes. Chemically probing a certain class of proteins helps to understand their biological function as a group, and discover new biosynthetic pathways for drug design.

This research describes two activity based proteomic methods that have been developed to probe and identify sulfur carrier proteins and vitamin B6 dependent proteins, respectively. Sulfur carrier proteins are small proteins (<10 kDa) involved in pathways for efficient sulfur delivery. A chemical probe with sulfonyl-azide functional group was designed to label and identify the sulfur carrier proteins through a thioacid-azide reaction. This method identified a new sulfur carrier protein in *Streptomyces coelicolor*. Further study of its biological function led to the discovery and characterization of a new pathway of homocysteine formation, which is probably another direct sulfurylation of methionine biosynthesis.

Vitamin B6 dependent proteins are a class of enzymes that cover a wide range of cellular functions such as transamination, racemization, and decarboxylation. Also, vitamin B6 dependent proteins have a critical role in human disease and the metabolic pathways of pathogens and plants. We used *Escherichia coli* as a model system to develop both radioactive and nonradioactive based methods to probe and identify vitamin B6 containing proteins in the bacterial proteome. This technique was then used to study how vitamin B6 proteins are regulated in response to cellular stress.

DEDICATION

This dissertation is dedicated to the people who have been so helpful to me in the past six years. I would not have endeavored such an undertaking without the great support from my husband Da, who was also a chemistry graduate student in Texas A&M University. Also, I would like to dedicate this to my parents, who have always given me solid support. Finally, I would also like to dedicate this to the friends that I have made in the past four years, both within the Begley group and within Texas A&M University.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Begley, who guided and encouraged me during the whole process. I would also like to thank my committee members, Dr. Raushel, Dr. Russell, Dr. Liu, and Dr. Hu, for helping me throughout the course of this research.

Thanks also go to Begley group members and former members, especially Dr. Kalyan Krishnamoorthy who taught me and helped me a lot for bioanalytical techniques in the lab, Dr. Sameh Abdelwahed, Dr. Dmytro Fedoseyenko, Dr. Dinesh Simkhada, and Brateen Shome who collaborated with me on those projects, and Dr. Benjamin J. Philmus and Dr. Lisa E. Cooper for giving me lots of suggestions during the past four years. I also want to thank the rest of the group members for making the group like a big family and leaving me a lot of great memories. In addition, I would also like to extend my gratitude to my friends and colleagues for making my time at Texas A&M University a great experience.

Finally, thanks to my mother and father for their encouragement and to my husband for his patience and love.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1 Proteomics

Proteins are essential parts of living organisms, playing essential roles in cellular structure, transport, storage, and metabolic pathways. According to the central dogma, proteins are manufactured by transcription of genomic DNA followed by translation of mRNA. The complexity of proteins, however, is much greater than that of genomes[1-3] due to post-translational modifications such as phosphorylation, acetylation, ubiquitination, and methylation[4-6]. Therefore, proteomics has emerged as an efficient tool to investigate proteins on a large scale[7-8].

Post-translational modification is an important research area of proteomics because of its essentiality to the protein's biological functions. For instance, many enzymes in the cell signaling process are phosphorylated via adding phosphate onto serine or threonine residues[9]. By identifying phosphorylated proteins and their modification sites, more proteins involved in cell signaling were discovered. Another example is ubiquitination, in which the target proteins' lysine residues are modified by a small conserved protein named ubiquitin, mostly followed by degradation in the proteasome[10-11]. Considering the difficulty of predicting protein ubiquitination by genomic study, proteomics enables the large scale identification of ubiquitinated proteins as a whole, which can then be individually investigated[12].

Apart from post-translational modification, proteomics also concerns itself with protein expression. Though genomics is a powerful tool for sequence analysis and

protein function prediction, the expression of each protein varies in accordance to different growth conditions and thus needs to be measured at the protein level. The use of biomarkers, a rising star in clinical proteomics, is a good example. Biomarkers are an excellent tool to develop indicators for specific biological states, such as the presence or stage of a disease[13-15]. Expression proteomics, on the other hand, studies differential protein expression as a function of stimulation and conditions including drug, disease, tissue, and time. This can be extremely useful in disease diagnosis and drug discovery.

Proteomics also has applications in other areas like high-throughput determination of protein structures[16], investigation and prediction of protein-protein interactions[17], mapping the location of proteins in whole cells during key cell events[18], and bioinformatics analysis on protein sequences and domain matching[19].

1.2 Techniques in Proteomics

To analyze the proteome, sample preparation is important to minimize the effect of interfering components and to enrich proteins of interest. Centrifugation can be used to remove insoluble substances; dialysis and size-exclusion chromatography are useful in removing salts and small molecules[20-22]. Protein precipitation using organic solvents such as methanol/chloroform precipitation also helps rid the sample of other components[23].

After a protein mixture is obtained, it still needs separation prior to protein visualization or mass spectrometry (MS) analysis. Gel electrophoresis has been playing a major role in the past several decades. Based on the separation dimension, there are 1D-

PAGE (polyacrylamide gel electrophoresis) and 2D-PAGE. The former achieves separation by different protein sizes which cause the migration difference in polyacrylamide, whereas the latter first separates proteins on a pH strip according to their pI values, and then completes the second dimension of separation by protein size. Nowadays, gel electrophoresis is a mature technique and can either be used for protein visualization and quantitation or may be interfaced with other techniques such as western blotting and mass spectrometry[24-25]. For visualization, proteins in the gel can be stained with a variety of methods like silver staining[26-27], Coomassie brilliant blue staining[28-29], radioactive gel visualization[30], and fluorescence detection[31-32]. After staining, proteins in the same gel can be quantitated through measurement of the darkness of each band.

In modern proteomics, gel electrophoresis is used before detection by other techniques. With western blotting, proteins are transferred into a membrane and stained with antibodies that are specific towards certain protein targets[33-34]. With mass spectrometry, proteins stained in the gel are cut off, digested with proteases and then subjected to mass spectrometry[25, 35] and exhibits a much higher detection limit than mass spectrometry on its own[35-36]. This is largely due to dynamic range in the staining process and poor recovery of proteins from the gel. As a result, scientists are making efforts to improve SDS-PAGE and find alternative methods.

Chromatography, especially liquid chromatography (LC), has been developed rapidly, making it a competitive alternative to gel electrophoresis. Coupled with mass spectrometry, LC-MS/MS (tandem mass spectrometry) is a highly effective technique

for peptide identification and profiling. Compared with gel based techniques, LC-MS/MS exhibits tremendous advantages in detection speed, sensitivity, reproducibility, and applicability to a wide range of samples[37-39].

Mass spectrometry provides highly accurate measurements for protein/peptide molecular weight and charge. The fast analysis speed, low detection limit, and high mass accuracy have made it an essential technique in modern proteomics. Mass spectrometry based proteomics is divided into top-down proteomics and bottom-up proteomics depending on the sample preparation and data analysis process. For top-down proteomics, intact proteins are separated and subjected directly to mass spectrometry[40-41]. The proteins are fragmented into peptides by colliding their amide bonds, and peptides are further broken down to obtain sequence information. In this way, each protein provides sufficient fragment information, which enables the analysis on peptides or post-translational modifications with low abundance. However, the low ionization of large proteins (100 kDa) makes it difficult to generate a good signal. In this case, bottom-up proteomic techniques are usually used, in which proteins are first digested into peptides by proteases such as trypsin, chymotrypsin, LysC, and LysN. The peptide mixture is then separated by LC and detected by tandem mass spectrometry in which the peptides are fragmented to obtain sequence information[42-43]. Bioinformatics and database are often needed for sequence analysis based on tandem mass spectra. The bottom-up approach is widely used in protein identification and modification identification. The drawback of this technique is that the small portion of detectable digested peptides

results in only a fractional yield of useful fragments. Therefore, the top-down and bottom-up methods complement each other and should be selected appropriately.

In addition to protein primary structure study, mass spectrometry has also been used for protein quantitation with the development of various isotope methods. For example, SILAC (stable isotope labeling by amino acids in cell culture) incorporates isotopes (usually deuterium, ^{13}C and ^{15}N) by feeding isotope labeled amino acids into cell culture[44-46]. And ICAT (isotope coded affinity tag) uses an isotope incorporated chemical labeling reagent to label the side chains of amino acids. The "light" peptides (natural peptides) and the "heavy" peptides (isotope labeled peptides) have hardly any chemical difference, so the migration of both peptides in chromatography and the ionization efficiency in mass spectrometry are the same. As a consequence, the intensity ratio of the mass spectra reflects the actual ratio of the corresponding peptides/proteins in each culture.

In recent years, label free quantitation has also emerged and become more and more popular in quantitative proteomics. SRM (selected reaction monitoring) and MRM (multiple reaction monitoring) are two representative techniques and use triple quadruple mass spectrometry for quantitation. This technique can be used for constructing calibration curves and providing absolute quantitation[47-48].

Gel electrophoresis, chromatography, and mass spectrometry are all powerful techniques in proteomics, which when combined, provide insight into both a protein's primary structure and the quantitative information of the proteome.

1.3 Activity Based Protein Profiling

Though proteomics provides methods to globally analyze protein primary structure and quantitate protein abundance, it lacks the capacity to profile proteins according to biological activity or functional state. This has prompted the development of alternative strategies for targeting and identifying proteins with specific activities in a complex proteome.

Activity based protein profiling shows promise in addressing these issues[49-52]. Active site based covalent probes are designed to target specific subsets of enzymes. Since the 1990s, various probes have been designed to target different classes of enzymes, such as glycosidases[53], kinases[54], phosphatases[55], and proteases[56-59]. The basic procedure for activity based protein profiling is shown in figure 1. Since the probe molecule is designed for a certain class of enzymes, the targeted proteins can be visualized after separating by gel electrophoresis, or they can be enriched by affinity chromatography and digested by proteases in preparation for mass spectrometry analysis.

The essential part of activity based protein profiling is the chemical probe. It is usually composed of three parts: 1) a reactive group for enzyme active site probing; 2) a flexible linker to increase water solubility and create a proper distance between the two termini; 3) a reporter group like a fluorescent molecule for visualization or a tag such as biotin for affinity isolation and purification (figure 2). In most cases, PEG (polyethylene glycol) is selected as the linker and biotin or fluorescent molecules such as rhodamine and dansyl derivatives are used for the reporter group.

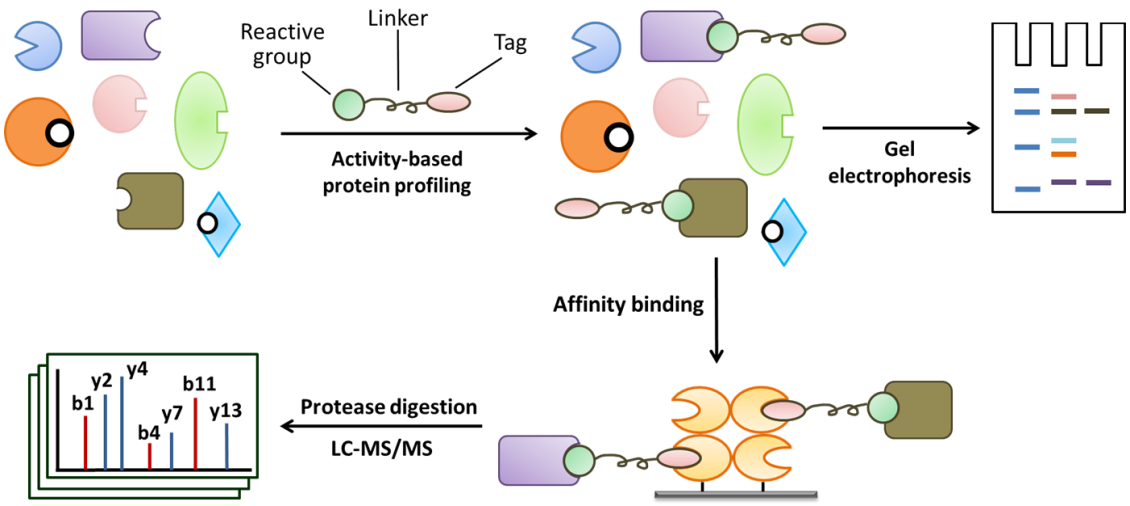


Figure 1. Schematic diagram of activity based protein profiling. Proteins of interest are targeted by molecular probe followed by visualization or affinity chromatography enrichment and mass spectrometry identification.

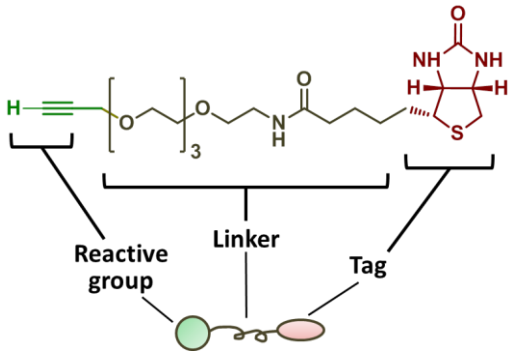


Figure 2. The typical configuration of a chemical probe. It contains a reactive group (such as alkyne for click reaction), a linker (such as PEG (polyethylene glycol)), and a tag (such as biotin).

The reactive group component varies according to need. In general, a biorthogonal reaction needs to be selected for the approach and both the probe molecules and proteins are engineered with complementary chemical entries. One typical way is to mimic substrates or inhibitors of the proteins of interest. A click reaction is commonly

selected as the probing reaction due to its high specificity, fast reaction speed, and compatibility to aqueous solution[60-62]. Copper (I)-catalyzed Azide-Alkyne Cycloaddition (CuAAC), also known as Huisgen's 1,3-dipolar cycloaddition, is a well-known click-chemistry approach between alkyne and azide reaction groups. Staudinger ligation is another example which covalently links a phosphane and azido group. Besides click chemistry, photoaffinity labeling is another widely used reaction. It labels the proximal residues in an enzyme active site after UV (ultraviolet) irradiation[63-64].

With a specifically designed chemical probe, activity based protein profiling can be carried out using the analytical techniques and strategies described. These approaches have been applied to target discovery[65], inhibitor discovery[66-67], and characterization of enzyme active sites[68-69]. The aim of our research is to develop activity based protein profiling techniques to investigate sulfur carrier proteins and PLP (pyridoxal-5'-phosphate)-dependent enzymes in the bacterial proteome.

2. METHOD DEVELOPMENT AND VALIDATION OF PROBING AND IDENTIFYING SULFUR CARRIER PROTEINS

2.1 Introduction

Sulfur carrier proteins are a family of small proteins participating in sulfur trafficking in prokaryotic metabolism[70]. As indicated by the name, sulfur carrier proteins carry sulfur in their floppy C-termini and transfer it efficiently to the target molecule, enabling the receptors to continue biosynthesis of corresponding sulfur-containing metabolites. In this process, the sulfur carrier proteins are first activated at the C-termini by ATP (adenosine triphosphate) to form a protein-AMP (adenosine monophosphate) complex, catalyzed by an activating enzyme; secondly, with the participation of a sulfur donor (HS^-, RSSH)[71], the activated C-terminus is converted into the thiocarboxylate form in preparation for the sulfur transfer (figure 3). This process is similar to protein ubiquitination, and the structures of sulfur carrier proteins and ubiquitin are much alike. However, unlike ubiquitination, which usually leads to protein proteolysis, sulfur carrier proteins are designed for sulfur transfer and are only observed in prokaryotes.

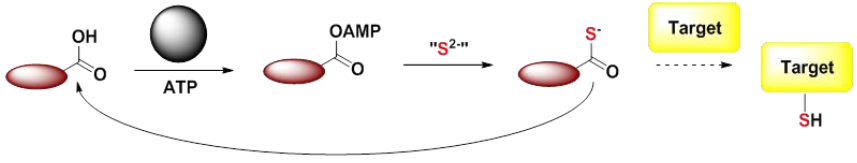


Figure 3. Formation of the thiocarboxylate group in sulfur carrier proteins. In the first step, the sulfur carrier proteins in natural form are activated by ATP to form protein-AMP; secondly, the activated protein is converted into protein-thiocarboxylate with the participation of a sulfur donor.

To date, a few sulfur carrier proteins have been discovered in different biosynthetic pathways widely existing in bacteria. The most famous ones are ThiS, MoaD, CysO, and QbsE, which participate in the sulfur incorporation step in biosynthesis of thiamin, molybdopterin, L-cysteine and thioquinolobactin, respectively[72-74]. Figures 4 and 5 show the structure of these compounds and the gene clusters of ThiS, MoaD and CysO, respectively.

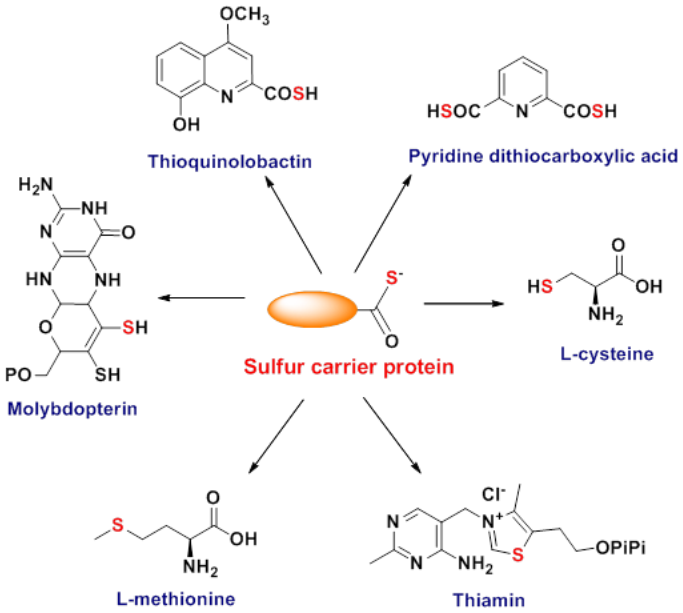


Figure 4. Biosynthetic pathways involving sulfur carrier proteins[75]. The activated protein in the thiocarboxylate form donates sulfur to the target molecule. The sulfur atoms which come from the sulfur carrier proteins are marked in red.

Interestingly, while the structures of these metabolites are quite different, the sulfur carrier proteins have some common features. First, they're usually small proteins of less than 10 kDa (around 100 amino acid residues) and hold conserved structures with

a high similarity with ubiquitin. Second, the C-termini of sulfur carrier proteins are flexible and usually contain a glycine-glycine motif. During sulfur transfer, the floppy C-terminus inserts into the active site of the corresponding catalyzing enzyme to achieve sulfur donation[76].

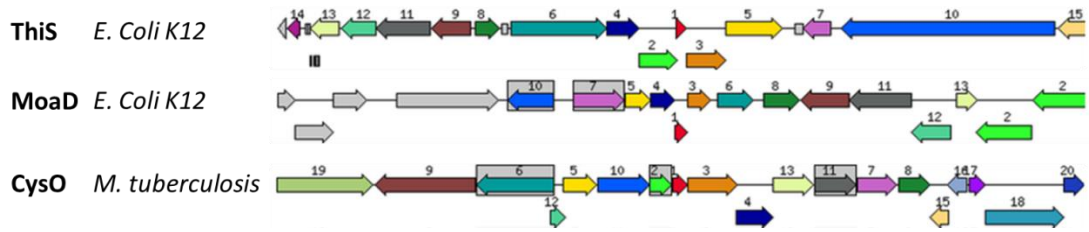


Figure 5. Gene clusters of ThiS, MoaD and CysO in *E. coli* and *M. tuberculosis*[77]. Each arrow represents a gene. The sulfur carrier proteins are shown in red arrow and numbered "1".

Pfam database analysis suggests that sulfur carrier proteins are widely distributed and may participate in uncharacterized biosynthetic pathways, making it very important to understand their functions. Genomics has become a useful tool for gene annotation and function prediction by comparing genes for sequence similarity and structural homologies. Nevertheless, there are still numerous uncharacterized hypothetical proteins and misannotated proteins. Moreover, it's difficult to anticipate the expression level of each protein under different growth conditions, which is fundamental to cell metabolism. Developing an activity based chemical probe for targeting and identifying protein thiocarboxylates in the proteome can help to discover new sulfur carrier proteins and new biosynthetic pathways, which assist efforts in drug design or pathway engineering

for sulfur containing compounds. The probe itself could potentially be developed as a drug that targets sulfur carrier proteins for disease treatment.

Though sulfur carrier proteins are not enzymes, the thiocarboxylate group acts as a *de facto* "active site" since sulfur donation requires C-terminal thiocarboxylate formation. In this way, the idea and strategy of activity based protein profiling can be applied to this system; selective labeling of sulfur carrier proteins, however, is a challenging goal. First, the concentration of the thiocarboxylate functional group is very low since it only presents in the "active" sulfur carrier proteins. Also, there are many other nucleophiles in the cytosol which may interfere with the specific probing. After surveying various possibilities, we decided to explore the click reaction between thiocarboxylates and electron deficient sulfonyl-azides to form N-acyl-sulfonamides as our labeling strategy (figure 6A). The following probe molecule was thus designed: biotin sulfonyl-azide (figure 6B)[78-80].

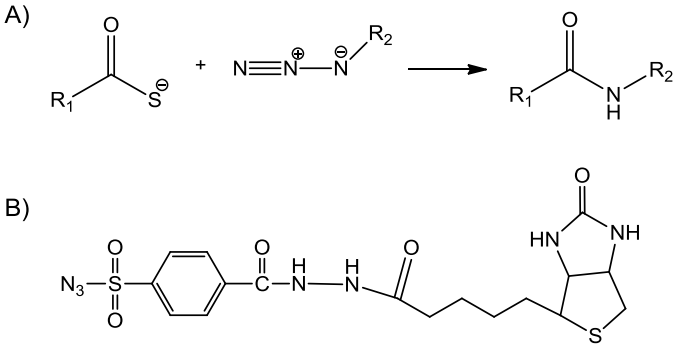


Figure 6. Thioacid-azide reaction and structure of the probe molecule. A) Thioacid-azide reaction; B) Structure of biotin sulfonyl-azide.

2.2 Experimental Methods

Chemicals and Reagents

4-carboxybenzyl-sulfonazide, biotin hydrazide, N-hydroxysuccinimide, *O*-phthalaldehyde, iodoacetic acid, urea, adenosine triphosphate (ATP) were purchased from Sigma Aldrich (St. Louis, MO). Isopropyl-D-thiogalactopyranosid (IPTG) was from Lab Scientific Inc. (Livingston, NJ). Luria-Bertani medium (LB) was from EMD Biosciences (Gibbstown, NJ). *Escherichia coli BL21(DE3)* (ATCC BAA-1025) was purchased from the American Type Culture Collection (Manassas, VA). Chitin beads and the pTYB1 vector were obtained from New England Biolabs (Ipswich, MA). The streptavidin resin was from GenScript (Piscataway, NJ) and streptavidin-R-phycoerythrin (PE) was from QIAGEN (Valencia, CA). Protein concentrations were determined using the Bradford assay[81]. Econo-pac 10 DG desalting columns were from Biorad (Hercules, CA), and the Novagen D-tube Maxi dialyzer with 3.5 kDa molecular weight cutoff were obtained from EMD Biosciences. The polyvinylidene difluoride (PVDF) membrane and trypsin gold were purchased from Promega (Madison, WI).

All fluorescence gel images were scanned using a Typhoon trio instrument (excitation, 532 nm green laser; emission, 580 nm band-pass filter (580 BP 30)) from GE Healthcare Biosciences (Piscataway, NJ). UV absorbance of protein samples was tested on a Varian Cary 300 Bio UV-Visible Spectrophotometer (Palo Alto, CA). The sonicator 3000 from Misonix Inc. (Farmingdale, NY) was used for sonication. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis was done using a Hoefer SE 250 mini-vertical gel electrophoresis unit.

Overexpression and Purification of *Thermus thermophilus* ThiS-COSH

The *Thermus thermophilus thiS* gene was constructed and cloned into the pTYB1 vector and transformed into *Escherichia coli BL21(DE3)* as previously described[75]. The cells were grown in 3 L Luria-Bertani (LB) medium at 37 ℃ to reach OD_{600} of 0.6. Protein overexpression was then induced by adding 500 μM IPTG at 15 ℃ and incubated overnight. The resulting culture was harvested using an Avanti JE centrifuge from Beckman Coulter and the cell extract was obtained by sonicating at 4 ℃ in 20 mM Tris, 500 mM NaCl, 1 mM EDTA, 0.1% Triton X-100, pH 7.8. After centrifugation at 14000 rpm for 30 minutes, the cell extract was isolated from the cell pellet and loaded onto 30 ml chitin beads which had been pre-equilibrated in 20 mM Tris, 500 mM NaCl, 1 mM EDTA, pH 7.8. The cell extract was allowed to pass through the chitin beads at a flow rate of 0.5 ml/min. The resulting chitin beads with bound *Thermus thermophilus* ThiS protein (TtThiS) were washed with 300 ml 20 mM Tris, 500 mM NaCl, 1 mM EDTA, pH 7.8 and TtThiS was eluted with the C-terminal thiocarboxylate group (TtThiS-COSH) by incubating in 30 ml of 50 mM NaHS at 4 ^{o}C for 48 hours. The eluted proteins were then concentrated and dialyzed twice using 3.5 kDa molecular weight cut off (MWCO) in 1 L 50 mM K_2HPO_4, pH 8.0) and store at -80 ℃ in 30% glycerol. The protein concentration was tested using a Bradford assay.

Labeling and Trapping of Purified *Thermus thermophilus* ThiS-COSH

TtThiS-COSH (100 μM, 50 μl) was buffer exchanged into 9 M urea, 50 mM KPi pH 6.0 using a Biorad P6 desalting column. The protein solution was then treated with

0.5 µl 250 mM biotin sulfonyl-azide to a final concentration of 2.5 mM. The labeling reaction was carried out at room temperature for 30 minutes, followed by the addition of 6 µl 250 mM TCEP (final concentration 30 mM) and 30 minutes incubation. The resulting solution was then buffer exchanged into 100 mM PBS buffer (100 mM K_2HPO_4, 300 mM NaCl, pH 7.2) by Biorad P6 size-exclusion chromatography. An aliquot of 5 µl sample was saved for gel electrophoresis and the rest was loaded onto 150 µl of streptavidin resin slurry (pre-equilibrated with 100 mM PBS buffer) in a 1.5 ml eppendorf tube shaking for 30 minutes at room temperature. The resin was washed with 10 ml PBS buffer.

Elution, On-resin Digestion, SDS-PAGE, and LC-MS/MS Analysis of *Thermus thermophilus* ThiS Protein

Streptavidin resin with bound biotinylated TtThiS-COSH was divided into two aliquots. One aliquot was added to 50 µl SDS sample buffer (2% w/v SDS, 2 mM β-mercaptoethanol, 4% w/v glycerol, 40 mM Tris-HCl pH 6.8, 0.01% w/v bromophenolblue) and heated at 95 ℃ for 5 min. After cooling down, the slurry was centrifuged and the proteins in the supernatant were separated by 12% SDS-PAGE, transferred to a PVDF membrane, and detected with streptavidin-PE. The stained blot was imaged by a Typhoon trio instrument (excitation, 546 nm green laser; emission, 580 nm band-pass filter) from GE Healthcare Biosciences (Piscataway, NJ).

The other aliquot was resuspended in 50 µl of 50 mM ammonium bicarbonate buffer and heated at 70 ℃ for 1 hour. After cooling, 1 µl of 1 µg/µl trypsin in 50 mM

ammonium bicarbonate buffer was added and incubated at 37 ℃ overnight. The slurry was then centrifuged and the supernatant containing digested peptides was subjected to LC-MS/MS analysis using the Agilent 1200 capillary HPLC system interfaced Bruker ESI-QTOF II instrument (Bruker, Billerica, MA).

