X-RAY CRYSTAL STRUCTURES OF ENZYMES IN PURINE/PYRIMIDINE SALVAGE PATHWAY, SULFUR ASSIMILATION PATHWAY AND STRUCTURE-BASED DESIGN AND ORGANIC SYNTHESIS OF A LIGHT-ACTIVATED INHIBITOR FOR THIAMIN BIOSYNTHETIC ENZYMES

by Timothy Hoang Tran

X-RAY CRYSTAL STRUCTURES OF ENZYMES IN PURINE/PYRIMIDINE SALVAGE PATHWAY, SULFUR ASSIMILATION PATHWAY AND STRUCTURE-BASED DESIGN AND ORGANIC SYNTHESIS OF A LIGHT-ACTIVATED INHIBITOR FOR THIAMIN BIOSYNTHETIC ENZYMES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Timothy Hoang Tran

August 2010

X-RAY CRYSTAL STRUCTURES OF ENZYMES IN PURINE/PYRIMIDINE SALVAGE PATHWAY, SULFUR ASSIMILATION PATHWAY AND STRUCTURE-BASED DESIGN AND ORGANIC SYNTHESIS OF A LIGHT-ACTIVATED INHIBITOR FOR THIAZOLE BIOSYNTHETIC ENZYMES

Timothy Hoang Tran, Ph. D.

Cornell University 2010

The structure of *Streptococcus pyogenes* uridine phosphorylase (UP) in the salvage pathway reveals that it uses different residues to stabilize the transition state. Sequence alignment of the enzyme uncovers a subclass of uridine phosphorylases that contain these active site residues. These data, together with the specificity loop of uridine phosphorylases, may be used to distinguish between UP and purine nucleoside phosphorylase (PNP) enzymes at the level of primary sequence, and thus may allow proper annotation. In *Bos taurus*, the active site of its purine nucleoside phosphorylase can accommodate the major by-product of DNA damage, pyrimidopurinone 2'-deoxyribose, and cleave its glycosidic bond. The base-binding site of the enzyme helps explain why the reaction cannot proceed in the reverse direction.

The structure of a protein complex, ThiS-ThiG, in the thiazole biosynthetic pathway was used to rationally design and synthesize a light-activated analog. Preliminary characterization of the analog demonstrates that it could be covalently linked to ThiS using intein chemistry. We anticipate that this analog will be of use for the structural and mechanistic characterization of thiazole synthase.

The crystallization and diffraction screenings of selected enzymes in the methionine biosynthetic and sulfur assimilation pathways of *Wolinella succinogenes*

were carried out. The crystallization conditions for the MetY/homocysteine complex have been optimized from 8 Å to 4 Å resolution by screening with different cryoreagents. Detergent screens for the apo-MetY generated small crystals, which were used for X-ray diffraction experiments.

The MetY structure, which is the first o-acetylhomoserine sulfhydrylase, has been determined. A search for structurally homologous proteins reveals that MetY has the same fold as cystathionine gamma lyase and methionine gamma lyase. The active sites of these enzymes, which contain PLP, share a high degree of structural similarity, suggesting that MetY belongs to the γ-elimination subclass of Cys/Met metabolism PLP-dependent family of enzymes. The structure of MetY, together with biochemical data, provides useful insight to the mechanism of sulfur transfer to a small molecule *via* a protein thiocarboxylate.

BIOGRAPHICAL SKETCH

Timothy Hoang Tran attended Sexton High School in Lansing, MI and college at the University of Michigan in Ann Arbor. After graduating from college, he received an NSF fellowship to explore research opportunities at Los Alamos National Laboratory. About a year later, he returned to the University of Michigan for a master's degree in biomedical engineering. During that time he began to develop an interest in X-ray crystallography. He returned to Los Alamos every subsequent summer to work for Tom Terwilliger in collaboration with Geoffrey Waldo to determine the structure of superfolder green fluorescent protein. After the first master's degree at Michigan, his curiosity in an Ivy League education propelled him to enroll in another professional master's degree in engineering biotechnology at the University of Pennsylvania, Philadelphia. At Penn, he learned different technologies to screen for potent drugs. From this learning experience, together with X-ray crystallography experience in Los Alamos, he decided to come to Cornell for a PhD in structure-based drug design and synthesis. He joined both Begley and Ealick groups at the beginning in order to achieve this objective.

To my family and Anna

ACKNOWLEDGMENTS

For the rest of my life I will always remember and be grateful for Professor Steve Ealick, who truly understands my strengths and weaknesses to instruct me accordingly. He always treats me with respect since the first year I co-joined his lab and he knows exactly what I mean even when I say the opposite of what I really mean. He taught me how to think scientifically in his own little way, and how to be a responsible citizen and a decent human being. I literally would not be here today without his help and guidance.

I would like to thank my committee members, Professor Ailong Ke and Professor Richard Cerione, for their time and their willingness to evaluate my work. Their insightful suggestions and questions help me think critically about my own work. I also would like to thank Professor Tadhg Begley for teaching me how to design and synthesize enzyme inhibitors. I am indebted to Dr. Tom Terwilliger for his continual support and encouragement even long after I left his lab. I really appreciate all the members of Ealick lab who were very supportive and helpful during my learning process. Of course, I have to mention Leslie Kinsland for her excellent work to keep the lab in order, and Dr. Cynthia Kinsland for her chemist's explanations of molecular biology and her reliable cloning work, which saved me a lot of time. I am thankful to the staff at NE-CAT and CHESS for their technical assistance with data collection and processing. Finally, I am grateful for my family for their support throughout the years, in particular, my fiancée for her cooking and spending many hours with me in the lab during my dissertation writing.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

The adage 'A picture is worth a thousand words' is true to the core of structural biology. No biologists would deny the importance of the structural approach to the field of biology. Indirect experimental proofs can be quite clever and provide useful insight to biology. Alfred Gilman's novel work on adenylyl cyclase represents some of the most creative experiments (1, 2). However, the element of uncertainty and scientific skepticism always remain high in the absence of direct evidence. Gilman did not receive the Nobel Prize until the advent of structures that validated his novel experiments. 'Seeing is believing' applies to structural biology as well. Structures themselves can be direct experimental evidence. The high-resolution structure of the potassium channel lends support to its mechanism of potassium transport (3). Structure biology has revolutionized the way biologists think about problems. Structures serve as a guide in designing biological experiments.

The major tool in structural biology is X-ray crystallography due to its capacity to determine large protein structures at high resolution. Approximately 85-90% of all the structures deposited in the PDB were determined by X-ray crystallography. The rest include nuclear magnetic resonance, and electron microscopy. Especially, in the field of enzymology, X-ray crystallography has helped elucidate many interesting, complex, 'impossible' biochemical transformations that have no precedents in the traditional chemical synthesis. Atomic resolution of substrates in the active sites of enzymes have led to design of analogs for various drug targets as well as mechanistic studies, such as those for thiamin biosynthesis (4-6) and purine salvage pathway (7). Recently, the structure of a neurotransmitter transporter has helped shed light on the mechanism of antidepressants in humans (8). Structures of the sirtuin family of proteins revealed the chemical basis of their function, which were proposed to play a

major role in aging (9-12). Without X-ray crystallography, rational drug design would not have been possible. The old combinatorial method of random screens among millions of organic compounds for each particular disease would still be the only method of choice.

In spite of many success stories in the field of protein crystallography, many challenges face crystallographers today. Crystallization of membrane proteins and macromolecular complexes has been very difficult due to the unique environment in which membrane proteins reside and the conformational heterogeneity of protein complexes. Approximately 3% of all the proteins in the PDB (13) are membrane proteins to date. However, this percentage, albeit small, represents significant achievements in the field considering that the crystallization of these proteins were once thought to be impossible: "They said it can't be done"--Roderick MacKinnon--a membrane protein pioneer who worked out the structure of the potassium channel that won him the 2003 Nobel Prize. G-protein coupled receptor (GPCR) has been studied for many years and its structure had eluded researchers, due to the fluidity of seven tranmembrane helices with intervening extracellular and intracellular loops, until approximately three years ago (14). The crystal structure of GPCR enabled biologists to better understand the mechanism in which the extracellular hormone epinephrine induces conformational changes in GPCR, which transmits signal into the cytosol and triggers various cancer-signaling pathways. Thus, the future looks promising for those who study membrane proteins and macromolecular complexes. Approximately, 60-70% of drugs in the market target membrane proteins. The number of determined structures of membrane proteins has increased exponentially in recent years. Considering the number of membrane proteins and protein complexes in the human genomes and other species, research opportunities in these areas are seemingly endless.

The work presented in this dissertation is an application of X-ray crystallography to the field of mechanistic enzymology and is categorized into three research areas. The first area focuses on X-ray crystal structures of enzymes in the purine/pyrimidine salvage pathway. The second area utilizes a solved crystal structure of a protein complex to design and synthesize substrate analogs to further understanding of the thiamin biosynthetic pathway. The third area also includes enzymes in the sulfur assimilation pathway and methionine biosynthesis. These three areas are divided into five chapters.

Chapter 2 explores the structure of *Streptococcus pyogenes* uridine phosphorylase (SpUP). This is a key enzyme in the uridine nucleoside salvage pathway that catalyzes the reversible phosphorolysis of uridine (2'-deoxyuridine) nucleoside into uracil and ribose (2'-deoxyribose) 1-phosphate. The reverse reaction generates nucleosides, which are precursors for RNA and DNA synthesis. The salvage pathway is an alternative pathway to the de novo biosynthetic pathway, which consumes much more energy to synthesize RNA and DNA's precursors. Although uridine phosphorylase is present in all organisms, its structure is unique between prokaryotes and eukaryotes. In addition, some parasites, such as *Plasmodium falciparum*, rely entirely on this salvage pathway; thus, it has been an attractive target for antimalarial drugs. Primary sequence alignment (15) of SpUP against the database from a BLAST search (16) revealed that the two conserved active site residues of a homolog *Escherichia coli* uridine phosphorylase (EcUP) are not present in SpUP. This led to questions regarding the conservation and catalytic mechanism of SpUP relative to the proposed one based on known structures of the superfamily. Thus, the crystal structure of SpUP was determined and enzymatic assays of the wildtype and mutants were performed to address the questions.

Chapter 3 relates to Chapter 2 in the sense that it also focuses on the salvage

pathway. However, it examines a purine nucleoside phosphorylase (PNP) structure of a higher eukaryote, specifically bovine purine nucleoside phosphorylase (bPNP). Structural studies of PNPs are particularly important since the absence of PNP activity is linked to T-cell immunodeficiency, due to the accumulation of dGTP, while other tissues remain unaffected (17, 18). Since T-cells are known to mediate autoimmune disorders, some cancers, and tissue transplant rejection, PNP has become an important drug target. This chapter examines the crystal structure of the complex containing phosphate/sulfate and a guanosine nucleoside analog, pyrimidopurinone 2-deoxyribose (M_1G-dR), which is a major by-product of DNA damage. The synthesis of this analog facilitates biochemical studies of the mechanism of DNA repairs. However, its chemical synthesis is challenging due to the lability of glycosydic bond under acidic conditions, which are critical for adding and removing protecting groups. Thus, PNPs are used in the last step by coupling a sugar moiety and base to generate a nucleoside analog. Interest in studying this complex originates from the difficulty of performing the coupling step of this analog using bPNP, while EcPNP can carry out the coupling reaction giving 61-64% yield. Therefore, crystal structures of M_1G-dR/PO_4/bPNP and M_1G-dR/SO_4/bPNP complexes, in which SO_4 acts as an unreactive mimic of PO_4, were determined in order to rationalize the biochemical observations.

Chapter 4 utilizes a known crystal structure of a protein complex to design and synthesize a light-activated inhibitor involved in the biosynthesis of the thiazole moiety of thiamin. Thiamin is an essential cofactor in all living organisms (19). Unlike most prokaryotes and some eukaryotes, humans cannot biosynthesize thiamin, and thus have to intake 1.4 mg/day from dietary or supplemental sources (20). The interest in synthesizing this inhibitor originates from recent studies of thiazole biosynthesis, which suggest that thiazole synthase (ThiG) is covalently linked to a sulfur carrier protein (ThiS) by a novel reaction intermediate (21, 22). This is so far

the only known intermediate identified that covalently links two different proteins. A light-activated inhibitor is the ultimate goal because this light activation bypasses the removal of protecting groups by chemical means that would precipitate out the proteins. Thus, the whole process of inhibitor design and the actual synthesis were carried out and discussed. In addition, preliminary characterization of the inhibitor was performed to demonstrate the proof-of-concept experiments, which utilize intein chemistry to covalently link small molecules to ThiS.

Chapter 5 presents the preliminary crystallographic studies of enzymes in a gene cluster containing genes for proteins in methionine biosynthesis and sulfur assimilation pathway of *Wolinella succinogenes*. This gene cluster was discovered from a search for a family of sulfur carrier proteins. The presence of a sulfur carrier protein, HcyS, in this cluster suggests that it somehow links to the sulfur assimilation pathway and methionine biosynthesis. In fact, from our preliminary biochemical studies, it was hypothesized that the sulfur from sulfate in the sulfur assimilation pathway is transferred into the sulfur carrier protein in the form of a protein thiocarboxylate. Eventually, the sulfur from the protein thiocarboxylate is proposed to incorporate into methionine. In order to further support this hypothesis, initial attempts to crystallize selected proteins and protein complexes in the proposed pathway were carried out.

Chapter 6 describes the crystal structure of *Wolinella succinogenes* o-acetylhomoserine sulfhydrylase (OAHS), or MetY, one of the enzymes in the gene cluster mentioned in chapter four. MetY is the first structure of an OAHS. Based on our preliminary data, it was proposed that MetY catalyzes the transfer of sulfur from the protein thiocarboxylate form of HcyS to o-acetylhomoserine to generate the HcyS-homocysteine adduct. From the structure of MetY, together with biochemical data, a mechanism of the sulfur transfer catalyzed by MetY has been proposed.

REFERENCES

1. Gilman, A. G., and Nirenberg, M. (1971) Regulation of adenosine 3',5'-cyclic monophosphate metabolism in cultured neuroblastoma cells, *Nature 234*, 356-358.
2. Seeds, N. W., and Gilman, A. G. (1971) Norepinephrine stinulated increase of cyclic AMP levels in developing mouse brain cell cultures, *Science 174*, 292.
3. Doyle, D. A., Morais Cabral, J., Pfuetzner, R. A., Kuo, A., Gulbis, J. M., Cohen, S. L., Chait, B. T., and MacKinnon, R. (1998) The structure of the potassium channel: molecular basis of K^+ conduction and selectivity, *Science 280*, 69-77.
4. Agyei-Owusu, K., and Leeper, F. J. (2009) Thiamin diphosphate in biological chemistry: analogues of thiamin diphosphate in studies of enzymes and riboswitches, *FEBS J. 276*, 2905-2916.
5. Sudarsan, N., Cohen-Chalamish, S., Nakamura, S., Emilsson, G. M., and Breaker, R. R. (2005) Thiamine pyrophosphate riboswitches are targets for the antimicrobial compound pyrithiamine, *Chem. Biol. 12*, 1325-1335.
6. Jurgenson, C. T., Begley, T. P., and Ealick, S. E. (2009) The structural and biochemical foundations of thiamin biosynthesis, *Annu. Rev. Biochem. 78*, 569-603.
7. Ho, M. C., Shi, W., Rinaldo-Matthis, A., Tyler, P. C., Evans, G. B., Clinch, K., Almo, S. C., and Schramm, V. L. (2010) Four generations of transition-state analogues for human purine nucleoside phosphorylase, *Proc. Natl. Acad. Sci. U. S. A. 107*, 4805-4812.
8. Singh, S. K., Yamashita, A., and Gouaux, E. (2007) Antidepressant binding site in a bacterial homologue of neurotransmitter transporters, *Nature 448*,

952-956.

9. Avalos, J. L., Bever, K. M., and Wolberger, C. (2005) Mechanism of sirtuin inhibition by nicotinamide: altering the NAD^+ cosubstrate specificity of a Sir2 enzyme, *Mol. Cell 17*, 855-868.

10. Marmorstein, R. (2004) Structure and chemistry of the Sir2 family of NAD+-dependent histone/protein deactylases, *Biochem. Soc. Trans. 32*, 904-909.

11. Chang, J. H., Kim, H. C., Hwang, K. Y., Lee, J. W., Jackson, S. P., Bell, S. D., and Cho, Y. (2002) Structural basis for the NAD-dependent deacetylase mechanism of Sir2, *J. Biol. Chem. 277*, 34489-34498.

12. Lin, H. (2007) Nicotinamide adenine dinucleotide: beyond a redox coenzyme, *Org. Biomol. Chem. 5*, 2541-2554.

13. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank, *Nucleic Acids Res. 28*, 235-242.

14. Rasmussen, S. G., Choi, H. J., Rosenbaum, D. M., Kobilka, T. S., Thian, F. S., Edwards, P. C., Burghammer, M., Ratnala, V. R., Sanishvili, R., Fischetti, R. F., Schertler, G. F., Weis, W. I., and Kobilka, B. K. (2007) Crystal structure of the human beta2 adrenergic G-protein-coupled receptor, *Nature 450*, 383-387.

15. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res. 22*, 4673-4680.

16. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res. 25*, 3389-3402.

17. Giblett, E. R., Ammann, A. J., Wara, D. W., Sandman, R., and Diamon, L. K. (1975) Nucleoside-phosphorylase deficiency in a child with severely defective T-cell immunity and normal B-cell immunity, *Lancet 1*, 1010-1013.

18. Hershfield, M. S., and Mitchell, B. S. (1995) The Metabolic Basis of Inherited Disease, 1725-1768.

19. Butterworth, R. F. (2003) Thiamin deficiency and brain disorders, *Nutr. Res. Rev. 16*, 277-284.

20. Soriano, J. M., Molto, J. C., and Manes, J. (2000) Dietary Intake and Food Pattern among University Students, *Nutr. Res. (NY) 20*, 1249-1258.

21. Dorrestein, P. C., Zhai, H., McLafferty, F. W., and Begley, T. P. (2004) The biosynthesis of the thiazole phosphate moiety of thiamin: the sulfur transfer mediated by the sulfur carrier protein ThiS, *Chem. Biol. 11*, 1373-1381.

22. Dorrestein, P. C., Huili Zhai, H., Taylor, S. V., McLafferty, F. W., and Begley, T. P. (2004) The biosynthesis of the thiazole phosphate moiety of thiamin (vitamin B1): the early steps catalyzed by thiazole synthase, *J. Am. Chem. Soc. 126*, 3091-3096.

CHAPTER 2

CRYSTAL STRUCTURE OF STREPTOCOCCUS PYOGENES URIDINE PHOSPHORYLASE REVEALS A SUBCLASS OF THE NP-I SUPERFAMILY THAT UTILIZES DIFFERENT RESIDUES TO STABLIZE THE TRANSITION STATE

Section 2.1 Introduction

Uridine phosphorylase (EC 2.4.2.3) is a key enzyme in the pyrimidine-salvage pathway. The enzyme is present in all organisms and is essential for extracting carbon sources from nucleosides (*1*). This salvage pathway is an alternative route to the energetically expensive *de novo* biosynthetic pathway, which requires approximately ten different biochemical transformations to make precursors for DNA an RNA biosynthesis (*2*). The enzyme salvages free bases generated by other processes, such as apoptosis, to regenerate DNA and RNA precursors as illustrated in Figure 2.1. Specifically, uridine phosphorylase catalyzes the reversible phosphorylation of ribonucleosides or 2′-deoxynucleosides of uracil (Ura) and their analogs to the corresponding nucleobases and (2′-deoxy) ribose-1-phosphate (R1P) as shown in Figure 2.2. Some species, particularly parasites such as *Plasmodium falciparum*, lack the *de novo* biosynthetic pathway (*3, 4*). As a result, these organisms must rely entirely on the salvage pathway as a source of nucleosides, making these salvage pathways a potential antiparasitic drug target.

Nucleoside phosphorylases are also known to inactivate certain pyrimidine and purine nucleoside analogs that possess anti-tumor properties (*5, 6*). Thus, inhibitors for these enzymes could enhance the efficacy of the nucleoside analogs. The mechanisms of inhibition by analogs and binding sites in uridine phosphorylases have been proposed (*7, 8*). Current knowledge of uridine phosphorylase has led to the

clinical application of 5-fluorouridine as a chemotherapeutic agent for the treatment of advanced stage colorectal cancer (*9*). In spite of these studies, the details of the chemical mechanism and the role of active site residues are not fully understood. For example, the sulfate ion has been previously used as an unreactive mimic of phosphate ion (*10*) was recently discovered to cleave uridine into two pieces, the ribose intermediate and the base (*11*). Interestingly, the proposed ribose intermediate exists as a glycal, a phenomenon that has not been observed previously in nucleoside phosphorylases.

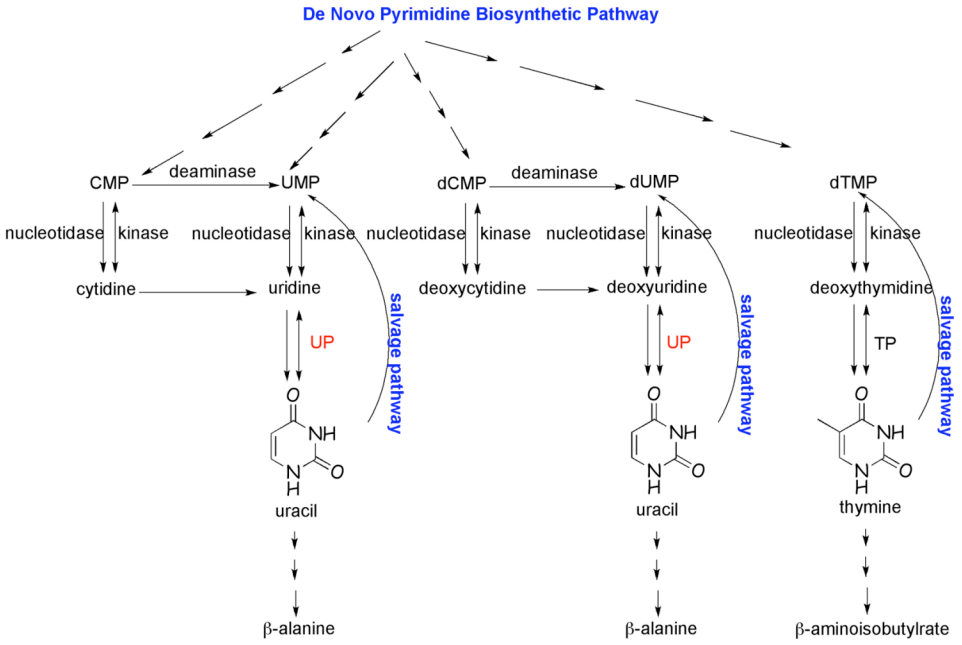


Figure 2.1. Generalized de novo pyrimidine biosynthetic and salvage pathways. TP stands for thymidine phosphorylase; CMP, UMP, dCMP, dUMP, and dTMP are cytidine, uridine, deoxycytidine, deoxyuridine, and deoxythymidine monosphosphate, respectively.

In addition to drug design applications, UP can be used as an alternative chemical tool in making nucleoside analogs for certain drugs through enzymatic synthesis. Specifically, the glycosidic bond is highly labile in acidic organic solvents.

Thus, traditional chemical synthesis of nucleoside analogs becomes quite challenging, particularly due to the necessity of adding and removing protecting groups, which lead to very low yields.

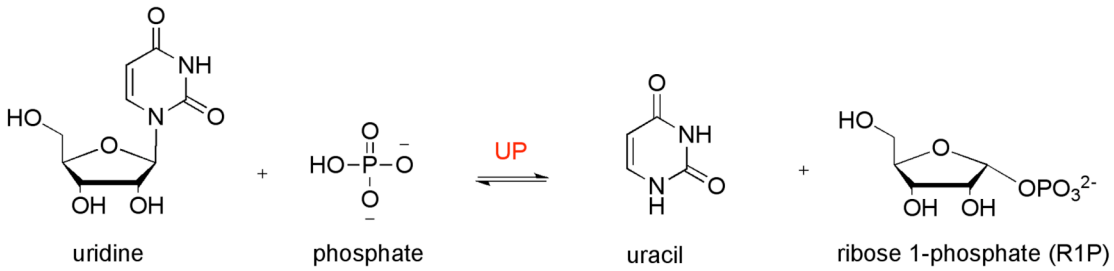


Figure 2.2. Phosphorolysis reaction catalyzed by uridine phosphorylase.