The LC separation was carried out on a Phenomenex Synergi™ column (50 x 2 mm, 2.5 μm particle size, purchased from Phenomenex, Torrance, CA), using the following gradient at flow rate 0.4 ml/min: solvent A is water with 0.1% formic acid; solvent B is 100% acetonitrile with 0.1% formic acid. 0 min: 100% A; 5 min: 100% A; 38 min: 35% A, 65% B; 40 min: 100% B; 41 min: 100% B; 42.5 min: 100% A; 44 min: 100% A. The data was analyzed by Bruker DataAnalysis 4.0 and searched in the NCBI database using Bruker Daltonics BioTools 3.2.

In vitro* Reconstitution, Detection, and Identification of Overexpressed ThiS Protein in *Escherichia coli.

The plasmid pCAC111 containing *E. coli thiFS* gene was constructed and transformed into *E. coli BL21(DE3)* as previously described[82]. A single colony of *E. coli BL21(DE3)* containing pCAC111 was grown at 37 ℃ in 10 ml LB medium to reach an OD_{600} of 0.6. The ThiFS protein overexpression was induced by adding 500 μM IPTG at 15 ℃ and incubated overnight. The culture was then harvested and the cell extract was obtained by sonication in 3 ml 100 ml PBS buffer followed by centrifugation. The thiocarboxylate group of the thiS protein was reconstituted by adding 6 mM $MgCl_2$, 9 mM Na_2S, and 18 mM ATP and allowed to react at room temperature for 6 hours. After

the reaction, the lysate was taken to 20% ammonium sulfate saturation at 4 ℃ by slowly adding solid ammonium sulfate and gently stirring for 45 min. The mixture was then centrifuged at 15000 rpm for 20 min, and the supernatant was slowly added to ammonium sulfate to make the final concentration 50%. The mixture was again centrifuged after gently stirring for 45 min and the resulting pellet was stored at -80 ℃.

The protein pellet was dissolved in 600 μl PBS buffer and 100 μl was removed to add 2.5 mM biotin sulfonyl-azide. After reacting for 30 min, 30 mM TCEP was added and allowed to react for another 30 min. The mixture was then incubated with 300 μl streptavidin resin slurry for 30 min at room temperature. The resin was washed with 10 ml 100 ml PBS buffer, 10 ml 50 mM ammonium bicarbonate buffer and then incubated at 70 ℃ for 1 h to denature the bound proteins. After cooling, 5 μg trypsin in 5 μl 50 mM ammonium bicarbonate buffer was added to the resin and incubated at 37 ℃ in a shaker overnight. The resulting peptide solution was collected by centrifugation and subjected to LC-MS/MS analysis as previously described.

2.3 Result and Discussion

Procedure Design for Identification of Sulfur Carrier Proteins

After the thioacid-azide reaction was selected as the biorthogonal probe reaction, biotin sulfonyl-azide was designed as the probe molecule. The sulfonyl-azide group covalently attaches to the sulfur carrier protein by reacting with its C-terminal thiocarboxylate functional group. For the reporter group, biotin was chosen as the tag molecule due to its high affinity for streptavidin[83-85]. Similar to the basic strategy for

activity based protein profiling, a protocol was designed to target and identify sulfur carrier proteins as illustrated in figure 7. In it, sulfur carrier proteins are first reconstituted to their active form by adding ATP and sulfide for the conversion of the C-terminus into the thiocarboxylate group. The probe molecule, biotin sulfonyl-azide, is then added to the proteome and specifically targets the sulfur carrier protein by reacting with the reconstituted C-terminal thiocarboxylate group. The bound proteins can then be separated by gel electrophoresis and detected by anti-biotin western blotting. Alternatively, these labeled proteins can also be isolated and enriched on streptavidin resin. For identification, on-resin tryptic digestion is usually carried out due to the difficulty in breaking the biotin-streptavidin interaction, and the resulting peptides are analyzed by LC-MS/MS. The proteins of interest can then be identified by a Mascot database search.

Synthesis of Biotin Sulfonyl-azide

This part of the work was done by Dr. Sameh Abdelwahed in Tadhg P. Begley's group in Texas A&M University. Biotin sulfonyl-azide was synthesized from 4-carboxybenzenesulfonazide and biotin hydrazide. The purified product was obtained in 44% yield and characterized by ^{1}H- NMR and high resolution ESI-MS.

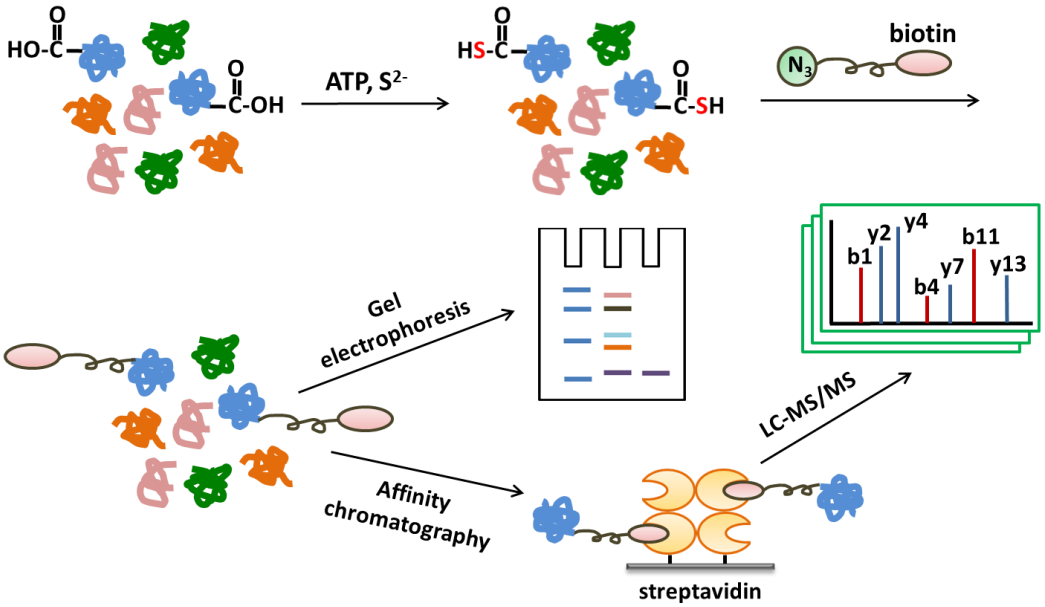


Figure 7. Schematic diagram of probing and identifying sulfur carrier proteins. After reconstitution and labeling steps, sulfur carrier proteins can either be detected by SDS-PAGE and anti-biotin western blotting, or enriched by affinity chromatography and identified by LC-MS/MS.

Method Validation with Purified *Thermus thermophilus* ThiS Protein

To validate this method, purified *Thermus thermophilus* ThiS protein (TtThiS) was used as the model system. The *thiS* gene from *Thermus thermophilus* was cloned into pTYB1 vector, transformed into *E. coli BL21(DE3)* and overexpressed. To generate the C-terminal thiocarboxylate functional group, purification was followed the IMPACT (Intein Mediated Purification with an Affinity Chitin-binding Tag) protocol as previously described[75]. The purified TtThiS-COSH protein was reacted with biotin-sulfonazide under denaturing conditions (9 M urea), ensuring that its C-terminal thiocarboxylate group was completely exposed to the reagent. TCEP was added after 30 min to disrupt non-specific couplings. After the reaction, a strong band showed up at 10

kDa, indicating that the TtThiS protein had been successfully labeled by biotin sulfonyl-azide (figure 8). The biotinylated TtThiS was desalted and then passed through streptavidin resin, where it was observed that all the proteins were retained on the column. The bound proteins were then washed with PBS buffer and eluted by heating in SDS sample buffer. However, the elution efficiency was observed to be very low (figure 8). Consequently, on-resin tryptic digestion was performed and the TtThiS protein was successfully identified through an NCBI database search with three peptides detected and 73% sequence coverage (table 1, figure 9, Appendix A).

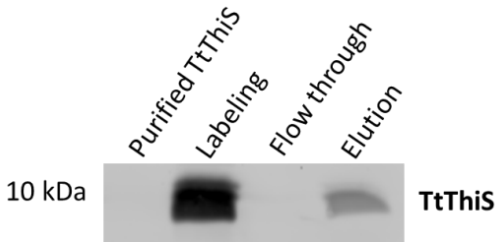


Figure 8. Labeling and elution of purified *Thermus thermophilus* ThiS protein. TtThiS (100 μM) was labeled with 2.5 mM biotin sulfonyl-azide in 9 M urea, 50 mM KPi buffer pH 6.0 for 30 min, followed by a 30 min reaction with 30 mM TCEP. Excess biotin sulfonyl-azide reagent was removed by Biorad P6 size-exclusion column and the resulting reaction mixture was loaded onto a 150 μl streptavidin resin slurry. Elution was carried out in SDS sample buffer and heated at 95 ^{o}C for 5 min. All the samples were separated by 12% SDS-PAGE and transferred to a PVDF membrane with anti-biotin staining by streptavidin-PE.

Table 1. Identified *Thermus thermophilus* ThiS protein

Name	GI Number	Organism	Sequence Coverage
Putative thiS protein	46198624	*Thermus thermophilus HB27*	73%

Figure 9. Sequence of the identified *Thermus thermophilus* ThiS protein. The detected peptides are highlighted in red and the cleavage site is marked in blue.

Detection and Identification of Overexpressed ThiS in the *E. coli* Proteome

To further validate the method with a more complex system, the *E. coli* ThiS protein was overexpressed with its activating enzyme ThiF in *E. coli BL21(DE3)*. After lysing the cell, the *E. coli* proteome was treated with ATP, NaHS, and $MgSO_4$ to reconstitute the sulfur carrier proteins for the formation of the C-terminal thiocarboxylate (figure 3). The proteome solution was then fractionated by different percentages of ammonium sulfate to avoid insolubility[82]. By adding biotin-4-carboxybenzenesulfonyl-azide, the fraction precipitated between 20% and 50% ammonium sulfate was found to be labeled. In the anti-biotin western blot, a strong band was detected below 10 kDa (actual molecular weight: 7.4 kDa) corresponding to the ThiS protein (figure 10). This indicates that *E. coli* ThiS has been successfully reconstituted and labeled with biotin sulfonyl-azide, demonstrating the feasibility of the labeling protocol. Nevertheless, a small portion of ThiS protein was not retained on the streptavidin resin, which is probably because excess labeling reagent competed with the biotinylated ThiS proteins for binding with streptavidin.

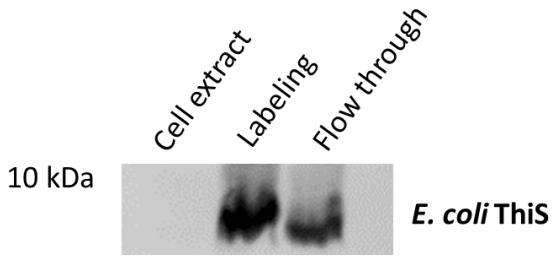


Figure 10. Labeling of ThiS protein in ThiFS overexpressed *E. coli* proteome. The ThiS protein was reconstituted *in vitro* by reacting with 6 mM $MgCl_2$, 9 mM Na_2S, and 18 mM ATP for 6 h at room temperature. The 20%-50% ammonium sulfate precipitation fraction of the reconstituted cell extract was reacted with 2.5 mM biotin sulfonyl-azide in 9 M urea, 50 mM KPi pH 6.0. After reaction, the reaction mixture was loaded on streptavidin resin and the flow through was collected. All the samples were separated by 12% SDS-PAGE and transferred to a PVDF membrane with anti-biotin staining with streptavidin-PE.

The bound *E. coli* ThiS proteins were washed with PBS buffer and put through on-resin tryptic digestion, as in the previous procedure on purified TtThiS. *E. coli* ThiS protein was successfully identified with one peptide and 22% sequence coverage (table 2, figure 11, Appendix A). Here, the relatively low sequence coverage is mainly due to the limited number of tryptic cleavage sites in the *E. coli* ThiS protein: there are only two cutting sites which theoretically will generate 3 peptides. The C-terminal peptide, however, is expected to be retained on the resin and won't be detected. Also, since the excess labeling reagent competes with the biotinylated ThiS protein in binding the streptavidin resin, the amount of bound protein may not be enough for detection of the N-terminal peptide as some peptides are difficult to ionize and thus do not provide good response to mass spectrometry.

Table 2. Identified *Escherichia coli* ThiS protein

Name	GI Number	Organism	Sequence Coverage
Sulfur carrier protein ThiS	38704219	*Escherichia coli*	22%

Figure 11. Sequence of identified *Escherichia coli* ThiS protein. The detected peptides are highlighted in red and the cleavage site is marked in blue.

2.4 Conclusion

In this chapter, a chemical probe biotin sulfonyl-azide was designed to specifically target sulfur carrier proteins through the thioacid-azide reaction. The method was developed with purified *Thermus thermophilus* ThiS-COSH protein which was successfully detected in western blotting. The protein was further identified by on-resin tryptic digestion and LC-MS/MS analysis with high sequence coverage. Using the *E. coli* ThiFS overexpressed system, this method was modified with the addition of thiocarboxylate reconstitution and proteome fractionation. The success of labeling and identifying the *E. coli* ThiS protein validates the use of this method in a complex proteome.

3. IDENTIFICATION AND FUNCTIONAL STUDY OF A NEW SULFUR CARRIER PROTEIN IN *STREPTOMYCES COELICOLOR*

3.1 Introduction

In chapter 2, we developed and validated a method to specifically probe, enrich, and identify sulfur carrier proteins in the proteome. Now this method can be used to explore other bacterial sulfur carrier proteins. Back in 2010, Kalyan Krishnamoorthy from the Begley lab developed a fluorescent reagent to target sulfur carrier proteins and detected one protein in *Streptomyces coelicolor*[75], however, due to the high sensitivity of fluorescence and the poor protein recovery from gel-based methods, the detected protein was visualized as a strong band but was not identified. Therefore, the fluorescence and gel-based technique is a good detection method but it has limited capacity to identify sulfur carrier proteins in the proteome, especially those in low abundance. The method developed in chapter 2 requires an additional western blotting step in detection, but is superior at identifying the target protein by enriching them through high-throughput affinity chromatography.

In this chapter, we explored *Streptomyces coelicolor* for the unidentified sulfur carrier protein. After successful identification, this protein was further investigated by bioinformatics study and we unraveled its reaction with another protein annotated as threonine synthase. The product of the two proteins reaction was characterized and a mechanism was proposed for the discovered sulfur carrier protein.

3.2 Experimental Methods

Chemicals and Reagents

All chemicals were purchased from Sigma Aldrich (St. Louis, MO) unless otherwise mentioned. Isopropyl-D-thiogalactopyranosid (IPTG) was from Lab Scientific Inc. (Livingston, NJ). Luria-Bertani medium (LB) was from EMD Biosciences (Gibbstown, NJ) and M9 salts were obtained from Becton Dickinson (Franklin Lakes, NJ). *Streptomyces coelicolor* (ATCC 10147) was purchased from the American Type Culture Collection (Manassas, VA). Streptavidin resin was from GenScript (Piscataway, NJ) and streptavidin-R-phycoerythrin (PE) was from QIAGEN (Valencia, CA). The Novagen D-tube Maxi dialyzer with 3.5 kDa molecular weight cutoff was obtained from EMD Biosciences. The PVDF membrane and trypsin gold was purchased from Promega (Madison, WI). The Fe-NTA phosphopeptide enrichment kit was from Pierce (Rockford, IL).

All fluorescence gel images were scanned using a Typhoon trio instrument (excitation, 532 nm green laser; emission, 580 nm band-pass filter (580 BP 30)) from GE Healthcare Biosciences (Piscataway, NJ). A sonicator 3000 from Misonix Inc. (Farmingdale, NY) was used for sonication. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis was done using a Hoefer SE 250 mini-vertical gel electrophoresis unit.

Analytical HPLC (Agilent 1260 instrument) was carried out using a Supelcosil LC-18 column (150 mm x 4.6 mm, 3 µm particle size). LC-MS (Agilent 1260 instrument and Bruker microTOF-Q II) was carried out using a Supelcosil LC-18

column (150 mm X 4.6 mm, 3 μm particle size) for small molecule analysis and Phenomenex Synergi™ LC Column (50 x 2 mm, 2.5 μm particle size) for peptide separation. The solvents for HPLC and LC-MS were HPLC grade and LC-MS grade respectively and were obtained from EMD Biosciences (Gibbstown, NJ).

In vitro Reconstitution, Detection, and Identification of a New Sulfur Carrier Protein in *Streptomyces coelicolor*

Streptomyces coelicolor was maintained on an LB agar plate at 4 ^{o}C. A colony was inoculated in 30 ml LB medium and grown untill OD_{600} reached 0.8. The cells were harvested, washed with M9 medium and then resuspended in 1.5 L M9 minimal medium supplemented with 2 mM $MgSO_4$, 100 μM $CaCl_2$, and 0.4% glucose. The culture was grown for one week and then harvested followed by resuspension in 15 ml PBS buffer. Cell extract was obtained by sonication and centrifugation. The C-terminal thiocarboxylate group of the sulfur carrier protein was reconstituted by adding 6 mM $MgCl_2$, 9 mM Na_2S, and 18 mM ATP and reacted at room temperature for 6 hours. After the reaction, the lysate was precipitated by different concentrations of ammonium sulfate and the fraction between 20% and 50% was collected. The protein pellet was then dissolved in 600 μl PBS buffer from which a 100 μl aliquot was removed and added to 2.5 mM biotin sulfonyl-azide. The mixture was allowed to react for 30 min; then 30 mM TCEP was added and reacted for another 30 min. The excess biotin sulfonyl-azide was removed by desalting using Biorad P6 size-exclusion chromatography and the labeled protein solution was incubated with 300 μl streptavidin resin slurry for 30 min at room

temperature. The resin was then washed with 100 mM PBS buffer and digested with 5 μg trypsin. The resulting peptides were subjected to LC-MS/MS as previously described for TtThiS protein identification and the data was searched in the NCBI database using Bruker BioTools 3.2.

Overexpression and Purification of *Streptomyces coelicolor* Sulfur Carrier Protein (SCO4294) and Threonine Synthase (SCO4293)

The gene of the detected sulfur carrier protein SCO4294 in *Streptomyces coelicolor* was cloned into pTYB1 and the neighboring gene threonine synthase (SCO4293) was cloned into pET28b by Dr. Kinsland (Cornell, protein facility) and both vectors were transformed into *E. coli BL21(DE3)*. Overexpression and purification of the *S. coelicolor* sulfur carrier protein SCO4294 was achieved using the IMPACT system, the same procedure as for the purification of *Thermus thermophiles* ThiS-COSH described in chapter 2. For threonine synthase, the transformed *E. coli BL21(DE3)* cells were grown in 3 L Luria-Bertani medium at 37 ^{o}C to an OD_{600} of 0.6. Protein overexpression was induced by adding 500 μM IPTG at 15 ^{o}C and incubating overnight. The resulting culture was harvested using the Avanti JE centrifuge from Beckman Coulter and cell extract was obtained by sonication in 50 ml lysis buffer (50 mM Tris, 300 mM NaCl, 10 mM imidazole, 1 mM EDTA, 2 mM TCEP, pH 8.0) and centrifugation at 14000 rpm for 0.5 hours. The cell extract was then passed through a 5 ml HisTrap HP column and threonine synthase was eluted with 100 mM imidazole in

lysis buffer. The protein was concentrated using 10 kDa MWCO, desalted by a 10 DG desalting column, and stored in 100 mM PBS buffer with 30% glycerol.

Reaction between *Streptomyces coelicolor* Sulfur Carrier Protein (SCO4294) and Threonine Synthase (SCO4293)

O-phospho-L-homoserine was enzymatically synthesized by homoserine kinase (thrB) with L-homoserine as the substrate. Plasmid containing the *thrB* gene was obtained from the *E. coli* ASKA collection and was transformed into *E. coli BL21(DE3)*. For purification, cells were grown in 3 L Luria-Bertani medium at 37 ^{o}C to an OD_{600} of 0.6. Protein overexpression was induced by adding 500 µM IPTG at 15 ^{o}C and incubating overnight. The resulting culture was harvested using an Avanti JE centrifuge from Beckman Coulter and cell extract obtained by sonication in 50 ml lysis buffer (50 mM Tris, 300 mM NaCl, 10 mM imidazole, 1 mM EDTA, 2 mM TCEP, pH 8.0) and centrifugation at 14000 rpm for 0.5 hours. The cell extract was then passed through a 5 ml HisTrap HP column and threonine synthase was eluted with 100 mM imidazole in lysis buffer. The protein was concentrated using a 10 kDa MWCO filter, desalted by a 10 DG desalting column, and stored in 100 mM PBS buffer with 30% glycerol. The reaction for enzymatically synthesizing *O*-phospho-L-homoserine was performed in 100 mM HEPES/KOH pH 8.0 with 20 mM KCl, 10 mM $MgCl_2$, 8 mM ATP, and 10 mM L-homoserine in a total volume 50 ml. The reaction was triggered after adding 2 µg homoserine kinase and incubated at 37 ^{o}C for 30 hours. *O*-phospho-L-homoserine was purified by anion exchange chromatography using 5 g Biorad AG1 X8 resin and eluted

with 12 mM HCl. The ninhydrin positive fractions were collected, concentrated, and the concentration was quantitated using a malachite green assay.

ESI-MS Analysis and Western Blot Assay of the Reaction between the *Streptomyces coelicolor* Sulfur Carrier Protein (SCO4294) and Threonine Synthase (SCO4293)

The purified *O*-phospho-L-homoserine was reacted (final concentration 100 μM) with 10 μM threonine synthase, 200 μM ThiS-like protein and 1 mM TCEP at 37 ^{o}C for 2 hours. For ESI-MS analysis, the reaction mixture was buffer exchanged into 20 mM NH_4OAc using Biorad P6 size-exclusion chromatography to remove the small molecules. It was then diluted in a 1:1 solution of acetonitrile and water to adjust the concentration of the sulfur carrier protein to 10 μM. Each sample was added with 0.1% formic acid and directly infused into ESI-QTOF II at a flow rate of 300 μl/h. The spectra were collected for 5 min and then averaged and deconvoluted using Bruker Daltonics DataAnalysis 4.0. For the western blot assay, the reaction mixture was separated by 12% SDS-PAGE and then transferred onto a PVDF membrane by Biorad Trans-Blot Turbo System at 100 V for 30 min. The blot was washed with TBS buffer (2 mM Tris-HCl, 50 mM NaCl, pH 7.5), blocked with 3% BSA (dissolved in TBS buffer), washed with TBS buffer again and then stained by streptavidin-PE at 1:3000. The stained blot was stored in TBS buffer and imaged using a Typhoon trio instrument (excitation, 546 nm green laser; emission, 580 nm band-pass filter (580 BP 30)) from GE Healthcare Biosciences (Piscataway, NJ).

HPLC Analysis of the Reaction between the *Streptomyces coelicolor* Sulfur Carrier Protein (SCO4294) and Threonine Synthase (SCO4293)

To detect small molecule products in the reaction, the solution was filtered with a 3.5 kDa MWCO filter and the filtrate was collected. Twenty microliters of 50 mM iodoacetic acid (dissolved in 100 mM sodium borate, pH 11.5) was added to 20 µl filtrate and allowed to react for 0.5 h in the dark at room temperature. The mixture was then added to 16.7 µl fresh OPA reagent (50 µl of 37 mM *O*-phthalaldehyde + 200 µl of 400 mM sodium borate pH 10.5 + 8 µl β-mecaptoethanol) and left in the dark at room temperature for 10 min. The resulting solution was loaded onto HPLC or LC-MS after passing through a 3.5 kDa MWCO filter, with the signal detected by FLD in HPLC (extinction: 340 nm; emission: 455 nm) or by DAD in LC-MS at 340 nm absorbance. Alternatively, 50 µl filtrate was mixed with 1 µl 100 mM bromobimane (final concentration 2 mM) to react for 1 hour. After filtration with the 3.5 kDa MWCO filter, the mixture was loaded on HPLC or LC-MS, and signal was detected by UV absorbance at 390 nm.

HPLC separation was carried out on a Supelcosil LC-18 column (150 mm x 4.6 mm, 3 µm particle size), using the following gradient at flow rate 1.0 ml/min: solvent A is water; solvent B is 100 mM KH_2PO_4, pH 6.6; solvent C is methanol. 0 min: 100% B; 5 min: 45% A, 40% B, 15% C; 8 min: 40% A, 25% B, 35% C; 16 min: 40% A, 15% B, 45% C; 28 min: 25% A, 10% B, 65% C; 32 min: 100% B; 38 min: 100% B.

For LC-MS analysis, the LC separation was carried out on a Supelcosil LC-18 column (150 mm x 4.6 mm, 3 µm particle size), using the following gradient at flow rate

0.4 ml/min: solvent A is 5 mM NH_4OAc pH 6.6; solvent B is 75% methanol. 0 min: 100% A; 5 min: 100% A; 20 min: 70% A, 30% B; 22 min: 70% A, 30% B; 24 min: 100% A; 29 min: 100% A.

Detection of the PLP binding site of *Streptomyces coelicolor* "Threonine Synthase" (SCO4293) by IMAC (Immobilized Metal Affinity Chromatography) Based Phosphopeptide Enrichment

To enrich the PLP binding peptide of "threonine synthase" SCO4293, 100 μM SCO4293 was thawed and desalted with a Biorad P6 size-exclusion column to remove the glycerol. After desalting, the sample buffer was Tris-HCl buffer pH 7.5. Around 1 mg $NaBH_4$ (dissolved in water) was then added and the mixture was left on ice for 0.5-1 h to reduce the imine bond formed between PLP and the lysine residues. The solution was then buffer exchanged to 25 mM ammonium bicarbonate, pH 8.0 using a Biorad P6 size-exclusion column.

For trypsin digestion, the protein was heat denatured at 70-80 ^{o}C for 1 hour. After cooling, trypsin was added in a 1:50 (trypsin:protein) ratio and the mixture was incubated overnight at 37 ^{o}C. Solvent from the resulting peptide solution was removed by SpeedVac (Thermo Fisher Scientific Inc., Waltham, MA) and then the protocol provided in the Pierce Fe-NTA phosphopeptide enrichment kit was followed for PLP bound peptide enrichment. The enriched peptide solution was injected into LC-MS/MS as previously described and the result was analyzed by Bruker DataAnalysis 4.0.

3.3 Result and Discussion

Detection and Identification of a Sulfur Carrier Protein in *Streptomyces coelicolor*

The *S. coelicolor* proteome was first treated with ATP, NaHS, and $MgCl_2$ to reconstitute the thiocarboxylate functional group of the sulfur carrier protein. The proteome solution was then precipitated by different percentages of ammonium sulfate to fractionate the protein[82]. By adding biotin 4-carboxybenzene sulfonyl-azide, the fraction that precipitated between 20% to 50% ammonium sulfate was found to be labeled showing a strong band at 9.8 kDa in anti-biotin western blotting (figure 12). This molecular weight is in accordance with the reported small size of sulfur carrier proteins, and the labeling pattern is similar as that in Dr. Kalyan Krishnamoorthy's paper[75].

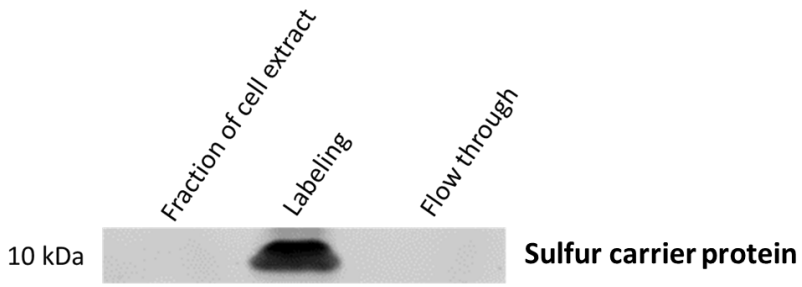


Figure 12. Detection of sulfur carrier protein(s) in *S. coelicolor*. Cell extract was *in vitro* reconstituted by 6 mM $MgCl_2$, 9 mM NaHS, and 18 mM ATP. The 20%-50% ammonium sulfate precipitation fraction of the reconstituted cell extract was labeled by reacting with 2.5 mM biotin sulfonyl-azide in 9 M urea, 50 mM KPi buffer pH 6.0. After the reaction, the mixture was desalted with a Biorad P6 size-exclusion column and passed through streptavidin resin.

Meanwhile, non-specific labeling was observed in the high molecular weight region of the gel (Appendix A). Though these bands are much weaker than that of the

sulfur carrier protein, they could still cause interference in the protein identification step. To remove the non-specifically labeled proteins, the protein bound resin was heated in SDS sample buffer at 95 ^{o}C. It was observed that the large non-specifically biotinylated proteins were eluted out under this condition but the sulfur carrier protein was retained on the resin (Appendix A). This may be because small biotinylated proteins interact more strongly with streptavidin due to less steric hindrance.

It should be noted that there is another sulfur carrier protein in *S. coelicolor*, annotated as ThiS. It was not picked up using our strategy, probably due to the low expression level under the growth condition. Also, we used sulfide ion to reconstitute the thiocarboxylate group in the sulfur carrier protein, but it is possible that this ThiS protein needs other sulfur source for reconstitution. Furthermore, if a sulfur carrier protein's C-terminal is blocked by other amino acid residues such as alanine, it cannot be labeled without removal of the blocking residue.

Table 3. Identified *Streptomyces coelicolor* sulfur carrier protein

Name	GI Number	Organism	Sequence Coverage
Hypothetical protein SCO4294	21222687	*Streptomyces coelicolor A3(2)*	50%

Figure 13. Sequence of identified *S. coelicolor* sulfur carrier protein. The detected peptides are highlighted in red and the cleavage site is marked in blue.

Using this approach, the detected sulfur carrier protein was identified as hypothetical protein SCO4294 (table 3, figure 13, appendix A) with 50% sequence coverage. The C-terminal characteristic glycine-glycine sequence further confirms its membership in the sulfur carrier protein family. Also, it should be noted that the C-terminal peptide was not detected in LC-MS/MS because it remained bound to the solid phase after on-resin tryptic digestion.

By experimenting with purified TtThiS and the *E. coli* proteome with overexpressed ThiS, a method of probing, enriching, and identifying sulfur carrier proteins has been established and validated. Using this method, a sulfur carrier protein was detected in *Streptomyces coelicolor* and identified as hypothetical protein SCO4294. Compared to the previous fluorescence based detection method[75], this affinity chromatography based method can enrich and identify sulfur carrier proteins, enabling the detection and identification of sulfur carrier proteins even at low expression levels. As for the identified hypothetical protein SCO4294, further study is needed to unravel its cellular function.

Bioinformatics Study of *Streptomyces coelicolor* Sulfur Carrier Protein (SCO4294)

To elucidate the function of sulfur carrier protein SCO4294, the gene cluster of *S. coelicolor* was investigated on the SEED database. Interestingly, unlike other identified sulfur carrier proteins, SCO4294 conservatively neighbors with the gene SCO4293 which was annotated as threonine synthase (figure 14). To the best of our knowledge, sulfur carrier proteins are usually associated with a ThiF-like protein for its activation and another non-conserved protein for transferring the sulfur from the thiocarboxylate group to corresponding targets. However, in the case of SCO4294, only one protein, threonine synthase, was found to be conservatively close to the identified sulfur carrier protein. Moreover, threonine synthase is an enzyme that converts *O*-phospho-L-homoserine to threonine, which does not include a sulfur atom. Hence, it is highly possible that the threonine synthase is mis-annotated and that sulfur carrier protein SCO4294 may be involved in a novel metabolic pathway.

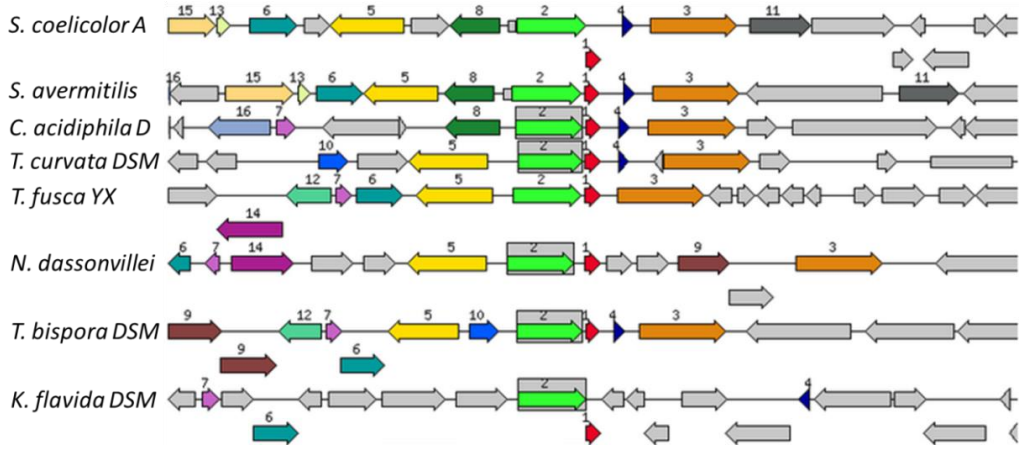


Figure 14. Gene cluster of SCO4294 in *S. coelicolor* and analogs in other species[77]. 1: ThiS/MoaD-family protein; 2: threonine synthase; 3: heat shock protein; 4: cold shock protein.

Study of the Reaction between the *Streptomyces coelicolor* Sulfur Carrier Protein (SCO4294) and Threonine Synthase (SCO4293)

Genes of the hypothetical protein SCO4294 and threonine synthase SCO4293 from *Streptomyces coelicolor* were cloned into the pTYB1 and pET28b vectors, respectively, by Dr. Kinsland from the Cornel Protein Facility. The plasmids were transformed into *E. coli BL21(DE3)* and overexpressed. In order to generate the C-terminal thiocarboxylate functional group, the purification of the sulfur carrier protein was followed by the IMPACT (Intein Mediated Purification with an Affinity Chitin-binding Tag) protocol as previously described[75]. The threonine synthase was cloned with an N-terminal His-tag, purified by Ni-NTA chromatography, and eluted in 100 mM imidazole.

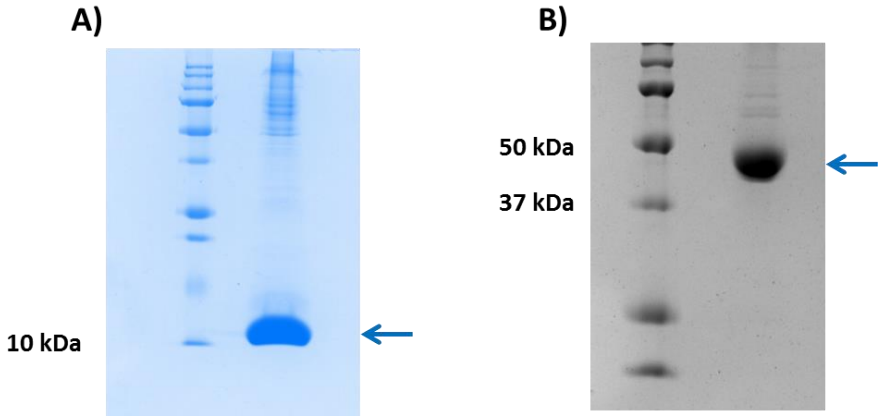


Figure 15. Purification of *S. coelicolor* SCO4294 and SCO4293. (A) The sulfur carrier protein SCO4294 was purified by the IMPACT system. (B) Threonine synthase SCO4293 was purified by Ni-NTA chromatography. The blue arrows refer to the position of the proteins of interest.

To find out whether the two proteins react with each other, *O*-phospho-L-homoserine was selected as the substrate according to the mechanism of reported threonine synthases[86]. The substrate was enzymatically synthesized and purified by anion exchange chromatography. After incubating *O*-phospho-L-homoserine with sulfur carrier protein SCO4294 and threonine synthase SCO4293 at 37 ^{o}C for 2 h, the mixture was desalted and the protein fraction was tested by ESI-MS. Here, we hypothesized that the molecular weight of the sulfur carrier protein would change due to sulfur donation if a reaction occurred between the substrate and the two proteins. The result is shown in figure 16. The mass of the sulfur carrier protein decreased by 16.0 Da (from 9455.0 to 9439.0), after incubating with threonine synthase and *O*-phospho-L-homoserine. The 16.0 Da decrease corresponds with the hydrolysis of the thiocarboxylate group (-COSH) to a carboxylate group (-COOH).

To further confirm that the mass decrease of 16.0 Da attributes to the loss of sulfur from the sulfur carrier protein, the reaction mixture was treated with biotin sulfonyl-azide followed by SDS-PAGE separation and anti-biotin western blot assay. Clearly, the sulfur carrier protein donates its sulfur in the full reaction but not in the absence of the substrate *O*-phospho-L-homoserine or the enzyme threonine synthase (figure 17). Other substrates, including O-phospho-L-serine, O-acetyl-L-homoserine, and O-acetyl-L-serine, were also tested by the same method but the sulfur carrier protein didn't hydrolyze with any of these molecules as the substrate (data not shown).

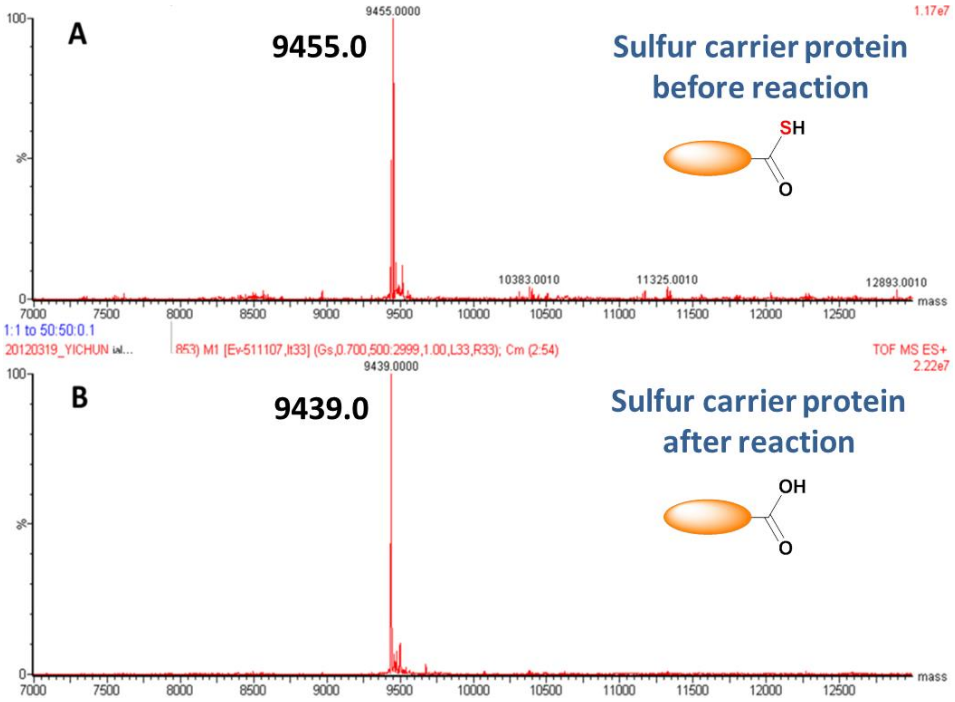


Figure 16. ESI-MS detection of the *S. coelicolor* sulfur carrier protein before and after the reaction with *O*-phospho-L-homoserine and threonine synthase. The reaction mixture was desalted and buffer exchanged into 20 mM NH_4OAc. A) ESI-MS spectrum of the purified *S. coelicolor* sulfur carrier protein before reaction; B) ESI-MS spectrum of the *S. coelicolor* sulfur carrier protein after reaction.

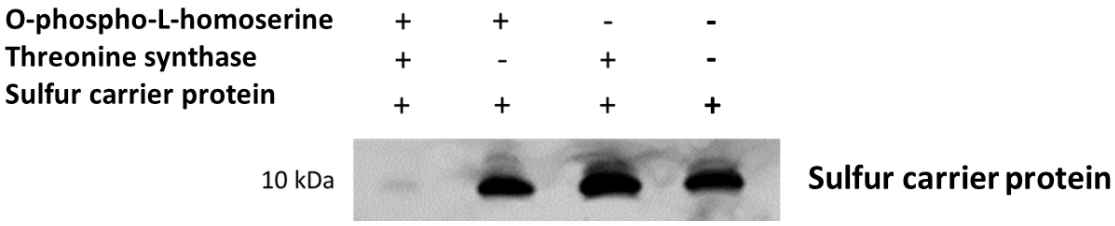


Figure 17. Western blot detection of the reaction between the *S. coelicolor* sulfur carrier protein and threonine synthase with *O*-phospho-L-homoserine as the substrate. All the samples were treated with biotin sulfonyl-azide reagent to target the thiocarboxylate group.

Product Characterization of the Reaction between the *Streptomyces coelicolor* Sulfur Carrier Protein (SCO4294) and Threonine Synthase (SCO4293)

After the sulfur carrier protein was confirmed to react with *O*-phospho-L-homoserine and threonine synthase, we needed to characterize the product of this reaction. We proposed a mechanism (figure 18) which went through PLP based catalysis and generated thiothreonine as the product.

Since thiothreonine is not a UV-active compound, it had to be derivatized before detection. For amino acids, derivatization of the free amino group with *O*-phthalaldehyde (OPA) and β-mecaptoethanol is a highly sensitive method and is widely used. However, for thiol-containing compounds, the thiol group interferes with the derivatization reaction and causes very poor signal[87-89]. Therefore, iodoacetic acid is used to protect the free thiol group, and then the compound is derivatized by OPA and β-mecaptoethanol (figure 19).

Meanwhile, to confirm the product contains a free thiol group, the reaction mixture was also derivatized by bromobimane, which reacts specifically with the –SH functional group (figure 20).

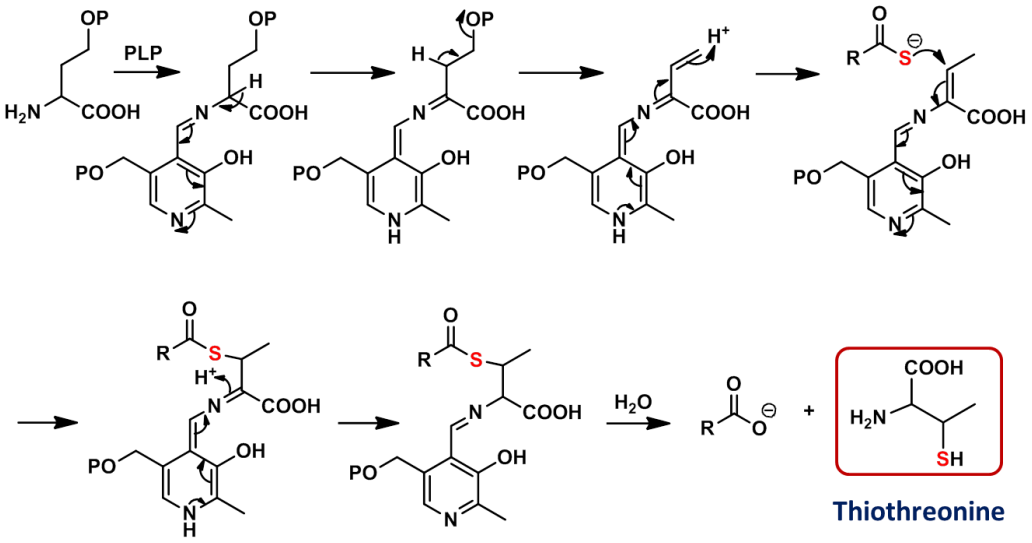


Figure 18. Proposed reaction mechanism of SCO4294. The reaction is proposed to use a PLP mediated mechanism and generates thiothreonine as the product.

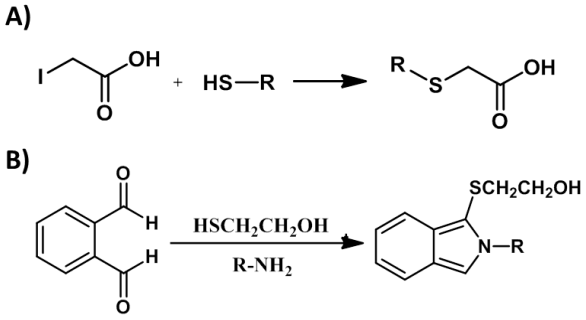


Figure 19. Derivatization reactions of amino thiol-containing compounds. A) The thiol group is protected by iodoacetic acid. B) The amino group is then derivatized by *O*-phthalaldehyde with β-mecaptoethanol. The derivative is a fluorescent molecule with excitation at 340 nm and emission at 455 nm.

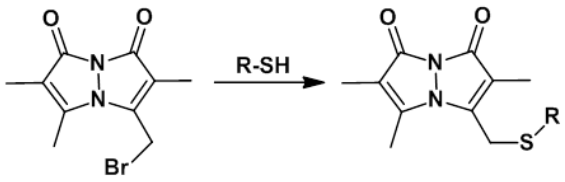


Figure 20. Derivatization reaction of thiol-containing compounds by bromobimane. The derivative has a UV absorbance of 390 nm.

After derivatization, the mixtures were subjected to HPLC and LC-MS analysis. For OPA derivatized samples, HPLC spectra were collected by FLD at excitation 340 nm and emission 455 nm whereas signals in LC-MS were monitored by DAD at 340 nm. For bromobimane derivatized reactions, signals in both HPLC and LC-MS were obtained at UV 390 nm. The results are shown in figure 21. It is clearly seen that a new peak is generated in the full reaction, which confirms that the product has both $–NH_2$ and –SH groups since it was detected by both derivatization methods. By comparing with reference molecules including cysteine, homocysteine, methionine, threonine, and homoserine, the major product had the same retention time as homocysteine in HPLC. Two side products, threonine and homoserine, were also observed but in very low abundance (data not shown). Unlike the major product, homocysteine requires the presence of both the sulfur carrier protein and threonine synthase, the side product threonine is only dependent on threonine synthase, and homoserine is a non-enzymatic product.

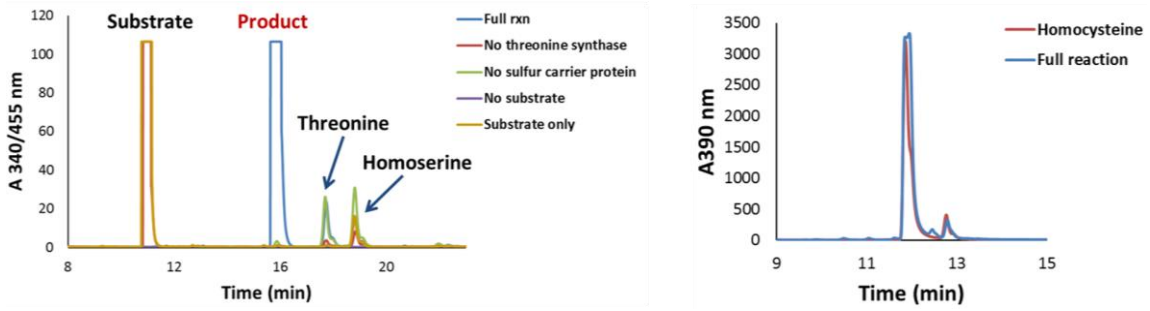


Figure 21. HPLC spectra of the SCO4294 reaction after derivatization. *S. coelicolor* sulfur carrier protein was reacted with threonine synthase and *O*-phospho-L-homoserine. A) All the samples were derivatized by iodoacetic acid and OPA reagent (OPA+β-mecaptoethanol). Blue trace: full reaction; red trace: reaction without threonine synthase; green trace: reaction without sulfur carrier protein; purple trace: reaction without substrate; yellow trace: substrate only. B) The full reaction (blue trace) and homocysteine (red trace) were derivatized by bromobimane respectively and detected by DAD at 390 nm.

To further confirm that homocysteine is the product, the product fractions in each HPLC spectrum were collected and analyzed by LC-MS. Both fractions showed a single UV absorption peak in liquid chromatography, and the corresponding masses are presented in figure 22. The *m/z* 370.1 and *m/z* 326.1 peaks correspond to the OPA derivatized and bromobimane (mBBr) derivatized homocysteine, respectively.

As homocysteine, rather than thiothreonine, was confirmed as the product, the following two-step mechanism was proposed (figure 23). In the first step, *O*-phospho-L-homoserine loses the phosphate group and forms a thioester with the sulfur carrier protein, catalyzed by threonine synthase. In the second step, the product homocysteine is released through hydrolysis and the C-terminus of the sulfur carrier protein returns to being the carboxylate group. This mechanism is simple and independent of PLP.

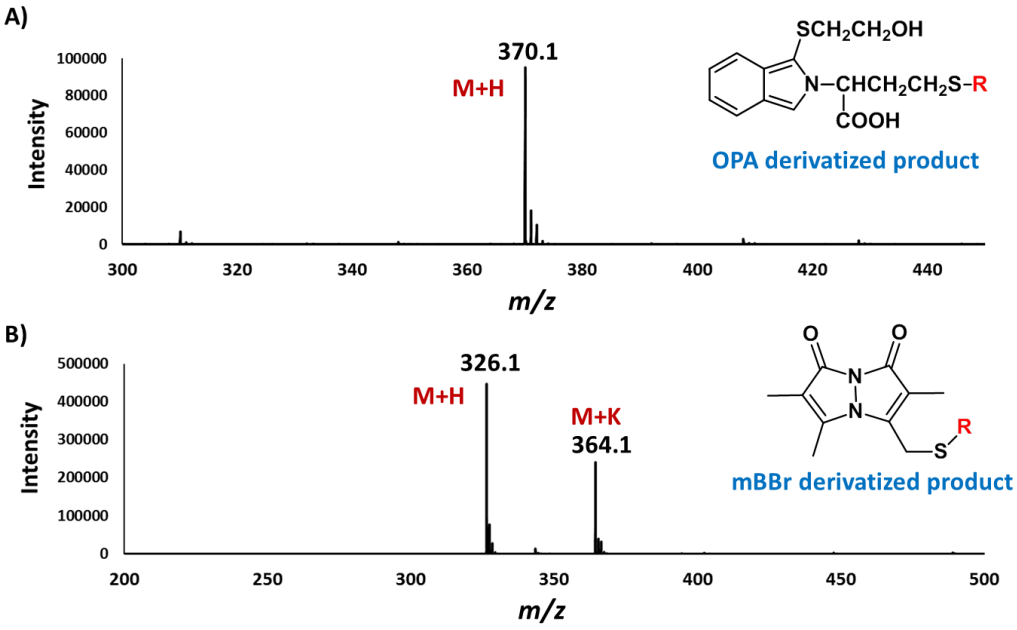


Figure 22. LC-MS spectra of the collected HPLC fractions. A) Mass spectrum of the OPA derivatized product; B) mass spectrum of the bromobimane derivatized product.

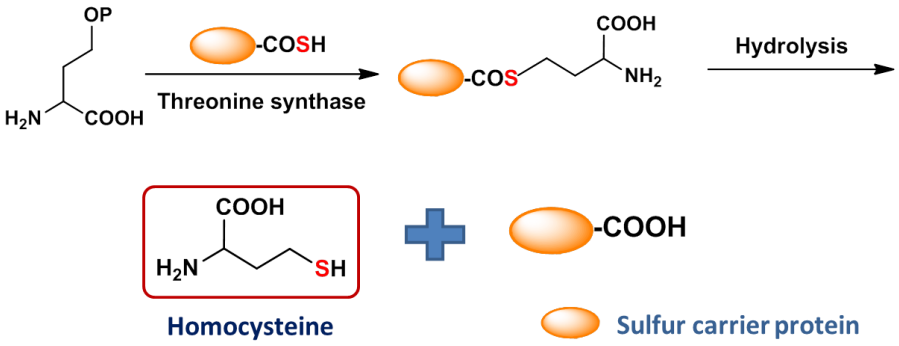


Figure 23. Proposed mechanism for the homocysteine formation. The sulfur carrier protein is proposed to form a covalent adduct with *O*-phospho-L-homoserine and then hydrolyze to produce homocysteine.

Homocysteine has been discovered to be a metabolite in the methionine biosynthetic pathway, and its formation is from homoserine. In *E. coli*, homoserine is

first converted to O-succinylhomoserine and then goes through trans-sulfurylation with cysteine to form homocysteine, with cystathionine as the intermediate in this process. In many other bacterial organisms, however, the sulfurylation step is achieved by directly reacting with hydrogen sulfide or methanethiol, without the formation of cystathionine. This process is also known as a direct sulfhydrylation pathway for methionine biosynthesis.

Homocysteine formation in *S. coelicolor* is different from all other reported organisms. First, in the homoserine activation step, it uses *O*-phosphohomoserine whereas all other bacteria reported use O-succinylhomoserine or O-acetylhomoserine[90]. The enzymes for catalyzing these two reactions are MetA (homoserine O-succinyltransferase) and MetX (homoserine O-acetyltransferase), respectively. Nevertheless, only a homoserine kinase was found in *S. coelicolor*, but not MetA or MetX. *S. coelicolor* uses *O*-phosphohomoserine to synthesize both threonine and methionine, as plants do. In addition, the homoserine kinase gene in *S. coelicolor* is right beside another threonine synthase gene (SCO5307), indicating the latter is a "real" threonine synthase in the threonine biosynthetic pathway. As for the discovered "threonine synthase" SCO4293, it is more likely to be an *O*-phosphohomoserine thiolase though it shows high sequence similarity as threonine synthases.

Second, since there is no cystathionine formation as the intermediate, this pathway can be considered as a direct sulfhydrylation pathway. But unlike the other direct sulfhydrylation pathways which use hydrogen sulfide or methanethiol, the *S. coelicolor* pathway uses a sulfur carrier protein. The advantage of using the sulfur carrier

protein is probably due to its high efficiency in sulfur transfer, especially when the organism lives in a low sulfur environment. Back in 2011, a new methionine biosynthetic pathway in *Wolinella succinogenes* was reported, which is also dependent on a sulfur carrier protein[91]; it uses O-acetylhomoserine instead of *O*-phosphohomoserine. Also, the *Wolinella succinogenes* pathway is more complicated and requires more enzymes including MetY as the O-acetylhomoserine thiolase, and HcyD for hydrolyzing the homocysteine sulfur carrier protein complex.

Therefore, compared with the reported sulfurylation in methionine biosynthesis, the homocysteine formation pathway in *S. coelicolor* is novel in three aspects: 1) homoserine is activated as *O*-phosphohomoserine, which has not been found in other bacteria or archaea; 2) in the direct sulfurylation step, it uses a sulfur carrier protein instead of hydrogen sulfide or methanethiol, but requires fewer enzymes; 3) SCO4293, the enzyme that acts as an *O*-phosphohomoserine thiolase has high sequence similarity to threonine synthases and actually shows minor catalytic activity for threonine formation.