Streptococcus pyogenes is a gram positive bacteria that cause streptococcal illnesses that vary based on the location of infection (*12*). SpUP contains 259 amino acids per monomer with a molecular weight of 27.8 kD. Multiple sequence alignments (*13*) of SpUP against the nonredundant BLAST (*14*) database revealed a surprising result. The two critical active-site arginine residues found in EcUP, which has 39% sequence identity with SpUP, were not present in SpUP or a number of other UP's from diverse branches of the phylogenetic tree of the UP superfamily. The sequence alignment suggests that SpUP belongs to the NP-I family of proteins and thereby shares a common α/β fold with hexameric and trimeric purine nucleoside phosphorylases, adenosine monophosphate nucleosidase, methylthioadenosine nucleosidase, and methylthioadenosine phosphorylases (*6, 15*).

In this report we present the X-ray structures of SpUP in complex with the products (R1P/Ura) and the substrate uridine (Urd) at 1.8 Å and 2.2 Å, respectively. The three-dimensional structure shows that the overall monomeric fold is similar to nucleoside phosphorylase I (NP-I) superfamily. The structures of these complexes

shed light on the mechanism of catalysis in the absence of two critical active site arginine residues, which are hypothesized to stabilize the transition state of EcUP. In addition, these studies provide useful insight to the relationship of SpUP to other members of the NP-I superfamily.

Section 2.2 Experimental

Molecular cloning of native SpUP. The *up* gene was PCR amplified from *S*treptococcus *pyogenes* (ATCC 12344) genomic DNA using the following primers: upstream primer 5'-GGG TAG CAT ATG CAA AAT TAT TCA GGT GAA GTC GG-3' (inserts an *Nde*I site at the start codon of the *up* open reading frame); downstream primer 5'-CCC TAC TCG AGT TAT TGT GAT TTA TCA TTT TCA ATA AG-3' (inserts an *Xho*I site after the end of the *up* open reading frame). The purified PCR product was digested with *Nde*I and *Xho*I, purified and ligated into similarly digested pTHT (which is a pET-28 derived vector which allows attachment of a modified 6xHisTag followed by a TEV protease cleavage site onto the N-terminus of the expressed protein). Colonies were screened for presence of the insert and a representative plasmid was designated pSpUP.THT. The PCR-derived DNA was sequenced and shown to contain no errors.

Standard methods were used for DNA restriction endonuclease digestion, ligation and transformation of DNA (*16, 17*). Automated DNA Fluorescence Sequencing was performed at the Cornell BioResource Center. Plasmid DNA was purified with the GeneJet miniprep kit (Fermentas, Glen Burnei, MD). DNA fragments were separated by agarose gel electrophoresis, excised and purified with the Zymoclean gel DNA recovery kit (Zymo Research, Orange, CA). *E. coli* strain MachI (Invitrogen, Madison, WI) was used as a recipient for transformations during plasmid construction and for plasmid propagation and storage. An Eppendorf Mastercycler and Phusion DNA polymerase (New England Biolabs, Ipswich, MA) were used for

PCR. All restriction endonucleases and T4 DNA ligase were purchased from New England Biolabs (Ipswich, MA). *E. coli* strain BL21(DE3) and the pET overexpression system were purchased from Novagen (Madison, WI).

Expression and purification of native SpUP. The plasmid described above was transformed into expression strain BL21(DE3) *E. coli* cells. An overnight culture of 10 mL was grown in lysogeny broth media at 37°C supplemented with 50 μg/mL kanamycin, and then introduced into 1 L volume containing 50 μg/mL kanamycin. The 1L cell culture was shaken at 37°C and 200 rpm until the OD_{600} reached 0.6, at which point the temperature was reduced to 15°C and 1 mM of isopropyl-1-β-D galactopyranoside (IPTG) was added. Approximately 16 hours later, the cells were harvested by centrifugation at 7459 *g* for 20 minutes. The pellet was stored at –80 °C.

The frozen cell pellet was thawed overnight at 4°C. Approximately 30 mL of lysis buffer (20 mM Tris pH 8, 10 mM imidazole, and 300 mM NaCl) were added to the pellet to solubilize it. The cell suspension was sonicated and then centrifuged at 47, 488 *g* for 1 hour at 4°C to remove the cell debris. The clarified lysate was loaded onto a pre-equilibrated Ni-NTA gravity column, after which the column was rinsed with 20 column volumes of the same lysis buffer. SpUP was then eluted with 10 mL of elution buffer (20 mM Tris pH 8, 300 mM imidazole, 10% glycerol, and 300 mM NaCl). The eluted protein was loaded directly onto a size exclusion column (Pharmacia G200 superdex, hiload) for further purification. The protein fractions from the column were pooled together and concentrated to 25-30 mg/mL using an Amicon Ultra centrifugal filter. The SpUP concentration was determined by the method of Bradford (*16*). The protein was confirmed to be at least 95% pure by SDS-PAGE. The pure protein was buffer exchanged into 20 mM Tris pH 8, 50 mM NaCl, and 1 mM DTT, flash frozen with liquid nitrogen and stored at –80 °C.

Crystallization experiments. The frozen protein was thawed at 4°C and

incubated for 12 hours with 8 mM of either Urd plus ammonium sulfate (A.S.). in Tris buffer pH 8 or 8 mM R1P in Tris buffer pH 8 with 2 mM of Ura in 5% DMSO. Co-crystallization trials were initially carried out by vapor diffusion hanging drop method at 22°C, using sparse matrix screening solutions (Hampton Research, Emerald Biosystems). 1 µl of the protein solution was combined with an equal volume of well solution. The R1P/Ura/SpUP complex crystallized in 0.1 M sodium citrate pH 5.2, 18% (w/v) PEG 4000, and 16% isopropanol. The sizes of these crystals vary from (0.03 mm x 0.1 mm x 0.2 mm) to (0.1 mm x 0.2 mm x 0.45 mm). For the Urd/A.S./SpUP complex, crystals grew in 0.1 M Hepes, pH 7.8, 20% isopropanol, and 18% (w/v) PEG 4000. The crystal sizes vary within the same range as those of the product bound complex. The crystals for both complexes, which were plate like or blocky in shape took 3-7 days to grow. Although crystals were grown for other ranges of pH or other precipitant conditions, these crystals were either very small or grew in clusters due to fast nucleation. Since one of the precipitant reagents, isopropanol, is very volatile, it was later replaced by 2-methyl-2,4-pentanediol. As a result, the crystals became much smaller and lower in quality. Other crystallization conditions gave needles or needle clusters. Nevertheless, the electron density for the uridine of the Urd/A.S. complex was very poor. Therefore, crystals from a different crystallization condition, (0.1 M Tris pH 7.3 and 31% (w/v) PEG 1000), were soaked with 8 mM of uridine.

Data Collection and Processing. Both of the product-bound (R1P/Ura) and the substrate-bound (Urd) complexes were collected at the 24-ID-C station of NE-CAT beam line at the Advanced Photon Source (APS) at Argonne National Laboratory using an ADSC (Area Detector system Corp.) Quantum 315 detector. The data were collected at 0.97849 Å over 360° using a 0.5° oscillation range. A cryoprotectant solution of 10% glycerol in the mother liquor was used for the crystals

of the uridine complex. All the data sets were indexed, integrated, and scaled using the HKL2000 program suite (*18*). The data collection statistics are shown in Table 2.1.

Structure Determination. Using the EcUP structure (1K3F) (*19*) as the search model, the SpUP structure was solved by molecular replacement using REFMAC5 program of the CCP4 program suits (*20*). CHAINSAW in the CCP4 suite was used to remove all the side chains of the molecular replacement. After one around of tight restrained refinement, the R_{factor} and R_{free} dropped significantly to 32.4% and 34.7%, respectively. In subsequent rounds of refinements, the restraints were gradually relaxed. Most of the side chains were built in during these rounds of refinement. Further refinements were performed in CNS (*21*), starting with rigid body refinement, stimulated annealing, B-factor refinement, and finally energy minimization. Difference Fourier map, $1F_o$-F_c, $2F_o$-F_c, and composite omit maps were calculated from models with all ligands removed after each round of refinement. Manual model building was done successively using COOT (*22*) and the NCS-averaged composite omit map to minimize model bias. Only after the R_{factor} and R_{free} converged were water molecules added. The ligands were modeled into the active site of each chain of their corresponding SpUP complexes. The ligands were generated using PRODRG (*23*). The geometry of SpUP was validated using PROCHECK (*24*). Figures were generated using Pymol and ChemBioDraw (*25*). The refinement statistics are tabulated in Table 2.2.

Table 2.1. Data Collection Statistics for Uridine Phosphorylase Complexes.

	Urd	R1P/Ura
beamline	APS NE-CAT 24-ID-C	APS NE-CAT 24-ID-C
wavelength (Å)	0.97849	0.97849
resolution (Å)	2.2	1.8
space group	$P2_1$	$P1$
molecules / a.s.u.	1	3
a (Å)	90.9	90.0
b (Å)	81.7	91.7
c (Å)	123.4	169.3
α (°)	90.0	78.5
β (°)	99.2	82.3
γ (°)	90.0	60.1
measured reflections	220163	793731
unique reflections[a]	82371 (8372)	404352 (16027)
average I / σ[a]	13.5 (2.9)	19.3 (4.1)
redundancy[a]	2.7 (2.6)	2.0 (1.9)
completeness (%)[a]	91.2 (93.6)	94.6 (74.7)
Rsym (%)[a,b]	9.5 (37.5)	4.9 (20.5)

[a]Values for the highest-resolution shell are given in parentheses.
[b]Rsym = $\Sigma\Sigma_i \mid I_i - \langle I \rangle \mid / \Sigma \langle I \rangle$, where $\langle I \rangle$ is the mean intensity of the N reflections with intensities I_i and common indices h,k,l.

Table 2.2. Data Refinement Statistics for Uridine Phosphorylase Complexes

	Uridine	Ura/R1P
resolution (Å)	2.2	1.8
no. of protein atoms	10915	33759
no. of ligand atoms	102	396
no. of water atoms	344	2750
reflections in working set	78069	383726
reflections in test set	4249	20277
R-factor[a] (%)	19.6	17.7
R_{free}[b] (%)	22.4	20.3
rmsd from ideals		
bonds (Å)	0.007	0.005
angles (°)	1.328	1.283
avg B factor (Å^2)	25.5	12.0
Ramachandran Plot		
most favored (%)	90.4	92.4
additionally allowed (%)	8.8	7.1
generously allowed (%)	0.8	0.5
disallowed (%)	0	0

[a]R-factor = Σ_{hkl}| $|F_{obs}| - k |F_{cal}|$ | / $\Sigma_{hkl} |F_{obs}|$ where F_{obs} and F_{cal} are observed and calculated structure factors, respectively.
[b]For R_{free}, the sum is extended over a subset of reflections (5%) excluded from all stages of refinement.

Section 2.3 Results

Quaternary Structure of SpUP. As expected from its 39% sequence identity with EcUP, the biological unit of SpUP is a hexamer having 32 symmetry (Figure 2.3a). The toroidal hexamer has a diameter of approximately 100 Å and a thickness of 40 Å. The hexamer of SpUP also has a central channel of 18 Å in diameter. To date all the known prokaryotic members of the NP-1 superfamily have the hexameric quaternary structure. The oligomeric state is also confirmed by the size exclusion chromatography data, which estimated the molecular weight of the SpUP oligomer to be approximately 167 kDa, the weight of six SpUP protomers. Three hexamers are found in the asymmetric unit of the R1P/Ura complex and one hexamer for the Urd complex. While each protomer contains an active site, the active site is located at a dimer interface and contains two residues from the adjacent monomer. Thus, the minimal functional unit is a dimer. The distance between two active sites, which are located on opposite faces of the hexamer, is about 21 Å. The interactions between each dimer of the hexamer are primarily hydrophobic. The 170-182 loop contains several residues that participate in polar dimer-dimer interactions.

Structure of SpUP Protomer. As in the case of EcUP (*26*), the SpUP protomer shown in Figure 2.3b adopts an α/β fold. The core contains a mixed, twisted eight-stranded β-sheet in the following orientation: $\beta_2\uparrow\beta_3\downarrow\beta_4\uparrow\beta_1\uparrow\beta_5\uparrow\beta_{10}\uparrow\beta_8\downarrow\beta_6\downarrow$. The β-sheet is flanked on both sides by six α-helices with α_1, α_4 and α_9 on one side and α_2, α_3 and α_7 on the other. β_2, β_5 and β_{10} are relatively longer than the rest of the β strands, containing nine, ten, and eight residues, respectively. Alpha helices, α_5, α_6 and α_8 as well as a 3_{10} helix (η_1) are clustered on one side of the active site. Four α helices, α_2, α_4, α_6, and α_9 are much longer than the rest. α_9 capping the C-terminus, is the longest consisting of 17 residues whereas helices α_2, α_4, and α_6 contain 11, 13, and

13 residues, respectively. The loop between α_8 and α_9, residues 234-238, is generally disordered. The accompanying topology diagram is shown in Figure 2.3c.

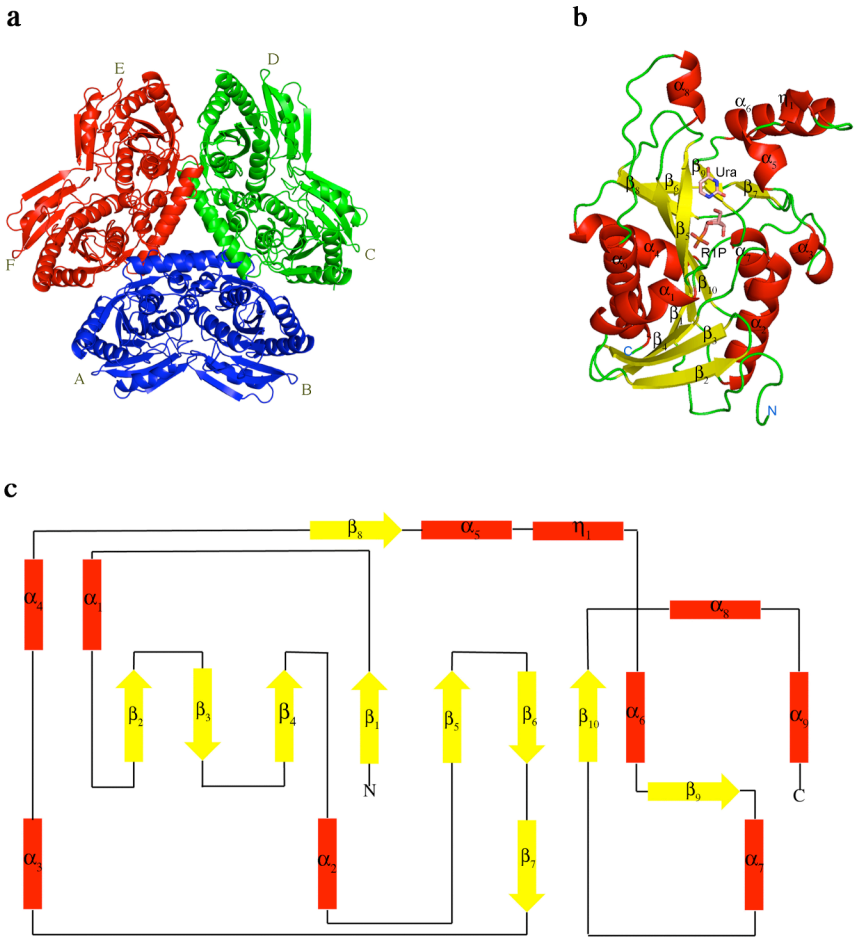


Figure 2.3. Structures of SpUP. (a) Hexameric quaternary structure of SpUP color-coded to emphasize the 32 point symmetry or a trimer of dimers. Six subunits are labeled A to F. (b) SpUP protomer with the bound products, R1P and Ura, shown in the active site. The helices are labeled red, the strands yellow, and the loops green, respectively. The N-terminus and the C-terminus are also indicated with letters N and C. (c) Topology diagram of SpUP protomer. The colors of the helices and strands are the same as Figure 2.1b, except the loops is colored as black lines. The beginning and end residues of each secondary structure are indicated by numbers.

Phosphate Binding Site. The stereo figure of the R1P/Ura complex in Figure 2.4 reveals the details of the active site at 1.82 Å resolution. The most significant

feature of the phosphate binding site is the presence of three highly conserved arginine residues, Arg33, Arg51, and Arg 94. Arg51 comes from a neighboring monomer. The three arginine residues hold the phosphate moiety of R1P in place with a strong ionic network. Two other conserved residues, Gly29 and Thr97, also hydrogen bond to the phosphate group. In addition, one of the oxygen atoms of the phosphate ion closest to the ribose also forms a hydrogen bond to O3′ of the sugar.

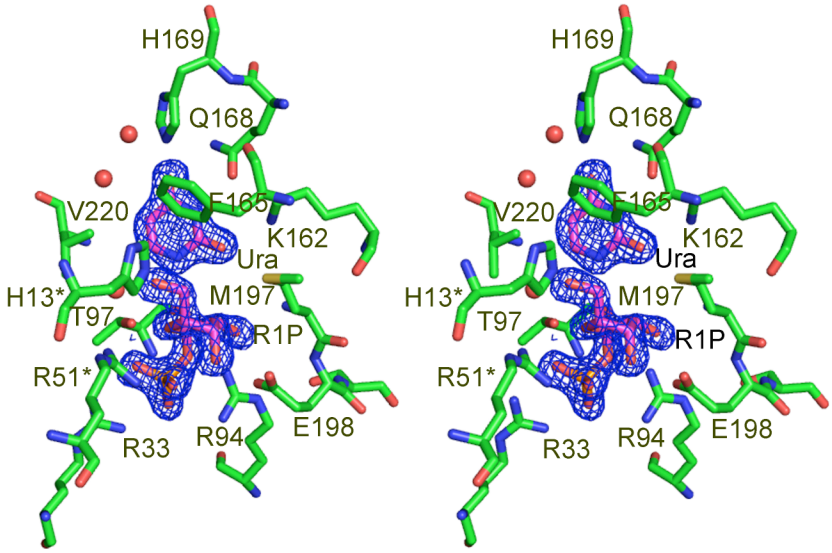


Figure 2.4. Stereo diagram of the active site containing R1P and ura inside the 2 F_o-F_c map contoured at 2.5 σ. The red spheres represent water molecules.

Ribose Binding Site. Figures 2.4 and 2.5 show the ribose binding site containing another conserved residue His13 from the adjacent monomer. This residue hydrogen bonds to the hydroxyl group of C5′ from the upper face, or beta face, of the ring. Thr97 makes a second hydrogen bond with the O4′ of the ribose ring. The hydroxyl group at C2′ forms two hydrogen bonds to Arg94 and Glu198, which involve the main interactions with the sugar by forming a bidentate with the 2′ and 3′ hydroxyl groups. The sulfur atom of the highly conserved Met197 interacts hydrophobically

with the hydrophobic beta face of the ribose, perhaps positioning the substrate in the active site. In addition, the main-chain nitrogen of Met197 hydrogen bonds to the 2′ hydroxyl group. A conserved water molecule nearby also hydrogen bonds to the O4′ of the sugar. The water molecule in the Urd complex shown in Figure 2.5 is not conserved. In the absence of the phosphate substrate, the water masks the position of one of the oxygen atoms of the phosphate ion.

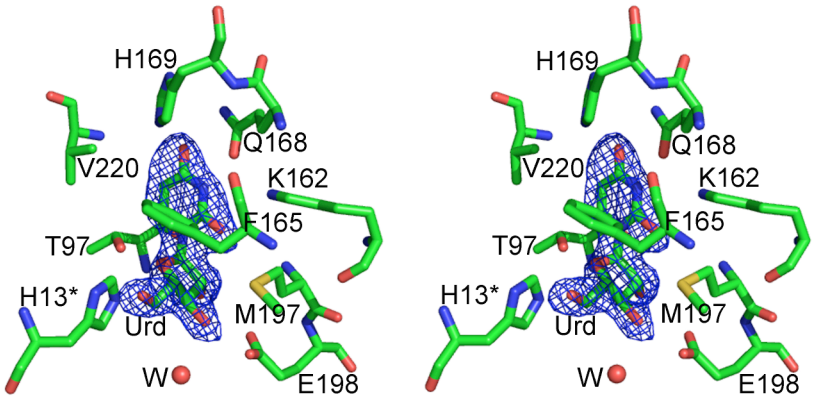


Figure 2.5. Stereo diagram of SpUP in complex with the substrate uridine showing the 2 F_o-F_c electron density map contoured at 2.5 σ.

Uracil Binding Site. It is within the base binding site that we observe His169 and Lys162 in the active site. These residues have not been observed in any known structures of the UP superfamily. His169 is located approximately 3 Å away from the O4 of the uracil and approximately 60° above the plane of the uracil ring. A Lys162 residue hydrogen bonds to O2 and N3 of the uracil. Gln168 is strictly conserved among all the known NP-I structures and makes hydrogen bonds with the N3 hydrogen and O4 of the uracil base. Phe165 forms a herringbone interaction with the aromatic ring of the uracil. Val220 interacts with the hydrophobic portion of the uracil ring.

Uridine Phosphorolysis Assay. Based on the crystal structure of SpUP, several point mutations were made to probe the critical residues in the base-binding region of the active site. Lys162 was mutated to alanine, and His169 was mutated to alanine, asparagine, and aspartate. In addition, Val220 was mutated to aspartate and glutamate to see if the enzyme could accommodate a purine base in the presence of these charged groups. The results are shown in Table 2.3.

Table 2.3. Uridine Phosphorolysis Assay Results for SpUP and Mutants

SpUP	**Uridine cleavage activities (nmoles / mg / hr)**
Wildtype	1,500,000
K162A	20
H169A	112,000
H169D	9,000
H169N	42,000
V220D	550
V220E	3,400

Section 2.4 Discussion

Overall structure of SpUP. The biological unit of SpUP is a hexamer as expected for a prokaryotic uridine phosphorylase (Figure 2.3a). Only one hexamer is present in the asymmetric unit of the Urd complex. However, three hexamers are

found in the asymmetric unit of the R1P/Ura complex. One of these hexamers is slightly twisted in relation to the others such that the packing results in the lowest symmetry P1 space group (Table 2.1). Since each active site of the hexamer contains two conserved residues from an adjacent monomer at each dimer interface, the minimal functional unit is a dimer.

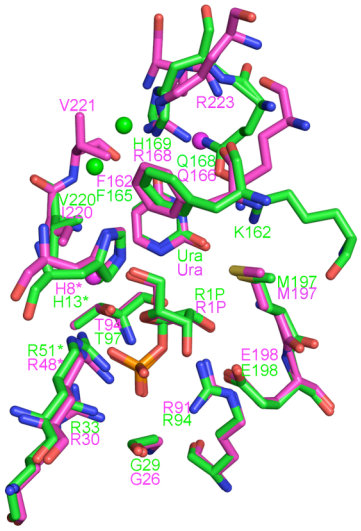


Figure 2.6. Superposition of the active sites of SpUP (green) and EcUP (magenta).

The SpUP protomer superposes very well on the EcUP protomer, especially in the active site as shown in Figure 2.6. This nearly identical superposition of the active sites is interesting considering only 39% sequence identity between the two UPs. Furthermore, like EcUP, SpUP adopts the same universal monomeric α/β fold of all known nucleoside phosphorylases. Hence, SpUP most likely possesses the same mechanism of phosphorolysis, in which the uridine undergoes a unimolecular substitution via the formation of a high-energy ribosyloxocarbenium ion intermediate. Residues 170-182 including the 3_{10} helix (η_1) and part of α_6 constitute the signature loop (*27*), which allows only nucleosides containing one-ring bases such as uridine, thymidine, and their derivatives to bind to the active site. The primary sequence

alignment demonstrating this so-called specificity loop is shown in Figure 2.7. As shown in the figure, SpUP has two less residues in the specificity region compared with EcUP. However, this loop is missing in EcPNP. Thus, the space occupied by this loop allows PNP to accommodate large-ring purine bases.

Uridine Phosphorolysis Assay. As seen in Table 2.4, the most noticeable feature is the activity of K162A mutant, which is near the assay's detection limit. Clearly, Lys162 is required for SpUP activity. However, the enzyme still maintains a low level of activity with histidine mutants that do not carry a negative charge, H162A and H162N. Hence, His169 is not as critical as Lys162. Furthermore, the presence of a negative charge at Val220 is also detrimental to the enzyme as seen in V220D and V220E mutants. These mutational data suggest that during the transition state, the negative charge delocalizes over C2, C5, and C6 of the uracil ring as expected. This negative charge repels any nearby negatively charged residues, and destabilizes or ultimately inhibits the phosphorolysis reaction. Accordingly, positively charged residues are predicted to enhance the rate of the phosphorolysis, or at least increase the binding affinity of the enzyme for the substrate. However, this coulombic interaction in the active site of SpUP would require a delicate balance because too strong of an electrostatic interaction would hinder the product release step. Accordingly, the reaction would then prefer to shift toward the synthesis direction.

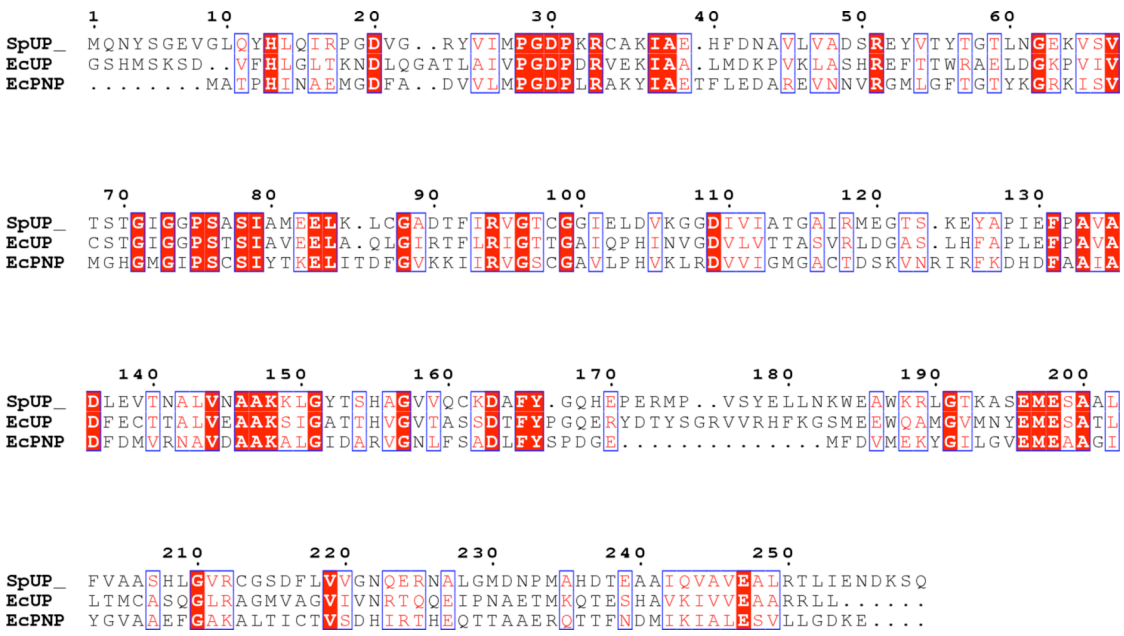


Figure 2.7. Specificity region (residues 170-182) of SpUP is revealed by the alignment of SpUP against EcUP and EcPNP.

Comparison of the Active Sites of SpUP and EcUP. Having the same fold as that of EcUP (*8*) is translated into nearly identical active site residues in this particular case, despite sharing only 39% sequence identity (Figure 2.6). The structure of EcUP crystallized with R1P/Ura (1TGY) was used for comparison. As seen in Figure 2.4, the active sites and, in particular, the ligands of the two structures superimpose very well. The individual active-site residue comparison is shown in Table 2.4. In addition, the active site of EcUP has an equivalent arginine (Arg48) and a histidine (His8) from the neighboring monomer, which correspond to Arg51 and His13 of SpUP. Single-site mutagenesis studies by Oliva *et. al.* (*28*) showed that the His8Ala mutation in EcUP reduced the enzyme's activity to 21% compared to the wild type (Table 2.5). This mutation indicates that His8 or His13 in SpUP binds and stabilizes the ribose moiety in addition to its role in subunit interaction.

Table 2.4. Comparison of active site residues between EcUP and SpUP.

	EcUP	**SpUP**
Phosphate	Gly26	Gly29
	Arg30	Arg33
	Arg48*	Arg51*
	Arg91	Arg94
Ribose	His8*	His13*
	Arg91	Arg94
	Thr94	Thr97
	Met197	Met197
	Glu198	Glu198
Uracil	--------	Lys162
	Phe162	Phe165
	Gln166	Gln168
	Arg168	His169
	Ile220	--------
	Val221	Val220
	Arg223	--------

* = residues from the neighboring monomer

As indicated in Table 2.4 and Figure 2.6, the two base binding sites differ to some extent. SpUP has conserved residue Gln168 which hydrogen bonds with N3H and O4 of the uracil base. Mutation of its equivalent, Gln166 in EcUP, to alanine inactivated the enzyme (Table 2.5). However, the Gln166Asp mutation lowered the activity to 30% of the wild type while Gln166Ala mutation completely inactivated the enzyme. This substitution demonstrates that Gln166 in EcUP or Gln168 in SpUP is required for hydrogen bonding to the base. Phe165 of SpUP, similar to Phe162 in EcUP, forms a herringbone stacking interaction with the aromatic ring of the uracil. The Phe162Ala substitution in EcUP abolished the activity indicating that Phe162 (Phe165 in SpUP) is required to position and stabilize the base. Val220 interacts with the nonpolar side of the uracil ring. No equivalent of Ile220 in EcUP exists in SpUP. Nevertheless, two water molecules are present nearby, which most likely mark the solvent path leading to the surface of the protein. Both Glu198Asp and Glu198Gln mutations completely inactivate EcUP (Table 2.5). These data imply that hydrogen bonding to the hydroxyl groups of the sugar cannot account for the function of Glu198 in either EcUP or SpUP. Thus, the correct chain length is critical and the negative charge is required for catalysis. These mutations clearly support the hypothesis for the transition state, which was proposed to contain a continuum of electrostatic interactions for both EcUP and SpUP, as shown in Figures 2.8a and 2.8b.

Table 2.5. Characterization of single amino acid mutants of uridine phosphorylase.

Expressed proteins[a]	Activity (U/mL extract)	Total protein (mg/ml)	Specific activity (U/mg total proteins)	Percent of wild type activity
Wild-type UP	540	24	22.5	100
--	0	12	0.0	0
His8Ala	150	31	4.8	21
Arg30Ala	5	15	0.3	1
Arg30Lys	17	14	1.2	5
Ile69Ala	900	36	25.0	111
Arg91Ala	0	13	0.0	0
Arg91Lys	11	23	0.5	2
Thr94Ala	5	14	0.4	2
Phe162Ala	0	33	0.0	0
Tyr195Ala	378	14	27.0	120
Tyr195Gly	5	12	0.4	2
Gln196Ala	0	37	0.0	0
Gln196Asp	84	11	7.6	34
Met197Ala	150	37	4.1	18
Met197Ser	32	14	2.3	10
Glu198Gly	0	23	0.0	0
Glu198Asp	0	13	0.0	0
Glu198Gln	0	32	0.0	0

[a]The activity of wild-type UP produced by pGM679 is reported. UP produced by pGM870 showed less than 2% difference. The negative control is pUC18.

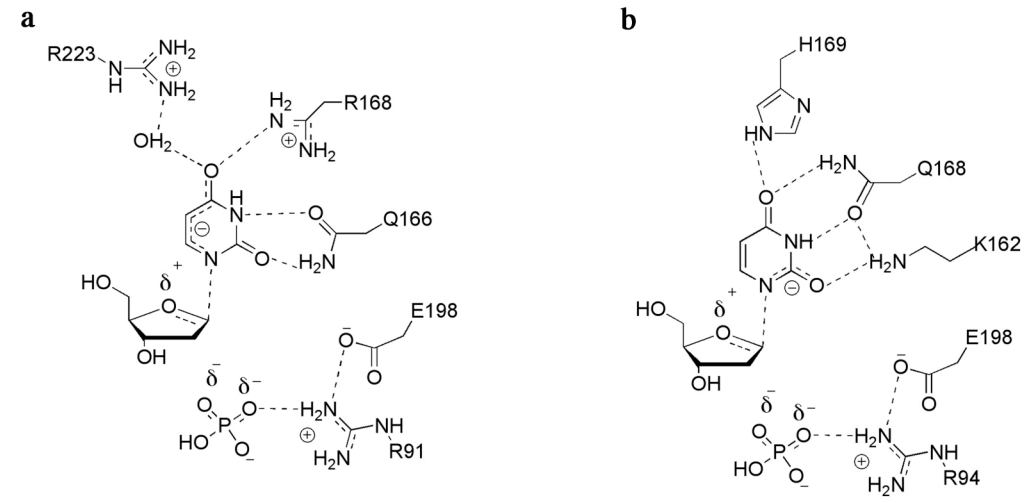


Figure 2.8. Transition state stabilization of high energy intermediates of the phosphorolysis reaction in EcUP (a) and SpUP (b) by a continuum of electrostatic and hydrogen bond interactions as illustrated by the dotted line.

The most obvious difference between the two active sites is the superposition of His169 of SpUP with Arg168 of EcUP. In addition to this residue, Lys162 hydrogen bonds to O2 of the uracil base. These two residues that directly interact with a base have not been observed in any known UP structures. This raises two questions. Are these two residues conserved, and if so what role do they play in the transition state of the phosphorolysis reaction? The fact that Lys162 hydrogen bonds to O2 of the uracil suggests that the negative charge resulting from the glycosidic bond cleavage may reside mostly on the O2. Arg168 in EcUP was proposed to be involved in the binding to the uracil base (*27*). The authors suggested that Arg168 might control the substrate specificity. The proposed transition state for the phosphorolysis reaction of EcUP is shown in Figure 2.8a. Similarly to Arg168, the imidazole ring of His169 in the active site of SpUP probably possesses a positive charge and plays the same role as Arg168 by controlling the substrate specificity (Figure 2.8b). The imidazole ring of His169 cannot form a hydrogen bond with the O4 of the uracil to

any appreciable extent due to its unfavorable geometry at approximately 60 degrees above the plane of the uracil ring. Arg168 in EcUP is actually more able to make the hydrogen bond due to its greater degree of freedom of the sp^3 hybridized nitrogen. Instead, His169 most likely interacts electrostatically with the O4 of the uracil. However, the electrostatic interaction of His169 would be weaker than Arg168 of EcUP because the developing negative charge presumably delocalizes primarily on the O2 of the uracil ring in SpUP, whereas the negative charge resides primarily on the O4 of the uracil ring in EcUP. If this electrostatic stabilization hypothesis is correct, then a R168E mutation in EcUP or a H169E mutation in SpUP would severely destabilize the transition state and thus would inactivate the corresponding enzyme due to negative charge repulsion between the new glutamate side chain and the developing negative charge on the uracil base. This hypothesis is supported by the results of the uridine phophorolysis assay (Table 2.3) for the His169 mutants. The H162D mutant exhibits at least an order of magnitude lower activity than the H162A mutant. In addition, the R168A mutation in EcUP and the H169A mutation in SpUP are expected to be less detrimental than the corresponding R168E mutation in EcUP and H169E mutation in SpUP due to the lack of negative charge repulsion. Nevertheless, the R168A mutation is expected to be more detrimental to EcUP than the equivalent H169A mutation in SpUP because the negative charge resides mostly on the O4 of the uracil ring in EcUP. Therefore, Lys162 is, in all probability, the critical residue in stabilizing the transition state and perhaps donates a hydrogen atom to the leaving uracil base.

Caradoc-Davies and colleagues (*27*) also speculated that the reason that EcUP does not cleave cytidine is because of the assumed positive charge on N4 of the cytosine base will be repelled by Arg168. This hypothesis seems less plausible because the pKa of the N4 of cytidine is 4.2. At physiological pH, the N4 is unlikely

to carry a positive charge unless the local pH drops below 4 during catalysis. Incidentally, this would also support the protonated form of the aspartic acid residue found in the base binding pocket of most structurally characterized NP-I enzymes. Hence, a more conclusive explanation for the lack of reactivity of cytidine remains to be discovered.

Another interesting feature regarding the active site of SpUP is observed based on structural superposition of the R1P/Ura complex, Urd complex, and sulfate ion complex (not shown). Assuming that the product bound complex takes on the closed conformation, the active site is in the closed form upon the binding of any of either the phosphate or Urd or R1P/Ura. In other words, the active site must be occupied by at least one ligand in order for SpUP to assume a closed conformation.

Significance of the SpUP Structure. Comparing of the active sites of SpUP and EcUP suggests that the two enzymes have the same transition state. However, SpUP utilizes different residues to stabilize the negative charge on the base. Hence, the next question to ask is whether this difference is unique in SpUP or it also exists in other species. A multiple sequence alignment of SpUP and EcUP with sequences found using a BLAST search against the nonredundant database was performed using CLUSTAL W (*13*) is shown in Figure 2.9. Among all the representative species in the alignments, 22 out of 53 sequences contain the Lys and His residues equivalent to K162 and H169 of SpUP. None of these have the equivalents of Arg168 and Arg223 of EcUP. The other 31 sequences, including EcUP, show the reverse. These 31 enzymes possess the two equivalent arginine residues, but none of them have the equivalent K162 and H169 of SpUP. The phylogenetic tree of all the sequences from the alignments is shown in Figure 2.10. With the exception of three species, *Sebaldella termitidis*, *Desulfonatronospira thiodismutans*, and *Francisella tularensis*, all the species colored red, or SpUP-like species, are clustered together. The same is

true for those in blue, or EcUP-like species. This suggests a new way of clustering of UP's.

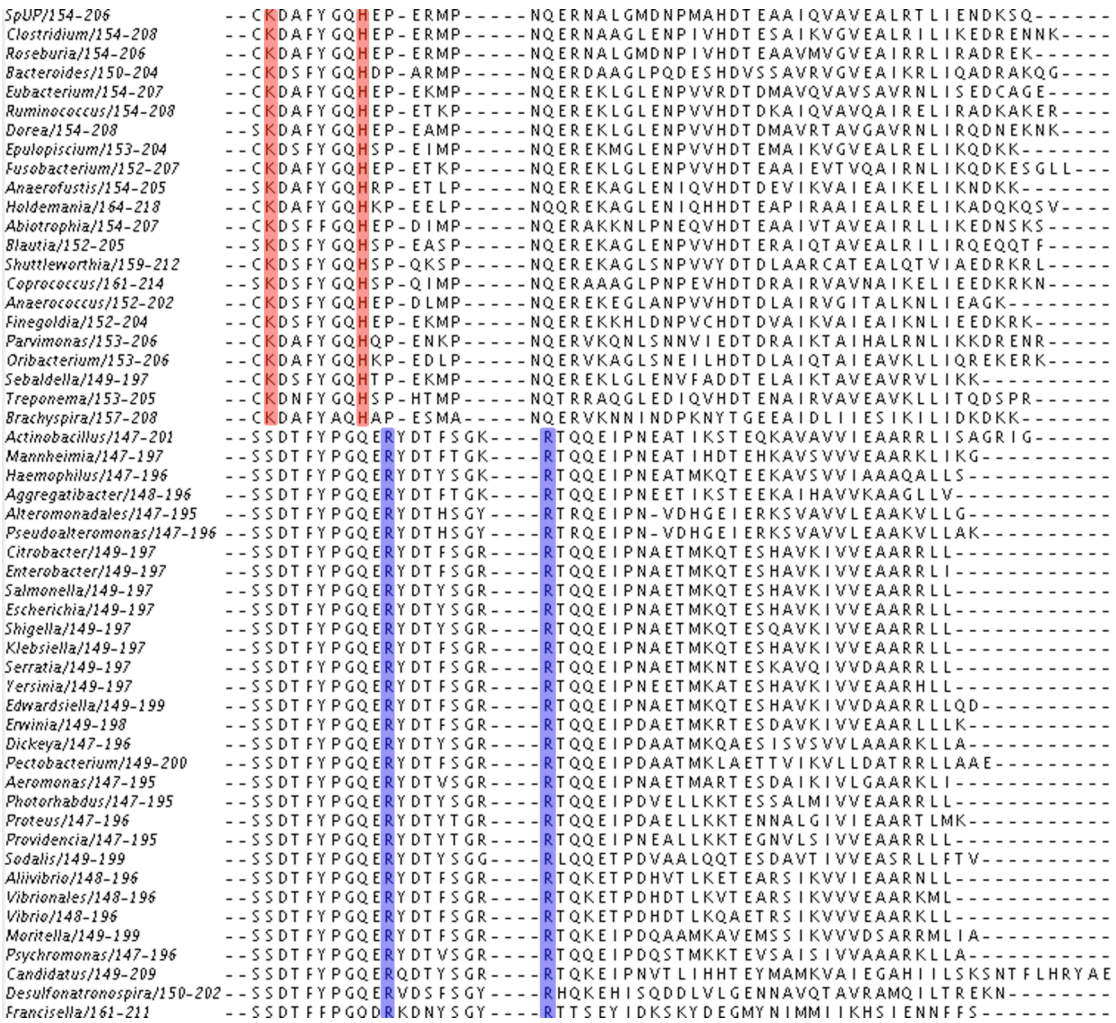


Figure 2.9. Multiple sequence alignment of uridine phosphorylase superfamily. The amino acids highlighted in red and blue are critical active site residues found in SpUP and EcUP, respectively.

These structural studies of SpUP led to the discovery of a subclass of UP's that utilize different active-site residues to stabilize the transition state. Furthermore, the two discriminating residues, Lys162 and His169, together with UP's specificity loop

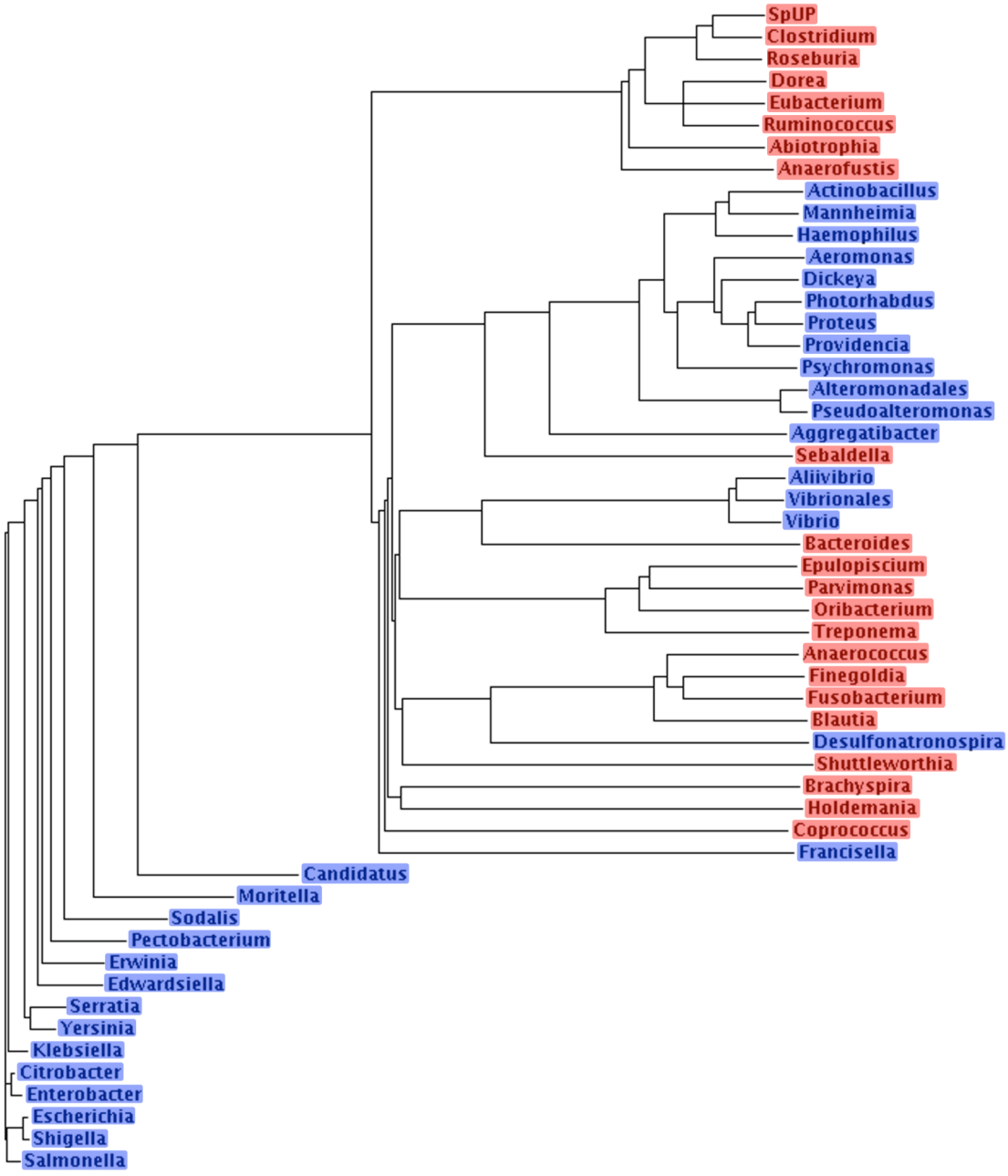


Figure 2.10. Phylogenetic tree of uridine phosphorylase superfamily from Figure 2.6 applying the same color-coding scheme.

may be used to distinguish between UP and PNP enzymes at the level of primary sequence, and thus may allow proper annotation.

REFERENCES

1. Leer, J. C., Hammer-Jespersen, K., and Schwartz, M. (1977) Uridine phosphorylase from *Escherichia coli*. Physical and chemical characterization, *Eur. J. Biochem. 75*, 217-224.
2. Zhang, Y., Morar, M., and Ealick, S. E. (2008) Structural biology of the purine biosynthetic pathway, *Cell Mol. Life Sci. 65*, 3699-3724.
3. Hassan, H. F., Coombs, G. H. (1988) Purine and Pyrimidine Metabolism in Parasitic Protozoa, *FEMS Microbiology Reviews 4*, 47-83.
4. Schnick, C., Robien, M. A., Brzozowski, A. M., Dodson, E. J., Murshudov, G. N., Anderson, L., Luft, J. R., Mehlin, C., Hol, W. G. J., Brannigan, J. A., Wilkinson, A. J. (2005) Structures of *Plasmodium falciparum* Purine Nucleoside phosphorylase Complexed with Sulfate and its Natural Substrate Inosine, *Acta Cryst. D 61*, 1245-1254.
5. Morgunova, E., Mikhailov, A. M., Popov, A. N., Blagova, E. V., Smirnova, E. A., Vainshtein, B. K., Mao, C., Armstrong Sh, R., Ealick, S. E., Komissarov, A. A., and et al. (1995) Atomic structure at 2.5 Å resolution of uridine phosphorylase from *E. coli* as refined in the monoclinic crystal lattice, *FEBS Lett. 367*, 183-187.
6. Pugmire, M. J., and Ealick, S. E. (2002) Structural analyses reveal two distinct families of nucleoside phosphorylases, *Biochem. J. 361*, 1-25.
7. Niedzwicki, J. G., el Kouni, M. H., Chu, S. H., and Cha, S. (1983) Structure-activity relationship of ligands of the pyrimidine nucleoside phosphorylases, *Biochem. Pharmacol. 32*, 399-415.
8. Bu, W., Settembre, E. C., el Kouni, M. H., and Ealick, S. E. (2005) Structural basis for inhibition of *Escherichia coli* uridine phosphorylase by 5-substituted

acyclouridines, *Acta Crystallogr. D Biol. Crystallogr. 61*, 863-872.

9. Longley, D. B., Harkin, D. P., Johnston, P. G. (2008) 5-Florouracil: Mechanisms of Action and Clinical Strategies, *Nature 3*, 330-338.

10. Federov, A., Shi, W., Kicska, G., Tyler, P. C., Furneaux, R. H., Hanson, J. C., Gainsford, G. J., Larese, J. Z., Schramm, V. L., and Almo, S. C. (2001) Transition State Structure of Purine Nucleoside Phosphorylase and Principles of Atomic Motion in Enzymatic Catalysis, *Biochemistry 40*, 853-860.

11. Paul, D., O'Leary, S. E., Rajashankar, K., Bu, W., Toms, A., Settembre, E. C., Sanders, J. M., Begley, T. P., and Ealick, S. E. (2010) Glycal formation in crystals of uridine phosphorylase, *Biochemistry 49*, 3499-3509.

12. Ryan, K. G. (2003) Sherris Medical Microbiology: An Introduction to Infectioius Diseases, 1-992.

13. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res. 22*, 4673-4680.

14. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res. 25*, 3389-3402.

15. Zhang, Y., Parker, W. B., Sorscher, E. J., Ealick, S. E. (2005) PNP Anticancer Gene Therapy, *Current Topics in Medicinal Chemistry 5*, 1259-1274.

16. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Vol. 3, Cold Spring Harbor Laboratory Press, Plainview, New York.

17. Ausubel, F. M., and Brent, F. (1987) *Current Protocols in Molecular Biology*,

John Wiley and Sons, New York.

18. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, *Methods Enzymol. 276*, 307-326.

19. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank, *Nucleic Acids Res. 28*, 235-242.

20. Collaborative Computational Project-Number 4. (1994) The CCP-4 suite: programs for protein crystallography, *Acta. Crystallogr. D 50*, 760-763.

21. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr. D 54*, 905-921.

22. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr. D Biol. Crystallogr. 60*, 2126-2132.

23. Schuettelkopf, A. W., van Aalten, D. M. F. (2004) PRODRG: a Tool for High-Throughput Crystallography of Protein-Ligand Complexes, *Acta Crystallogr. D60*, 1355-1363.

24. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283-291.

25. DeLano, W. L. (2002) Unraveling hot spots in binding interfaces: progress and challenges, *Curr. Opin. Struct. Biol. 12*, 14-20.

26. Burling, F. T., Kniewel, R., Buglino, J. A., Chadha, T., Beckwith, A., and Lima, C. D. (2003) Structure of *Escherichia coli* uridine phosphorylase at 2.0 Å, *Acta Crystallogr. D Biol. Crystallogr. 59*, 73-76.

27. Caradoc-Davies, T. T., Cutfield, S. M., Lamont, I. L., and Cutfield, J. F. (2004) Crystal structures of *Escherichia coli* uridine phosphorylase in two native and three complexed forms reveal basis of substrate specificity, induced conformational changes and influence of potassium, *J. Mol. Biol. 337*, 337-354.

28. Oliva, I., Zuffi, G., Barile, D., Orsini, G., Tonon, G., De Gioia, L., and Ghisotti, D. (2004) Characterization of *Escherichia coli* uridine phosphorylase by single-site mutagenesis, *J. Biochem. 135*, 495-499.

CHAPTER 3

CRYSTAL STRUCTURE OF BOVINE PURINE NUCLEOSIDE PHOSPHORYLASE IN COMPLEX WITH PYRIMIDOPURINONE 2'-DEOXYRIBOSE AND PHOSPHATE/SULFATE ION

Section 3.1 Introduction

Purine nucleoside phosphorylase (PNP; EC 2.4.2.1) is an indispensable enzyme in the purine salvage pathway that catalyzes the reversible phosphorolysis of ribonucleosides or 2'-deoxyribonucleosides to ribose or 2'-deoxyribose 1-phosphate (R1P or dR1P) and the free base (Figure 3.1).

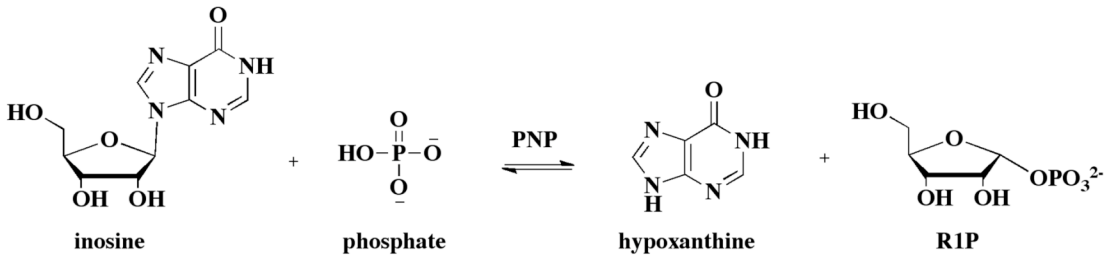


Figure 3.1. Phosphorolysis reaction catalyzed by purine nucleoside phosphorylase .

The salvage pathway is an alternative pathway to the *de novo* purine biosynthetic pathway, which undergoes approximately ten different biochemical transformations to make precursors for DNA and RNA biosynthesis (Figure 3.2). This enzyme catalyzes an important first step in all organisms and various tissue types by extracting carbon source from free purine bases and nucleosides (Parks & Agarwal, 1972, Leer *et al*., 1977). Some parasites, such as *Plasmodium falciparum*, lack the *de novo* biosynthetic pathway (Hassan, 1988, Schnick, 2005). As a result, these organisms must rely entirely on the salvage pathway as a source of nucleosides, making these salvage pathways potential antiparasitic drug targets. Furthermore, the

absence of PNP activity is linked to T-cell immunodeficiency due to the accumulation of dGTP, while other tissues remain unaffected (Giblett *et al.*, 1975, Hershfield & Mitchell, 1995). Since T-cells are known to mediate autoimmune disorders, some cancers, and tissue transplant rejection, PNP has become an important drug target.

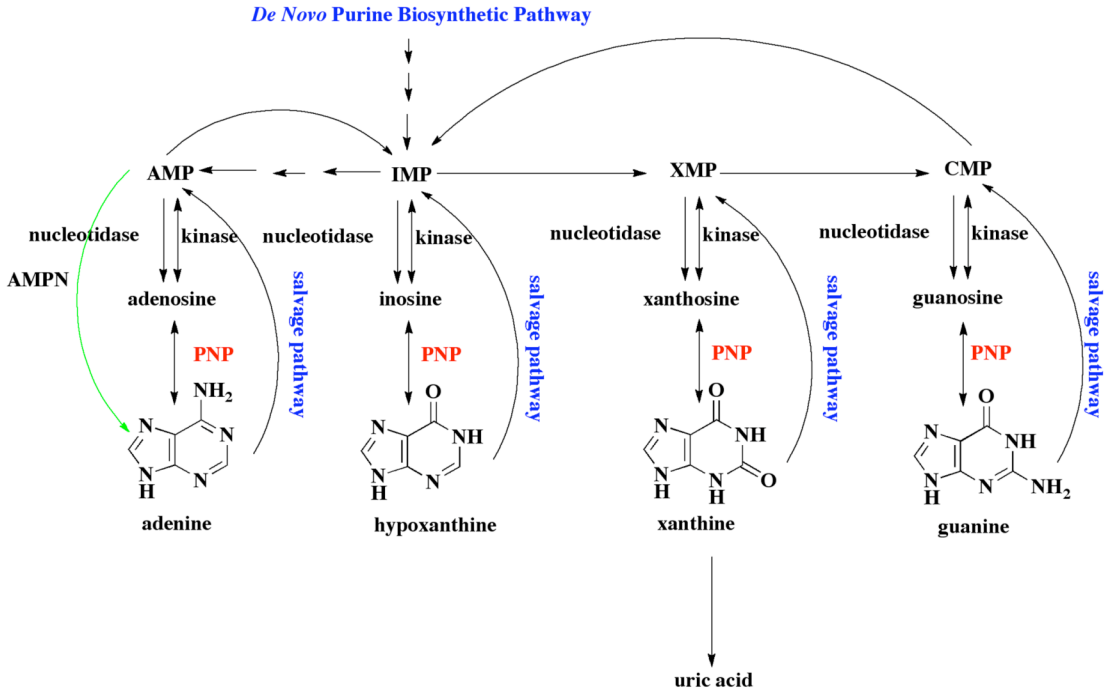


Figure 3.2. Generalized de novo purine biosynthetic and salvage pathways. AMPN stands for adenosine monophosphate nucleosidases, and AMP, IMP, XMP, CMP adenosine, inosine, xanthine, and cytosine monophosphate, respectively. The green arrow indicates that the transformation is unique to prokaryotes.

Although PNP enzymes have been studied more extensively than uridine phosphorylase (UP) enzymes, the details of the chemical mechanism and the role of active site residues is not fully understood. For example, the sulfate ion, which has been previously used as an unreactive mimic of phosphate ion (Federov *et al.*, 2001), was recently discovered to cleave uridine into two pieces in bovine UP, the ribose intermediate and the base (Paul *et al.*, 2010). Interestingly, the proposed ribose

intermediate exists as a glycal, a phenomenon that has not been observed previously in nucleoside phosphorylases.

The structure of bovine PNP (bPNP) is trimeric with molecular weight of 31 kDa (Mao *et al*., 1998). Its primary sequence is 86% identical to that of human PNP. Previous work (Ealick *et al*., 1991) suggested that in the phosphorolytic direction the order of binding of phosphate and inosine is random. The product release step, however, is sequential as R1P must be released before hypoxanthine. In addition, it was demonstrated that in the synthetic direction hypoxanthine must bind before R1P can bind.

The lability of the glycosidic bond in aqueous solvent as well as in most common organic solvents makes traditional chemical synthesis of nucleoside analogs quite challenging. This is complicated by the necessity of adding and removing protecting groups, which lead to very low yields (e.g. 2%). One way to overcome this challenge is to chemically synthesize the modified base and/or modified sugar separately and then couple the base and the sugar together enzymatically (Figures 3.3A, 3.3B). Therefore, nucleoside phosphorylases can serve as a useful synthetic tool for this purpose. Recent biochemical data (Chapeau & Marnett, 1991) indicate that the glycosidic bond of the most abundant by-product of DNA damage, M_1G-dR, could be generated using *Escherichia coli* purine nucleoside phosphorylase (EcPNP) by coupling the guanine analog, M_1G, with the sugar moiety 1 (Figure 3.3A) to give a yield of 64%. However, bPNP cannot couple the same substrates. Therefore, structures of bPNP containing M_1G-dR and phosphate or sulfate are needed in order to rationalize the observation.

Here we present two X-ray structures of bPNP in complex with M_1G-dR and phosphate or sulfate, which was initially thought to be an unreactive mimic of phosphate. A comparison of the structure of EcPNP/inosine/SO_4 complex (Bennett *et*

al., 2003) with the structures of the two bPNP complexes suggests that the hydrophilic residue Glu201 sterically constrained M_1G and induced the glycosidic bond cleavage rather than making the glycosidic bond with dR1P. This structural evidence supports our biochemical assay results in which the sugar moiety 1 (Figure 3.3) was successfully coupled with M_1G using EcPNP but not bPNP. In addition, we unexpectedly observed that the sulfate ion cleaved M_1G-dR into a glycal intermediate and the M_1G base.

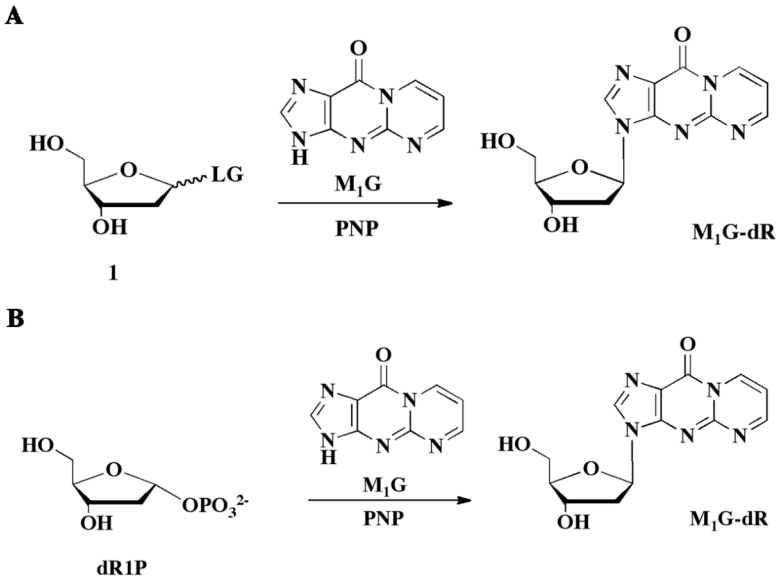


Figure 3.3. Coupling reaction to synthesize purine nucleoside analog catalyzed by PNP. LG of species 1 in panel A represents a leaving group (usually a thymine) and B the synthesis reaction where the leaving group is phosphate.

Section 3.2 Experimental

Synthesis of M_1G-dR. Pyrimidopurinone-2'-deoxyribose was synthesized according to the previously published procedures (Schnetz-Boutaud *et al.*, 2000).

Protein Purification. Partially purified purine nucleoside phosphorylase from calf spleen was purchased from Sigma Chemical Co. This protein sample was further purified using a reactive dye matrix column. All steps were performed at 4 °C unless otherwise noted. The Sigma protein was dialyzed against 10 mM MES, pH 6.0 and 1 mM dithiothreitol (DTT) (loading buffer). The dialyzed sample was then concentrated to a volume of 5 mL, and directly loaded onto a 5 mL green-19 dye matrix column pre-equilibrated in loading buffer. The column was then washed with excessive loading buffer until the eluate had no detectable protein concentration. A two-step wash gradient was performed with loading buffer supplemented with 0.4 M and 0.7 M NaCl. The protein was eluted with loading buffer supplemented with 2 M NaCl and 20 mM ATP. The eluted protein was dialyzed against 10 mM Tris, pH 7 and 1 mM DTT, and concentrated to 8 mg/mL based on a Bradford assay.

Enzyme Assay. PNP activity was measured with inosine as substrate (check above reference for reference for assay). Hyphoxanthine formed in the reaction was oxidized to uric acid by xanthine oxidase. The uric acid concentration was monitored spectrophotometrically at 293 nm to reflect the amount of PNP product concentration With 0.5 mM inosine in 50 mM potassium phosphate buffer at pH 7.5 and xanthine oxidase 0.02 μ/mL, 1 unit of PNP activity was defined as the amount of PNP that phosphorolyzes 1 μmole of inosine per minute under a temperature of 25 °C. Specific activity is expressed as units per milligram of protein.

Crystallization. The purified bovine PNP was crystallized using the hanging drop vapor diffusion method at room temperature. Inhibitors were cocrystallized with the enzyme. The reservoir solution contains 100 mM Tris buffer at pH 7.2-7.6, 10-12% polyethylene glycol 2000 monomethylether (PEG2000 MME), and 100 mM

magnesium acetate with 1 mM potassium phosphate. Each drop was set up with 2 μL of protein solution plus 2 μL of reservoir solution with additional 3mM of inhibitor. Crystals appeared from clear drop after about 24 h of equilibrating against the reservoir solution, and reached their maximum size over the course of an additional few weeks.

Data Collection. The X-ray diffraction data were collected at the Cornell High Energy Synchrotron Source (CHESS) station F1. The data were recorded with a Quantum 4 mosaic CCD-based X-ray detector made by ADSC. The detector was placed 150 mm from the sample. The crystals were flash frozen in a gaseous nitrogen stream immediately before data collection. One crystal was used to collect each data set. Each frame was measured with a 1° oscillation angle and a 20 s exposure time. 53° of 2.4 Å data for the bPNP/M_1G-dR/PO_4 complex and 60 ° of 2.4 Å data for the bPNP/M_1G-dR/SO_4 complex were collected with nearly 90% completion.

Data Processing. X-ray diffraction data were processed using the Collaborative Computational Project, Number 4 (CCP4) suite of programs (Collaborative Computational Project-Number 4, 1994) Measured intensity were converted to structure factors using program TRUNCATE from CCP4. Data collection and processing statistics are shown in Table 3.1.

Structure Determination. Using a coordinate file of a bovine PNP (4PNP) from the protein databank (Federov *et al.*, 2001) as the search model (Berman *et al.*, 2000), the bPNP complexes were solved by molecular replacement using CNS (Brünger *et al.*, 1998), starting with rigid body refinement, simulated annealing, B-factor refinement, and finally energy minimization. All the ligands and water molecules in the search model were removed. Several iterative rounds of refinement

were performed and alternated by manual model building in COOT (Emsley & Cowtan, 2004). Most of the side chains were adjusted and a few were removed during several rounds of refinement. σ_A-weighted difference Fourier maps (Read, 1986), $1F_o$-F_c and $2F_o$-F_c, as well as composite omit maps were calculated from models without ligands after each round of refinement. Manual model building was done using the composite omit map to minimize model bias. Once the R_{factor} and R_{free} converged, water molecules were added. The ligands were modelled into the active site of each chain of their corresponding bPNP complex. The ligands were generated using PRODRG (Schuettelkopf, 2004). The geometry of bPNP was validated using PROCHECK (Laskowski *et al.*, 1993). All the figures were produced using PyMOL and ChemBioDraw (DeLano, 2002). The refinement statistics are tabulated in Table 3.2.

Table 3.1. Data Collection Statistics for bPNP Complexes.

	bPNP/M_1G-dR/SO_4	bPNP/ M_1G-dR/PO_4
source	F1 CHESS	F1 CHESS
wavelength (Å)	0.916	0.916
resolution (Å)	2.4	2.4
space group	$P2_13$	$P2_13$
molecules / a.s.u.	1	1
a (Å)	92.59	92.59
b (Å)	92.59	92.59
c (Å)	92.59	92.59
α (°)	90°	90°
β (°)	90°	90°
γ (°)	90°	90°
unique reflections[a]	9139	9055
completeness (%)	86.8 (69.1)	89.7 (70.0)

[a]Values for the highest-resolution shell are given in parentheses.
[b]Rsym = $\Sigma\Sigma_i | I_i - \langle I \rangle | / \Sigma \langle I \rangle$, where $\langle I \rangle$ is the mean intensity of the N reflections with intensities I_i and common indices h,k,l.

Table 3.2. Data Refinement Statistics for bPNP Complexes.

	bPNP/M_1G-dR/SO_4	bPNP/ M_1G-dR/PO_4
resolution (Å)	2.4	2.4
no. of protein atoms	2199	2177
no. of ligand atoms	27	27
no. of water atoms	139	160
reflections in working set	8620	8820
reflections in test set	682	690
R-factor[a] (%)	23.6	23.3
R_{free} [b] (%)	27.7	28.3
rmsd from ideals		
bonds (Å)	0.009	0.008
angles (°)	1.21	1.13
avg B factor ($Å^2$)	34.0	13.3
Ramachandran Plot		
most favored (%)	89.0	89.0
additionally allowed (%)	11	11
generously allowed (%)	0.0	0.0
disallowed (%)	0	0

[a]R-factor = $\Sigma_{hkl}| |F_{obs}| - k |F_{cal}| | / \Sigma_{hkl} |F_{obs}|$ where F_{obs} and F_{cal} are observed and calculated structure factors, respectively.
[b]For R_{free}, the sum is extended over a subset of reflections (5%) excluded from all stages of refinement.

Section 3.3 Results

Overall Structure. The overall structure of bPNP in complex with ligands resembled those previously described (Mao *et al.*, 1998, Luic *et al.*, 2004, Koellner *et al.*, 1997, Luic *et al.*, 2001, Toms *et al.*, 2005). The enzyme is a trimer of three identical subunits, each of which has the typical PNP α/β fold containing a central mixed eight-stranded β sheet and a smaller five-stranded sheet (Figures 3.4A, B). The twisted β-barrel is flanked on both sides by a total of nine α-helices and two short 3_{10} helices. The fold of the bPNP complexes are highly compact with a Matthews number of 2.07 corresponding to a solvent content of 40%.

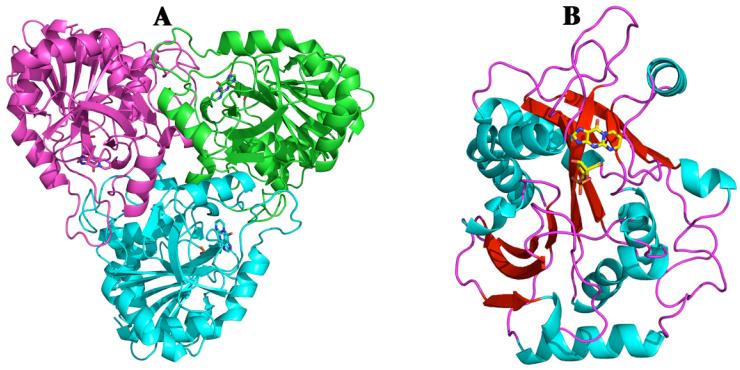


Figure 3.4. Structures of bPNP. Panel A contains the biological unit of bPNP and panel B contains the monomer found in the asymmetric unit. The ligands are represented by the stick model.

As seen in Figure 3.4, the active site is located near the monomer-monomer interface and buried deeply inside the protein fold. The active site is divided into three parts: the phosphate binding site, the sugar-binding site, and the base-binding site. The phosphate-binding pocket is very hydrophilic and contains Ser220, His86, Arg84, Hi64, Ala116, and Ser33. The β face of the sugar moiety interacts hydrophobically with M219. His257 and Tyr88 form hydrogen bonds with O5' and O3' of the sugar

respectively. Both of the phosphate and the sugar binding sites are well conserved among the nucleoside phosphorylase I (NP-I) superfamily. However, the base-binding pocket varies within the members of the superfamily. Residues Thr242, Asn243, Val245, Glu201, V217, and Phe200 surround the base region. Half of these residues are hydrophobic and the other half hydrophilic.

Structure of the bPNP/M_1G-dR/PO_4 Complex. Figure 3.5 shows the stereo diagram of the active site of the bPNP/M_1G-dR/PO_4 complex. The phosphate ion reacted with M_1G-dR to generate M_1G and dR1P, both of which are present in the active site. Most of the side chains are well ordered except for a few Lys, Glu, Gln, and Arg residues on the surface of the protein. The electron density from the difference map or composite omit map for all the active sites residues as well as for the M_1G and the phosphate ion are relatively strong. However, the density for the sugar is very weak as is generally the case for these enzymes. Although the density for the sugar is weak, it is clear that the glycosidic bond is broken. The phosphate ion makes hydrogen bonds with Ser220, His86, Arg84, His64, Ala116, and Ser33. One of the oxygen atoms of the phosphate also hydrogen bonds to the 3' OH of the sugar. The 2'-deoxyribose makes three hydrogen bonds with His257, Tyr88, and the phosphate. In addition, Met219 interacts hydrophobically with the β face of the sugar ring. Asn243 is still able to make a hydrogen bond with N7 of M_1G. An ordered water molecule in the base binding site hydrogen bonds to O6 of the base as well as to Glu201. Another ordered water molecule is present between the phosphate and sugar binding sites and makes a hydrogen bond with the sulfate, Ser220. Phe200 forms a herringbone interaction with M_1G.

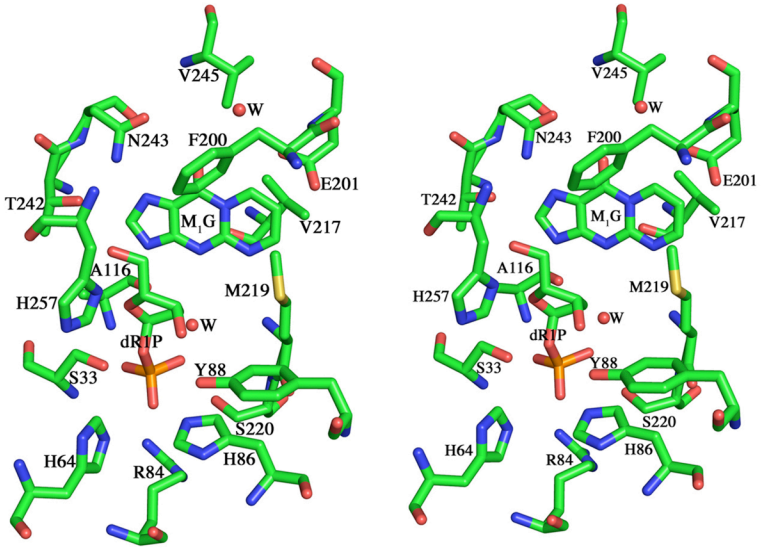


Figure 3.5. Stereo view of the active site of bPNP/M_1G-dR/PO_4 complex, where W denotes water (red sphere).

Structure of the bPNP/M_1G-dR/SO_4 Complex. The active site of the bPNP/M_1G-dR/SO_4 complex remains largely unchanged relative to other bPNP complexes (Federov *et al.*, 2001, Koellner *et al.*, 1997, Luic *et al.*, 2004, Toms *et al.*, 2005) as shown in Figure 3.6. All the residues in the phosphate, sugar, and base binding sites did not show any movement beyond expected coordinate errors. The ordered water molecule between the phosphate and sugar binding sites is still present but the one in the base binding pocket is not observed. One unusual observation is the appearance of a glycal intermediate in the active site (Figure 3.7). As mentioned previously, the densities for M_1G and the sulfate ion were relatively strong but weak for the sugar. The glycosidic bond was not observed in the density. The structures and electron density for M_1G, glycal, and the sulfate are shown in Figure 3.7B. The glycal, together with the sulfate, forms the same hydrogen bonding network with some of the active site residues as discussed in Figure 3.5.

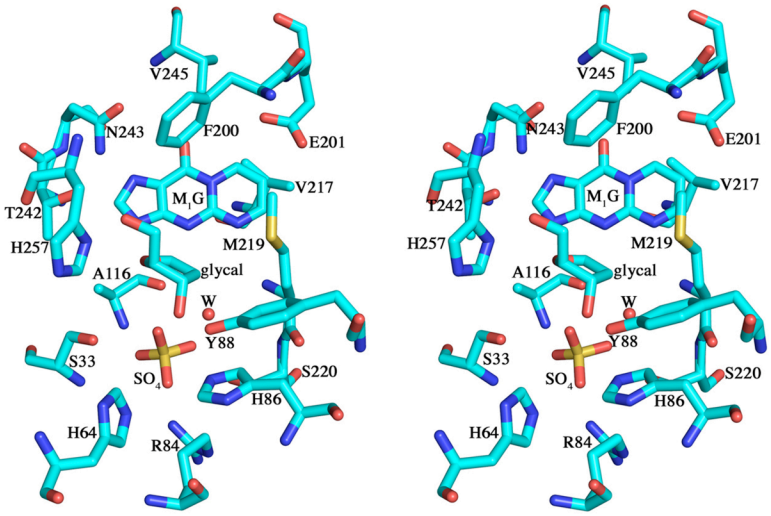


Figure 3.6. Stereo view of the active site of bPNP/M_1G-dR/SO_4 complex, where W denotes water (red sphere).