In the following, SCO4293 is still noted as threonine synthase for consistency but is written in quotation marks to indicate it's not a real threonine synthase.

Comparison of Sulfur Donation Efficiency of the Sulfur Carrier Protein (SCO4294) with Sulfide Ion

As mentioned above, hydrogen sulfide is the sulfur source in direct sulfurylation pathway in many bacterial organisms. Therefore, we tested whether the sulfide ion could

be used as the sulfur source for homocysteine formation. Through HPLC analysis, it was found that addition of the sulfide ion did lead to product formation but with lower yield than when using the sulfur carrier protein.

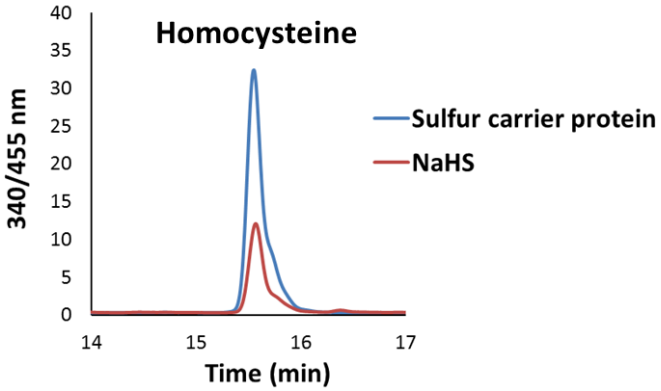


Figure 24. HPLC spectrum of SCO4293 reaction with different sulfur sources. Both sulfur carrier protein and NaHS were added at a concentration of 100 μM. The product was derivatized by iodoacetic acid and OPA reagent (*O*-phthalaldehyde and β-mecaptoethanol) and was detected by FLD at excitation 340 nm, emission 455 nm. Blue trace: *S. coelicolor* sulfur carrier protein SCO4294 as the sulfur source; red trace: NaHS as the sulfur source.

To quantify the product ratio and find out how much more efficient the sulfur carrier protein is, the sulfur carrier protein SCO4294-CO^{32}SH was mixed with NaH^{34}S (NaH^{34}S was prepared by Dr. Bekir Eser) at a 1:1 ratio and reacted with "threonine synthase" SCO4293 and *O*-phospho-L-homoserine. After filtering the proteins, the reaction mixture was derivatized by iodoacetic acid and OPA reagent, followed by LC-MS analysis. From the relative intensity in mass spectrum (figure 12), the ratio of ^{32}S-homocysteine versus ^{34}S-homocysteine is easily calculated. Since the natural isotope distribution of the derivatized homocysteine is M: (M+2) = 100%:11.90%, the actual

^{34}S-homocysteine generated from NaH^{34}S is 25.67% - 11.90% = 13.77%. Hence, the ratio of homocysteine produced from the sulfur carrier protein and sulfide is 100:13.77 = 7.3:1. This demonstrates the higher reaction efficiency of the sulfur carrier protein as the sulfur donor compared with a sulfide ion alone.

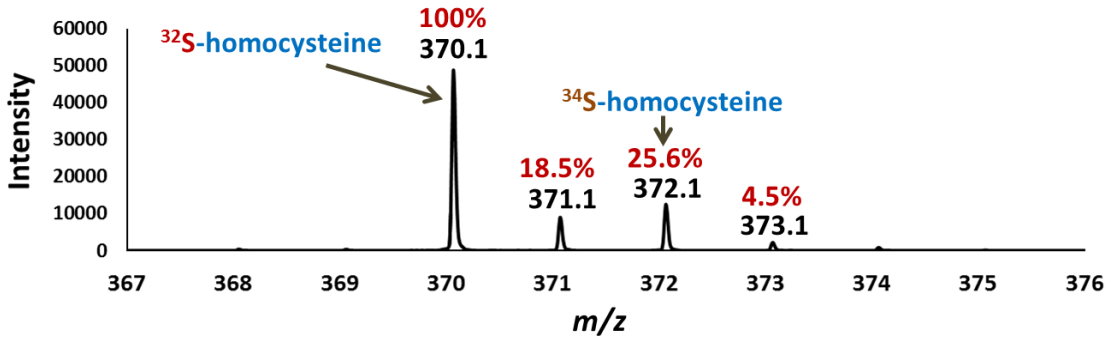


Figure 25. LC-MS spectrum of ^{32}S and ^{34}S labeled homocysteine. Homocysteine was produced in the reaction catalyzed by *S. coelicolor* "threonine synthase" SCO4293 with 1:1 mixed sulfur carrier protein SCO4294-CO^{32}SH and NaH^{34}S as the sulfur source (total concentration: 100 μM). The product was derivatized by iodoacetic acid and OPA reagent prior to LC-MS analysis and was detected by DAD at 340 nm.

Investigation of the PLP Binding Site of *Streptomyces coelicolor* "Threonine Synthase" (SCO4294)

For PLP-dependent enzymes, the PLP binding site is usually the active site of the enzyme. Hence, we digested "threonine synthase" SCO4293 with trypsin and analyzed the fragments by LC-MS/MS to find the PLP binding site. PLP bound peptides, however, suffer from lower ionization efficiency and fewer fragments in tandem mass spectrometry. Compared to non-modified peptides, phosphopeptides are more difficult to ionize during mass spectrometry analysis. Furthermore, the phospho group causes

difficulty in tandem mass spectrometry fragmentation. Tandem mass spectrometry has been widely used for peptide identification, the most common of which is collision activated dissociation (CAD) or collision induced dissociation (CID). In the case of phosphopeptides, however, CAD or CID promotes elimination of phosphoric acid instead of breaking the amide bond[92]. This also causes difficulty in identifying phosphopeptides, especially those with low abundance.

Consequently, it is necessary to enrich phosphopeptides prior to identification of the PLP binding site. Immobilized metal ion affinity chromatography (IMAC) and metal oxide (TiO_2) based affinity chromatography are the most commonly used methods[93-96]. In these methods, the negatively charged phosphate group coordinates with the metal ion (Fe^{3+}, Ni^{2+}, Cu^{2+}, Zn^{2+}, etc) or metal oxide (TiO_2) and is consequently retained on the column while non-phosphorylated peptides are washed away. Among the metal ions, Fe^{3+} has a high affinity towards phosphate groups and a relatively low affinity for carboxyl and phenolic groups, so it is most frequently used in the IMAC technique[97]. The enriched phosphopeptides can be eluted out with ammonium hydroxide solution and then subjected to LC-MS/MS analysis. Though the enrichment technique is not highly specific, it still increases the yield of phosphopeptides for identification[92]. Here, we used the Fe-NTA kit from Pierce to enrich the PLP bound peptides after protein digestion.

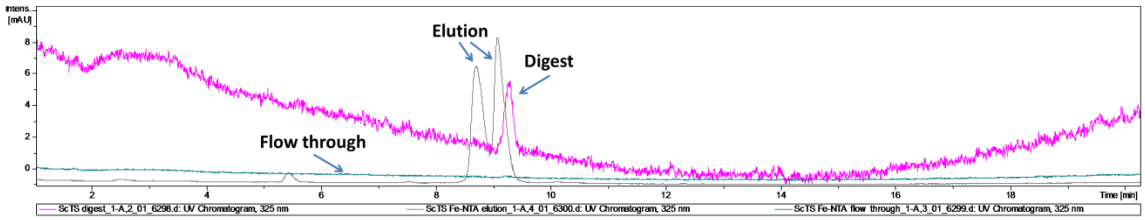


Figure 26. LC chromatogram of threonine synthase PLP bound peptides. The tryptic digested peptides of *S. coelicolor* "threonine synthase" were collected before and after phosphopeptide enrichment by IMAC chromatography. The chromatogram traces were collected at 325 nm, corresponding to the absorbance of reduced PLP bound peptides. Purple trace: *S. coelicolor* "threonine synthase" tryptic digested peptides without IMAC enrichment; cyan trace: flow through during IMAC enrichment; gray trace: elution after IMAC enrichment.

Prior to trypsin digestion of SCO4293, the PLP molecule was first "fixed" to the protein's lysine residue by adding $NaBH_4$ to reduce the imine bond. The PLP bound protein was then digested and enriched by the Fe-NTA kit. The UV absorbance of the reduced PLP peptide is 325 nm, enabling detection by DAD in liquid chromatography. From figure 26, it is obvious that there is a weak signal in the digested peptide mixture, which increases markedly in the elution trace after enrichment. Also, the averaged mass spectrum after enrichment (figure 27) is much cleaner than the digested sample, which contains more than 20 peptide peaks (data not shown). It should be noted that there are two distinct peaks in the elution chromatogram, which are probably different species of the same peptide (i.e. COOH and COO^-) since they give the same mass spectra.

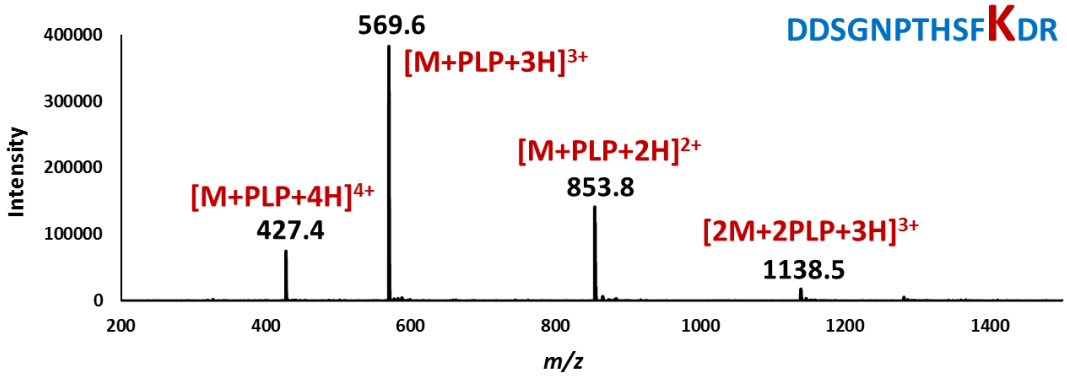


Figure 27. Mass spectrum of detected PLP bound peptide in threonine synthase. The spectrum was averaged corresponding to the major UV absorption peaks in the LC chromatogram. The peak (*m/z* 569.6) with highest intensity corresponds to the PLP bound peptide with three positive charges. Peptides with four and two positive charges were also observed. The mass of the detected peptide is in accordance with peptide DDSGNPTHSFKDR in *S. coelicolor* "threonine synthase" SCO4293, with PLP bound to the lysine residue K128.

The obtained PLP bound peptide mass spectrum (figure 27) was analyzed and compared with the theoretically generated peptides after trypsin digestion. The peaks observed in the spectrum with *m/z* 427.4, *m/z* 569.6, and *m/z* 853.8 were found to belong to the PLP bound peptide DDSGNPTHSFKDR with +4, +3, and +2 positive charges, respectively. A dimer formed by this peptide was also observed with a +3 positive charge at *m/z* 1138.5.

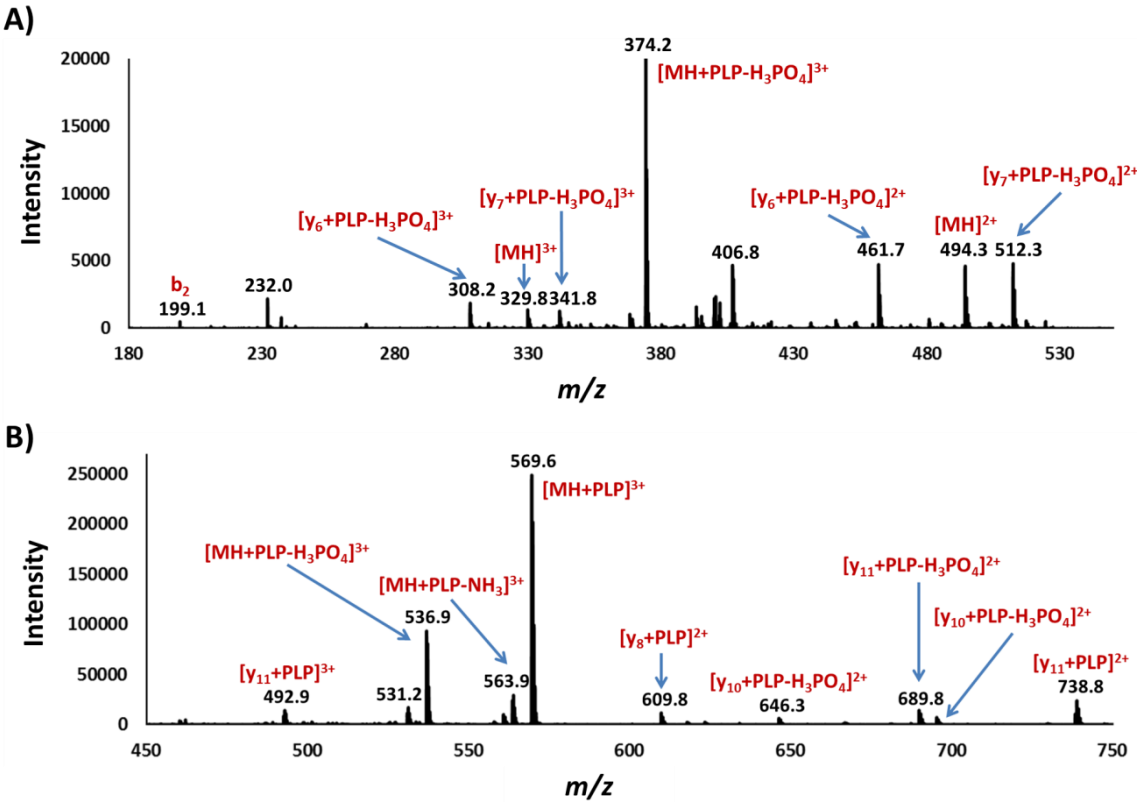


Figure 28. Tandem mass spectra of the detected PLP bound peptide peaks. (A) Fragmentation of peak *m/z* 427.4; (B) fragmentation of peak *m/z* 569.6. The major loss of the parent ions is attributed to the loss of PLP or phosphoric acid. For peptide fragments, b_2, y_6, and y_7 ions were detected in the fragmentation of peak *m/z* 427.4 whereas y_8, y_{10} and y_{11} ions were detected in the fragmentation of peak *m/z* 569.6.

In addition, two peaks, *m/z* 427.4 and *m/z* 569.6, were fragmented to confirm the identity of peptide DDSGNPTHSFKDR. In both tandem mass spectra, the major loss of the parent ion attributes to the loss of PLP or phosphoric acid. Fragments corresponding to breaking of a peptide amide bond were also observed but with low intensity. In the tandem mass spectrum of the parent peak *m/z* 427.4, b_2, y_6, and y_7 ions were detected (figure 28A), whereas y_8, y_{10}, and y_{11} ions were observed in the fragmentation of *m/z* 569.6 (figure 28B).

As a result, DDSGNPTHSFKDR was found to be the PLP bound peptide, and K128 was identified as the PLP binding site, which is probably also the enzyme active site.

3.4 Conclusion

In this chapter, we utilized the developed probing method to detect, enrich, and identify an unknown sulfur carrier protein as hypothetical protein SCO4294 in *Streptomyces coelicolor*. Unlike other sulfur carrier proteins that associate with ThiF-like proteins or ThiG-like proteins, SCO4294 was found to conservatively cluster with a protein annotated as "threonine synthase" (SCO4293). The reaction between sulfur carrier protein SCO4294 and "threonine synthase" SCO4293 was studied with *O*-phospho-L-homoserine as the substrate. The product generated in this reaction was homocysteine. Herein, the protein annotated as "threonine synthase" is actually an *O*-phosphohomoserine thiolase, though it possesses a high sequence similarity to threonine synthases. This pathway is another direct sulfurylation pathway for homocysteine formation, which is probably involved in methionine biosynthesis. This is also the first time that *O*-phosphohomoserine is shown to be the activation product of homoserine in bacterial homocysteine formation.

Sodium hydrosulfide can be an alternative sulfur donor in this reaction but with much lower efficiency. By using ^{34}S-labeld NaHS, the sulfur transfer efficiency of the sulfur carrier protein was quantified to be 7.3 times greater than the sulfide ion. This demonstrates the higher efficiency of the sulfur carrier protein, and may also be an

explanation for *S. coelicolor* use of a sulfur carrier protein instead of a sulfide ion. Moreover, the PLP binding site of the "threonine synthase" SCO4293 was studied by IMAC based phosphopeptide enrichment and PLP was shown bound to K128, which is probably the active site of this enzyme.

4. DEVELOPMENT OF A RADIOACTIVITY BASED PROTEOMIC METHOD FOR PROBING AND IDENTIFYING PLP-DEPENDENT ENZYMES

4.1 Introduction

Pyridoxal 5'-phosphate (PLP) is one of the active forms of vitamin B6. In cells, it acts as a cofactor and assists the catalysis of a large variety of reactions. Since its identification in 1951[98], the diverse catalyzing styles of PLP-dependent enzymes have been studied, including transamination[99], decarboxylation[100], racemization[101], elimination, and replacement[102]. The versatility attributes to the ability of PLP to covalently bind with substrates and stabilize different types of cabanionic intermediates. Almost all PLP-dependent enzymes are involved in biomedical pathways that contain amino compounds.

Apart from functional diversity, PLP-dependent enzymes are also classified by five different fold types, in accordance with their different structures[103]. Fold type I usually functions as a homodimer, with two active sites per dimer. It is the most common structure and is found in aminotransferases and decarboxylases. Fold type II enzymes mainly includes enzymes that catalyze β-elimination reactions. Fold type III features a $(\beta/\alpha)_8$ barrel structure and is found in alanine racemase as well as some amino acid decarboxylases. The fold type IV group mainly contains D-alanine aminotransferases, whereas fold type V includes glycogen and starch phosphorylases. The structural classification helps identify PLP-dependent enzymes through genomic methods[104].

The functional and structural diversity of PLP-dependent enzymes makes it difficult to investigate them as part of a whole proteome. Most PLP-dependent enzymes are studied individually after recombination, overexpression, and purification, allowing each enzyme to be characterized in detail, such as its catalysis mechanism, kinetic behavior, and cellular function. However, this method is time consuming and is unsuitable for high-throughput purposes like searching for new PLP-dependent enzymes. Proteomics, a high-throughput technique, can solve this problem to some extent by separating and identifying all the proteins in the proteome. But since we are only interested in PLP proteins, activity based protein profiling is more suitable in that it only focuses on a certain class of enzymes. Proteins are profiled and distinguished according to their biological activity and functional states in the cell. As a consequence, there is much less interference from other proteins and enzymes with low abundance can be more easily discovered.

As described in chapter 1, chemical probes are the key factors in activity base protein profiling. These probes are usually composed of a reactive group for the active site probe, a flexible linker, and a tag for visualization or isolation. Pyridoxal-5'-phosphate itself is an ideal probe for activity based protein profiling. It forms aldimine with the active site lysine residue in the enzyme, which can be "fixed" by sodium borohydride through a Schiff base reduction. Therefore, PLP forms the reactive and linker group and only the reporter group needs to be designed.

Radioactive labeling, though challenging in operation, is advantageous in its sensitivity and specificity. The phosphate group of the PLP molecule can be easily

labeled by ^{32}P. This can be achieved through enzymatic conversion of pyridoxal to pyridoxal 5'-phosphate with ^{32}P-ATP, catalyzed by pyridoxal kinase. The ^{32}P-PLP bound enzymes are then able to be specifically visualized through radiation detection. This allows PLP-dependent enzymes to be studied and investigated as a whole family in the proteome. Considering that *E. coli* is able to biosynthesize PLP by itself, a PLP auxotroph strain, *E. coli ΔpdxJ* was used to encourage sufficient uptake of ^{32}P-PLP from the media.

Since radioactive samples cannot be loaded onto LC-MS for detection, a non-radioactive gel can be run under identical conditions to the radioactive experiments to identify the PLP-dependent enzymes. The radiation screened image and the non-radioactive scanned gel can be overlapped to determine the PLP enzymes' positions. The bands in non-radioactive gel will then be excised and digested, followed by peptide analysis for protein identification.

4.2 Experimental Methods

Chemicals and Reagents

[γ-^{32}P]ATP (10 Ci/mmol 2 mCi/ml, 250 μCi) was obtained from Perkin Elmer (Waltham, MA). Pyridoxal hydrochloride, pyridoxal-5'-phosphate, adenosine triphosphate disodium salt, and sodium azide were purchased from Sigma Aldrich (St. Louis, MO). B-Per solution was obtained from Pierce (Rockford, IL). Streptavidin resin was from GenScript (Piscataway, NJ) and streptavidin-R-phycoerythrin (PE) was from QIAGEN (Valencia, CA). The PVDF membrane and trypsin gold were purchased from

Promega (Madison, WI). M9 salts were obtained from Becton Dickinson (Franklin Lakes, NJ). The *E. coli ΔpdxJ* strain was from the Coli Genetic Stock Center in Yale University (New Haven, CT).

All radioactive gels were dried with a Biorad 583 gel dryer (Bio-Rad Laboratories Inc, Hercules, CA) and then incubated with a storage phosphor screen in an exposure cassette, both from Molecular Dynamics (Sunnyvale, CA). The storage phosphor screen was then scanned using a Typhoon trio instrument from GE Healthcare Biosciences (Piscataway, NJ). UV absorbance of protein samples was tested on a Varian Cary 300 Bio UV-Visible Spectrophotometer (Palo Alto, CA). Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was carried out using a Hoefer SE 250 mini-vertical gel electrophoresis unit.

Enzymatic Synthesis of Pyridoxal-5'-[^{32}P]-phosphate

The pyridoxal kinase (pdxK) plasmid was prepared by Dr. Cynthia Kinsland at the Cornell Protein Facility. The plasmid was transformed into *E. coli BL21(DE3)* cells and grown overnight at 37 ^{o}C on an LB agar plate with 100 μg/ml ampicillin. A colony was inoculated in 3 L LB medium at 37 ^{o}C to reach an OD_{600} of 0.6. Protein overexpression was induced by adding 500 μM IPTG at 15 ^{o}C and followed by overnight incubation. The resulting culture was harvested using Avanti JE centrifuge from Beckman Coulter and cell extract was obtained by sonication in 50 ml lysis buffer (50 mM Tris, 300 mM NaCl, 10 mM imidazole, 1 mM EDTA, 2 mM TCEP, pH 8.0), followed by centrifugation at 14000 rpm for 0.5 hours. The cell extract was then passed

through a 5 ml HisTrap HP column and washed with lysis buffer. Pyridoxal kinase was eluted with 300 mM imidazole in lysis buffer. The protein was concentrated using a 10 kDa MWCO filter, desalted by a 10 DG desalting column, and stored in 100 mM PBS buffer with 30% glycerol. The protein concentration was determined to be 300 µM by Bradford assay.

For synthesis of pyridoxal-5'-[^{32}P]-phosphate, pyridoxal (20 µl of 50 mM stock) was added to 230 µl 100 mM PBS buffer pH 7.2 to a final concentration of 2 mM. $MgCl_2$ (0.8 µl of 1 M stock) and ATP (120 µl [γ-^{32}P]ATP (10 Ci/mmol 2 mCi/ml, 250 µCi) and 100 µl 10 mM "cold" stock) were added. The reaction was triggered by adding 50 µl 300 µM pdxK and incubated at 37 ^{o}C for 1 hour. The mixture was transferred to a 10 kDa MWCO filter and centrifuged at 5000 rpm for 15 min. The filtrate was collected and stored at -20 ^{o}C for use in *E. coli* growth experiments.

Preparation and Analysis of *E. coli* Proteome

The PLP auxotroph *ΔpdxJ* strain was grown on an LB-kanamycin-agar plate at 37 ^{o}C for 12 hours. A colony was added to sterile M9 minimal media containing 0.4% glucose, 2 mM $MgSO_4$, 100 µM $CaCl_2$, 50 µM $Fe(NH_4)_2(SO_4)_2$, 40 µg/ml kanamycin, and 15 µM pyridoxal-5'-phosphate or 15 µM of enzymatically synthesized radioactive ^{32}P-PLP, each filtered through a sterile syringe drive with a 0.23 µm filter (Millipore, Bedford, MA). All samples were grown in a 37 ^{o}C shaker for 14 h for non-time course experiments. The cultures were transferred into 2 ml eppendorf tubes and centrifuged multiple times to pellet the cells. The resulting pellet was resuspended in 200 µl B-Per

solution for protein extraction and allowed to sit at room temperature for 15 min. The resulting protein solution was obtained by centrifugation and treated with $NaBH_4$ (1 mg/ 200 µl lysate) for 10 min at room temperature. The frothy mixture was desalted with a methanol/chloroform precipitation as previously described[23]. The resulting protein pellet was dissolved in 20 µl SDS sample buffer and 10 µl of each sample was loaded onto a 12% SDS gel and run at 90 V in a Biorad Mini-Protein Tetra Cell (Bio-Rad Laboratories Inc, Hercules, CA). The gels were dried on a Biorad 583 gel dryer and developed on a storage phosphor screen overnight. The storage phosphor screen was then scanned using a Typhoon trio instrument from GE Healthcare Biosciences (Piscataway, NJ).

Growth Conditions of *E. coli ΔpdxJ* under Different Stress Conditions

For nutrient starvation stress conditions, *E. coli ΔpdxJ* was grown in sterile M9 minimal media containing 0.4% glucose, 2 mM $MgSO_4$, 100 µM $CaCl_2$, 50 µM $Fe(NH_4)_2(SO_4)_2$, 15 mM radioactive ^{32}P-PLP (40 µCi), and 40 µg/ml kanamycin. In starvation experiments (including amino acids starvation, glucose starvation, nitrogen starvation, and PLP starvation), the culture was grown in the media described above except for the amino acid starvation experiment, in which 0.1% casamino acids were supplemented. After 14 hours of growth, cells were harvested, washed three times with M9 minimal media and then resuspended in the same media depleted of one component (casamino acids for amino acid starvation, glucose for glucose starvation, NH_4Cl in M9 media for nitrogen starvation, and PLP for PLP starvation). The resuspended cells were

incubated at 37 ^{o}C for 20 h and a 1.5 ml culture was taken out at different time points. The culture was harvested and the cell pellet was stored at -80 ^{o}C for further use.

For all other stress conditions, *E. coli ΔpdxJ* was grown in sterile M9 minimal media containing 0.4% glucose, 2 mM $MgSO_4$, 100 µM $CaCl_2$, 50 µM $Fe(NH_4)_2(SO_4)_2$, 15 µM radioactive ^{32}P-PLP (40 µCi), and 40 µg/ml kanamycin at 37 ^{o}C for 14 hours. Stress conditions were implemented afterward. For cold shock, the temperature was decreased to 15 ^{o}C and cells were incubated for 6.5 hours. Heat shock was achieved by incubating at 42 ^{o}C for 2 min. In acid shock, 20 µl 34% HCl was added to 10 ml culture to reduce the pH to around 4.5. For osmotic stress, sucrose was added to a 40% final concentration and incubated at 37 ^{o}C. Oxidative stress was triggered by adding 1 mM H_2O_2 to the medium, and the UV irradiation experiment was completed by exposing cells to UV irradiation (254 nm, 1 J/m2/sec) for 1 min.