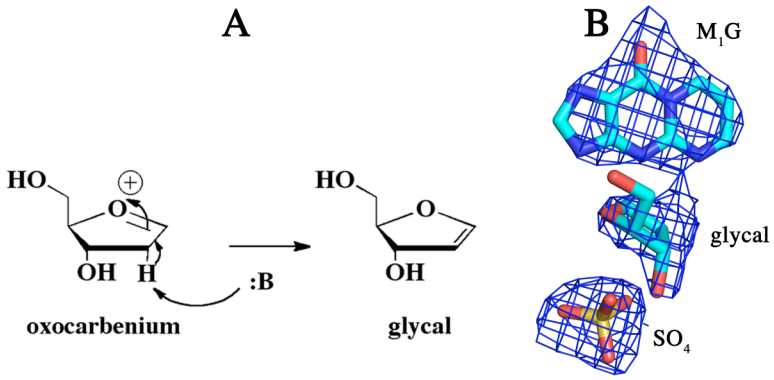


Figure 3.7. (A) The formation of the glycal intemediate from the oxocarbenium by some base :B. (B) The bound ligands shown in a composite omit map contoured at 1.7 σ from bPNP/M_1G-dR/SO_4 complex.

Section 3.4 Discussion

Comparison of bPNP/M1G-dR/PO4 and bPNP/M1G-dR/SO4 Complexes.

Superposition of the active sites of the two complexes is shown in Figure 3.8. Since the two complexes differ only by the identity of the sulfate or phosphate ion, no major

conformational change is expected, especially in the active site. The only observed structural difference is a shift in the position of the sugar moiety. This is due to the difference in the outcome of the reaction. However, the orientation of the sugar remains unchanged in general because, in both cases, the sugar moiety interacts hydrophobically with Met219. In both complexes, the same hydrogen bonds are observed with His257, Tyr88, and the sulfate ion. The phosphate actually forms a covalent bond with the sugar to become dR1P. The reason for the weak electron density observed for the sugar can be explained by the loss of three hydrogen bonds (2' OH) with the sulfate (phosphate), the amide backbone of Met219, and Tyr88 (Figures 3.10A, B). Although the hydrophilic residue Glu201 could still hydrogen bond to the conserved water observed in the bPNP/M_1G-dR/PO_4 complex, it lost an important hydrogen bond with M_1G. The distance between the Glu201 side chain and the closest atoms on M_1G fluctuated between 2.6 and 3 Å during different refinement.

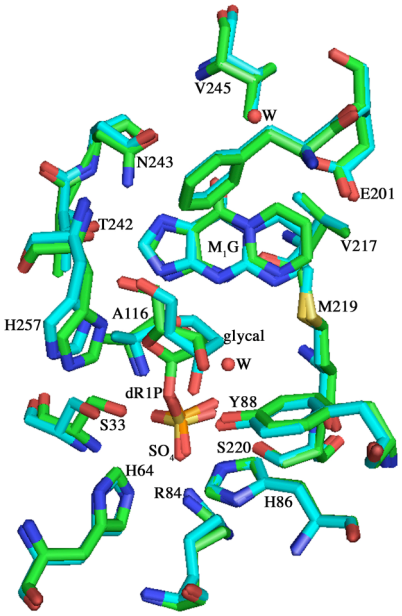


Figure 3.8. Superposition of the active sites of bPNP/M_1G-dR/SO_4 (cyan) and bPNP/M_1G-dR/PO_4 (green) complexes, where W denotes water (red sphere).

An obviously unfavourable interaction between the charged side chain of the Glu201 and M_1G is observed in the structure. This interaction is explained by the weak electron density near the edge of M_1G (Figure 3.7B). In fact, the averaged B-factor for Glu201 side chain is approximately 14 but for M_1G it is about 45. Nevertheless, adding an extra ring to the natural hypoxanthine base surprisingly did not affect the ability of M_1G to bind to the active site. The size of the base does not preclude Asn243 from making the critical hydrogen bond to N7 of M_1G.

The biggest difference between the two bPNP complexes is the observation that the sulfate ion unexpectedly cleaved M_1G-dR but did not form a covalent bond with the sugar. The bond formation between a sulfate's oxygen and C1´of the sugar is not necessarily impossible even though this phenomenon has not been reported in any biological transformations. However, an S-O-C bond does have precedent in traditional organic synthesis (Portoghese & Telang, 1971). In any event, the electron density of the difference map as well as the composite omit map do not support the existence of the bond between the sulfate and C1´ of the sugar. As mentioned above, the electron density for the sugar, particularly the bPNP/M_1G-dR/SO_4 complex, is weak; however, the geometry of M_1G-dR does not fit the observed density. The composite omit map contoured at 1.7 σ shows the density for the three separate pieces (Figure 3.7B). The possible mechanism for the glycal formation is shown in Figure 3.7A. The identity of the base responsible for proton abstraction is not yet known. The existence of the glycal intermediate has a precedent in bovine uridine phosphorylase from a recent publication (Paul *et al.*, 2010), although the structure of that particular glycal is slightly different from one observed here. In addition, Paul *et al*. observed that the glycal intermediate only formed for uridine, 5-fluouridine, and 5-fluoro-2'-deoxyribose but not 2'-deoxyribose. However, as shown in Figure 3.7A, a high-energy intermediate like the oxocarbenium has a very short lifetime. Thus,

essentially any base would easily remove one of the hydrogen atoms at C2' to spontaneously relieve the positive charge on the O4'. This brings up the next question regarding the identity of the base. Examining the active site of the complex (Figure 3.6) reveals that no plausible residues nearby can function as the base. However, a water molecule is present about 4.5 Å away from the sugar. This water appeared to be well ordered by making three hydrogen bonds with Ser220, the sulfate ion, and Tyr88. Thus, if the water was to serve as a base, the deprotonation would most likely occur at the same time as the active site undergoes conformational change.

Comparison with the Complex of Escherichia coli PNP/inosine/SO_4. Since *Escherichia coli* PNP (EcPNP) was able to carry out the coupling reaction of M_1G with a sugar substrate (Figure 3.3), it is necessary to compare the active site of EcPNP with that of the bPNP complexes. Ideally, an active site of EcPNP in complex with M_1G-dR and SO_4 would be best for comparison. However, because the structure of this ideal complex is not available the closest structure of EcPNP/inosine/SO_4 complex (Bennett *et al*., 2003) is used as shown in Figure 3.9 and Figure 3.10, which is redrawn in ChemDraw for comparison with bPNP. The stereo diagram (Figure 3.9) reveals that EcPNP has a Ser203 which hydrogen bonds to Asn204 in the base binding pocket which is equivalent to Thr242 and Asn243 of bPNP. EcPNP also has a conserved water hydrogen bonding to the O6 of hypoxanthine, and Ile206, Val178 and Phe159 that interact hydrophobically with hypoxanthine, which are equivalent to Val245, Val217, and Phe200 in bPNP. However, EcPNP does not have any equivalent of Glu201 in bPNP. The space presumably occupied by a glutamate residue as in bPNP is filled with a few water molecules leading to the widely open solvent path. Thus, EcPNP can easily accommodate at least an extra ring as in M_1G base. The presence of the destablizing Glu201-M_1G interaction most likely explains why EcPNP can couple the reaction of M_1G with a sugar moiety (Figure 3.3) to

produce M_1G-dR analog, whereas bPNP cannot. The structures of the two bPNP complexes also demonstrate that both sulfate and phosphate can cleave the M_1G-dR though they did not give the exact same products. Van der Waals repulsion between Glu201 and M_1G might have facilitated the glycosidic bond cleavage as the active site closed up. The glycal intermediate might have started to form during this time and finished by the time the conformational change completed.

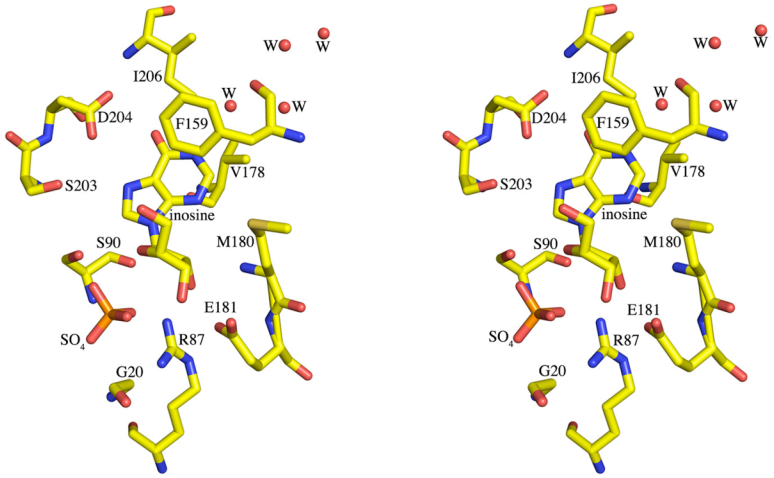


Figure 3.9. Stereo view of the active site of EcPNP/inosine/SO_4 complex reproduced from 1PR0 (Bennett *et al.*, 2003), where W denotes water (red sphere).

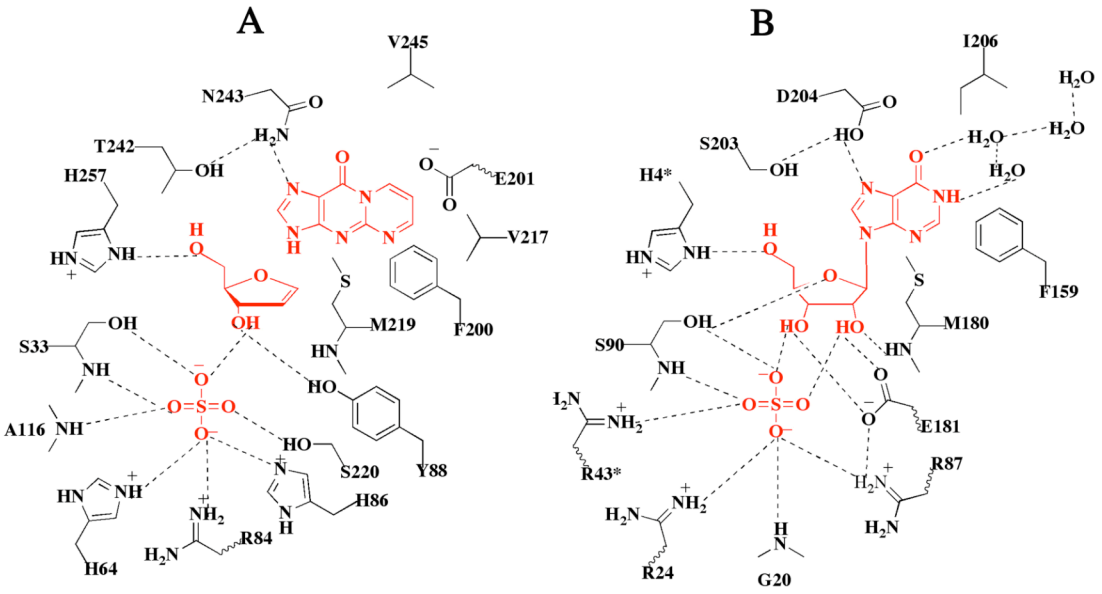


Figure 3.10. ChemDraw® of the active sites of bPNP/M_1G-dR/SO_4 (B) and EcPNP/inosine/SO_4 (A) complexes. Residues with an asterisk come from an adjacent monomer. The ligands are highlighted in red.

Section 3.5 Conclusion

The crystal structures of the bPNP/M_1G-dR/PO_4 and bPNP/M_1G-dR/SO_4 complexes were determined at 2.4 Å resolution. In addition to the observation of the glycal intermediate in the bPNP/M_1G-dR/SO_4 complex, these structures provide useful insight to an intriguing question regarding the enzymatic coupling reaction to produce M_1G-dR, the most abundant by-product of DNA damage in humans at physiological pH (Chaudhary, Nokubo, Marnett *et al.*, 1994, Chaudhary, Nokubo, Reddy *et al.*, 1994, Leuratti *et al.*, 1998). For the M_1G-dR analog in bPNP, the phosphorolytic reaction using a phosphate or a sulfate nucleophile should be spontaneous due to the strain-induced facilitation by Glu201 on the M_1G base. However, the synthesis of M_1G-dR is not possible because of the instability of M_1G caused by Glue201 to form a glycosidic bond with the sugar moiety. Thus, EcPNP would be an excellent candidate for coupling various nucleoside analogs to study their mutagenic potential in humans.

REFERENCES

1. Bennett, E. M., Li, C., Allan, P. W., Parker, W. B. & Ealick, S. E. (2003). J. Biol. Chem. 278, 47110-47118.
2. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). Nucleic Acids Res. 28, 235-242.
3. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Acta Crystallogr. D 54, 905-921.
4. Chapeau, M. C. & Marnett, L. J. (1991). Chem. Res. Toxicol. 4, 636-638.
5. Chaudhary, A. K., Nokubo, M., Marnett, L. J. & Blair, I. A. (1994). Biol. Mass. Spectrom. 23, 457-464.
6. Chaudhary, A. K., Nokubo, M., Reddy, G. R., Yeola, S. N., Morrow, J. D., Blair, I. A. & Marnett, L. J. (1994). Science 265, 1580-1582.
7. Collaborative Computational Project-Number 4 (1994). *Acta. Crystallogr. D* 50, 760-763.
8. DeLano, W. L. (2002). Curr. Opin. Struct. Biol. 12, 14-20.
9. Ealick, S. E., Babu, Y. S., Bugg, C. E., Erion, M. D., Guida, W. C., Montgomery, J. A. & Secrist, J. A., III (1991). Proc. Nat. Acad. Sci. U.S.A. 88, 11540-11544.
10. Emsley, P. & Cowtan, K. (2004). Acta Crystallogr. D Biol. Crystallogr. 60, 2126-2132.
11. Federov, A., Shi, W., Kicska, G., Tyler, P. C., Furneaux, R. H., Hanson, J. C., Gainsford, G. J., Larese, J. Z., Schramm, V. L. & Almo, S. C. (2001).

Biochemistry 40, 853-860.

12. Giblett, E. R., Ammann, A. J., Wara, D. W., Sandman, R. & Diamon, L. K. (1975). Lancet 1, 1010-1013.
13. Hassan, H. F., Coombs, G. H. (1988). FEMS Microbiology Reviews 4, 47-83.
14. Hershfield, M. S. & Mitchell, B. S. (1995) The Metabolic Basis of Inherited Disease, 1725-1768.
15. Koellner, G., Luic, M., Shugar, D., Saenger, W. & Bzowska, A. (1997). J. Mol. Biol. 265, 202-216.
16. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). J. Appl. Crystallogr. 26, 283-291.
17. Leer, J. C., Hammer-Jespersen, K. & Schwartz, M. (1977). Eur. J. Biochem. 75, 217-224.
18. Leuratti, C., Singh, R., Lagneau, C., Farmer, P. B., Plastaras, J. P., Marnett, L. J. & Shuker, D. E. (1998). Carcinogenesis 19, 1919-1924.
19. Luic, M., Koellner, G., Shugar, D., Saenger, W. & Bzowska, A. (2001). Acta Crystallogr. D 57, 30-36.
20. Luic, M., Koellner, G., Yokomatsu, T., Shibuya, S. & Bzowska, A. (2004). Acta Crystallogr. D 60, 1417-1424.
21. Mao, C., Cook, W. J., Zhou, M., Federov, A. A., Almo, S. C. & Ealick, S. E. (1998). Biochemistry 37, 7135-7146.
22. Parks, R. E., Jr. & Agarwal, R. P. (1972). Editor. Purine Nucleoside Phosphorylase, 3rd ed. New York: Academic Press.
23. Paul, D., O'Leary, S. E., Rajashankar, K., Bu, W., Toms, A., Settembre, E. C., Sanders, J. M., Begley, T. P. & Ealick, S. E. (2010). Biochemistry 49, 3499-

3509.

24. Portoghese, P. S. & Telang, V. G. (1971). Tetrahedron 27, 1823-1829.

25. Read, R. J. (1986). Acta Cryst. A 42, 140-149.

26. Schnetz-Boutaud, N. C., Mao, H., Stone, M. P. & Marnett, L. J. (2000). Chem. Res. Toxicol. 13, 90-95.

27. Schnick, C., Robien, M. A., Brzozowski, A. M., Dodson, E. J., Murshudov, G. N., Anderson, L., Luft, J. R., Mehlin, C., Hol, W. G. J., Brannigan, J. A., Wilkinson, A. J. (2005). Acta Cryst. D 61, 1245-1254.

28. Schuettelkopf, A. W., van Aalten, D. M. F. (2004). Acta Crystallographica D60, 1355-1363.

29. Toms, A. V., Wang, W., Li, Y., Ganem, B. & Ealick, S. E. (2005). Acta Crystallogr. D 61, 1449-1458.

CHAPTER 4

DESIGN AND SYNTHESIS OF A LIGHT-ACTIVATED INHIBITOR AND PRELIMINARY CHARACTERIZATION OFA NOVEL PROTEIN-PROTEIN CROSS-LINKED INTERMEDIATE INVOLVED IN THIAMIN BIOSYNTHESIS

Section 4.1 Introduction

Thiamin pyrophosphate (TPP) or thiamin, also known as vitamin B_1, is an indispensable cofactor in all living systems. Specifically, it plays a critical role in stabilizing acyl carbanion intermediates involved in amino acid and carbohydrate metabolism[1, 2]. Unlike most prokaryotes and some eukaryotes (plants and fungi), humans cannot biosynthesize TPP, and thus need 1.4 mg/day from dietary or supplemental sources [3]. In prokaryotes, thiamin is formed by a condensation reaction of thiazole and HMP-PP moieties catalyzed by ThiE and ThiL as shown Figure 4.1.

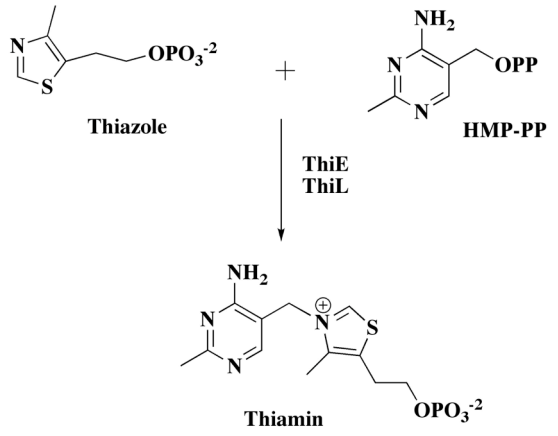


Figure 4.1. Coupling reaction of two biosynthetic products thiazole and HMP-PP to produce thiamin.

The thiazole and HMP-PP moieties themselves are biosynthesized by two different biosynthetic pathways. In *Bacillus subtilis*, thiazole synthase (ThiG),

together with ThiS and three other enzymes (ThiF, ThiO and cysteine desulfurase), catalyzes the formation of the thiazole moiety of thiamin using 1-deoxy-D-xylulose-5-phosphate (DXP), cysteine, and glycine as substrates [4-7] (Figure 4.2).

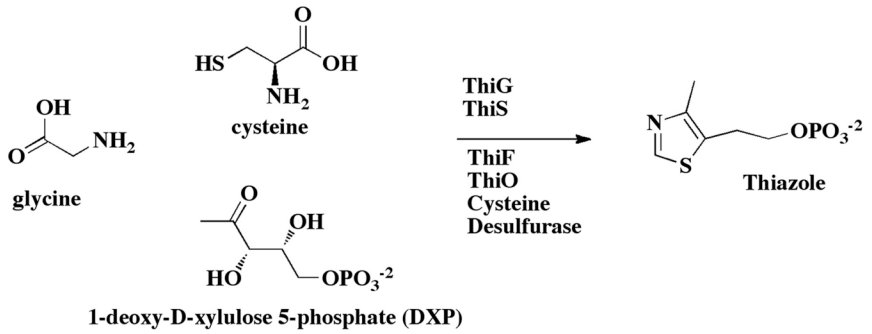


Figure 4.2. Biosynthesis of thiazole in *B. subtilis.*

Recent mechanistic studies [5, 7] suggest that thiazole synthase is covalently linked to a sulfur carrier protein (ThiS) by a novel reaction intermediate (**4**) as shown in Figure 4.3. This is so far the only known intermediate identified that covalently links two different proteins. In this mechanistic proposal, the substrate DXP (**1**) migrates to the active site of ThiG to form an imine linkage with Lys96 and tautomerizes to enamine **2**, which is further tautomerized to amino-ketone **3**. ThiS thiocarboxylate adds to **3** to form intermediate **4**, which undergoes an S-O acyl shift to give **5**. Dehydration and tautomerization of **5** generates **6**. Elimination of ThiS carboxylate from **6** generates **7**, which adds to dehydroglycine to form **8**. Species **8** undergoes transamination to release Lys96 of ThiG and decarboxylation to give carboxylated thiazole. As seen in the mechanistic proposal for the biosynthesis of thiazole, all the proposed biochemical transformations occur in the active site of ThiG and involve mainly ThiS-ThiG complex. The other enzymes in this pathway (ThiF,

ThiO and cysteine desulfurase) only play a 'peripheral' role in modifying ThiS, glycine, and cysteine substrates.

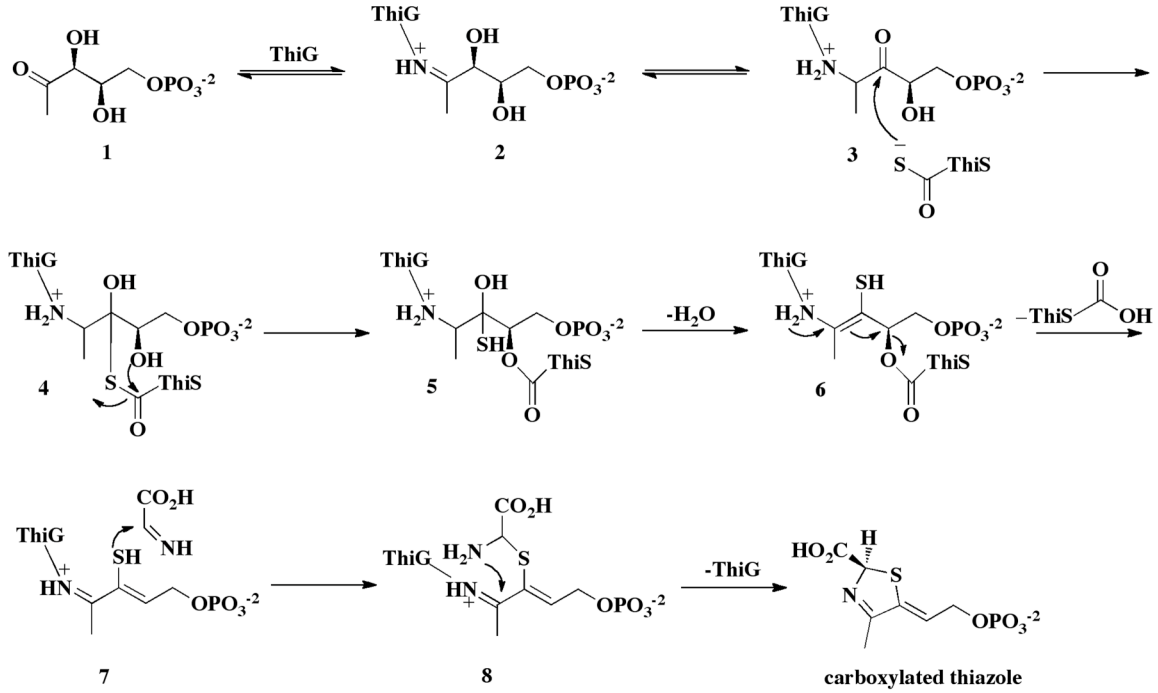


Figure 4.3. The formation of the thiazole proceeds via a novel protein-protein crosslinking intermediate (**4**).

The X-ray crystal structure of the ThiS-ThiG complex at 3.2 Å resolution has been recently determined [6] (Figure 4.4). The red ball-and-stick molecule occupying the phosphate-binding pocket in the active site is the phosphate ion from the crystallization buffer. This pocket presumably binds to the phosphate group of the substrate DXP (**1**).

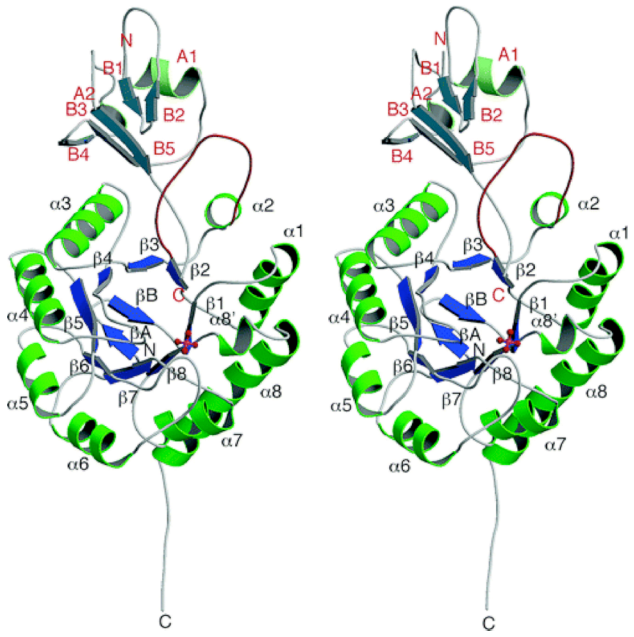


Figure 4.4. Stereo view of the ThiS-ThiG complex with ThiS containing secondary elements labeled in red and ThiG in black. The carboxy terminus of ThiS is labeled with the red letter C proximal to the phosphate ion, which is shown as the red ball-and-stick model.

The ThiS-ThiG heterodimer above exists as a tetramer of heterodimers (Figure 4.5) in solution as measured by dynamic light scattering and size exclusion chromatography experiments. Underlying the static X-ray crystal structure is a dynamic binding equilibrium. The four ThiS monomers bind to the ThiG tetramers with an $K_{on,off}$ rate that translates into a K_d of approximately 70 nM (unpublished data), which is a reasonably tight binding. Nevertheless, the X-ray crystal structure of the complex shows that the C-terminus of ThiS in the active site of ThiG is disordered. This probably affects the rigidity of the complex, and thus may explain the observed low-resolution crystal structure.

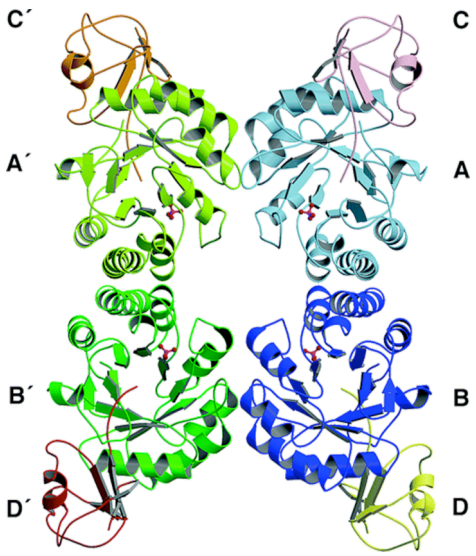


Figure 4.5. ThiS-ThiG tetrameric complex at 3.15 Å resolution with ThiS subunits labeled as **C**, **C'**, **D**, **D'** and ThiG as **A**, **A'**, **B**, **B'**. The phosphate ion is shown in ball-and-stick.

To identify the catalytically important residues, the proposed intermediate **4** was modeled into the active site. This modeling revealed that in addition to the Lys96 involved in imine formation, a Glu98 and an Asp182 are also present in the active site (Figure 4.6A).

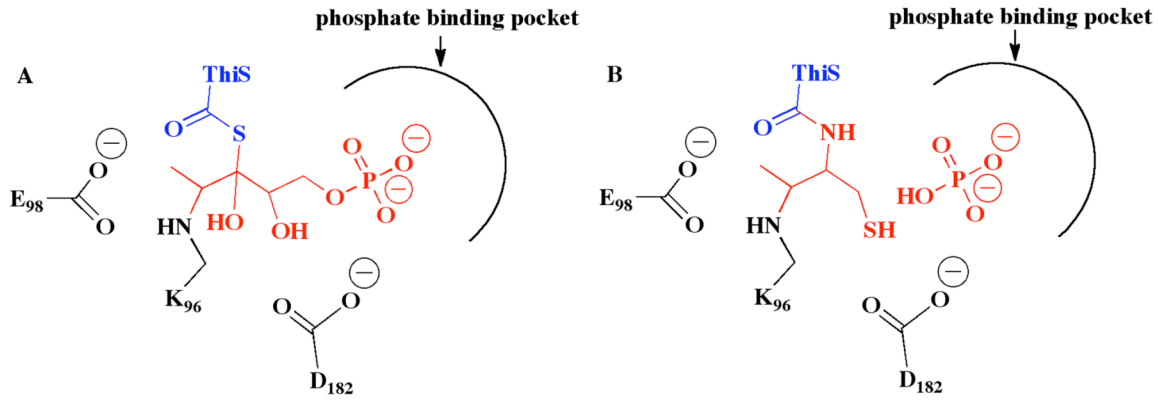


Figure 4.6. Comparison of the structure of the ThiS-ThiG crosslinking intermediate (**A**) with that of the proposed analog (**B**).

Since the complex was solved at a relatively low resolution without any bound ligand and the functions of Glu98 and Asp182 remain ambiguous by mutagenesis studies, the focus of this project is to synthesize an analog of the putative intermediate **4**, and to solve the X-ray crystal structure of the analog complex. The analog complex is anticipated to give higher resolution crystals, because the flexible carboxy terminus of ThiS-COOH will be anchored at the active site.

The proposed analog is shown in Figure 4.B. This analog is similar to the intermediate **4** (Figure 4.3 or Figure 4.6A) with the following differences. 1) ThiS is linked to the analog via a stable amide bond as opposed to the thioester bond of the intermediate. 2) The thiol group at C_4 of the analog replaces the hydroxyl group at C_4 of the intermediate. 3) The phosphate-binding pocket of the analog will be occupied by a phosphate ion presumably from the crystallization buffer.

Section 4.2 Results and Discussion

The strategy for making the analog **4** (Figure 4.6B) is shown in Figure 4.7. The thiol moiety of cysteine (**9**), being a powerful nucleophile as well as a strong oxidizer, needs to be protected to form species **10**. The amino group is next protected with a compatible protecting group to produce **11**. The carboxylic acid is then stepwise converted to a methyl ketone **12**. The carbonyl oxygen is finally protected with an appropriate protecting group. Now the amino group is deprotected first to generate species **14**. Then the thiol moiety is deprotected in order to carry out intein chemistry[8] in the next step. Chemical ligation using intein chemistry gives **16**. After an S-to-N acyl shift, a more thermodynamically stable species **17** is generated. The carbonyl oxygen is finally deprotected to free up the methyl ketone moiety as shown as **18**,

which is inserted into the active site of ThiG to form a Schiff base linkage with Lys96 of ThiG as in **19**. Following reduction of **19**, a stable N-C bond of species **20** is formed.

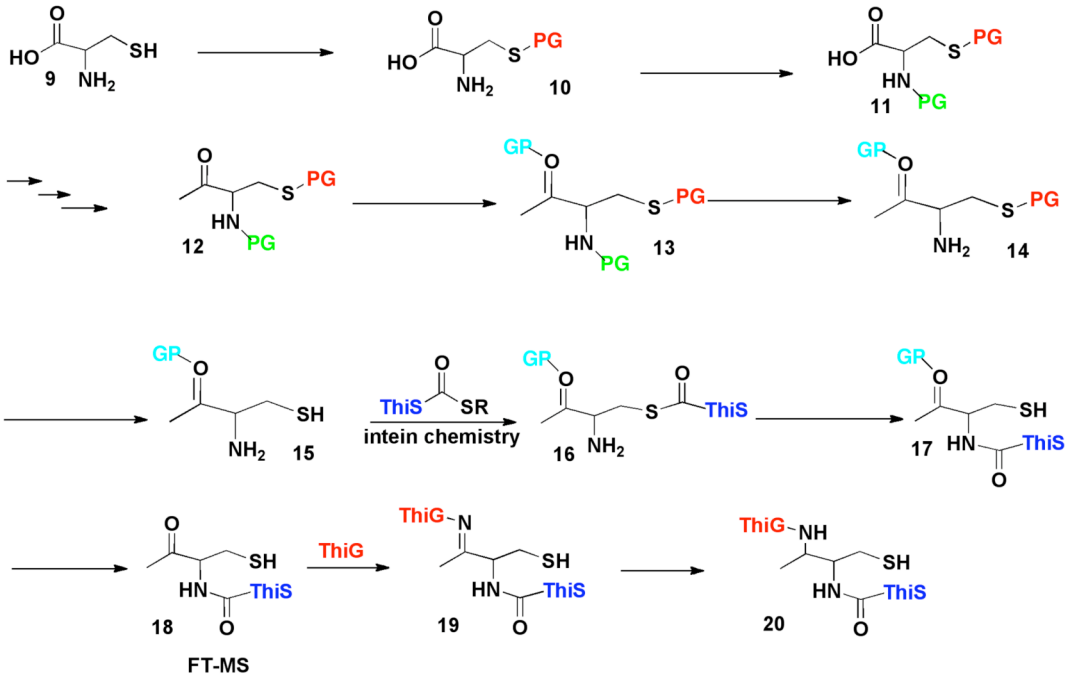


Figure 4.7. Strategy for making analog **4**, where 'PG' represents a protecting group.

The intein chemistry employed in one of the steps above is, in general, a powerful way to ligate together different protein segments, or a protein and a small cysteine-like molecule containing a sulfhydryl and an amino groups that are three bonds apart, to generate a fully functional protein in a process known as protein semisynthesis (Figure 4.8). This technique is very useful for producing a large quantity of proteins that are difficult to express by conventional biochemical methods. More specifically for this particular system, ThiS is attached to an intein construct that contains a chitin binding domain (CBD, **21**), which undergoes an equilibrium N-S acyl shift. The resulting adduct **22** undergoes 2-mercaptoethanesulfonic acid (MESNA)-

induced cleavage from the CBD column to give a thioester intermediate **23**. A neighboring thiol group of a cysteine on ThiS participates in the formation of the cyclic intermediate **24** by attacking **23** to displace the MESNA from ThiS. The cyclic species **24** then can react with any amino-thiol containing compounds to form the ligated ThiS.

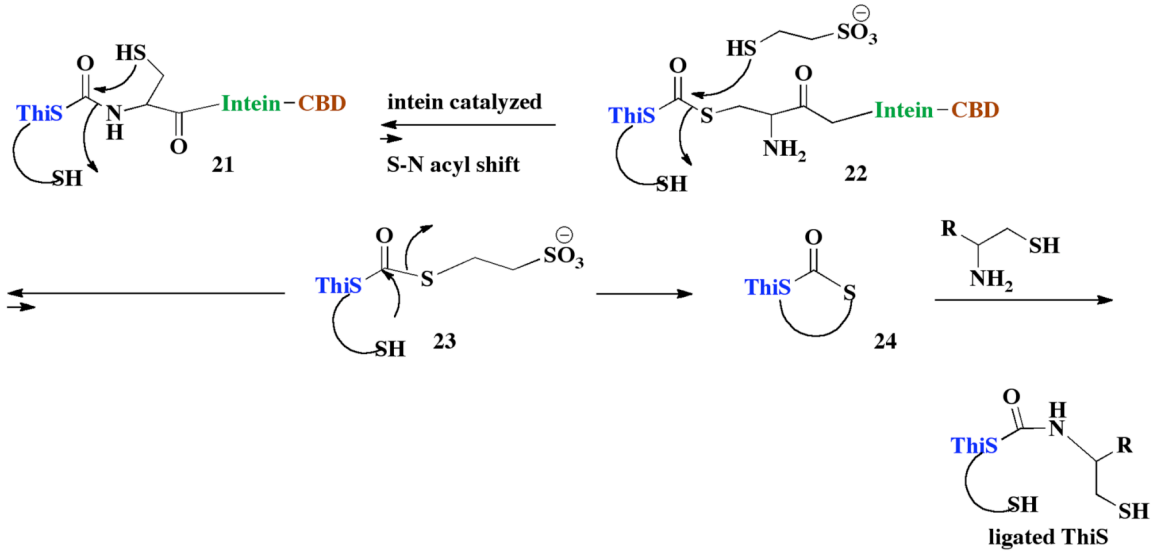


Figure 4.8. Intein chemistry to ligate ThiS to a cysteine-like molecule containing a sulfhydryl and an amino groups that are three bonds apart; CBD stands for chitin binding domain.

As indicated above, the analog must have the amino-thiol functionality for intein chemistry to work. In addition, it has to contain a methyl ketone in order to form an imine bond with Lys96 of ThiG in the active site. Thus, the general strategy for making the proposed analog is shown in Figure 4.9. Accordingly, the amino-thiol-ketone **25** is ligated to ThiS by intein chemistry. Following rearrangement via **26**, the ligated adduct **27** is reacted with ThiG to form **28,** and finally the proposed analog **4** after sodium borohydride reduction.

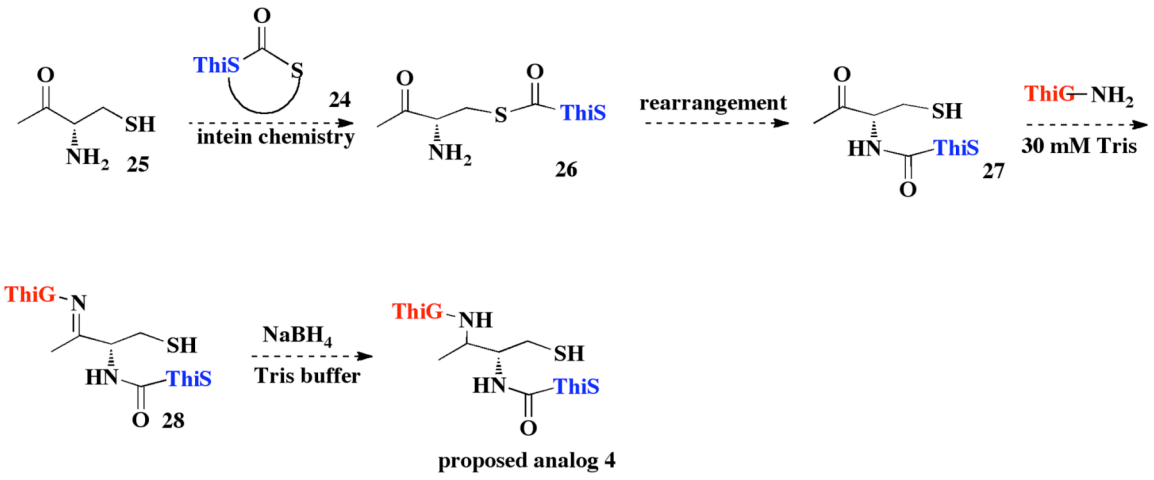


Figure 4.9. Design of the proposed analog **4**.

The chemical synthesis of the amino-ketone **25** is shown in Figure 4.10. The thiol moiety of cysteine (**9**) is protected with the tert-butyl group under acidic conditions to form **30**. The amino group is next protected with Boc_2O under basic conditions to produce **31**. The carboxylic acid is then converted to Weinreb's [9] amide **32** using 1-(3-dimethylaminopropyl)-3-ethyl carbodiimide hydrochloride (EDCI) as an activator, and 4-dimethylaminopyridine (DMAP) and hydroxybenzotriamole (HOBt) as catalytic nucleophiles. Finally, the tert-butyl group is removed with mercuric acetate in trifluoroacetic acid to afford **25**.

Compound **25** was synthesized, but intein chemistry could not be carried out because the molecule is very unstable and degraded at neutral and basic pHs. Intein chemistry requires a pH of 7 or higher.

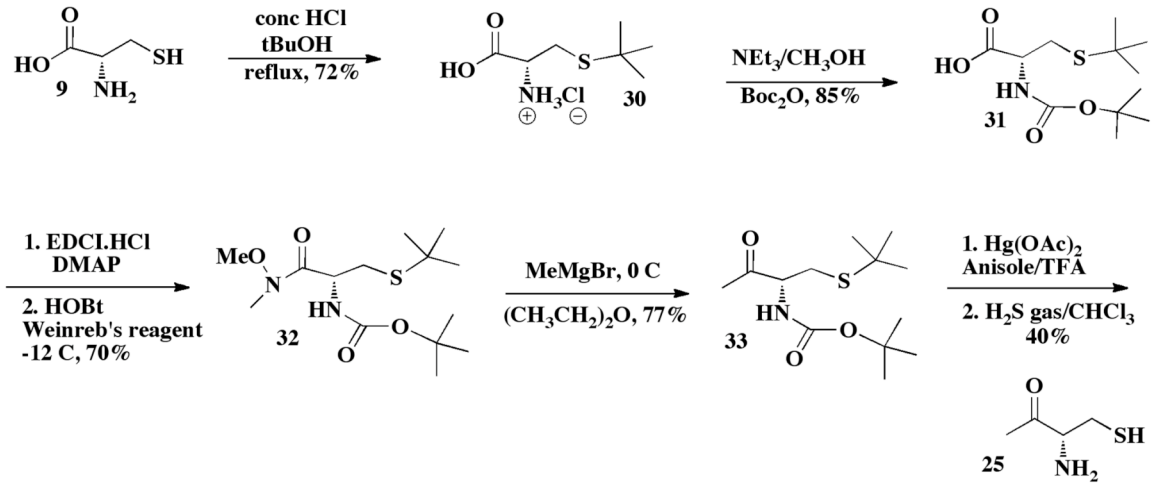


Figure 4.10. Chemical synthesis of proposed amino-ketone **25**.

Thus, one solution to this problem is to find a way to synthesize an analog containing the amino-thiol functionality and a protected methyl ketone. This thinking led to the modified strategy is shown in Figure 4.11. One modification in the strategy is that the ketone protecting group in shown in **34, 35,** and **36** somehow must be removed before insertion into the active site of ThiG.

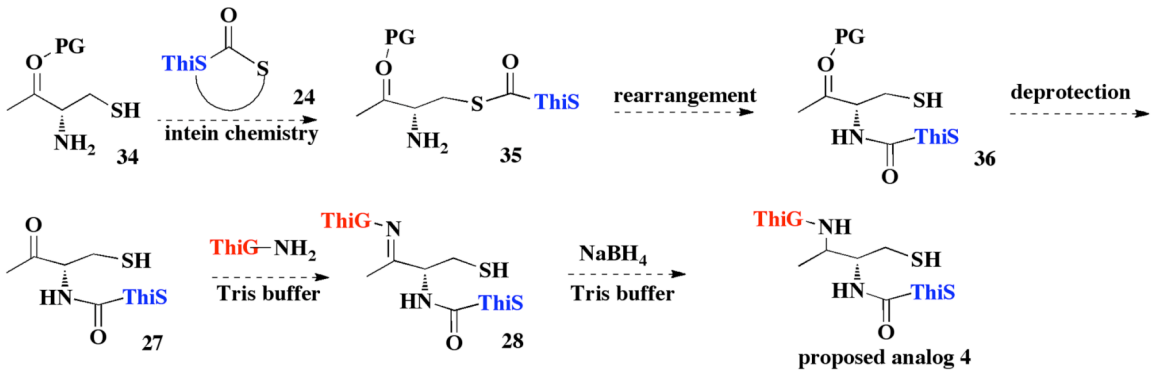
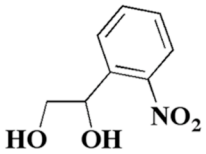


Figure 4.11. Modified strategy to make the proposed analog **4** with a protecting group (PG) on the ketone.

Since any protecting group used must be compatible with intein chemistry, the choices are severely limited—no acids, bases, nor oxidizing/reducing reagents. The most logical protecting group is the photolyzable nitrophenyl ethylene glycol as shown below.



nitrophenyl ethylene glycol

Unfortunately, this compound is quite expensive for initial trial reactions; thus, ethylene glycol was used as a model protecting group while a workable synthetic route as well as intein chemistry was designed, before using the photolyzable protecting group. The synthetic scheme for model analog **37** to generate **39** via **38** is shown in Figure 4.12 and the actual synthesis is presented in Figure 4.13.

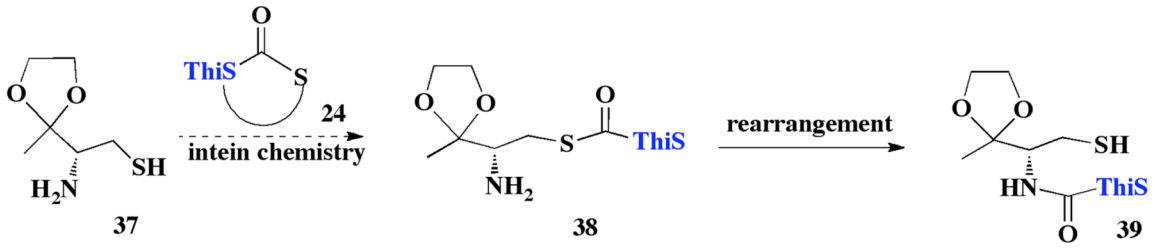


Figure 4.12. Model analog **37** for a workable synthetic route and intein chemistry.

The first three steps to generate **32** are the same as before (**Figure 10**). In the next step, the Boc protecting group is replaced with trifluoroacetamide to give **41** via **40** because Boc interferes with the protection of the ketone with ethylene glycol in a later step. Next, the Weinreb's amide is converted to ketone **42** by a Grignard's reaction. The trifluoroacetamide protecting group on the amine is simultaneously

removed with excess Grignard's reagent. The ketone moiety is then protected as an acetal (**43**) using ethylene glycol. Finally, the tert-butyl group is removed with mercuric acetate to afford **37.**

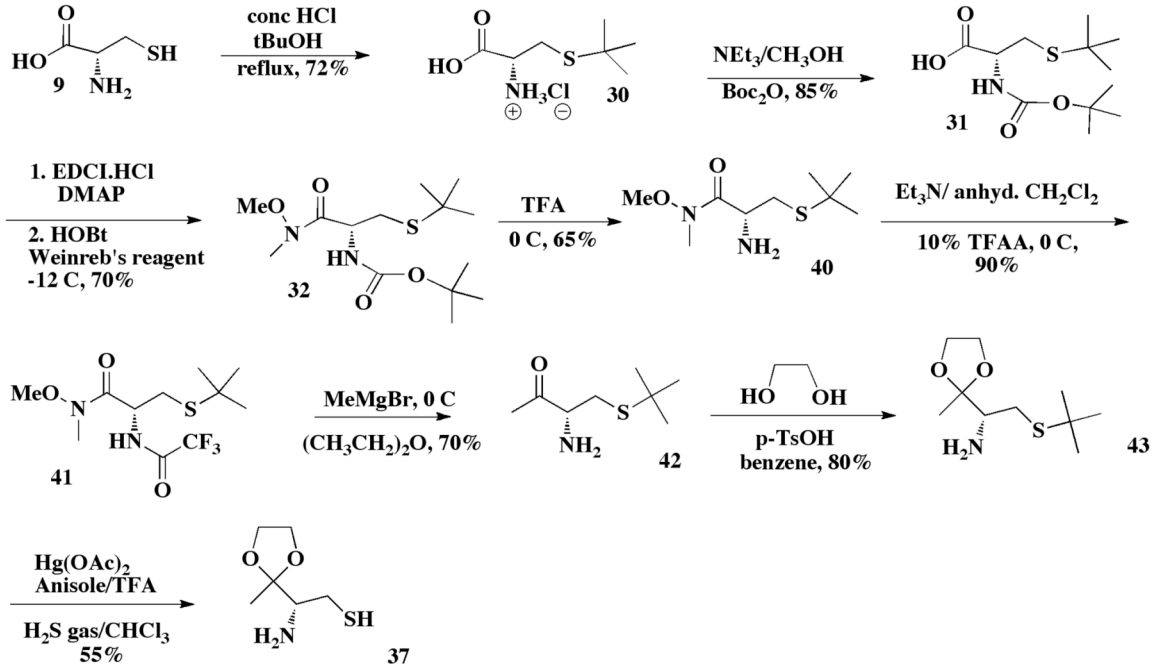


Figure 4. 13. Synthetic route for the model analog **37**.