Identification of PLP Enzymes in *E. coli* Proteome with Radioactive Detection

E. coli ΔpdxJ was grown in 10 ml sterile M9 minimal media containing 0.4% glucose, 2 mM $MgSO_4$, 100 µM $CaCl_2$, 50 µM $Fe(NH_4)_2(SO_4)_2$, 40 µg/ml kanamycin, and 15 µM ^{32}P-PLP or non-radioactive PLP. After 14 hours of growth at 37 ^{o}C, cells growing with ^{32}P-PLP were cold shocked at 15 ^{o}C and incubated for 3 hours. Non-radioactive PLP supplemented *E. coli* cells were grown in four different conditions: 7 h after glucose starvation, 6 h PLP after starvation, 3 h after osmotic stress, and 3 h after cold shock. The cells were harvested and the cell pellet was resuspended in 1 ml B-Per solution. After a 30 min incubation, the protein mixture was obtained by centrifugation

and treated with 5 mg $NaBH_4$. The reaction was left for 15 min and then desalted using methanol/chloroform precipitation. The resulting protein pellet was dissolved in 200 µl SDS sample buffer and 100 µl of the suspension was loaded onto a 20 cm x 20 cm 12% SDS gel. The gel was run at 120 V on an Owl Dual-Gel Vertical Electrophoresis System (ThomasScientific, Swedesboro, NJ) overnight. The radioactive gel was dried and incubated with a storage phosphor screen for imaging. The non-radioactive gel was Coomassie stained and the resulting gel image was overlapped with the radioactive gel screen. Bands shown in the radioactive gel were excised from the corresponding positions in the non-radioactive gel and placed in clean eppendorf tubes. The gel pieces were covered with acetonitrile for 15 min at room temperature and then dried with a SpeedVac (Thermo Fisher Scientific Inc., Waltham, MA). The dried gel pieces were re-swelled in buffer containing 1 µg trypsin in 50 mM NH_4HCO_3 at 4 ^oC for 45 minutes, then incubated at 37 ℃ overnight. The supernatant containing the digested peptides were collected, and 25 mM NH_4HCO_3 was added to the gel pieces. After soaking for 25 min, the supernatant was collected and peptides were further extracted by soaking the gel pieces in a solution of 50% acetonitrile, 45% water, and 5% formic acid for 30 minutes. This process was repeated 3 times and all the supernatants were collected, combined, and dried by SpeedVac to a volume of 15 µl. The peptide solution was then subjected to LC-MS analysis.

LC separation was carried out on a Phenomenex Synergi™ column (50 x 2 mm, 2.5 µm particle size), using the following gradient at a flow rate 0.4 ml/min: solvent A, water with 0.1% formic acid; solvent B, 100% acetonitrile with 0.1% formic acid. 0 min:

100% A; 5 min: 100% A; 38 min: 35% A, 65% B; 40 min: 100% B; 41 min: 100% B; 42.5 min: 100% A; 44 min: 100% A. The data was analyzed by Bruker DataAnalysis 4.0 and the NCBI database was searched using Bruker Daltonics BioTools 3.2.

4.3 Result and Discussion on the ^{32}P Radioactive Labeling Method

Visualization of PLP-dependent Enzymes in *E. coli* Proteome by ^{32}P Incorporation

The plasmid containing the pyridoxal kinase (pdxK) gene was transformed into *E. coli BL21(DE3)* and overexpressed in 3 L LB media. The His-tagged protein was purified by Ni-NTA chromatography and eluted in 300 mM imidazole (figure 29).

To create the reporter group, radioactive ^{32}P must be incorporated into the 5' position of PLP. This was accomplished by reacting the purified pdxK with pyridoxal, $MgCl_2$, and both radioactive and non-radioactive ATP (figure 30). After a 1 h incubation at 37 ^{o}C, the color of the reaction mixture changed to bright yellow, indicating the production of PLP.

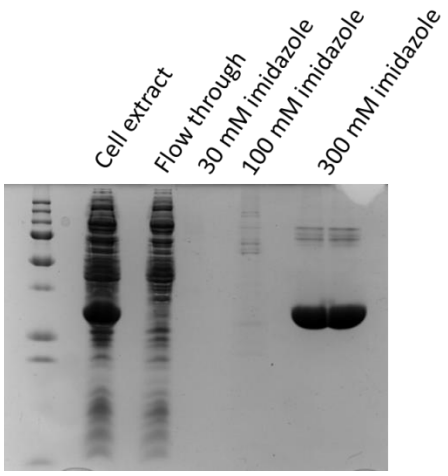


Figure 29. Purification of pyridoxal kinase. The protein was overexpressed in *E. coli BL21(DE3)* in 3 L LB media and purified by Ni-NTA chromatography. After washing with 30 mM and 100 mM imidazole, pdxK was eluted with 300 mM imidazole. 10 µl of the cell extract, flow through, 30 mM imidazole elution, 100 mM imidazole elution, and 300 mM imidazole elution were each mixed with 2 x SDS sample buffer, and then loaded onto a 12% SDS mini gel. The gel was run at 90 V and stained with Coomassie brilliant blue.

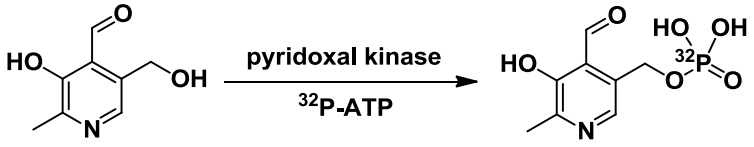


Figure 30. Enzymatic synthesis of ^{32}P-PLP. The pyridoxal (final concentration 2 mM) was mixed with 1.6 mM $MgCl_2$, 2 mM ATP ("cold" ATP + 250 µCi ^{32}P-ATP), and 30 µM pdxK. The reaction mixture was incubated at 37 ^{o}C for 1 h and then filtered by a 10 kDa MWCO filter. The filtrate was stored at -20 ^{o}C for use in the growth of *E. coli ΔpdxJ*.

The reaction mixture was filtered with a 10 kDa MWCO filter, and 1/6 of the filtrate (40 µCi) was used to grow 10 ml *E. coli ΔpdxJ*. Since PLP-dependent enzymes are mostly involved in amino acid biosynthetic pathways, amino acid starvation and amino acid supplemented conditions were first compared. In both conditions, *E. coli*

ΔpdxJ was grown in M9 minimal media containing glucose, $Fe(NH_4)_2(SO_4)_2$, $MgSO_4$, $CaCl_2$, kanamycin, ^{32}P-PLP, and casamino acids. After growing for 14 h at 37 ^{o}C, 1.5 ml culture was taken out at 0, 1, 2, 3, 4, 5, and 20 hours for amino acid (AA) supplemented conditions. For amino acid starvation, *E. coli* cells were harvested, washed with M9 media and then resuspended in the original growth media without casamino acids. At each time point, 0, 1, 2, 3, 4, 5, and 12 hours, 1.5 ml of culture was removed, harvested, and lysed with B-Per solution. The protein mixture was then treated with $NaBH_4$ to reduce the imine bond formed between the lysine residue and the PLP aldehyde group. Proteins were separated by 12% SDS-PAGE and the ^{32}P-labeled proteins were visualized through a phosphor storage screen (figure 31).

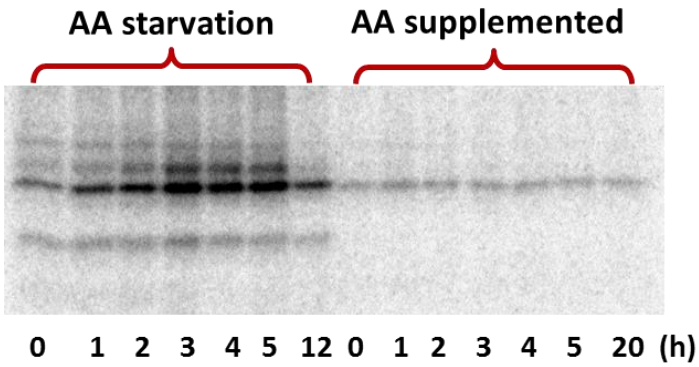


Figure 31. PLP enzymes in amino acid starved and supplemented conditions. *E. coli ΔpdxJ* was grown in minimal media with ^{32}P-PLP for 14 h and then aliquots removed at different time points. For amino acids starvation, 1.5 ml culture was taken out at 0, 1, 2, 3, 4, 5, and 12 hours; for amino acids supplemented condition, 1.5 ml culture was taken out at 0, 1, 2, 3, 4, 5, and 20 hours.

Clearly, the bands in the radioactive gel associated with the amino acid starved cells are significantly greater in intensity, indicating a greater amount of PLP enzymes

under the same conditions. Also, after the starvation starts, the PLP-dependent enzymes show significant increase, demonstrating their essentiality in the amino acid biosynthetic pathways. Considering vitamin B6 (PLP) is essential in amino acid biosynthesis, it is expected that PLP is expressed at significantly higher amounts under amino acid starvation. Therefore, for better visualization of PLP enzymes, no casamino acids were added in the following experiments. This method for visualizing PLP-dependent enzymes enables us to compare the expression patterns of PLP utilizing enzymes.

Response of *E. coli* PLP-Dependent Enzymes under Different Stress Conditions

It is well known that PLP-dependent enzymes play essential roles in amino acid biosynthesis, but in recent years, PLP was discovered to have an anti-oxidation function in singlet oxygen stress[105-107]. We carried out experiments with eight different stress conditions to explore new biological functions of PLP-dependent enzymes.

The eight conditions are categorized into three types: starvation conditions, shock conditions, and other stress conditions. In the three starvation conditions, *E. coli* was first grown in M9 minimal media with full nutrient supplementation. After overnight growth, the cells were harvested and washed with M9 minimal media to deplete the nutrient for which the cells would be starved in the experiment. The cell pellet was resuspended in the same media as it was previously growing, but with the removal of the starved nutrient: glucose as the carbon source, NH_4Cl as the nitrogen source, or PLP as the essential vitamin. Aliquots were taken out at different time points after starvation and

the expression change of the PLP-dependent enzymes were visualized by the radioactive labeling method previously described.

During glucose starvation (figure 32A), the expression of the PLP-dependent enzymes increases from 1-7 hours. At 17 h, proteins in the upper bands show a decrease, whereas the protein(s) with the lowest molecular weight are still increasing. This increase in the earlier part of the time frame indicates that the expression of the PLP-dependent enzymes is increased when the cells are experiencing carbon source starvation. The purpose of the expression increase may be to more efficiently use the limited carbon source, which is up-regulated by the bacterial stringent response. But at the 17 hour time point, when the limited amount of glucose has been depleted, PLP-dependent enzyme expression is less due to decreased need for the enzyme. The only protein (band 4) that still remains in a relatively high level at this time point may play an important role in glucose metabolism.

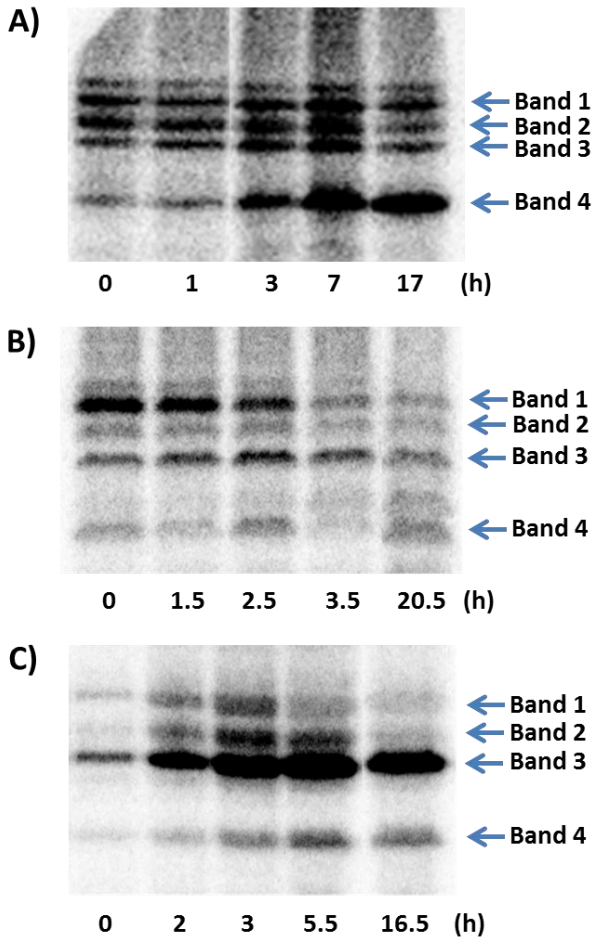


Figure 32. SDS-PAGE gels following growth of *E. coli ΔpdxJ* in minimal media containing ^{32}P-PLP under starved conditions. A) Glucose starvation, where glucose was depleted after 14 h of growth; B) nitrogen starvation, where NH_4Cl was depleted after 14 h of growth; C) PLP starvation, where PLP was depleted after 14 h of growth.

During nitrogen starvation (figure 32B), in which NH_4Cl was depleted as the nitrogen source, all PLP-dependent enzymes decrease over time. This is understandable since all amino acids rely on nitrogen metabolism, in which ammonia is assimilated into glutamate and transferred to other metabolites by deamination. When the nitrogen source is limited, cells must down regulate the production of enzymes in the amino acid

biosynthetic pathways. As a result, the PLP-dependent enzymes dramatically decrease their expression when the bacteria are nitrogen starved.

As for PLP starvation (figure 33C), interestingly, different PLP-dependent enzymes exhibit much difference in expression. While the other three bands show a weak expression level, one protein band is heavily expressed, indicating the corresponding protein(s) are somehow involved in mitigating PLP starvation. In bacterial stringent response, when the cells face nutrient starvation, the stable RNA transcription is ceased, releasing RNA polymerase for transcription of appropriate response genes[108]. As a result, some metabolite pathways are up-regulated. Here, it is highly possible that the PLP biosynthetic pathway is up-regulated for PLP production. PdxF, also known as 3-phosphoserine aminotransferase, is a PLP-dependent enzyme in PLP biosynthesis. It catalyzes the transamination between (*3R*)-3-hydroxy-2-oxo-4-phosphonooxybutanoate and glutamate to form 4-hydroxy-L-threonine phosphate[109]. From the protein identification result in later chapters, the increased band during PLP starvation should be 3-phosphoserine aminotransferase.

The second series are shock conditions. Here, instead of depletion of certain nutrients, cells were "shocked" with cold temperatures, heat, or acid. This stimulates the bacteria at the proteome level and allows us to observe the reaction of the PLP-dependent proteins. For cold shock, cells were grown in a 37 ^{o}C shaker for 14 hours and then cold shocked by incubation at 15 ^{o}C. Aliquots (1 ml) at time points 0, 0.5, 1, 2.5, 3, 4, and 6.5 hours were taken out and the previously developed protocol was followed to visualize the PLP-dependent enzymes. The gel clearly shows that the level of all PLP-

dependent enzymes decreased after the initial cold shock (0.5 h) (figure 33A). After that, the expression amount increased back and didn't show much change (only a slight increase from 1-3 h and decrease from 3-6.5 h). The sharp decrease of all the PLP-dependent enzymes at 0.5 h is probably due to the stringent response, which down-regulates most of the protein production in the cell. After shock, all the PLP enzymes returned to a normal level. This indicates that PLP-dependent proteins are not directly involved in the regulation of cold shock conditions.

As for heat shock, 10 ml *E. coli* cells were incubated at 42 ^{o}C for 2 min as the "heat shock". The culture was then returned to 37 ^{o}C and incubated to observe the change of the expression of the PLP-dependent enzymes. The result is shown in figure 33B. Similar to the cold shock condition, all PLP-dependent enzyme levels decrease initially after the shock condition was applied, but the expression amount returned to normal and remained approximately constant.

Under acid shock conditions, the expression level of the PLP enzymes fluctuates over time (figure 33C), which is possibly due to the unstable pH environment in the cell. Therefore, in all three shock conditions, the PLP-dependent enzymes don't seem to have any correlation with the cell regulation process.

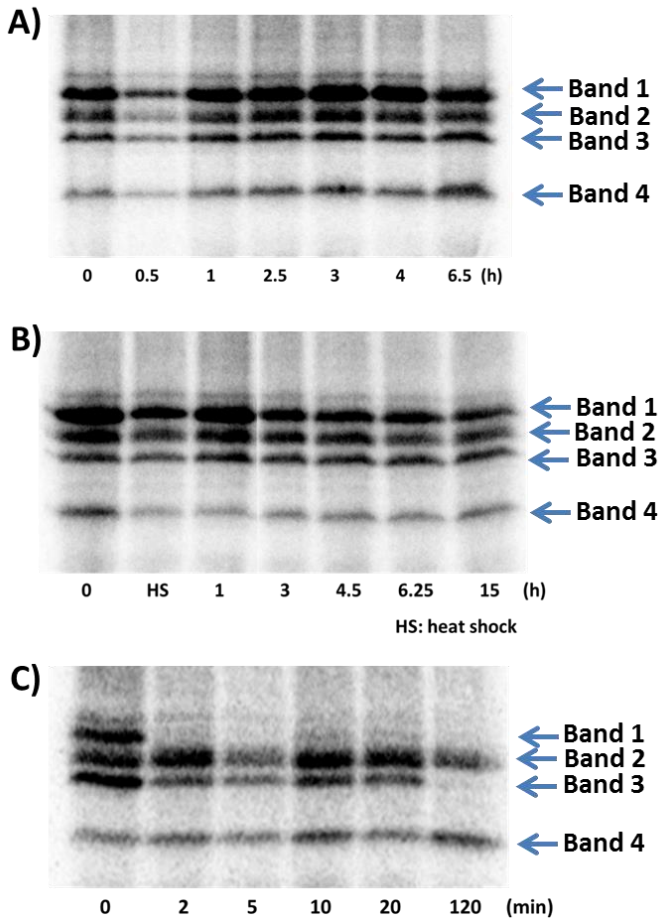


Figure 33. SDS-PAGE gels following the growth of *E. coli ΔpdxJ* in minimal media containing ^{32}P-PLP under shock conditions. A) Cold shock, where cells were incubated at 15 ^{o}C after 14 h of growth at 37 ^{o}C; B) heat shock, where cells were incubated at 42 ^{o}C for 2 min after 14 h of growth at 37 ^{o}C, and then back at 37 ^{o}C; C) acid shock, where 20 µl 34% HCl was added to 10 ml culture to drop pH down to around 4.5.

Finally, we carried out another two stress experiments: osmotic stress and UV irradiation. To create osmotic stress, sucrose was added to the media to a final concentration of 40%. The sudden increase of sucrose in the media around the *E. coli* cells causes a rapid change in the movement of water across the cell membrane. Under this condition, water is drawn out of the cells through osmosis, which inhibits the

transport of substrates and cofactors into the cell thus "shocking" the cell. Hence, osmotic stress is similar to starvation to some extent. Interestingly, the result is also close to that in the glucose starvation experiment. With the addition of a high concentration of sucrose, all PLP-dependent enzymes show an obvious increase at 4-6 h (figure 34A). As discussed before, the purpose of this up regulation may be to manage the use of limited nutrients including substrates and cofactors.

In the UV irradiation experiment, *E. coli* cells were irradiated with UV light for 1 minute. UV irradiation at around 250 nm can kill or inactivate microorganisms by destroying nucleic acids and disrupting DNA, crippling the organism's ability to perform vital cellular functions. Figure 34B, however, shows that there is no dramatic expression change in PLP-dependent enzyme levels. Therefore, PLP enzymes do not seem to participate in stringent response to UV irradiation, which probably focuses on DNA repair.

A clustered column graph has been made as a summary of all the experimented stress conditions. Here, the four major bands (band 1-4 from top to bottom) were semi-quantitated by measuring the band intensity and then plotted. From the results shown in figure 35, it is clear that different PLP enzymes respond differently under various stress conditions. For instance, the protein(s) in band 3 (red column) express at a significantly higher amount under the PLP starvation condition, whereas the protein(s) in band 1 (purple column) increase dramatically in response to cold shock. However, it should be noted that intensities among different conditions are not comparable since they were measured from different gels.

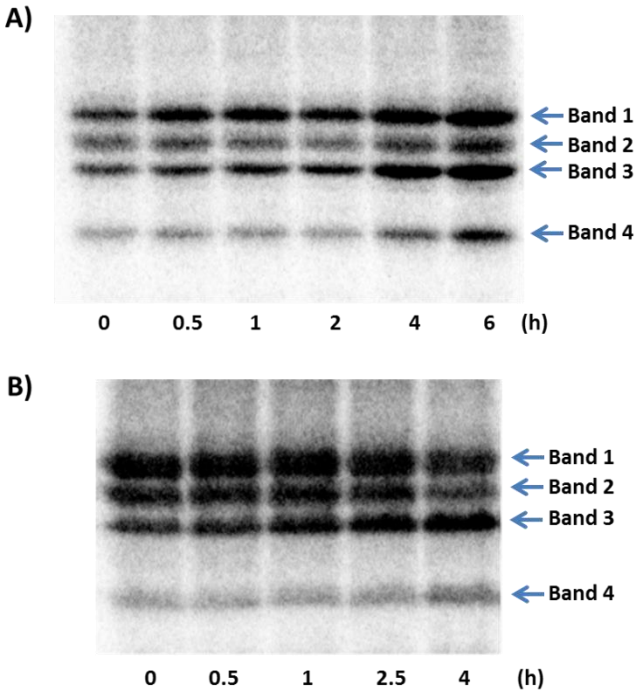


Figure 34. SDS-PAGE gels following growth of *E. coli ΔpdxJ* in minimal media containing ^{32}P-PLP under other stress conditions. A) Osmotic stress, where cells were incubated in 40% sucrose after 14 h of growth; B) UV irradiation, where cells were exposed to UV light (254 nm, 1 J/m2/sec) for 1 min after 14 h of growth.

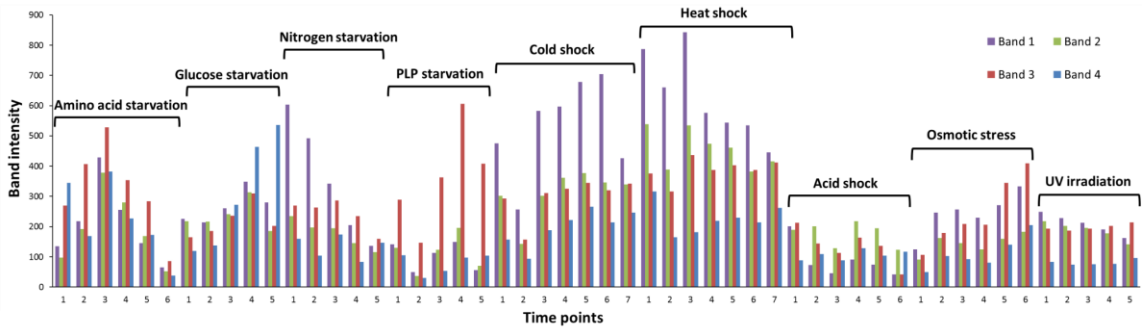


Figure 35. Summary of PLP enzymes under all the experimented conditions. The intensities of the four bands (band 1-4 from top to bottom) in each radioactive gel were measured and then plotted. Purple column: band 1; green column: band 2; red column: band 3; blue column: band 4.

Identification of PLP Enzymes in the *E. coli* Proteome by Radioactive Detection

Since PLP-dependent enzymes were shown to respond differently based on the specific stress condition, identifying them will allow us to explore new biological functions of PLP enzymes in cell regulation. However, considering that only non-radioactive samples are allowed on LC-MS, the radioactive gel cannot be analyzed by LC-MS for identification. Also, the protein ladder cannot be visualized in radiation imaging because it's not ^{32}P-labeled. To solve these problems, the radioactive gel was imaged by both radiation screening and normal gel scanning. The two images were then overlapped to determine the position of each radioactive band in comparison to the protein ladder. Meanwhile, samples prepared from four stress conditions were run on a non-radioactive gel and stained with Coomassie brilliant blue. Positions corresponding to the four radioactive bands were excised from the non-radioactive gel and we followed the procedure for in-gel tryptic digestion. The resulting non-radioactive peptides were then analyzed by LC-MS/MS and identified.

From previous results, the four bands are most clearly seen in the cold shock condition, so it was selected as the condition to determine molecular weights of the four bands. *E. coli ΔpdxJ* was grown in 10 ml sterile M9 minimal media containing glucose, Mg^{2+}, Ca^{2+}, Fe^{2+}, kanamycin, and ^{32}P-PLP. After 14 h of growth at 37 ^{o}C, cells were cold shocked at 15 ^{o}C and incubated for 3 h to maximize the expression of PLP-dependent enzymes. The 10 ml culture was harvested, lysed, and treated with $NaBH_4$ as previously described. After desalting, the radioactive protein mixture was separated with

a larger gel (20 cm x 20 cm) and visualized through a phosphor storage screen (figure 36).

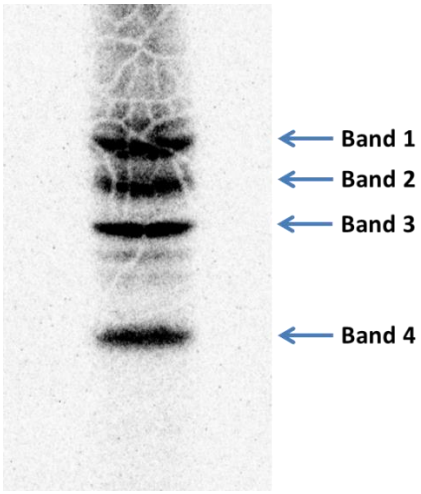


Figure 36. Larger SDS-PAGE gel of the *E. coli ΔpdxJ* proteome. *E. coli ΔpdxJ* was grown in 10 ml minimal media containing ^{32}P-PLP at 37 ^{o}C for 14 h and then cold shocked at 15 ^{o}C for 3 h to maximize the expression of PLP-dependent enzymes. The obtained proteome was reduced by $NaBH_4$ and loaded on a 12% SDS gel (20 cm x 20 cm).

Similar to previous results, the four major bands were clearly visible in the larger radioactive gel (figure 36, band 1-4, pointed by blue arrow). To determine the molecular weight range of each band, the gel was also scanned in the Biorad gel imager to see the protein ladder. The two images were overlapped (figure 37A) and the approximate region of each band was determined (red dash line): band 1 at 37-50 kDa, band 2 at ~37 kDa, band 3 at 25-37 kDa, band 4 at 20-25 kDa.

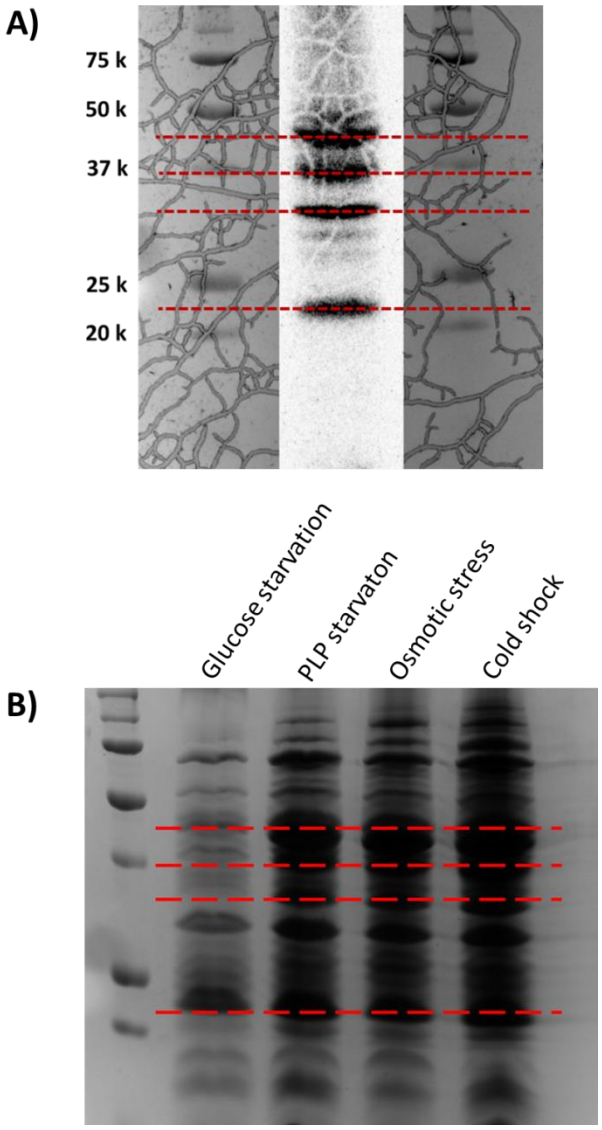


Figure 37. Larger SDS-PAGE gel for identification of PLP-dependent enzymes. The approximate regions of the four bands (figure 36, band 1-4, pointed by blue arrow) were marked with red dash lines. A) Overlap of radiation screening and normal gel scan of the ^{32}P-labeled radioactive gel of *E. coli ΔpdxJ* growing under cold shock conditions. B) Non-radioactive gel stained with Coomassie brilliant blue for identification of PLP-dependent enzymes. The five lanes from left to right correspond to *E. coli* cells growing under glucose starvation, PLP starvation, osmotic stress, and cold shock conditions.

For identification, *E. coli ΔpdxJ* was grown under four conditions: glucose starvation, PLP starvation, osmotic stress, and cold shock. Following the same procedure

but without using ^{32}P-PLP, the proteome was separated by SDS-PAGE and stained by Coomassie brilliant blue. With the determined approximate molecular weight region in figure 37A, the regions of the red dash lines in figure 37B were excised, digested with trypsin, and subjected to LC-MS/MS analysis. The resulting peptide data was searched in the NCBI database by Mascot, and five PLP-dependent enzymes were identified as threonine synthase (47.1 kDa), serine hydroxymethyltransferase (45.3 kDa), aspartate aminotransferase (43.6 kDa), 3-phosphoserine aminotransferase (39.8 kDa), and cysteine synthase A (34.5 kDa) (table 4).

Compared with the molecular weight of each protein, it is deduced that band 1 is a mixture of threonine synthase, serine hydroxymethyltransferase, and aspartate aminotransferase. Band 2 is 3-phosphoserine aminotransferase, and band 3 is cysteine synthase A. As for band 4, both threonine synthase and cysteine synthase A were detected in that region (~22 kDa) but not in the upper region (25-30 kDa). So it is highly possible that band 4 in the radioactive gel is a PLP-containing peptide fragment of threonine synthase and cysteine synthase A.

Also, the theoretical molecular weight of each protein does not perfectly match the band positions in the radioactive gel, which might be due to two reasons: 1) the overlapping method for determination of the band position is not very accurate because each lane migrates slightly differently; even the two lanes of protein ladders showed a position difference (figure 37A). 2) Due to PLP bonding or the interaction with other proteins in the cell environment and the possible adduct of small molecules, the

migration of each protein in the SDS gel may not correspond well with its molecular weight.

Table 4. Identified *Escherichia coli* PLP-dependent enzymes

GI number	Protein name	Sequence coverage
1579964	Threonine synthase	5%
485772118	Serine hydroxymethyltransferase	18%
16128874	Phosphoserine aminotransferase	27%
15802947	Cysteine synthase A	24%
78214801	Aspartate aminotransferase	7%

After identifying the enzymes, we can look back at the previous radioactive gel results. Here, all the experiments were carried out under stress conditions, which are regulated by the bacterial stringent response. In this process, the ribosome-associated gene *relA* synthesizes a signal molecule (p)ppGpp (guanosine pentaphosphate or tetraphosphate), which binds to RNA polymerase and ceases the transcription of stable RNAs. Stable RNAs represent only ~1% of the genome, but engage >60% of the RNA polymerase. When stringent response starts, the RNA polymerase constrained by stable RNAs is used for transcription of appropriate response genes. Consequently, different stress conditions lead to up or down regulation of different genes. Under amino acid starvation, amino acid biosynthetic pathways are up regulated[110-111]. Hence, all the detected PLP-dependent enzymes, which are involved in amino acid biosynthesis, increased after the starvation started (figure 31).

For glucose starvation, the signal molecule (p)ppGpp is also accumulated, triggering the stringent response. Though the carbon metabolism is not directly associated with amino acid biosynthesis, the translational pausing caused by carbon starvation can be sensed by RelA due to its physical attachment to the ribosome. The amino acyl-tRNA pool is thus monitored and the amino acid biosynthetic pathway is up regulated to prevent amino acid pool fluctuations[112]. As a result, all the PLP-dependent enzymes increased expression after glucose starvation started (figure 32A).

Nitrogen starvation triggers two regulation systems in *E. coli*: the nitrogen stress response and the stringent response. For the former, ~100 genes are expressed to scavenge for alternative nitrogen source. The transcription regulator in nitrogen stress response can further activate the transcription of *relA*, generating (p)ppGpp and changing the global transcription profiling[113]. In our experiment, we observed that the PLP-dependent enzymes involved in amino acid biosynthesis decreased along with time (figure 32B). This is probably due to the down regulation of the glutamate transamination, which is the major nitrogen source after nitrogen assimilation. This result is quite reasonable because organisms have to decrease the production of amino acids and proteins due to the inability to find alternative nitrogen source.

In the PLP starvation experiment (figure 32C), though not reported, we hypothesize that this nutrient starvation can also trigger the accumulation of (p)ppGpp, leading to bacterial stringent response. Similar as amino acid starvation, the biosynthetic pathway for the starved nutrient (here it is PLP) is supposed to be up regulated. Hence, the dramatically increased band is probably 3-phosphoserine aminotransferase, known as

both serC and pdxF, a bi-functional PLP-dependent enzyme that can convert (*3R*)-3-hydroxy-2-oxo-4-phosphonooxybutanoate and glutamate to form 4-hydroxy-L-threonine phosphate, an intermediate in the PLP biosynthesis.

Osmotic stress, a condition that is similar to nutrient starvation, also exhibits an increased trend for PLP-dependent enzymes. But for other stress conditions such as cold/heat/acid shock and UV irradiation, the PLP-dependent enzymes follow a global change in the transcription. For instance, under cold shock, only the cold shock proteins (proteins containing universally conserved "cold shock domain") are up regulated, whereas most other proteins appear to be repressed[114]. Hence, from these results, PLP-dependent enzymes participate more in the regulation of nutrient starvation rather than shock conditions.

4.4 Conclusion

In this chapter, we have developed a radioactivity based method for specifically detecting PLP-dependent enzymes by enzymatically synthesizing ^{32}P-PLP and incorporating it into the bacterial proteome. The ^{32}P-labeled PLP-dependent enzymes can thus be visualized by radiation screening. Using this method, we examined the effects of eight stress conditions – including three starvation conditions (glucose starvation, nitrogen starvation, and PLP starvation), three shock conditions (cold shock, heat shock, and acid shock), and two other stress conditions (osmotic stress and UV irradiation) – on PLP-dependent enzyme expression in *E. coli*. PLP-dependent enzymes

showed significantly different changes in expression, especially in the starvation conditions.

To identify PLP-dependent enzymes, non-radioactive SDS-PAGE was performed under identical conditions and the resulting gel was overlapped with the radioactive gel to determine the band positions. With in-gel tryptic digestion and LC-MS/MS analysis, five PLP-dependent proteins were identified including threonine synthase, serine hydroxymethyltransferase, 3-phosphoserine aminotransferase, cysteine synthase A, and aspartate aminotransferase.

The radioactivity base method provides a good way for specifically visualizing PLP-dependent enzymes as a group in the proteome. Apart from our study of stress conditions in *E. coli*, this method can also be used to explore PLP enzymes under other growing conditions in a variety of organisms.

5. DEVELOPMENT OF A NON-RADIOACTIVITY BASED PROTEOMIC METHOD TO PROBE AND IDENTIFY PLP-DEPENDENT ENZYMES

5.1 Introduction

Though PLP-dependent enzymes can be identified by doing non-radioactive experiments under the same conditions as the radioactive method, the procedure is complicated and time-consuming. Furthermore, the band position determined by overlapping the two gel images does not guarantee accuracy. Therefore, the development of a non-radioactive method to probe and identify PLP-dependent enzymes is of great benefit.

As previously discussed in activity based protein profiling, a chemical probe can be designed to target PLP and the active site of PLP-dependent enzymes. PLP itself, however, does not have good biorthogonal reactivity. Designing PLP analogs with reactive groups will allow the PLP analog bound enzymes to be targeted by a probe molecule.

After analyzing the structure of the PLP molecule, we decided to design halogenated analogs of the PLP molecule at its 6-position on the pyrimidine ring. The reason for the 6-position selection is a combination of stereo hindrance tolerance and the ease of organic synthesis. The halogenated PLP analogs are reactive with functional groups such as azide and thiol. After careful consideration and trial experiments (data not shown), we decided to explore the following scheme as our method development (figure 38). First, the imine bond between the PLP analogs and the bound protein is

reduced by $NaBH_4$ to prevent hydrolysis. The halogen group on the PLP analog is then converted to $–N_3$ by an excess amount of NaN_3. With the well-developed copper (I)-catalyzed alkyne-azide cycloaddition (CuAAC) (a "click reaction"), the azido-PLP can be covalently linked to a molecular probe with alkyne at one terminus and biotin at the other (figure 39). Using this approach, the PLP-dependent enzymes can be visualized or affinity-enriched by biotin as discussed in chapter 2 and 3.

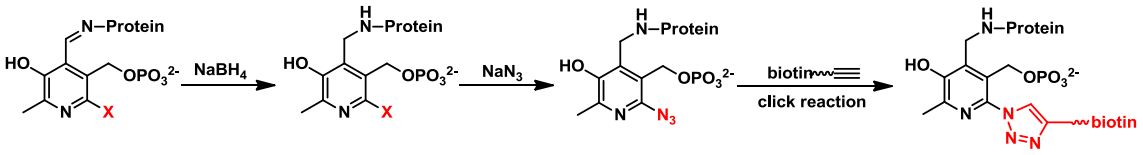


Figure 38. Strategy for probing PLP-dependent enzymes. The protein-bound halogenated PLP analog can be reduced by $NaBH_4$ and then converted to azido-PLP by an excess amount of sodium azide, followed by a copper (I) catalyzed click reaction using a biotin-alkyne reagent. The PLP-dependent enzymes are thus labeled with biotin and can be visualized or affinity-enriched for identification.

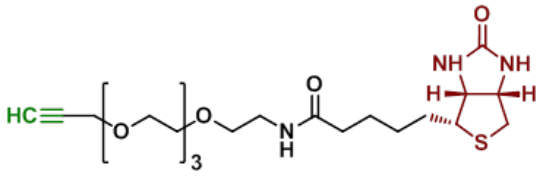


Figure 39. The biotin-$(PEG)_4$-alkyne reagent. The reactive alkyne functional group is marked in green and the reporter group biotin is labeled in red.

The labeled PLP-dependent enzymes can then be identified by a combination of affinity chromatography and mass spectrometry. The PLP enzymes with biotin tag are enriched on the streptavidin resin. After on resin digestion, the resulting peptides are

subjected to LC-MS/MS analysis and the tandem mass spectra are searched in NCBI database to identify PLP-dependent enzymes.

We still used *E. coli* as a model system and both PLP analog based and native PLP based methods were developed. We also treated the *E. coli* culture under amino acid starvation to discover the response of the PLP-dependent enzymes. PLP enzymes were identified under both amino acid supplemented and starved conditions.

5.2 Experimental Methods

Chemicals and Reagents

Pyridoxal hydrochloride, pyridoxal-5-phosphate, sodium azide, and amino acids were purchased from Sigma Aldrich (St. Louis, MO). B-Per solution was obtained from Pierce (Rockford, IL). Streptavidin resin was from GenScript (Piscataway, NJ) and streptavidin-R-phycoerythrin (PE) was from QIAGEN (Valencia, CA). The PVDF membrane and trypsin gold were purchased from Promega (Madison, WI). M9 salts were obtained from Becton Dickinson (Franklin Lakes, NJ). *E. coli ΔpdxJ* and the *E. coli -relA* strain was from Coli Genetic Stock Center at Yale University (New Haven, CT).

All fluorescence gel images were scanned using a Typhoon trio instrument (excitation, 532 nm green laser; emission, 580 nm band-pass filter (580 BP 30)) from GE Healthcare Biosciences (Piscataway, NJ). The UV absorbance of protein samples was tested on a Varian Cary 300 Bio UV-Visible Spectrophotometer (Palo Alto, CA). A sonicator 3000 from Misonix Inc. (Farmingdale, NY) was used for sonication. Sodium

dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was carried out using a Hoefer SE 250 mini-vertical gel electrophoresis unit.

Organic Synthesis of Three PLP Analogs

This part of the work was done by Dr. Dmytro Fedoseyenko and Brateen Shome in Tadhg P. Begley's laboratory.

Growth Study of *E. coli ΔpdxJ* Supplemented with Different PLP Analogs

E. coli ΔpdxJ was grown on an LB agar plate at 37 ^{o}C for 12 hours. A colony was added to 20 ml sterile M9 minimal media containing 0.4% glucose, 50 μM $Fe(NH_4)_2(SO_4)_2$, 2 mM $MgSO_4$, 100 μM $CaCl_2$, 40 μg/ml kanamycin, and 10 μM PLP analogs (6-chloropyridoxal, 6-bromopyridoxal, and 6-iodopyridoxal) each filtered through a sterile syringe drive 0.23 μm filter. The culture was allowed to grow at 37 ^{o}C for 25 h and absorbance at 600 nm was measured at 0, 3.6, 6.5, 8.5, 12, and 25 hours.

Labeling and SDS-PAGE of PLP Enzymes in *E. coli* Proteome Using PLP Analogs

E. coli ΔpdxJ was grown in 10 ml minimal media with PLP analogs supplemented as previously described. After 14 h of growth, 1 ml culture was harvested. The cell pellet was resuspended in 100 μl B-Per solution and incubated at room temperature for 15 min to lyse the cell. The mixture was centrifuged to obtain cell extract, of which 10 μl solution was removed to measure the total protein concentration by Bradford assay. $NaBH_4$ was added in an amount of 0.5 mg/100 μl and incubated at

room temperature for 10 min. Excess $NaBH_4$ was removed by methanol/chloroform precipitation, and the resulting protein pellet was dissolved in 100 µl 100 mM PBS buffer pH 7.2 with 8 M urea and 500 mM NaN_3. The solution was kept in the dark and allowed to react for 30 min at room temperature. Methanol/chloroform precipitation was then performed to remove the salts.

The protein pellet was dissolved in 95 µl 8 M urea in 100 mM PBS buffer pH 7.2, and the click reaction was started by sequentially adding $CuSO_4$ (0.4 µl of 25 mM stock), $NiSO_4$ (1 µl of 100 mM stock), Tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA, 1 µl of 50 mM stock), biotin-$(PEG)_4$-alkyne (1 µl of 100 mM stock), and sodium ascorbate (2 µl of 250 mM stock). The mixture was incubated in the dark at room temperature for 1 h, and then desalted by methanol/chloroform precipitation.

The protein pellet was added to 15 µl SDS sample buffer, 5 µl 8 M urea in 100 mM PBS buffer pH 7.2, and 0.3 µl β-mecaptoethanol. The mixture was heated at 60 ^oC for 5 min and then centrifuge at 13000 rcf for 5 min. A 5 µl sample of the solution was loaded onto a 12% SDS gel and run at 100 V on a SE 250 Mighty Small Gel runner (Hoefer, Hollisto, MA).

The resulting gel was either Coomassie stained, or transferred to a PVDF membrane by the Biorad Trans-Blot Turbo System at 100 V for 30 min. The blot was washed with TBS buffer (2 mM Tris-HCl, 50 mM NaCl, pH 7.5), blocked with 3% BSA (dissolved in TBS buffer), washed with TBS buffer again and then stained with streptavidin-PE at 1:3000. The stained blot was stored in TBS buffer and imaged using a

Typhoon trio instrument (excitation, 546 nm green laser; emission, 580 nm band-pass filter (580 BP 30)) from GE Healthcare Biosciences (Piscataway, NJ).

Labeling and Visualization of PLP Enzymes in *E. coli* Proteome Using Native PLP

E. coli ΔpdxJ was grown on an LB agar plate at 37 ^{o}C for 12 hours. A colony was added to 10 ml sterile M9 minimal media containing 0.4% glucose, 50 μM $Fe(NH_4)_2(SO_4)_2$, 2 mM $MgSO_4$, 100 μM $CaCl_2$, 40 μg/ml kanamycin, and 10 μM PLP, each filtered through a sterile syringe drive with 0.23 μm filter. After 14 hours of growth, 1 ml culture was harvested. The cell pellet was resuspended in 100 μl B-Per solution and incubated at room temperature for 15 min. The mixture was centrifuged to obtain protein solution, and the total protein concentration was measured by Bradford assay with 10 μl solution. The solution was then desalted by methanol/chloroform precipitation, and the resulting protein pellet was dissolved in 100 μl 100 mM PBS buffer pH 7.2 with 8 M urea and 500 mM NaN_3. The rest of the procedure is the same as labeling and visualizing PLP-dependent enzymes with PLP analogs.

Identification of PLP Enzymes in *E. coli* Proteome by the Developed Gel Free Method

E. coli –relA was grown in 50 ml sterile M9 minimal media containing 0.4% glucose, 2 mM $MgSO_4$, 100 μM $CaCl_2$, 50 μM $Fe(NH_4)_2(SO_4)_2$, and 40 μg/ml kanamycin. After 14 hours of growth at 37 ^{o}C, the cells were divided into four aliquots, harvested, and each aliquot was resuspended in 1 ml B-Per solution. After a 30 min

incubation, the protein mixture was obtained by centrifugation and then desalted using methanol/chloroform precipitation. The resulting protein pellet was dissolved in 1 ml 100 mM PBS buffer pH 7.2 with 8 M urea and 500 mM NaN_3. The solution was kept in the dark and allowed to react for 30 min at room temperature. Methanol/chloroform precipitation was then performed to remove the salts.

The protein pellet was then dissolved in 1 ml 8 M urea in 100 mM PBS buffer pH 7.2, and the click reaction was started by sequentially adding $CuSO_4$ (4 µl of 25 mM stock), $NiSO_4$ (1 µl of 100 mM stock), Tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA, 1 µl of 50 mM stock), biotin-$(PEG)_4$-alkyne (1 µl of 100 mM stock), and sodium ascorbate (2 µl of 250 mM stock). The mixture was incubated in the dark at room temperature for 1 h, and then desalted by methanol/chloroform precipitation.

The protein pellet was dissolved in 1 ml 8 M urea in 100 mM PBS buffer pH 7.2 and then diluted with 9 ml 100 mM PBS buffer pH 7.2 to decrease urea concentration to 0.8 M with a 10 ml final volume. The four aliquots were combined and cycled through 500 µl of streptavidin resin slurry (pre-equilibrated with 100 mM PBS buffer pH 7.2) at room temperature for 6 hours. The resin was then sequentially washed with 20 ml 100 mM PBS buffer pH 7.2 and 10 ml 50 mM NH_4HCO_3. Protein digestion was carried out by adding 10 µl 10 µg/µl trypsin gold and incubated at 37 ^oC overnight. The slurry was centrifuged and the supernatant was collected. The resin was then mixed with 200 µl 50 mM NH_4HCO_3 and the supernatant collected. This process was repeated 3 times and the supernatant was concentrated using a SpeedVac to 40 µl. The peptide solution was

subjected to LC-MS analysis as previously described and the data was analyzed by Bruker DataAnalysis 4.0 and BioTools 3.2.

Growth of *E. coli –relA* under Singlet Oxygen Stress

E. coli –relA was grown in 5 ml minimal media containing 0.4% glucose, 2 mM $MgSO_4$, 100 µM $CaCl_2$, 50 µM $Fe(NH_4)_2(SO_4)_2$, and 40 µg/ml kanamycin. After 14 h of growth, the 5 ml culture was divided into 5 aliquots with 1 ml per aliquot. The cercosporin was then supplemented to each culture at a final concentration of 0, 0.1, 0.3, 1, 5 µM. After incubating under light for another 3-4 h, cells were harvested and followed the established method for visualizing PLP-dependent enzymes.

Growth of *E. coli –relA* under Amino Acid Supplemented and Starved Condition

E. coli –relA was grown in 1 ml minimal media containing 0.4% glucose, 2 mM $MgSO_4$, 100 µM $CaCl_2$, 50 µM $Fe(NH_4)_2(SO_4)_2$, and 40 µg/ml kanamycin. The same media was used for the amino acid supplemented condition, except for addition of 0.1% casamino acids. After 14 h of growth, cells were harvested and followed the established method for visualizing PLP-dependent enzymes.

5.3 Result and Discussion of the Non-Radioactive Probing Method

Growth Study of *E. coli ΔpdxJ* Supplemented with Different PLP Analogs

To develop this PLP analog based method, we first had to confirm that *E. coli* cells could take up the PLP analogs. We tested different forms of PLP analogs including

6-bromo-pyridoxal-5'-phosphate (Br-PLP), 6-bromo-pyridoxal (Br-PL), and 6-bromo-pyridoxine (Br-pyridoxine) (figure 40) (all the molecules were synthesized by Dr. Dmytro Fedoseyenko and Brateen Shome). Of the three analogs, Br-PLP is the most difficult to synthesize but is the one most readily taken up by *E. coli* since it doesn't require any enzymatic modification. For Br-PL, it is easier to synthesize the molecule organically but it needs a kinase for phosphorylation in the cell. Br-pyridoxine is the easiest PLP analog to synthesize, but it requires two more steps in the cell: dehydrogenation and phosphorylation.

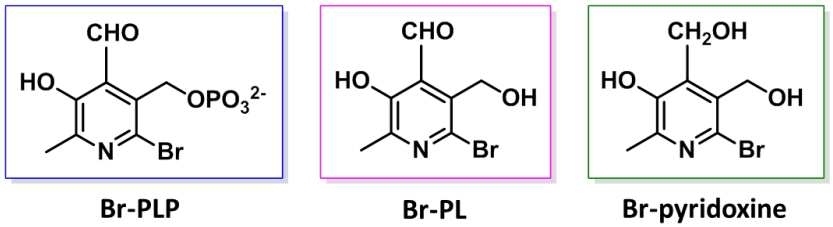


Figure 40. Structures of the three PLP brominated analogs: 6-bromo-pyridoxal-5'-phosphate (Br-PLP), 6-bromo-pyridoxal (Br-PL), and 6-bromo-pyridoxine (Br-pyridoxine).

To find out which molecule can support *E. coli* growth, we tested all three analogs by growth study. We found that both Br-PLP and Br-PL can support *E. coli*'s growth, but that Br-pyridoxine cannot (figure 41). This is probably because the kinase that phosphorylates pyridoxal tolerates the brominated substrate Br-PL but the dehydrogenase which turns pyridoxine into pyridoxal is more substrate specific and

doesn't recognize the brominated analog. Hence, the halogenated pyridoxal (X-PL, X=Cl, Br, I) was selected as the PLP analog for supplementing *E. coli ΔpdxJ*.

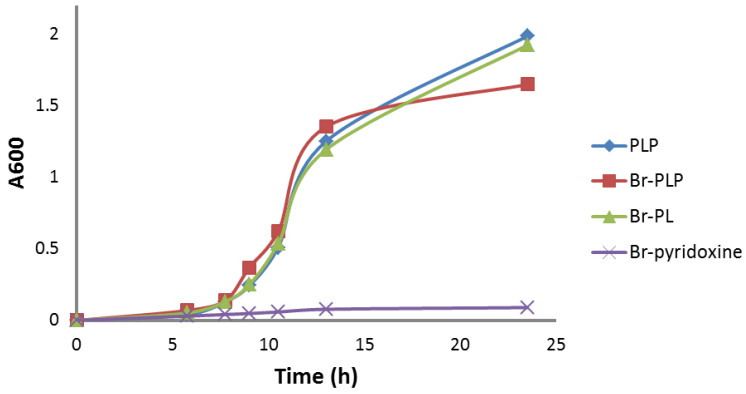


Figure 41. Growth curves of *E. coli ΔpdxJ* supplemented with PLP analogs. *E. coli ΔpdxJ* was grown in minimal media supplemented with 10 μM PLP (blue trace), 6-bromo-PLP (red trace), 6-bromo-pyridoxal (green trace), 6-bromo-pyridoxine (purple trace).

As a result, three PLP analogs with different halogen atoms were synthesized (by Dr. Dmytro Fedoseyenko and Brateen Shome): 6-chloro-pyridoxal, 6-bromo-pyridoxal, and 6-iodo-pyridoxal (short formed as Cl-PL, Br-PL, I-PL for convenience) (figure 42A). The three analog molecules were dissolved in water and fed to the *E. coli* PLP auxotroph strain *ΔpdxJ* at a concentration of 10 μM. The growth of *E. coli ΔpdxJ* in minimal media was recorded by measuring the absorbance at 600 nm at different time points from 0 h to 25 hours. Meanwhile, the growth of *E. coli ΔpdxJ* in minimal media supplemented with 10 μM native PLP was also tested as a positive control.

In the obtained growth curves (figure 42B), *E. coli* cells supplemented with the three PLP analogs all exhibit growth, similar to the control sample supplemented with native PLP. All three halogenated PLP analogs were found to be suitable to develop a method to probe PLP-dependent enzymes.

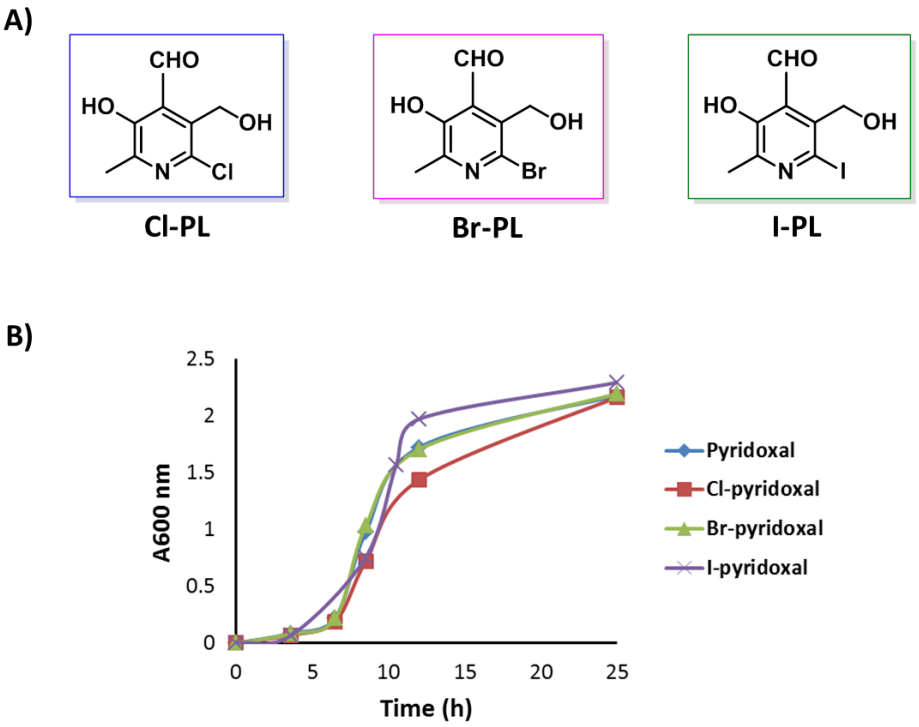


Figure 42. Structures of the three PLP halogen analogs and the growth curves of *E. coli ΔpdxJ* supplemented with PLP analogs or native PLP. A) The 6-position halogenated PLP analogs, including 6-cloro-pyridoxal (Cl-PL), 6-bromo-pyridoxal (Br-PL), 6-iodo-pyridoxal (I-PL). B) Growth curves of *E. coli ΔpdxJ* supplemented with Cl-PL, Br-PL, I-PL, and native PLP, respectively.