The synthesis was completed and a very good yield of **37** was obtained. Compound **37**, together with two other amino-ketone compounds, was ligated to ThiS to demonstrate the proof-of-concept for intein chemistry as shown in Figure 4.14. The 2-aminoethanethiol (**46**) is commercially available and is used as a control. Compound **44**, a close version of the proposed analog, was synthesized similarly to that of **37** (see Experimental Section). Compounds **37**, **44**, and **46** were ligated to ThiS (the cyclic intermediate is shown) to generate **39**, **45**, and **47**. The deconvoluted Electrospray Ionization Mass Spectrometry (ESI-MS) spectra for the cyclic ThiS (**24**) and one of the ligated ThiS adducts (**39**) are shown in Figures 4.15A and 4.15B. The

resolution of the ESI-MS is approximately one Dalton. Figure 4.15A shows the mass spectrum of the cyclic ThiS (**24**) with a molecular weight of 7609 Da for the expected molecular weight of 7607 Da. The expected molecular weight of **39** is 7770 Da and the mass spectrum in Figure 4.15B recorded a value of 7771 Da.

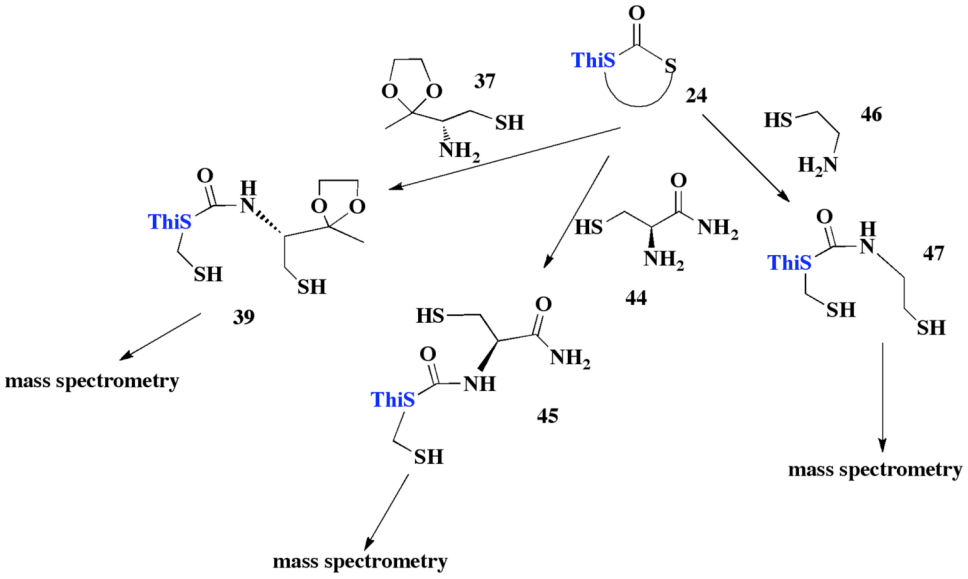


Figure 4.14. Intein chemistry to add amines to the C-terminus of ThiS-COOH.

In order to ascertain that the robustness of this ligation, compound **44** was synthesized and adduct **45** was made. The FT-MS for adduct **45** with the exact molecular weight of 7727 is shown in Figure 4.15C. These experiments turned out to be very robust and were measured repeatedly many times to ascertain the identity of the ligated adducts. Compound **45** was purchased from Sigma-Aldrich and used to generate **47**. The ESI-MS data for **47** was also taken to confirm successful ligation (data not shown).

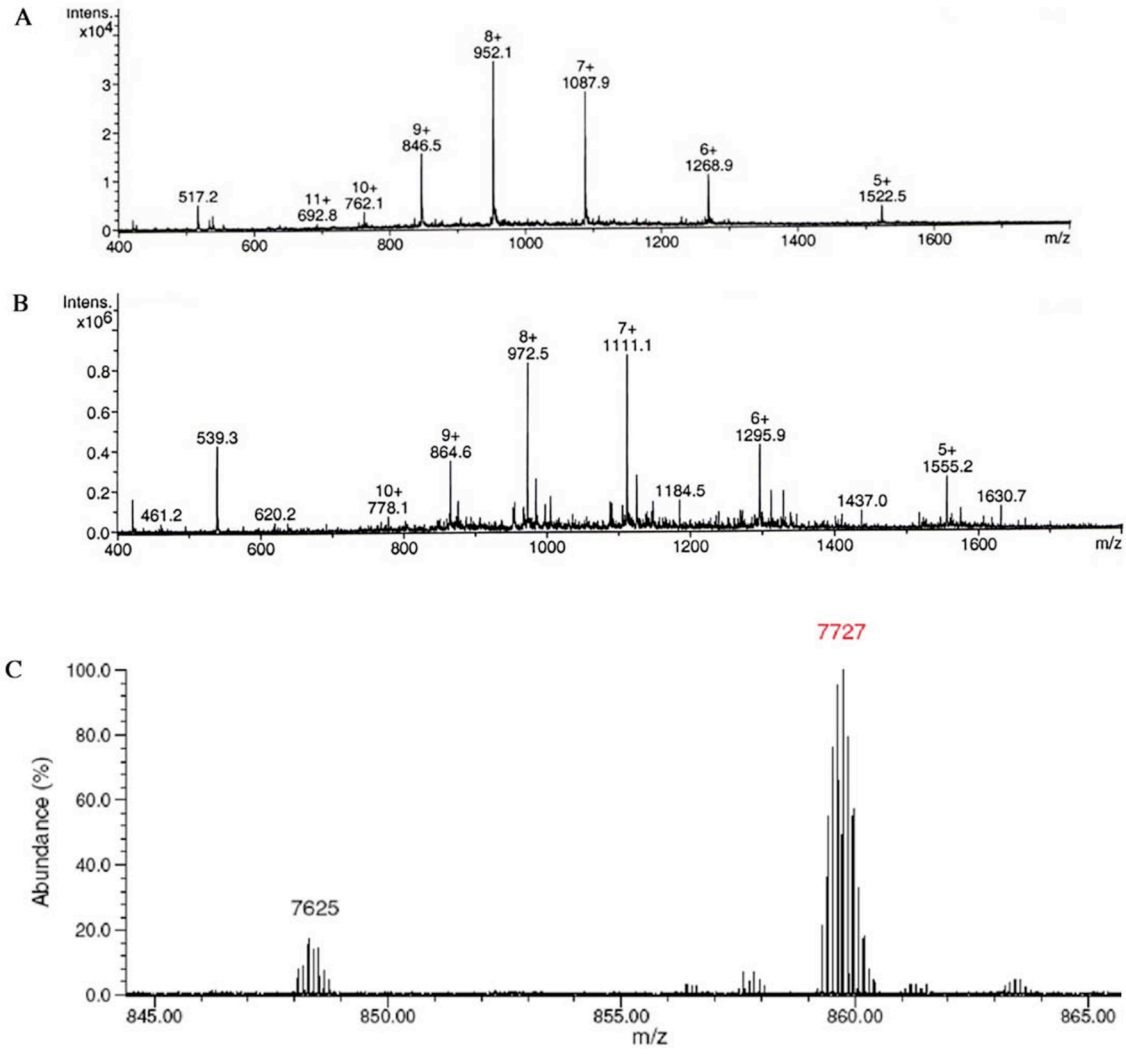


Figure 4.15. Deconvoluted ESI-MS spectra of the cyclic ThiS (**24**) in panel **A** and a ligated ThiS (**39**) in panel **B.** Panel **C** shows the FT-MS spectrum for adduct **45**. The peak at 7625 is the hydrolyzed product of the cyclic ThiS (**24**) or ThiS-COOH.

Since the synthesis and intein chemistry worked for the model analog **37**, the photolyzable analog **49** with nitrophenyl ethylene glycol protecting group was similarly synthesized as shown in Figure 4.16. The yield for **49** is not as high as **37** but sufficient for our purpose.

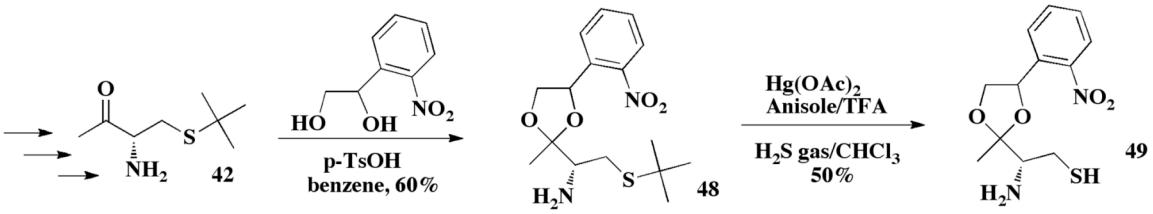


Figure 4.16. Synthesis of the photolyzable analog **49**.

The synthesized compound **49** was shown to be photolyzable in either methanol or water using **48** instead of **49** for simplicity (Figure 4.17). The deprotected compound **42** was formed in one hour as confirmed by NMR.

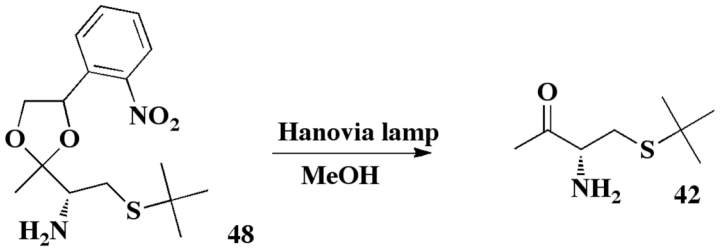


Figure 4.17. Model reaction for photo-deprotection.

Another unanticipated problem arose regarding the solubility of **49**. It was not soluble in 30 mM Tris buffer (pH 7.5), which is the buffer used for ThiS, and it precipitated out of the solution even with low level (2-10%) of organic solvents such as dimethyl formamide or dimethyl sulfoxide (Figure 4.18).

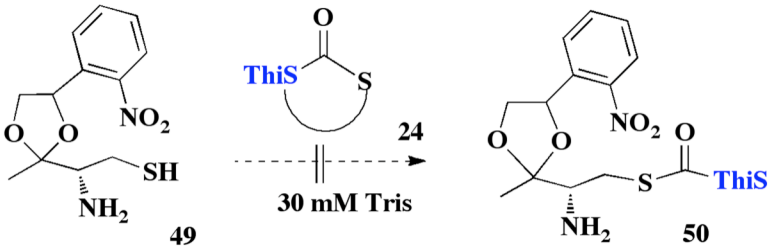


Figure 4.18. Intein chemistry to ligate **49** to ThiS.

To increase solubility, a hydrophilic functional group was put on **49**. Specifically, a carboxylic group was installed on the carbon of the primary diol as shown in Figure 4.19 (compound **54**). The nitrobenzaldehyde **51** undergoes a Wittig reaction to form alkene **52**, which is dihydroxylated to produce **53**. Ester hydrolysis of **53**, which is passed through the cation exchanged to give **54**.

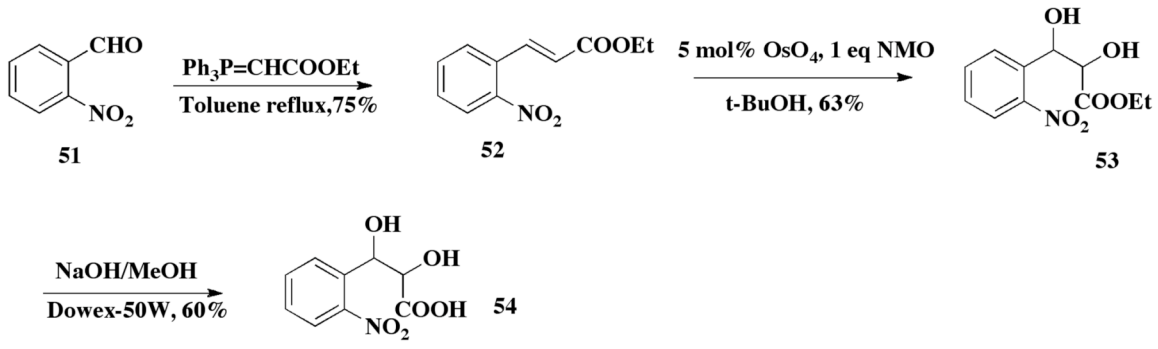


Figure 4.19. Synthesis of the modified photolyzable analog **54**.

Compounds **53** and **54** were then used to protect ketone **42** (Figure 4.20). However, both reactions failed because the ethyl group in **53** interfered with the protection step and compound **54** is not soluble in benzene.

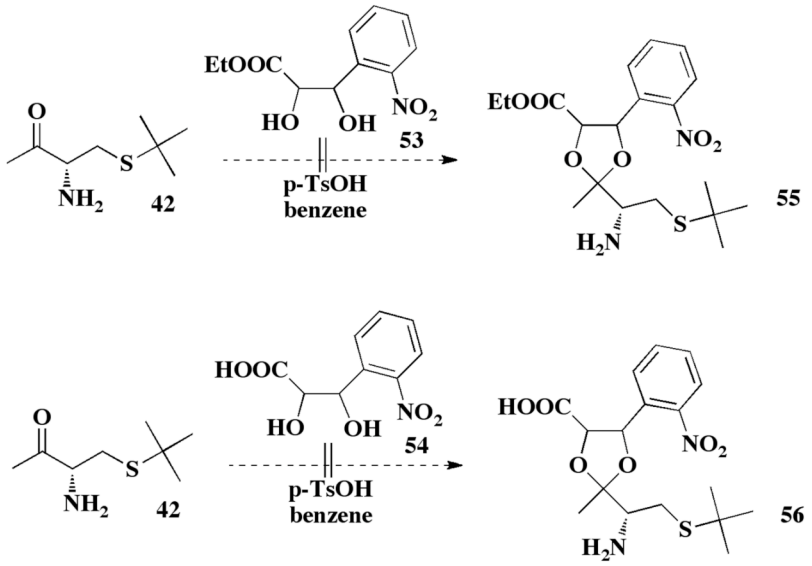


Figure 4.20. Protection of ketone **42** using **53** and **54**.

Therefore, several photolyzable analogs (**57**, **58**, **59**) with soluble functional groups added were proposed as shown in Figure 4.21. Compound **59** is most suitable for the intended purpose because ether linkage on the benzene ring is known to have no affect on the photodeprotection step. Its synthesis is shown in Figure 4.22.

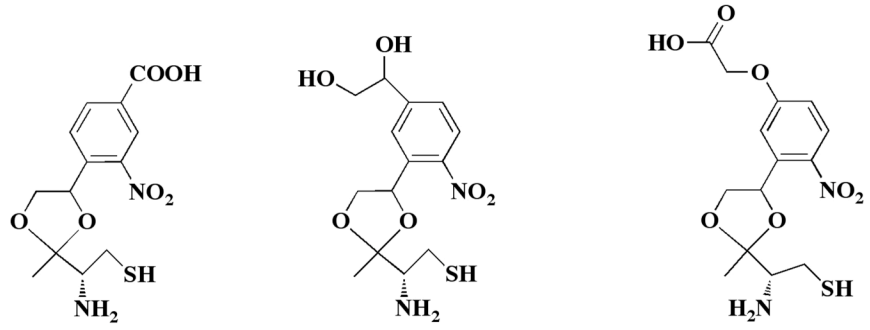


Figure 4.21. Proposed photolyzable analogs containing an additional soluble group.

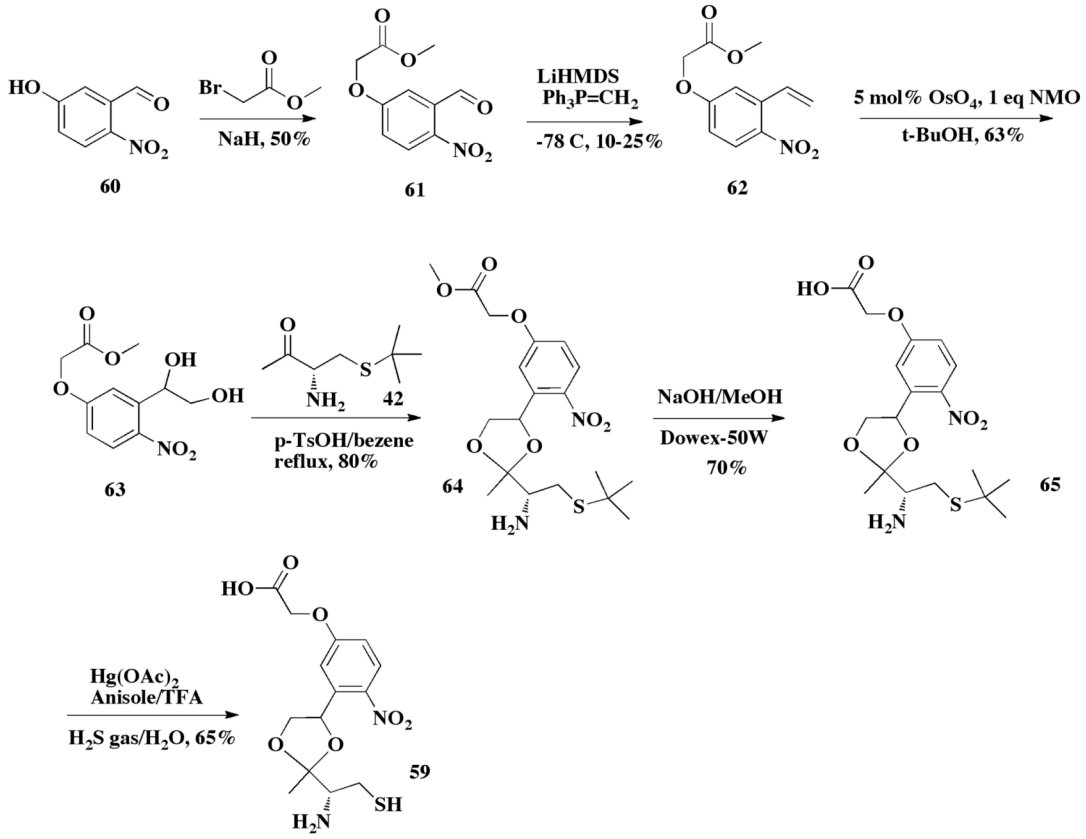


Figure 4.22. Synthesis of the photolyzable analog **59**.

The commercially available 5-hydroxy-2-nitrobenzaldehyde (**60**) is reacted with methyl bromoacetate in tetrahydrofuran to form **61**. The addition of the Wittig's reagent, methyl triphenyl phosphonium chloride, to **61** gives alkene **62**, which is dihydroxylated by osmium tetroxide to form **63**. The aminoketone **42,** which is synthesized as described in Figure 4.13, is protected as a photo-removable acetal **64** by a reaction with **63** under reflux condition. Compound **64** undergoes ester hydrolysis followed by cation exchange with Dowex-50W resin to give **65**. Finally, the tert-butyl group is removed by mercuric acetate in trifluoroacetic acid to afford the final product **59**. The synthesis for compound **59** was achieved with a good overall yield. Ellman reagent's assay [10] (Figure 4.23) was carried out to determine the concentration of the

free thiol of **59** using cysteine for the standard curve. Ellman's reagent **66** reacted with **59** to generate a disulfide **67** and **68**, which has the absorbance of visible light at 412 nm with an extinction coefficient of 14, 150 M^{-1} cm^{-1} [11]. The free thiol groups tend to form disulfide bonds relatively easily and thus may underestimate the actual concentration of **59**. Nevertheless, the concentration of **59** was estimated to be at least 6 mM by this assay.

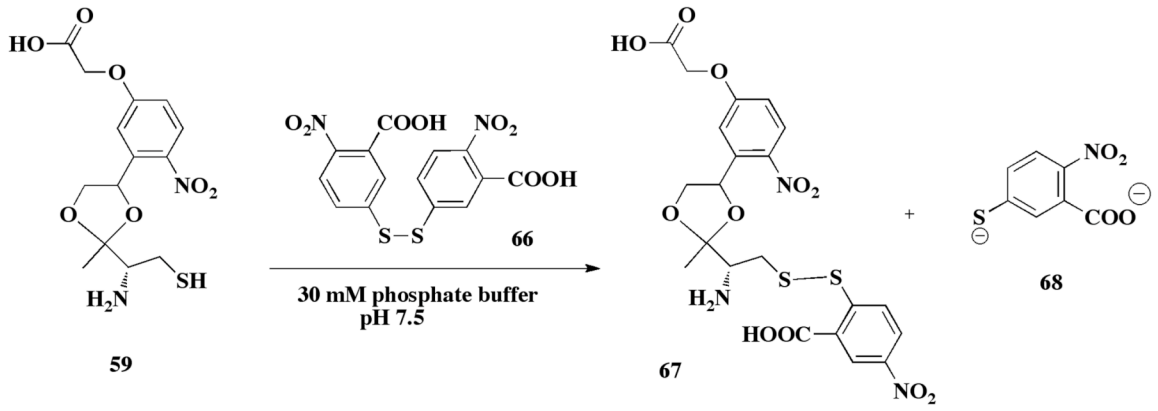


Figure 4.23. Ellman's assay for free thiol concentration.

Section 4.3 Conclusion

We have synthesized the analog in relatively good yield based on Ellman's assay and also demonstrated the proof-of-concept to covalently link an aminothiol ligand to ThiS using intein chemistry. In particular, we have shown that ThiS can be covalently linked to **37**, **44**, and **46** to give **39**, **45**, and **47**, respectively, and confirmed the resulting covalent adducts using ESI-MS and FT-MS. The objective of future studies is to attach **59** to ThiS using intein chemistry to generate **69** (Figure 4.24). Photolysis of **69** using Hanovia lamp in the presence of ThiG should regenerate methyl ketone **27** and facilitate the Schiff base linkage with Lys96 of ThiG to produce **28**. Sodium

borohydride reduction of **28** will trap the proposed analog **4**, which will be confirmed by SDS PAGE, FT-MS, and X-ray crystallography.

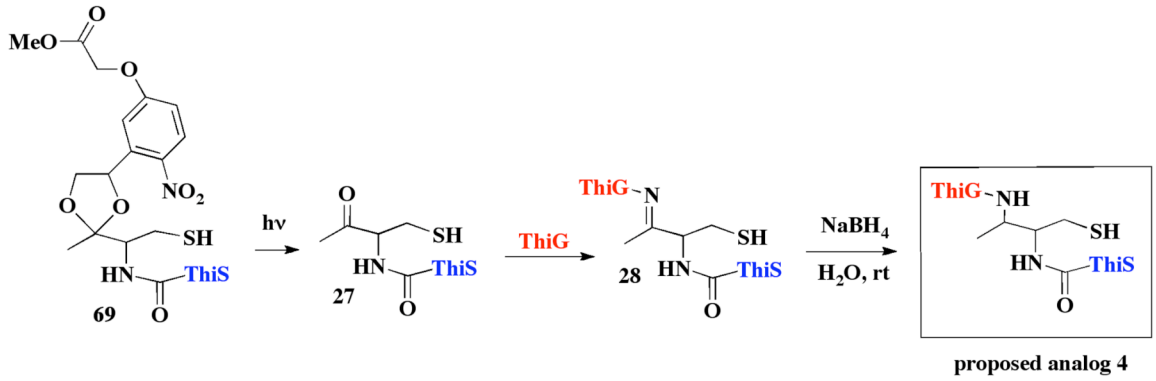


Figure 4.24. Reactions to covalently link ThiS-**59** adduct (**69**) to ThiG.

Section 4.4 Experimental Section

Mass Spectrometry. The molecular weight of small molecules and proteins were determined by an electrospray ioninzation mass spectrometer (ESI-MS) with a hybrid triple quadrupol/linear ion trap 4000 Q trap (Applied Biosystems/MDS Sciex). All samples were purified and resuspended in acetonitrile/water solvent system containing 0.1% of trifluoroacetic acid. Some of the protein samples were analyzed by FT-MS in collaboration with Professor Fred McLafferty's group.

Intein Chemistry to Ligate ThiS to Amino-Thiol Compounds. All the reagents in these procedures and subsequent synthesis were purchased from Sigma-Aldrich (St. Louis, MO), except that the chitin beads were purchased from New England BioLabs (Ipswich, MA). Unless noticed otherwise, all reactions were carried out at room temperature. The procedures are as follows. Two cell pellets of ThiS-chitin binding domain construct from 2 L cell culture, were resuspended in 50 mL cold lysis buffer (20 mM Tris pH 7.5, 500 mM NaCl, 1 mM EDTA, 0.1% (v/v) TritonX-100). The

suspension was lysed by sonication on ice. The lysate was centrifuged at 38, 000 *g* for 30 minutes to clarify. In the meantime, the chitin column was regenerated by soaking it with 0.3 M NaOH and then washing with 20 bed volumes of water followed by 5 bed volumes of column buffer (20 mM Tris pH 7.5, 500 mM NaCl, 1 mM EDTA). The clarified cell lysate was loaded on the column and was allowed to flow at a low rate (~1 mL/min). Then the column was washed with 20 bed volumes of column buffer. Finally column buffer containing 60 mM sodium 2-mercaptoethanesulphonate (MESNA) was loaded (3 x bed volume) and the column was allowed to sit for 20 hours at room temperature (20°C). After 20 hours the column was eluted with 20 mL of column buffer and was immediately desalted into 20 mM of Tris pH 7.0. The eluted protein or cyclic ThiS (**24**) was concentrated to 500 µL. To 150 µL of the concentrated protein an amino-thiol compound was added to get a final concentration of 10 mM. The mixture was incubated for 10 minutes and was immediately desalted to 20 mM of Tris pH 7.0 containing sodium azide. The sample was analyzed by ESI-MS.