Labeling and SDS-PAGE of PLP Enzymes in the *E. coli* Proteome Using PLP Analogs

Since all three PLP analogs were found to support the growth of *E. coli ΔpdxJ*, the next step was to find out which one provided the best labeling efficiency. An organic model experiment was carried out before applying it to the PLP-dependent enzymes. Here, 1 mM chloro-pyridoxal, bromo-pyridoxal, and iodo-pyridoxal were mixed with 2 M NaN_3 at pH 7, respectively (all reactions set up by Brateen Shome). After LC-MS analysis, it was found that the three PLP analogs exhibit significantly different reaction efficiencies with NaN_3. While I-PL generated the highest amount of product, Br-PL yielded about half compared with I-PL, and Cl-PL produced almost no product (figure 43).

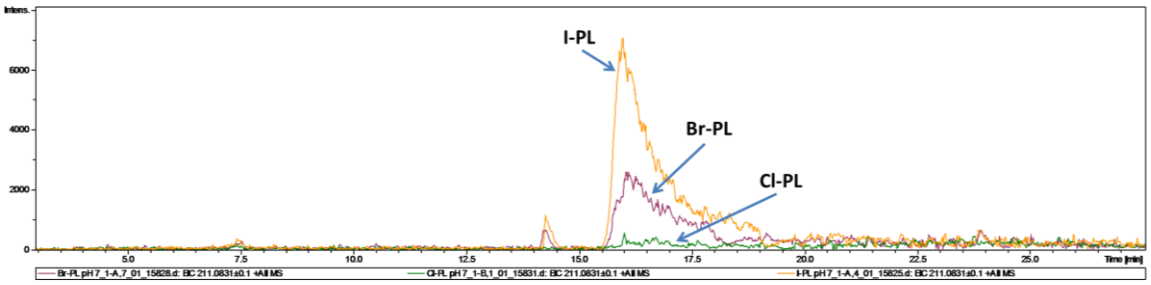


Figure 43. Extracted ion chromatogram of the reaction product between PLP analogs and sodium azide. Yellow trace: product of iodo-pyridoxal reacting with NaN_3; purple trace: product of bromo-pyridoxal reacting with NaN_3; green trace: product of chloro-pyridoxal reacting with NaN_3.

The experiment was repeated in the protein system. *E. coli ΔpdxJ* was fed each of the three PLP analogs and the resulting cells were harvested, lysed, and treated with $NaBH_4$. The reduced proteome was mixed with NaN_3 to convert the halogen group into

an azido group. After desalting, Cu^{2+}, Ni^{2+}, TBTA, and biotin-$(PEG)_4$-alkyne were added to the protein mixture. Sodium ascorbate was used to start the copper (I) catalyzed click reaction.

This reaction labeled the PLP analog-bound enzymes with biotin which was detected by western blot (anti-biotin). The result is shown in figure 44. As in the organic model reaction, iodo-pyridoxal supplemented *E. coli* culture gives the most intense labeling, followed by bromo-pyridoxal. Cells growing in choro-pyridoxal supplemented media show very weak labeling. This result is reasonable because iodo is the best leaving group among the three halogens. Therefore, iodo-pyridoxal was chosen to be the PLP analog molecule to probe PLP-dependent enzymes.

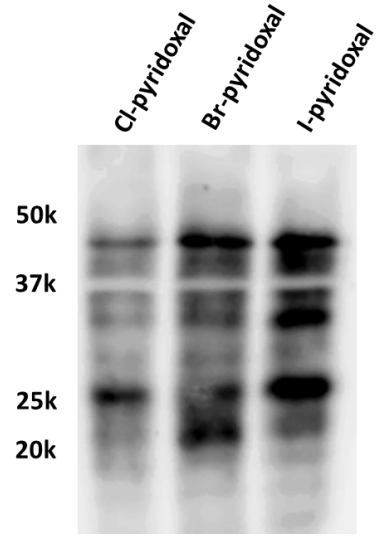


Figure 44. Western blot of *E. coli ΔpdxJ* proteome supplemented with PLP analogs. Chloro-pyridoxal, bromo-pyridoxal, and iodo-pyridoxal were added to the growth media of *E. coli ΔpdxJ*, respectively. The proteome was treated with $NaBH_4$, NaN_3, and copper (I) catalyzed click reaction. The resulting protein mixture was separated by a 12% SDS-PAGE, transferred to a PVDF membrane and stained by streptavidin-PE.

Labeling and Visualization of PLP Enzymes in the *E. coli* Proteome Using Native PLP

During the experiment, an interesting phenomenon was found: though iodo-pyridoxal supplementation shows more intense labeling than using native PLP, the latter seems to have the same labeling pattern as PLP analogs (figure 45A). If PLP bound enzymes can be directly labeled by NaN_3, the whole method is simplified. First, the synthesis of the PLP analogs can be omitted; second, it is not necessary to use the PLP auxotroph strain *ΔpdxJ*, which means this method can be further applied to other organisms without a readily available knockout.

However, the labeling mechanism of PLP reacting with NaN_3 might be different than PLP analogs since there is no good leaving group in the PLP molecule. In all previous experiments, PLP analogs were first reduced by $NaBH_4$ to be "fixed" on the proteins, in which the imine bond C=N was converted to C-N to prevent the PLP analogs from being hydrolyzed off the enzymes. In the case of native PLP, however, this imine bond may provide a better target for NaN_3 to label. To test our hypothesis, *E. coli ΔpdxJ* was grown in minimal media supplemented with native PLP. After 14 hours of growth at 37 ^{o}C, two aliquots of cells were harvested and the cell extract was obtained. One of the aliquots was treated with $NaBH_4$ as in previous experiments, and to the other was added the same amount of NaCl. The resulting proteome was again labeled by NaN_3 followed by a click reaction. Here, it is clearly seen that the sample without $NaBH_4$ treatment

shows much heavier labeling (figure 45B), indicating the imine bond does participate in the labeling process.

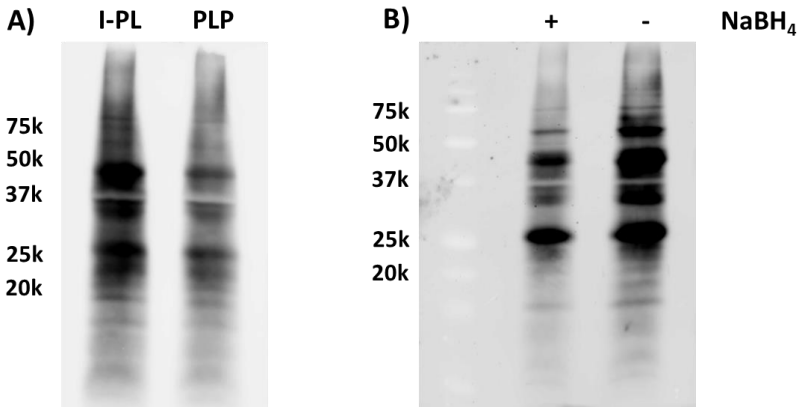


Figure 45. Western blot of native PLP or PLP analogs supplemented *E. coli ΔpdxJ* proteome with or without $NaBH_4$. A) Iodo-pyridoxal and pyridoxal-5'-phosphate (PLP) were added to the growth media, respectively; B) PLP supplemented in the growth media treated with or without $NaBH_4$. The proteome was labeled with NaN_3 followed by a copper (I) catalyzed click reaction. The resulting protein mixture was separated by 12% SDS-PAGE, transferred to a PVDF membrane, and stained with streptavidin-PE.

This result allows us to propose a mechanism for NaN_3 labeling of PLP-bound enzymes (figure 46). In it, the 6-position of the pyrimidine ring was attacked by $-N_3^-$. After a multi-step tautomerization, the imine bond C=N was finally converted to C-N and the azido group was covalently attached to the 6-position. This mechanism explains the result in the $NaBH_4$ effect experiment (figure 46B), but it needs to be confirmed by detecting the labeling product.

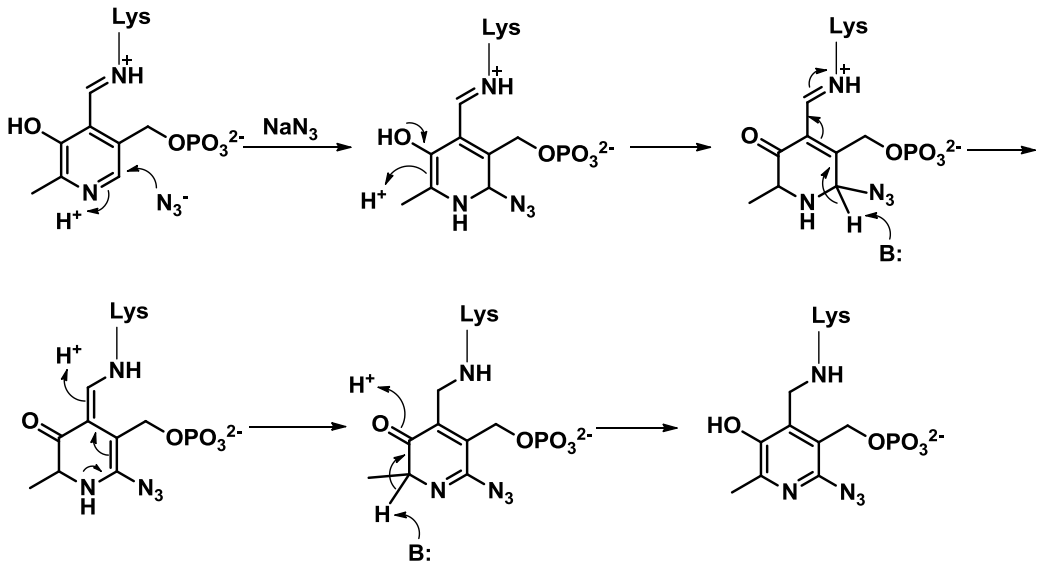


Figure 46. Proposed mechanism of NaN_3 labeling of PLP bound enzymes

Investigation of PLP-dependent Enzymes in the *E. coli* Proteome under Singlet Oxygen Stress

With the developed technique, we started to test the expression of PLP-dependent enzymes by applying stress conditions. Considering that the PLP-dependent enzymes were regulated by stringent response in the previous experiments, a relaxed response strain –*relA* was used to test the expression change of PLP-dependent enzymes.

In recent years, it has been reported that PLP has an anti-oxidation function against singlet oxygen[115-117]. However, most of the research focused on the PLP molecule instead of the PLP bound proteins. Here, with the ability to visualize PLP-dependent enzymes as a whole group, we applied singlet oxygen stress conditions to *E. coli* cells to detect the expression change of PLP-dependent enzymes.

Cercosporin was selected as the singlet oxygen generator as previously reported[118]. *E. coli -relA* was first grown at 37 ^{o}C for 14 h and then aliquoted into five cultures. Cercosporin was added to each sample in increasing concentrations (0, 0.1, 0.3, 1, and 5 μM) and incubated under light for another 3 hours. The resulting cultures were collected and analyzed as previously described. The expression of all PLP-dependent enzymes (four major bands pointed by blue arrows) was increased at higher concentrations of cercosporin (figure 47). This result indicates that PLP-dependent enzymes participate in the anti-oxidation process: they may directly regulate the cell as anti-oxidants or they assist the anti-oxidative PLP molecules by storing and transporting them. This process may be controlled by both oxidative stress response and the stringent response. Since those PLP enzymes have already been identified, they can be individually cloned and expressed to investigate their anti-oxidation function in future work.

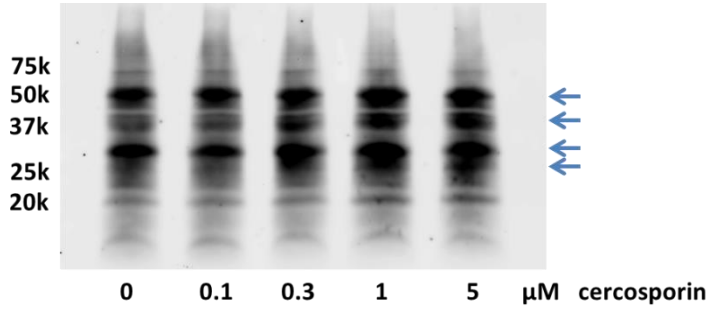


Figure 47. Western blot of *E. coli -relA* proteome with cercosporin supplemented. After 14 h of growth at 37 ^{o}C, *E. coli -relA* growth media was added cercosporin with concentrations of 0 μM, 0.1 μM, 0.3 μM, 1 μM, and 5 μM. The cultures were incubated for 3 h after adding cercosporin and then obtained cell extract. The proteome was labeled with NaN_3 followed by a copper (I) catalyzed click reaction. The resulting protein mixture was separated by 12% SDS-PAGE, transferred to a PVDF membrane and stained with streptavidin-PE.

Investigation of PLP-dependent Enzymes in the *E. coli* Proteome under Amino Acid Starvation

Since PLP-dependent enzymes are mostly involved in the amino acid biosynthetic pathways, we decided to focus on investigating PLP-dependent enzymes under amino acid starvation using the developed technique. *E. coli -relA* was grown in two different media: one is minimal media with glucose, $CaCl_2$, $MgSO_4$, $Fe(NH_4)_2(SO_4)_2$, kanamycin, and casamino acids supplemented; the other is minimal media with the same components except for the absence of casamino acids. Following the developed protocol, the PLP-dependent enzymes were visualized by western blot (figure 48).

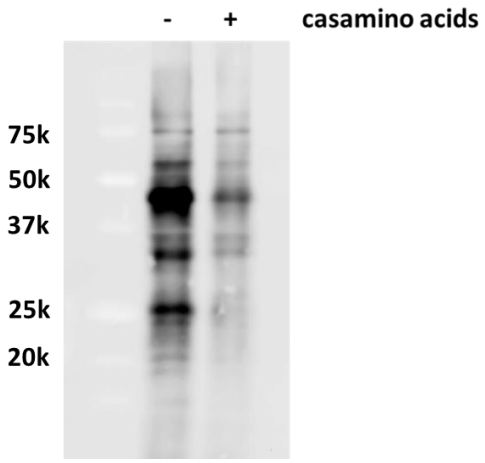


Figure 48. Western blot of *E. coli -relA* proteome with or without 0.1% casamino acids supplemented in the growth media. The proteome was labeled with NaN_3 followed by a copper (I) catalyzed click reaction. The resulting protein mixture was separated by 12% SDS-PAGE, transferred to a PVDF membrane and stained with streptavidin-PE.

It is obvious that the PLP-dependent enzyme level sharply decreases with the addition of casamino acids, especially the proteins in the lower molecular weight region. This is similar to the previous radioactive experiment and is reasonable. Though *relA* has been knocked out, *spoT* can synthesize a reduced amount of (p)ppGpp, triggering the relaxed stringent response. From our result, the -*relA* strain still has the ability to up regulate the amino acid biosynthesis. This is reasonable because the cells are expected to show no growth if they do not synthesize the amino acids that are unavailable from the environment. In the case of casamino acids supplementation, *E. coli* can obtain amino acids from the environment and thus the PLP-dependent enzymes are significantly depressed.

Identification of PLP Enzymes in the *E. coli -relA* Proteome by the Developed Gel Free Method

To verify that the developed gel free method could also identify PLP-dependent enzymes, a larger amount (50 ml) of *E. coli* culture was grown in both amino acid supplemented and amino acid starved conditions, following the same steps as the previously developed method except for the gel separation. After the click reaction, the protein mixture was desalted by methanol/chloroform precipitation and the protein pellet was dissolved in 8 M urea solution. After the solution became homogeneous, it was diluted into 2 M urea and loaded onto a streptavidin resin. Using a lower concentration of urea allows the protein mixture to remain in solution while the streptavidin is not denatured and still has high affinity towards biotin. To ensure sufficient binding, the

solution was cycled through a streptavidin resin for 6 h at room temperature. The protein bound resin was then washed and on-resin tryptic digestion was performed as discussed in chapters 2 and 3. Only one protein serine hydroxymethyltransferase (45.3 kDa) was identified in the amino acid supplemented condition, and five proteins were identified which were the same as those identified using the radioactive method in chapter 4: threonine synthase (47.1 kDa), serine hydroxymethyltransferase (45.3 kDa), aspartate aminotransferase (43.6 kDa), phosphoserine aminotransferase (39.8 kDa), and cysteine synthase A (34.5 kDa).

The identification result corresponds well with the gel image in figure 48. In amino acid rich media, *E. coli* cells obtain amino acids from the environment and thus repress the biosynthetic pathways. Serine hydroxymethyltransferase is the only identified PLP-dependent enzyme under this condition, indicating its essential function in the cell metabolism. This enzyme catalyzes the conversion between serine and glycine. Scission of the serine C_α-C_β bond provides the one-carbon unit, which is then used to synthesize purine, thymidylate, glycine, and other metabolites. Therefore, it is not unexpected to identify serine hydroxymethyltransferase under the amino acid rich condition.

For the amino acid starved condition, it is interesting to see that the same proteins are identified in both *relA*$^+$ (*E. coli ΔpdxJ* strain, previous experiments in chapter 4) and *relA*$^-$ strain. This is probably because the stringent response is regulated by both *relA* and *spoT*. The signal molecule, (p)ppGpp, is synthesized by RelA but can be hydrolyzed by spoT. Regulated by both proteins, (p)ppGpp is accumulated and maintained at a certain

level to trigger and regulate the stringent response. However, SpoT also has the ability to synthesize (p)ppGpp, though at a much lower level compared with RelA. Therefore, the *relA*$^-$ strain can still generate (p)ppGpp and thus is called a relax response strain, which also regulates the stringent response to some extent. Hence, the same enzymes identified in this research are deduced to play key roles in the amino acid biosynthesis.

Threonine synthase converts *O*-phospho-L-homoserine to L-threonine. The absolute concentration of threonine in the cytosol is 0.18 mM[119]. Threonine itself is incorporated into proteins and is also an intermediate in isoleucine biosynthesis. Also, the *thr* operon was found to be up regulated in both *relA*$^-$ and *relA*$^+$ strains[110]. Therefore, it is reasonable for threonine synthase to be expressed in a relatively high amount in the cell.

Serine hydroxymethyltransferase, as discussed above, is involved in the one-carbon metabolism, which is central for biosynthesizing a variety of metabolites. Under amino acid starvation, it has been reported that serine is heavily used as a carbon source[111]. The low concentration of serine in the cytosol (0.068 mM)[119] also indicates it is quickly converted after generation. And our experiment suggests that it is even expressed in the amino acid supplemented condition. Essential as it is, it is expected that this enzyme is identified.

Aspartate aminotransferase catalyzes the transamination reaction from glutamate to aspartate. After the nitrogen assimilation, glutamate is the nitrogen source for various metabolites. For this reason, glutamate has the highest concentration (96 mM)[119] among cell metabolites. Glutamate undergoes various transamination reactions to donate the

nitrogen to other metabolites, one of which is aspartate. Aspartate then becomes the precursor for many amino acids such as glycine, alanine, serine, and threonine. As expected, the concentration of aspartate is relatively high, which is 4.2 mM[119]. Consequently, the aspartate aminotransferase is supposed to be well expressed to produce aspartate. Also, aspartate aminotransferase has recently been found to involve in the cell replication. AspC mutant cells were smaller with fewer replication origins and had an increased doubling time[120]. Therefore, aspartate aminotransferase is another important enzyme coordinating both amino acid biosynthesis and cell growth.

For 3-phosphoserine aminotransferase, as discussed in chapter 4, it is known as both serC and pdxF. It is a bi-functional enzyme involved in both serine and PLP biosynthesis. Though serine doesn't represent a high concentration in the cell, it is one of the most intensively used amino acid. Computational investigation suggests that serine is low in energy cost and high in major codon usage[121-122]. Therefore, 3-phosphoserine aminotransferase in the serine biosynthetic pathway should be well expressed.

Cysteine synthase catalyzes the conversion of O-acetyl-L-serine and hydrogen sulfide to cysteine. After the free sulfur is assimilated into organic sulfur, cysteine acts as a central sulfur donor to a variety of sulfur containing metabolites. Here, we identified cysteine synthase with high sequence coverage (75%).

In this research, we have developed a non-radioactive technique that can use native PLP to work as an alternative for the radioactive method to probe and identify PLP-dependent enzymes. We identified five PLP enzymes that play important roles in amino acid biosynthesis or other cell functions.

Comparison of the Two Methods for Identification of PLP-Dependent Enzymes: Classical Gel Based Method and the Developed Gel Free Method

To compare the developed gel free method with the classical gel based technique, we used both methods to identify the PLP-dependent enzymes in *E. coli –relA* under amino acid starvation. More enzymes with higher sequence coverage were identified by the gel free method (table 5). This is because: 1) the efficiency to recover proteins from a gel is very low, so only a small percentage of proteins can be digested for detection. The gel free method doesn't have this problem. 2) Though the labeling efficiency in the gel free method is low, it is compensated in the affinity enrichment. By growing large culture of cells and cycling the biotin-labeled cell extract through the streptavidin resin, proteins can be enriched on the solid phase. Therefore, the enrichment helps to identify PLP-dependent enzymes with low concentrations.

It was observed that aspartate aminotransferase was identified in the previous gel based experiment (chapter 4, section 4.3 identification of PLP-dependent enzymes using radioactivity detection), but failed to be identified here. The reason is probably because different strain is used in the two experiments. The experiment in chapter 4 used *E. coli ΔpdxJ* and here the *E. coli –relA* is used. The $relA^-$ and $relA^+$ strains have some different regulations under amino acid starvation[110]. Hence, it's possible that aspartate aminotransferase was less expressed in the $relA^-$ than in the $relA^+$ strain. However, since the gel free method can enrich PLP-dependent enzymes on the resin, this enzyme was still identified using the developed gel free method.

It should be noticed that the biotin carboxyl carrier protein (BCCP) was not identified in both radioactivity based method and non-radioactive strategy. Failure for the BCCP identification may be due to two reasons: 1) the protein is expressed in low abundance under the condition we used; 2) in the non-radioactive method, copper cleaves the proteins at non-tryptic sites, leading to the miss-picking up by the database.

On the other side, the gel based method has no selectivity towards PLP-dependent enzymes. In the gel based experiment, we observed more than 80 proteins were co-identified with PLP-dependent enzymes. Therefore, the B6-dependent enzymes database was used to look for PLP-dependent enzymes in all the identified proteins. Nevertheless, only 4-8 non PLP proteins were observed in the gel free method, most of which are elongation factors. Detection of these non-PLP proteins is probably due to their high abundance in the proteome and thus retain non-specifically on the streptavidin resin. The good selectivity of the gel free method enables us to look for PLP-dependent enzymes in a proteome instead of cloning and purifying individual protein.

Therefore, if we want to focus on the behavior of PLP-dependent enzymes, the developed gel free technique possesses better sensitivity and selectivity in both visualization and identification.

Table 5. Comparison of the gel based and gel free methods

Proteins identified	MW	Sequence Coverage (Gel based)	Sequence Coverage (Gel free)
Threonine synthase	47 kDa	5%	16%
Serine hydroxy-methyltransferase	45 kDa	16%	18%
Aspartate aminotransferase	44 kDa	---	19%
3-Phosphoserine aminotransferase	40 kDa	3%	25%
Cysteine synthase A	34 kDa	24%	75%

5.4 Conclusion

In this chapter, we have developed non-radioactive methods to probe and identify PLP-dependent enzymes. A PLP analog based method was designed and tested by feeding an *E. coli* PLP auxotroph strain with PLP halogenated analogs, converting the analogs into azido-PLP, and finally labeling with biotin tag by a copper (I) catalyzed click reaction. Among 6-bromo pyridoxal-5'-phosphate, 6-bromo pyridoxal, and 6-bromo pyridoxine, the former two molecules were found to support the growth of *E. coli* whereas pyridoxine cannot. Also, 6-iodo-pyridoxal shows better labeling efficiency than both 6-bromo-pyridoxal and 6-chloro-pyridoxal in organic model reactions and *E. coli* proteome experiments. With 6-iodo-pyridoxal fed to *E. coli ΔpdxJ*, five PLP-dependent proteins were identified: threonine synthase (47.1 kDa), serine hydroxymethyltransferase (45.3 kDa), aspartate aminotransferase (43.6 kDa), phosphoserine aminotransferase (39.8 kDa), and cysteine synthase A (34.5 kDa), which are the same proteins identified in previous radioactive experiments.

It was found that native PLP fed *E. coli* proteome exhibited a weaker but similar labeling pattern as PLP analogs, and $NaBH_4$ helped to increase the labeling significantly. We proposed that the native PLP bound enzymes react directly with NaN_3, with the latter attacking at the 6-position.

With the developed non-radioactive technique, PLP-dependent enzymes were investigated under different stress conditions. In the singlet oxygen stress experiment, all PLP-dependent enzymes were found to increase at higher concentrations of cercosporin, the singlet oxygen generator. This indicates that the PLP-dependent enzymes participate in an anti-oxidation regulation, either directly or indirectly through activities such as storing and transporting PLP. Amino acid starvation experiment showed that PLP-dependent enzymes significantly increased expression compared with the amino acid supplemented conditions.

6. CONCLUSION AND PERSPECTIVE

6.1 Conclusion

In the first part of this research, we developed an activity based protein profiling method to specifically target and identify sulfur carrier proteins. The "active site" of the sulfur carrier proteins, the thiocarboxylate group, can be reconstituted *in vitro* and then reacted with the sulfonyl-azide group of the probe molecule through a thioacid-azide reaction. The other end of the probe molecule contained biotin, a tag that can either be used to visualize sulfur carrier proteins by western blot or for affinity enrichment by streptavidin resin, allowing the sulfur carrier proteins to be probed, enriched, and identified.

After validating the method with purified *Thermus thermophilus* thiS protein and the *E. coli* overexpressing ThiFS system, we moved on to explore *Streptomyces coelicolor*, in which an unknown sulfur carrier protein had been detected but remained unidentified. Using the developed method, the sulfur carrier protein was successfully identified as hypothetical protein SCO4294, possessing the characteristic glycine-glycine sequence at the C-terminus. A bioinformatics study found that sulfur carrier protein SCO4294 conservatively associates with enzyme SCO4293 annotated as threonine synthase. The reaction between sulfur carrier protein SCO4294 and "threonine synthase" SCO4293 was studied by western blot assay, HPLC, and LC-MS, with *O*-phospho-L-homoserine as the substrate. It was found that the sulfur carrier protein donates a sulfur atom to *O*-phospho-L-homoserine to generate homocysteine, and its C-terminal

thiocarboxylate group hydrolyzes into a carboxylate group. This is the first time that *O*-phospho-L-homoserine was found to be the substrate in the formation of homocysteine in bacteria. And this process probably represents a new direct sulfhydrylation pathway in methionine biosynthesis.

In another project, we developed a radioactivity based method to probe PLP-dependent enzymes. The ^{32}P-PLP was enzymatically synthesized and fed to the *E. coli* PLP auxotroph strain *ΔpdxJ* and the proteome was separated by SDS-PAGE. The expression levels of the ^{32}P-PLP labeled enzymes were examined under different stress conditions including amino acid starvation, glucose starvation, nitrogen starvation, PLP starvation, cold shock, heat shock, acid shock, osmotic stress, and UV irradiation. PLP-dependent enzymes showed different responses to the various stress conditions. Also, by overlapping the radioactive and non-radioactive gels, five PLP-dependent enzymes were identified using traditional in-gel tryptic digestion and LC-MS/MS database search.

Non-radioactive methods for probing PLP-dependent enzymes were also successfully developed. One way is to feed *E. coli* cells with 6-position halogenated PLP analogs. The PLP analog-bound enzymes can be reduced by $NaBH_4$ and then labeled with biotin-$(PEG)_4$-alkyne reagent after conversion with NaN_3 and a copper (I) catalyzed click reaction. Alternatively, enzymes bound with native PLP can also be labeled with a biotin tag without $NaBH_4$ reduction. The biotinylated PLP-dependent enzymes can be visualized by western blot or affinity enriched by streptavidin resin. Using this method, PLP-dependent enzymes under singlet oxygen stress were shown to increase, indicating that they participate in the anti-oxidation regulation process. Amino acid starvation

experiment found that PLP-dependent enzyme expression increased significantly when casamino acids were not supplemented. Furthermore, using the non-radioactive method, the five same PLP-dependent enzymes were identified in *E. coli* with higher sequence coverage.

6.2 Perspective

The detailed reaction mechanism between *S. coelicolor* sulfur carrier protein SCO4294 and "threonine synthase" SCO4293 remains to be unraveled. First, the proposed thioester formed by substrate *O*-phospho-L-homoserine and the sulfur carrier protein must be characterized. Second, the complex structure of the sulfur carrier protein and "threonine synthase" SCO4293 should be analyzed through crystallization, which will elucidate the catalytic mechanism of threonine synthase. Third, in our proposed mechanism, PLP didn't participate in the sulfur donation process, leading to the question of active site residues and the function of PLP.

Moreover, if the produced homocysteine is in the methionine biosynthetic pathway as proposed, then there is no reason for the bacteria to express the sulfur carrier protein in rich media since the cells can obtain abundant methionine directly from the environment. However, we have observed the expression of the sulfur carrier protein in both rich media and minimal media, indicating it has an important cellular function. Future isotope-based experiments can be designed to trace the shunt product *in vivo*.

Finally, the phenomenon of the sulfur carrier protein adjacent with "threonine synthase" SCO4293 widely exists in bacteria genomes. This combination can be

explored in other organisms. It is possible that the unnatural amino acid thiothreonine is the product in some organisms.

In the method development to probe and identify PLP-dependent enzymes, we used *E. coli* as the model system. The use of this method can be expanded to other organisms for probing and discovery of new PLP-dependent enzymes which were not predicted to be PLP-dependent. Also, PLP enzymes can be explored for their heat/cold resistant function in organisms capable of growing under extreme conditions such as *Thermus thermophilus* (optimal growth temperature 65 ^{o}C) and *Pseudomonas* species (growth temperature below 15 ^{o}C). Organisms like *Sulfolobus acidocaldarius* growing in sulfur-rich environments can be investigated to find out possible new PLP-dependent enzymes that are involved in sulfur metabolism.

Furthermore, this developed method can be used to investigate nutrient catabolic pathways. By comparing the change of PLP-dependent enzyme expression under different nutrient concentrations (from starvation to excess), the increase of one or several enzymes indicates possible roles in nutrient catabolism pathways.

Activity based protein profiling is a powerful tool to study a group of proteins based on their active site features. Developing methods for probing and identifying certain classes of proteins is very important in exploring new proteins and discovering novel pathways, which can be further studied in inhibitor development and drug design.

REFERENCES

1. Aebersold, R.; Goodlett, D. R., *Chem Rev* **2001,** *101* (2), 269-295.

2. Godovac-Zimmermann, J.; Brown, L. R., *Mass Spectrom Rev* **2001,** *20* (1), 1-57.

3. Rappsilber, J.; Mann, M., *Trends Biochem Sci* **2002,** *27* (2), 74-78.

4. Ideker, T.; Thorsson, V.; Ranish, J. A.; Christmas, R.; Buhler, J.; Eng, J. K.; Bumgarner, R.; Goodlett, D. R.; Aebersold, R.; Hood, L., *Science* **2001,** *292* (5518), 929-934.

5. Karlin, S.; Brocchieri, L.; Trent, J.; Blaisdell, B. E.; Mrazek, J., *Theor Popul Biol* **2002,** *61* (4), 367-390.

6. Shen, Y. F.; Zhao, R.; Berger, S. J.; Anderson, G. A.; Rodriguez, N.; Smith, R. D., *Anal Chem* **2002,** *74* (16), 4235-4249.

7. Taylor, J.; Anderson, N. L.; Scandora, A. E.; Willard, K. E.; Anderson, N. G., *Clin Chem* **1982,** *28* (4), 861-866.

8. Wilkins, M. R.; Sanchez, J. C.; Gooley, A. A.; Appel, R. D.; HumpherySmith, I.; Hochstrasser, D. F.; Williams, K. L., *Biotechnol Genet Eng* **1996,** *13*, 19-50.

9. Olsen, J. V.; Blagoev, B.; Gnad, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M., *Cell* **2006,** *127* (3), 635-648.

10. Kaiser, P.; Huang, L., *Genome Biol* **2005,** *6* (10), 233.

11. Pickart, C. M.; Fushman, D., *Curr Opin Chem Biol* **2004,** *8* (6), 610-616.

12. Peng, J. M.; Schwartz, D.; Elias, J. E.; Thoreen, C. C.; Cheng, D. M.; Marsischky, G.; Roelofs, J.; Finley, D.; Gygi, S. P., *Nat Biotechnol* **2003,** *21* (8), 921-926.

13. Li, J.; Zhou, J.; Chan, D. W., *Clin Chem* **2008,** *54* (6), A126-A126.

14. Srinivas, P. R.; Verma, M.; Zhao, Y. M.; Srivastava, S., *Clin Chem* **2002,** *48* (8), 1160-1169.

15. Rifai, N.; Gillette, M. A.; Carr, S. A., *Nat Biotechnol* **2006,** *24* (8), 971-983.

16. Norin, M.; Sundstrom, M., *Trends Biotechnol* **2002,** *20* (2), 79-84.

17. Dengjel, J.; Kratchmarova, I.; Blagoev, B., *Methods Mol Biol* **2010,** *658*, 267-278.

18. Uhlen, M.; Fagerberg, L.; Hallstrom, B. M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, A.; Kampf, C.; Sjostedt, E.; Asplund, A.; Olsson, I.; Edlund, K.; Lundberg, E.; Navani, S.; Szigyarto, C. A.; Odeberg, J.; Djureinovic, D.; Takanen, J. O.; Hober, S.; Alm, T.; Edqvist, P. H.; Berling, H.; Tegel, H.; Mulder, J.; Rockberg, J.; Nilsson, P.; Schwenk, J. M.; Hamsten, M.; von Feilitzen, K.; Forsberg, M.; Persson, L.; Johansson, F.; Zwahlen, M.; von Heijne, G.; Nielsen, J.; Ponten, F., *Science* **2015,** *347* (6220), 1260419.

19. Blueggel, M.; Chamrad, D.; Meyer, H. E., *Curr Pharm Biotechno* **2004,** *5* (1), 79-88.

20. Badock, V.; Steinhusen, U.; Bommert, K.; Otto, A., *Electrophoresis* **2001,** *22* (14), 2856-2864.

21. *Method Enzymol* **2009,** *463*, 1-851.

22. Issaq, H. J.; Conrads, T. P.; Janini, G. M.; Veenstra, T. D., *Electrophoresis* **2002,** *23* (17), 3048-3061.

23. Jiang, L.; He, L.; Fountoulakis, M., *J Chromatogr A* **2004,** *1023* (2), 317-320.

24. Sabounchi-Schutt, F.; Astrom, J.; Eklund, A.; Grunewald, J.; Bjellqvist, B., *Electrophoresis* **2001,** *22* (9), 1851-1860.

25. von Bredow, C.; Birrer, P.; Griese, M., *Eur Respir J* **2001,** *17* (4), 716-722.

26. Thompson, A. J.; Hart, S. R.; Franz, C.; Barnouin, K.; Ridley, A.; Cramer, R., *Anal Chem* **2003,** *75* (13), 3232-3243.

27. Smith, W. S.; Matthay, M. A., *Chest* **1997,** *111* (5), 1326-1333.

28. Patton, W. F.; Schulenberg, B.; Steinberg, T. H., *Curr Opin Biotech* **2002,** *13* (4), 321-328.

29. Neumann, M.; von Bredow, C.; Ratjen, F.; Griese, M., *Proteomics* **2002,** *2* (6), 683-689.

30. Montuschi, P.; Collins, J. V.; Ciabattoni, G.; Lazzeri, N.; Corradi, M.; Kharitonov, S. A.; Barnes, P. J., *Am J Resp Crit Care* **2000,** *162* (3), 1175-1177.

31. Stoeckli, M.; Chaurand, P.; Hallahan, D. E.; Caprioli, R. M., *Nat Med* **2001,** *7* (4), 493-496.

32. Strohman, R., *Bio-Technol* **1994,** *12* (4), 329-329.

33. Towbin, H.; Staehelin, T.; Gordon, J., *P Natl Acad Sci USA* **1979,** *76* (9), 4350-4354.

34. Renart, J.; Reiser, J.; Stark, G. R., *P Natl Acad Sci USA* **1979,** *76* (7), 3116-3120.

35. Courchesne, P. L.; Luethy, R.; Patterson, S. D., *Electrophoresis* **1997,** *18* (3-4), 369-381.

36. Figeys, D.; Ducret, A.; Yates, J. R.; Aebersold, R., *Nat Biotechnol* **1996,** *14* (11), 1579-1583.

37. Washburn, M. P.; Wolters, D.; Yates, J. R., *Nat Biotechnol* **2001,** *19* (3), 242-247.

38. Wolters, D. A.; Washburn, M. P.; Yates, J. R., *Anal Chem* **2001,** *73* (23), 5683-5690.

39. Wu, C. C.; MacCoss, M. J., *Curr Opin Mol Ther* **2002,** *4* (3), 242-250.

40. Whitelegge, J., *Expert Rev Proteomic* **2013,** *10* (2), 127-129.

41. Kelleher, N. L., *Anal Chem* **2004,** *76* (11), 197A-203A.

42. Chait, B. T., *Science* **2006,** *314* (5796), 65-66.

43. Zhang, Y. Y.; Fonslow, B. R.; Shan, B.; Baek, M. C.; Yates, J. R., *Chem Rev* **2013,** *113* (4), 2343-2394.

44. Edelmann, M. J.; Shack, L. A.; Naske, C. D.; Walters, K. B.; Nanduri, B., *Plos One* **2014,** *9* (12), 0114390.

45. Adhikari, J.; Fitzgerald, M. C., *J Am Soc Mass Spectr* **2014,** *25* (12), 2073-2083.

46. Zhang, H. Y.; Wu, P.; Chen, F. Y.; Hao, Y. W.; Lao, Y. X.; Ren, L. L.; Sun, L. Q.; Sun, W.; Wei, H. D.; Chan, D. W.; Jiang, Y.; He, F. C., *Proteomics* **2014,** *14* (17-18), 1977-1986.

47. Sano, S.; Tagami, S.; Hashimoto, Y.; Yoshizawa-Kumagaye, K.; Tsunemi, M.; Okochi, M.; Tomonagat, T., *J Proteome Res* **2014,** *13* (2), 1012-1020.

48. Afzal, V.; Huang, J. T.; Atrih, A.; Crowther, D. J., *Bmc Bioinformatics* **2011,** *12*, 338.

49. Speers, A. E.; Cravatt, B. F., *Chembiochem* **2004,** *5* (1), 41-47.

50. Paulick, M. G.; Bogyo, M., *Curr Opin Genet Dev* **2008,** *18* (1), 97-106.

51. Barglow, K. T.; Cravatt, B. F., *Nat Methods* **2007,** *4* (10), 822-827.

52. Evans, M. J.; Cravatt, B. F., *Chem Rev* **2006,** *106* (8), 3279-3301.

53. Vocadlo, D. J.; Hang, H. C.; Kim, E. J.; Hanover, J. A.; Bertozzi, C. R., *P Natl Acad Sci USA* **2003,** *100* (16), 9116-9121.

54. Patricelli, M. P.; Szardenings, A. K.; Liyanage, M.; Nomanbhoy, T. K.; Wu, M.; Weissig, H.; Aban, A.; Chun, D.; Tanner, S.; Kozarich, J. W., *Biochemistry-Us* **2007,** *46* (2), 350-358.

55. Lo, L. C.; Pang, T. L.; Kuo, C. H.; Chiang, Y. L.; Wang, H. Y.; Lin, J. J., *J Proteome Res* **2002,** *1* (1), 35-40.

56. Patricelli, M. P.; Giang, D. K.; Stamp, L. M.; Burbaum, J. J., *Proteomics* **2001,** *1* (9), 1067-1071.

57. Bogyo, M.; Verhelst, S.; Bellingard-Dubouchaud, V.; Toba, S.; Greenbaum, D., *Chem Biol* **2000,** *7* (1), 27-38.

58. Li, Y. M.; Xu, M.; Lai, M. T.; Huang, Q.; Castro, J. L.; DiMuzio-Mower, J.; Harrison, T.; Lellis, C.; Nadin, J. L.; Neduvelil, J. G.; Register, R. B.; Sardana, M. K.; Shearman, M. S.; Smith, A. L.; Shi, X. P.; Yin, K. C.; Shafer, J. A.; Gardell, S. J., *Nature* **2000,** *405* (6787), 689-694.

59. Chan, E. W. S.; Chattopadhaya, S.; Panicker, R. C.; Huang, X.; Yao, S. Q., *J Am Chem Soc* **2004,** *126* (44), 14435-14446.

60. Kolb, H. C.; Finn, M. G.; Sharpless, K. B., *Angew Chem Int Ed Engl* **2001,** *40* (11), 2004-2021.

61. Evans, R. A., *Aust J Chem* **2007,** *60* (6), 384-395.

62. Hoyle, C. E.; Bowman, C. N., *Angew Chem Int Edit* **2010,** *49* (9), 1540-1573.

63. Hangauer, D. G.; Smolinski, M.; Bu, Y. H.; Kazim, L.; Qu, J., *Abstr Pap Am Chem S* **2010,** *239*.

64. Nicholson, A. W.; Hall, C. C.; Strycharz, W. A.; Cooperman, B. S., *Biochemistry-Us* **1982,** *21* (16), 3797-3808.

65. Krysiak, J.; Breinbauer, R., *Top Curr Chem* **2012,** *324*, 43-84.

66. Adibekian, A.; Martin, B. R.; Wang, C.; Hsu, K. L.; Bachovchin, D. A.; Niessen, S.; Hoover, H.; Cravatt, B. F., *Nat Chem Biol* **2012,** *8* (3), 318-318.

67. Greenbaum, D. C.; Arnold, W. D.; Lu, F.; Hayrapetian, L.; Baruch, A.; Krumrine, J.; Toba, S.; Chehade, K.; Bromme, D.; Kuntz, I. D.; Bogyo, M., *Chem Biol* **2002,** *9* (10), 1085-1094.

68. Hemelaar, J.; Borodovsky, A.; Kessler, B. M.; Reverter, D.; Cook, J.; Kolli, N.; Gan-Erdene, T.; Wilkinson, K. D.; Gill, G.; Lima, C. D.; Ploegh, H. L.; Ovaa, H., *Mol Cell Biol* **2004,** *24* (1), 84-95.

69. Barglow, K. T.; Cravatt, B. F., *Angew Chem Int Edit* **2006,** *45* (44), 7408-7411.

70. Kessler, D., *Fems Microbiol Rev* **2006,** *30* (6), 825-840.

71. Mueller, E. G., *Nat Chem Biol* **2006,** *2* (4), 185-194.

72. Dorrestein, P. C.; Zhai, H. L.; Taylor, S. V.; McLafferty, F. W.; Begley, T. P., *J Am Chem Soc* **2004,** *126* (10), 3091-3096.

73. Leimkuhler, S.; Wuebbens, M. M.; Rajagopalan, K. V., *J Biol Chem* **2001,** *276* (37), 34695-34701.

74. Burns, K. E.; Baumgart, S.; Dorrestein, P. C.; Zhai, H. L.; McLafferty, F. W.; Begley, T. P., *J Am Chem Soc* **2005,** *127* (33), 11602-11603.

75. Krishnamoorthy, K.; Begley, T. P., *J Am Chem Soc* **2010,** *132* (33), 11608-11612.

76. Schmitz, J.; Wuebbens, M. M.; Rajagopalan, K. V.; Leimkuhler, S., *Biochemistry-Us* **2007,** *46* (3), 909-916.

77. The SEED Database. http://theseed.uchicago.edu/FIG/

78. Merkx, R.; Brouwer, A. J.; Rijkers, D. T. S.; Liskamp, R. M. J., *Org Lett* **2005,** *7* (6), 1125-1128.

79. Shangguan, N.; Katukojvala, S.; Greenburg, R.; Williams, L. J., *J Am Chem Soc* **2003,** *125* (26), 7754-7755.

80. Kolakowski, R. V.; Shangguan, N.; Sauers, R. R.; Williams, L. J., *J Am Chem Soc* **2006,** *128* (17), 5695-5702.

81. Noble, J. E.; Bailey, M. J. A., *Method Enzymol* **2009,** *463*, 73-95.

82. Taylor, S. V.; Kelleher, N. L.; Kinsland, C.; Chiu, H. J.; Costello, C. K.; Backstrom, A. D.; McLafferty, F. W.; Begley, T. P., *J Biol Chem* **1998,** *273* (26), 16555-16560.

83. Ueberle, B.; Frank, R.; Herrmann, R., *Proteomics* **2002,** *2* (6), 754-764.

84. Wilchek, M.; Bayer, E. A., *Methods in Enzymology* **1990,** *184*, 5-13.

85. Sun, R. Y.; Zhuang, H. S., *Food Agr Immunol* **2015,** *26* (5), 746-760.

86. Laber, B.; Gerbling, K. P.; Harde, C.; Neff, K. H.; Nordhoff, E.; Pohlenz, H., *Biochemistry-Us* **1994,** *33* (11), 3413-3423.

87. Benson, J. R.; Hare, P. E., *P Natl Acad Sci USA* **1975,** *72* (2), 619-622.

88. Pripis-Nicolau, L.; de Revel, G.; Marchand, S.; Beloqui, A. A.; Bertrand, A., *J Sci Food Agr* **2001,** *81* (8), 731-738.

89. Tcherkas, Y. V.; Denisenko, A. D., *J Chromatogr A* **2001,** *913* (1-2), 309-313.

90. Ferla, M. P.; Patrick, W. M., *Microbiol-Sgm* **2014,** *160*, 1571-1584.

91. Krishnamoorthy, K.; Begley, T. P., *J Am Chem Soc* **2011,** *133* (2), 379-386.

92. Chi, A.; Huttenhower, C.; Geer, L. Y.; Coon, J. J.; Syka, J. E. P.; Bai, D. L.; Shabanowitz, J.; Burke, D. J.; Troyanskaya, O. G.; Hunt, D. F., *P Natl Acad Sci USA* **2007,** *104* (7), 2193-2198.

93. Beltran, L.; Cutillas, P. R., *Amino Acids* **2012,** *43* (3), 1009-1024.

94. Lai, A. C. Y.; Tsai, C. F.; Hsu, C. C.; Sun, Y. N.; Chen, Y. J., *Rapid Commun Mass Sp* **2012,** *26* (18), 2186-2194.

95. Fila, J.; Honys, D., *Amino Acids* **2012,** *43* (3), 1025-1047.

96. Kanshin, E.; Michnick, S.; Thibault, P., *Semin Cell Dev Biol* **2012,** *23* (8), 843-853.

97. Muszynska, G.; Dobrowolska, G.; Medin, A.; Ekman, P.; Porath, J. O., *J Chromatogr* **1992,** *604* (1), 19-28.

98. Heyl, D.; Luz, E.; Harris, S. A.; Folkers, K., *J Am Chem Soc* **1951,** *73* (7), 3430-3433.

99. Toney, M. D., *Arch Biochem Biophys* **2005,** *433* (1), 279-287.

100. Muller, I. B.; Das Gupta, R.; Luersen, K.; Wrenger, C.; Walter, R. D., *Mol Biochem Parasit* **2008,** *160* (1), 1-7.

101. Stamper, C. G. F.; Morollo, A. A.; Ringe, D., *Biochemistry-Us* **1998,** *37* (29), 10438-10445.

102. Brzovic, P.; Holbrook, E. L.; Greene, R. C.; Dunn, M. F., *Biochemistry-Us* **1990,** *29* (2), 442-451.

103. Eliot, A. C.; Kirsch, J. F., *Annu Rev Biochem* **2004,** *73*, 383-415.

104. Mehta, P. K.; Christen, P., *Adv Enzymol Ramb* **2000,** *74*, 129.

105. Hsu, C. C.; Cheng, C. H.; Hsu, C. L.; Lee, W. J.; Huang, S. C.; Huang, Y. C., *Food Nutr Res* **2015,** *59*, 25702.

106. Endo, N.; Nishiyama, K.; Otsuka, A.; Kanouchi, H.; Taga, M.; Oka, T., *Brit J Nutr* **2006,** *95* (6), 1088-1093.

107. Anand, S. S., *J Appl Toxicol* **2005,** *25* (5), 440-443.

108. Jain, V.; Kumar, M.; Chatterji, D., *J Microbiol* **2006,** *44* (1), 1-10.

109. Mukherjee, T.; Hanes, J.; Tews, I.; Ealick, S. E.; Begley, T. P., *Bba-Proteins Proteom* **2011,** *1814* (11), 1585-1596.

110. Durfee, T.; Hansen, A. M.; Zhi, H.; Blattner, F. R.; Jin, D. J., *J Bacteriol* **2008,** *190* (3), 1084-1096.

111. Traxler, M. F.; Summers, S. M.; Nguyen, H. T.; Zacharia, V. M.; Hightower, G. A.; Smith, J. T.; Conway, T., *Mol Microbiol* **2008,** *68* (5), 1128-1148.

112. Traxler, M. F.; Chang, D. E.; Conway, T., *P Natl Acad Sci USA* **2006,** *103* (7), 2374-2379.

113. Brown, D. R.; Barton, G.; Pan, Z.; Buck, M.; Wigneshweraraj, S., *Nat Commun* **2014,** *5*, 4115.

114. Polissi, A.; De Laurentis, W.; Zangrossi, S.; Briani, F.; Longhi, V.; Pesole, G.; Deho, G., *Res Microbiol* **2003,** *154* (8), 573-580.

115. Huang, S. H.; Zhang, J. Y.; Wang, L. H.; Huang, L. Q., *Plant Physiol Bioch* **2013,** *66*, 63-67.

116. Bilski, P.; Li, M. Y.; Ehrenshaft, M.; Daub, M. E.; Chignell, C. F., *Photochem Photobiol* **2000,** *71* (2), 129-134.

117. Knockel, J.; Muller, I. B.; Butzloff, S.; Bergmann, B.; Walter, R. D.; Wrenger, C., *Biochem J* **2012,** *443*, 397-405.

118. Daub, M. E.; Hangarter, R. P., *Plant Physiol* **1983,** *73* (3), 855-857.

119. Bennett, B. D.; Kimball, E. H.; Gao, M.; Osterhout, R.; Van Dien, S. J.; Rabinowitz, J. D., *Nat Chem Biol* **2009,** *5* (8), 593-599.

120. Liu, F.; Qimuge; Hao, J.; Yan, H.; Bach, T.; Fan, L.; Morigen, *Plos One* **2014,** *9* (3), 92229.

121. Kaleta, C.; Schauble, S.; Rinas, U.; Schuster, S., *Biotechnol J* **2013,** *8* (9), 1105-1114.

122. Akashi, H.; Gojobori, T., *P Natl Acad Sci USA* **2002,** *99* (6), 3695-3700.

APPENDIX A

Protein View

Match to: **gi|46198624** Score: **54**
putative thiS protein [Thermus thermophilus HB27]
Found in search of DATA.TXT

Nominal mass (M_r): **6982**; Calculated pI value: **4.31**
NCBI BLAST search of gi|46198624 against nr
Unformatted sequence string for pasting into other applications

Variable modifications: Biotin (K)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **73%**

Matched peptides shown in **Bold Red**

```
     1 MVWLNGEPRP LEGKTLKEVL EEMGVELKGV AVLLNEEAFL GLEVPDRPLR
    51 DGDVVEVVAL MQGG
```

Database search result for identification of *Thermus thermophilus* ThiS.

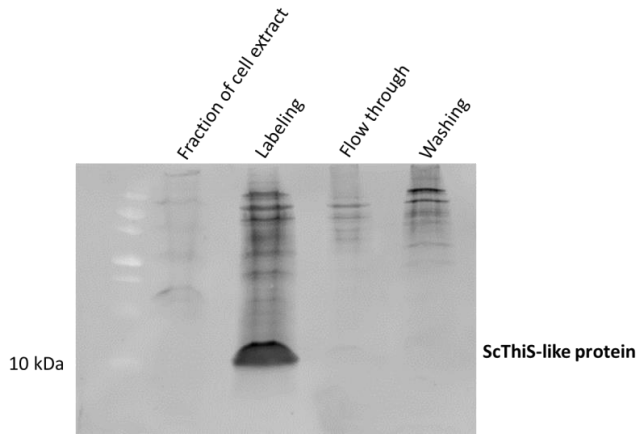
Protein View

Match to: **gi|38704219** Score: **90**
sulfur carrier protein ThiS [Escherichia coli O157:H7 str. Sakai]
Sequence Coverage: **22%**

Matched peptides shown in **Bold Red**

```
     1 MQILFNDQPM QCAAGQTVHE LLEQLVQRQA GAALAINQQI VPREQWTQHI
    51 VQDGDQILLF QVIAGG
```

Database search result for identification of *Escherichia coli* ThiS.



Labeling and washing of *S. coelicolor* sulfur carrier protein.

Protein View

Match to: **gi|21222687** Score: **156**
hypothetical protein SCO4294 [Streptomyces coelicolor A3(2)]
Found in search of DATA.TXT

Nominal mass (M_r): **9564**; Calculated pI value: **4.90**
NCBI BLAST search of gi|21222687 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Streptomyces coelicolor A3(2)
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|256786239 from Streptomyces lividans TK24
gi|289770130 from Streptomyces lividans TK24
gi|8248793 from Streptomyces coelicolor A3(2)
gi|289700329 from Streptomyces lividans TK24

Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **50%**

Matched peptides shown in **Bold Red**

1 MSVTVRIPTI LRTYTGGK**AE VSADGANLGE VISDLEKNHT GIAARVLDDQ**
51 **GK**LRR**FVNVY VNDDDVR**FEQ GLQTATPDGA GVSIIPAVAG G

Database search result for identification of *Streptomyces coelicolor* sulfur carrier protein.