*Synthesis of Compound **25**.* To 1.45 mmol of **33** in a 100 mL round-bottomed flask, 1.45 mmol of TFA and 0.790 mL or 7.25 mmol of anisole were added. The reaction mixture was cooled to 0º C before adding 1.45 mmol of $Hg(Oac)_2$. The reaction mixture was stirred for 30 min. Then, H_2S was bubbling in very slowly for 2 minutes until black precipitate (HgS) was formed. The whole reaction mixture was filtered through a celite bed and rinsed with 50% MeOH/TFA. The final mixture containing **25** was rotavaped and put on a vacuum overnight. Species **25** turned out to be unstable at neutral and basic pH. Thus, it was not purified and no NMR data were taken for this product.

*Synthesis of Compound **30***. In a 100 mL round-bottomed flask, 5 g cysteine, 5 mL of tBuOH, and 17 mL of 3 M HCl (36.6 g/mol) were added. The reaction mixture was refluxed overnight at 85°C. Thin layer chromatography (TLC) with Ninhydrin staining was carried out to monitor the reaction. The precipitate crystallized out and was collected by filtration with a Buchner funnel. The crystals was washed with ethyl ether and vacuum-dried. **30**: ^{1}H NMR (300 MHz, D_2O) δ 4.06 (m, 1H), 2.95 (m, 1H), 1.15 (s, 9H).

*Synthesis of Compound **31.*** In a 100 mL round-bottomed flask, 2 g of **30**, 20 mL of MeOH (anhydrous), 2.6 g of Boc_2O and 26.6 mmole of Et_3N was added. The whole reaction mixture was stirred for 3 hours at room temperature. TLC with ceric ammonium molybdate (CAM) was carried out to monitor the reaction progress. A milky semisolid product was extracted with CH_2Cl_2, saturated NH_4Cl, saturated $NaHCO_3$, and finally saturated NaCl. Then, it was dried over $MgSO_4$ and rotavaped to dryness. **31**: ^{1}H NMR (300 MHz, $CDCl_3$) δ 5.35 (s, NH), 4.52 (m, 1H), 3.01 (m, 1H), 3.48 (s, 1H), 1.30 (s, 9H), 1.43 (s, 9H).

*Synthesis of Compound **32.*** The yield from **31** above was stirred in a 100 mL round-bottomed flask and flushed with argon. 25 mL of anhydrous CH_2Cl_2 was added. The flask was transferred to a NaCl-containing -12°C ice bath and 4 mmol of DMAP were added. Five minutes later, 4 mmol of EDCD.HCl were added. After twenty minutes later, 4 mmol of HOBt and 4 mmol of Weinreb's amide, NHMe(OMe), were added simultaneously. The reaction mixture was allowed to react for about 2-3 hours. The reaction mixture was extracted with CH_2Cl_2, 1 M HCl, saturated NH_4Cl, saturated $NaHCO_3$, and finally saturated NaCl. Then, it was dried

over $MgSO_4$ and rotavaped to dryness. **32**: ^{1}H NMR (300 MHz, $CDCl_3$) δ 5.32 (s, NH), 4.84 (m, 1H), 3.22 (s, 3H), 3.78 (s, 3H), 2.88 (m, 1H), 1.30 (s, 9H), 1.42 (s, 9H).

*Synthesis of Compound **33***. 2.2 mmol of **32** dissolved in 12 mL of ethyl ether was cooled to 0º C. 24 mmol of The Grignard reagent, MeMgBr, was added dropwise over 10 min. The reaction mixture was allowed to run for 2 hours. Then, saturated NH_4Cl was added to quench MeMgBr until white precipitate of $Mg(OH)_2$ was observed to form. The liquid portion was extracted with CH_2Cl_2, saturated NH_4Cl, saturated $NaHCO_3$, and finally saturated NaCl. Then, it was dried over $MgSO_4$ and rotavaped to dryness. **33**: ^{1}H NMR (300 MHz, $CDCl_3$) δ 5.50 (s, NH), 2.10 (s, 3H), 4.73 (m, 1H), 2.81 (m, 1H), 2.95 (m, 1H), 1.33 (s, 9H), 1.42 (s, 9H).

*Synthesis of Compound **37***. In a 50 mL round-bottomed flask, the purified species **43** was added with 10 mL of 98% pure TFA and 9 mmol of anisole. The reaction mixture was cooled to 0°C before 2.2 mmol of $Hg(OAc)_2$ was added. The mixture was allowed to stir for 2 hours and then was rotavaped to complete dryness in 15 minutes to remove all the TFA. The dry solid was re-dissolved in ether and H_2S was bubbling in very slowly for 1 minute until black precipitate was formed. The whole reaction was filtered through a celite bed and washed with ethyl ether. The filtrate was rotavaped again for 15-20 minutes and **43** was purified by column chromatography with 5% MeOH/chloroform. **43**: ^{1}H NMR (300 MHz, $CDCl_3$) δ 6.75 (m, NH_2), 1.32 (s, 3H), 2.95 (m, 1H), 3.45 (m, 1H), 3.53 (m, 1H), 3.95 (m, 2H), 4.05 (m, 2H).

*Synthesis of Compound **40***. To the starting material **32** (2.2 mmol) in a 50-mL round-bottomed flask 3 mL of 98% pure TFA were added. The whole mixture was allowed to react for 2 hours at 0°C. The product was rotavaped for 20 minutes to get

rid some of TFA was extracted with CH_2Cl_2, saturated NH_4Cl, saturated $NaHCO_3$, and finally saturated NaCl. Then, it was dried over $MgSO_4$ and rotavaped to dryness. **40**: ^{1}H NMR (300 MHz, $CDCl_3$) δ 6.75 (m, NH_2), 3.67 (s, 3H), 2.71 (s, 3H), 3.53 (m, 1H), 3.05 (m, 1H), 3.45 (m, 1H), 1.25 (s, 9H).

*Synthesis of Compound **41***. To the starting material **40** (8 mmol) in a 50 mL round-bottomed flask 15 mL of anhydrous dichloromethane was added. Then 25 mmol of triethylamine was added and followed by 7.1 mmol of TFAA. The whole mixture was allowed to react with stirring at 0°C for 2 hours and then room temperature. The reaction mixture turned yellow after 0.5 hours. The product was extracted with saturated NH_4Cl, saturated $NaHCO_3$, and finally saturated NaCl. Then, it was dried over $MgSO_4$ and rotavaped to dryness. This reaction gave a high yield if left overnight. **41**: ^{1}H NMR (300 MHz, $CDCl_3$) δ 5.50 (s, NH), 3.67 (s, 3H), 2.71 (s, 3H), 4.81 (m, 1H), 3.12 (m, 1H), 2.95 (m, 1H), 1.33 (s, 9H).

*Synthesis of Compound **42***. To the starting material **41** (2.2 mmol), 12 mL of dry ether was added drop-wise to a 100 mL round-bottomed flask containing excess MeMgBr in ice (0ºC). The reaction mixture was allowed to run for 2 hours. 20 mL of saturated NH_4Cl was added to quench MeMgBr. The liquid portion was extracted with CH_2Cl_2, saturated NH_4Cl, saturated $NaHCO_3$, and finally saturated NaCl. Then, it was dried over $MgSO_4$ and rotavaped to dryness. **42**: ^{1}H NMR (300 MHz, $CDCl_3$) δ 6.75 (m, NH_2), 1.32 (s, 3H), 3.53 (m, 1H), 3.05 (m, 1H), 3.45 (m, 1H), 1.33 (s, 9H).

*Synthesis of Compound **43***. The starting material **42** (2 mmol) in a 100 mL round-bottomed flask containing 20 mL of benzene was stirred for a few minutes. 0.6 mmol of p-TsOH and excess of ethylene glycol were added to the flask. The reaction mixture was refluxed for 3 hours at 95ºC. The product was purified by flash column

with 1.25% methanol/chloroform solvent mixture. **43**: ^{1}H NMR (300 MHz, $CDCl_3$) δ 6.75 (m, NH_2), 2.10 (s, 3H), 1.33 (s, 9H), 3.53 (m, 1H), 3.05 (m, 1H), 3.45 (m, 1H), 3.95 (m, 1H), 4.05 (m, 1H), 4.16 (m, 1H), 4.26 (m, 1H).

*Synthesis of Compound **44.*** This compound was synthesized using the same synthetic route as shown in Figure 4.10 with two exceptions. First, in the 3rd step Weinreb's reagent is replaced with ammonia. Second, the 4th step is skipped and excess of TFA was used in the last step. **44**: ^{1}H NMR (300 MHz, D_2O) δ 4.13 (s, 1H), 3.32 (m, 2H).

*Synthesis of Compound **48**.* The procedures are the same as those of **43**. **48**: ^{1}H NMR (300 MHz, $CDCl_3$) δ 1.40 (s, 3H), 6.75 (m, NH_2), 5.75 (m, 1H), 2.95 (m, 1H), 3.12 (m, 1H), 1.12 (s, 9H), 3.65 (m, 2H), 4.75 (m, 2H), 7.50 (m, 2H), 7.81 (m, 1H), 8.18 (m, 1H).

*Synthesis of Compound **49**.* The procedures are the same as those of **37. 49**: ^{1}H NMR (300 MHz, $CDCl_3$) δ 1.40 (s, 3H), 6.60 (m, NH_2), 5.75 (m, 1H), 2.85 (m, 2H), 3.71 (m, 2H), 4.61 (t, 1H), 7.52 (m, 2H), 7.91 (m, 1H), 8.13 (d, 1H).

*Synthesis of Compound **52**.* In a 100 mL round-bottomed flask, 2 mmol of **51** and 1.5 mmol of Ph_3P=CHCOOEt were added. The mixture was refluxed in toluene for 3 hours. The adduct was then rotavaped and **52** was purified by column chromatography using 1.25% MeOH/$CHCl_3$ solvent mixture. **52**: ^{1}H NMR (300 MHz, $CDCl_3$) δ 1.36 (t, 3H), 4.29 (q, 2H), 8.36 (d, 1H), 6.38 (d, 1H), 8.00 (d, 1H), 7.79 (m, 1H), 7.89 (m, 1H), 8.21 (d, 1H).

*Synthesis of Compound **53**.* 2 mmol of **52**, 5 mol% of OsO_4 and 1equivalent of NMO were mixed together in a 100 mL round-bottomed flask containing t-BuOH solvent. The reaction mixture ran for 2 hours, after which **53** was purified by column

chromatography using 5% MeOH/chloroform solvent mixture. **53**: ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.64 (m, 1H), 8.01 (d, 1H), 7.77 (m, 1H), 7.62 (d, 1H), 5.22 (d, 1H), 4.52 (d, 1H), 3.65 (s, OH), 2.85 (s, OH), 4.21 (q, 2H), 1.29 (t, 3H).

*Synthesis of Compound **54***. 1.5 mmol of **53** was reacted with excess amount of NaOH in a 50 mL round-bottomed flask containing MeOH solvent. The reaction proceeded from 1 hour and **54** was purified by t-BuOH-AcOH-H2O (2:1:1, v/v) solvent system. **54**: ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.64 (m, 1H), 8.21 (d, 1H), 7.77 (m, 1H), 7.62 (d, 1H), 4.91 (d, 1H), 4.56 (d, 1H), 3.65 (s, 1H), 2.85 (s, 1H).

*Synthesis of Compound **59***. The procedures are the same as those of **37**. **59**: ^{1}H NMR (300 MHz, $CDCl_3$) δ 4.66 (s, 2H), 7.18 (d, 1H), 8.08 (d, 1H), 7.40 (s, 1H), 5.18 (m, 1H), 4.07 (m, 1H), 4.32 (m, 1H), 1.32 (s, 3H), 5.11 (d, NH_2), 3.53 (m, 1H), 2.90 (m, 1H), 2.65 (m, 1H).

*Synthesis of Compound **61***. 2.2 mmol of the commercially available **60** was dissolved in THF solvent in a 100 mL round-bottomed flask. One equivalent of NaH was added and allowed to run for about 1 hour. Then, 3 mmol of methyl 2-bromoacetate was added and the mixture was let to react for 3 hours. The product was filtered with Whatman paper and purified by column chromatography using 2.5% MeOH/chloroform solvent mixture. **61**: ^{1}H NMR (300 MHz, $CDCl_3$) δ 3.68 (s, 3H), 4.96 (s, 2H), 7.70 (s, 1H), 7.53 (d, 1H), 8.34 (d, 1H), 11.0 (s, 1H).

*Synthesis of Compound **62***. In a 100 mL round-bottomed flask, THF solvent was chilled to -78°C with dry ice. 2 mmol of the starting material **61** was added first to the flask. Then 4 mmol of $Ph_3P{=}CH_2$ was added and followed by 4 mmol of LiHMDS. The reaction ran for 3 hours and the product was purified by column chromatography using 2.5% MeOH/$CHCl_3$ solvent mixture. The yield for this

reaction is generally low, 10-25%. **62**: ^{1}H NMR (300 MHz, $CDCl_3$) δ 3.68 (s, 3H), 4.96 (s, 2H), 7.19 (s, 1H), 7.33 (d, 1H), 8.10 (d, 1H), 5.45 (m, 1H), 5.68 (m, 1H), 7.19 (m, 1H).

*Synthesis of Compound **63***. This dihydroxylation is a variation of Barry/Sharpless's procedure. 2 mmol of **62** was mixed with 5% mol of OsO_4, 1.1 equivalents of NMO in t-BuOH solvent. The reaction ran for about 5 hours and the adduct was filtered with Whatman paper. The filtrate was purified by column chromatography using 10% MeOH/$CHCl_3$ solvent mixture. **63**: ^{1}H NMR (300 MHz, $CDCl_3$) δ 3.68 (s, 3H), 4.96 (s, 2H), 7.40 (s, 1H), 7.18 (d, 1H), 8.08 (d, 1H), 4.59 (t, 1H), 3.88 (m, 1H), 4.13 (m, 1H), 3.65 (s, OH).

*Synthesis of Compound **64***. The procedures are the same as those of **43**. **64**: ^{1}H NMR (300 MHz, $CDCl_3$) δ 3.68 (s, 3H), 4.96 (s, 2H), 7.18 (d, 1H), 8.08 (d, 1H), 7.40 (s, 1H), 5.18 (m, 1H), 4.07 (m, 1H), 4.32 (m, 1H), 1.32 (s, 3H), 5.11 (d, NH_2), 3.53 (m, 1H), 2.79 (m, 1H), 2.54 (m, 1H), 1.33 (s, 9H).

*Synthesis of compound **65***. The procedures are the same as those of **54**. **65**: ^{1}H NMR (300 MHz, $CDCl_3$) δ 4.66 (s, 2H), 7.18 (d, 1H), 8.08 (d, 1H), 7.40 (s, 1H), 5.18 (m, 1H), 4.07 (m, 1H), 4.32 (m, 1H), 1.32 (s, 3H), 5.11 (d, NH_2), 3.53 (m, 1H), 2.79 (m, 1H), 2.54 (m, 1H), 1.33 (s, 9H).

REFERENCES

1. Butterworth, R. F. Thiamin deficiency and brain disorders. *Nutr. Res. Rev.* **2003,** 16, 277-84.
2. Jordan, F. Current mechanistic understanding of thiamin diphosphate-dependent enzymatic reactions. *Nat. Prod. Rep.* **2003,** 20, 184-201.
3. Soriano, J. M.; Molto, J. C.; Manes, J. Dietary Intake and Food Pattern among University Students. *Nutr. Res. (NY)* **2000,** 20, 1249-1258.
4. Park, J. H.; Dorrestein, P. C.; Zhai, H.; Kinsland, C.; McLafferty, F. W.; Begley, T. P. Biosynthesis of the thiazole moiety of thiamin pyrophosphate (vitamin B1). *Biochemistry* **2003,** 42, 12430-8.
5. Dorrestein, P. C.; Zhai, H.; McLafferty, F. W.; Begley, T. P. The biosynthesis of the thiazole phosphate moiety of thiamin: the sulfur transfer mediated by the sulfur carrier protein ThiS. *Chem. Biol.* **2004,** 11, 1373-81.
6. Settembre, E. C.; Dorrestein, P. C.; Zhai, H.; Chatterjee, A.; McLafferty, F. W.; Begley, T. P.; Ealick, S. E. Thiamin biosynthesis in *Bacillus subtilis*: structure of the thiazole synthase/sulfur carrier protein complex. *Biochemistry* **2004,** 43, 11647-57.
7. Dorrestein, P. C.; Huili Zhai, H.; Taylor, S. V.; McLafferty, F. W.; Begley, T. P. The biosynthesis of the thiazole phosphate moiety of thiamin (vitamin B1): the early steps catalyzed by thiazole synthase. *J. Am. Chem. Soc.* **2004,** 126, 3091-6.
8. Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. Synthesis of proteins by native chemical ligation. *Science* **1994,** 266, 776-9.

9. Nahm, S.; Weinreb, S. M. N-methoxy-n-methylamides as effective acylating agents. *Tetrahedron Lett.* **1981,** 22, 3815-3818.

10. Ellman, G. L. Tissue sulfhydryl groups. *Arch. Biochem. Biophys.* **1959,** 82, 70-7.

11. Collier, H. B. Letter: A note on the molar absorptivity of reduced Ellman's reagent, 3-carboxylato-4-nitrothiophenolate. *Anal. Biochem.* **1973,** 56, 310-1.

CHAPTER 5

PRELIMINARY CRYSTALLOGRAPHIC STUDIES OF ENZYMES IN THE METHIONINE BIOSYNTHETIC AND SULFUR ASSIMILATION PATHWAYS OF *WOLINELLA SUCCINOGENES*

Section 5.1 Introduction

One way proteins carry out sulfur transfer is via the sulfate assimilation pathway (Figure 5.1), in which a sulfate ion is somehow incorporated into adenosine monophosphate (AMP) at the 3' hydroxyl groups of the phosphate to form adenosine-5'-phosphosulfate (APS). APS kinase phosphorylates the 3' hydroxyl group of APS to produce 3'-phosphoadenosine-5'-phosphosulfate (PAPS). PAPS reductase, a thioredoxin-dependent protein, reduces PAPS to sulfite, which is taken up by ferridoxin sulfite reductase (FSR) to produce sulfide for incorporation into other proteins or amino acids such as cysteine and methionine. However, the mechanism of sulfur incorporation remains unknown. A BLAST (Altschul *et al.*, 1997) search for the superfamily of sulfur carrier proteins in the protein data base SEED identifies a gene cluster containing genes for proteins in the sulfate assimilation pathway, a sulfur carrier protein, as well as proteins in the methionine biosynthetic pathway of *Wolinella succinogenes*. This empirical observation has led us to hypothesize that the sulfur of methionine originates from the sulfate in the sulfate assimilation pathway. As shown in Figure 5.1, the sulfide generated from the sulfate assimilation pathway is hypothesized to transfer to the adenylated HcyS (a ThiS-like protein in *B. subtilis*), which has undergone a one-residue C-terminal cleavage from the full-length HcyS by some protein and adenylation by another protein. The gene cluster also contains,

among other unidentified proteins, ThiF-like protein (HcyF), QbsD-like protein (HcyD), and O-acetylhomoserine sulfhydrylase (MetY) protein.

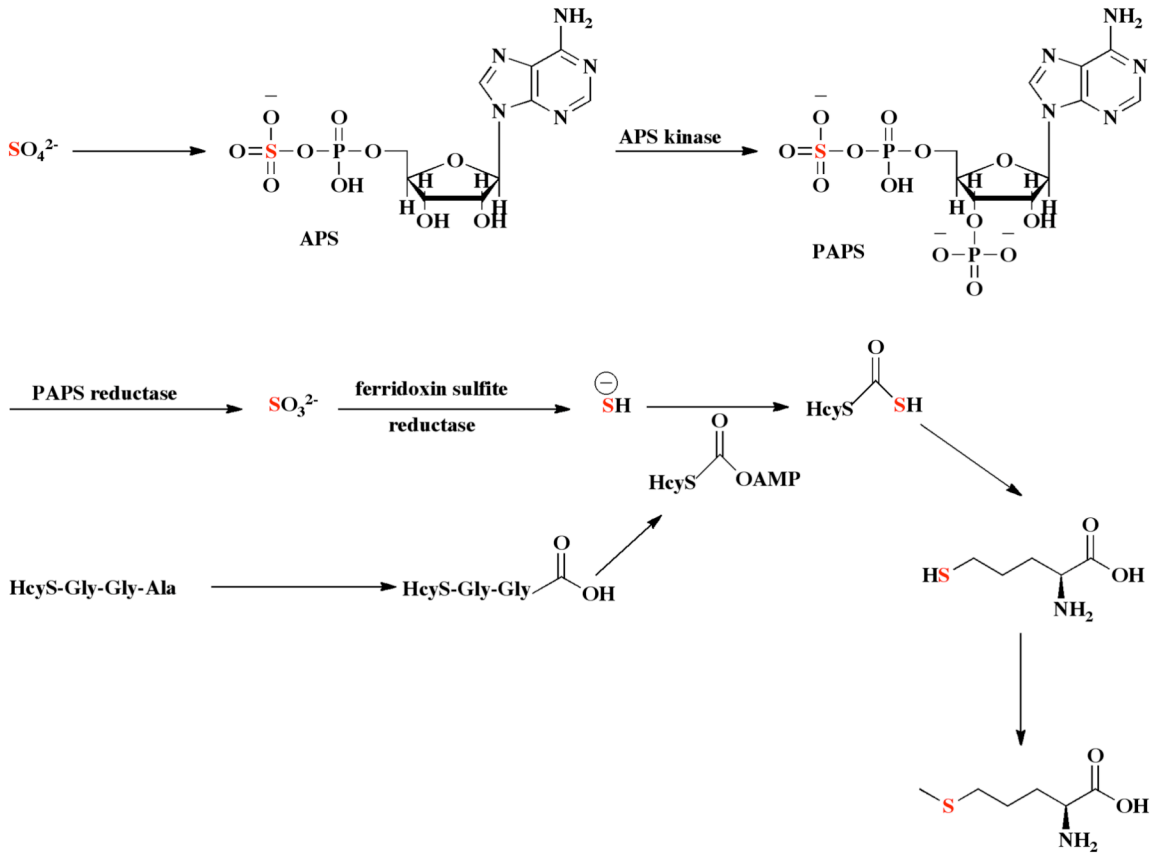


Figure 5.1. Putative sulfur source for methionine biosynthesis in *Wolinella succinogenes* with the full-length sulfur carrier protein HcyS-Gly-Gly-Ala. The transferred sulfur is labeled in red.

In *Escherichia coli* or *Bacillus subtilis*, ThiS was shown to be the sulfur carrier protein (Dorrestein *et al.*, 2004, Settembre *et al.*). QbsD from *Psuedomonas fluoescens* has been characterized recently (Godert *et al.*), although its structure has not been determined. It is a putative metal-dependent hydrolase with a JAMM motif (Burns *et al.*, 2005) based on a BLAST search (Altschul *et al.*, 1997). Thus, HcyD

probably cleaves the last terminal alanine of HcyS. The truncation exposes the gly-gly residue, which is the carboxy terminal signature motif for a family of sulfur carrier proteins like ubiquitin, MoeD, CysO, and ThiS (Godert *et al.*). After the truncation, the carboxy terminus of HcyS inserts into the active site of HcyF for activation or adenylation. *Escherichia coli* ThiF (EcThiF) is known to use ATP to activate EcThiS by adenylating its C-terminal carboxylic group, and thus turning it into a better leaving group (Duda *et al.*). Therefore, HcyF perhaps performs the same function as EcThiF by adenylating HcyS. The activated HcyS would then undergo further transformations to transfer sulfur to other substrates.

It is also known that the incorporation of sulfur into *Bacillus subtilis* ThiS is generally carried out by the enzyme cysteine desulfurase to generate ThiS thiocarboxylate (Dorrestein *et al.*, 2004). Cysteine desulfurase does not exist in this gene cluster. However, MetY is another protein found in this gene cluster. MetY belongs to the family of O-acetylserine sulfhydrylases, which are known to convert O-acetylserine to cysteine using a sulfide. Accordingly, MetY should convert O-acetylhomoserine to homocysteine using a sulfide. Alternatively, MetY in this gene cluster might participate in sulfur transfer from HcyS thiocarboxylate to generate homocysteine. The objective of this project is to obtain crystal structures of key proteins and protein complexes in order to shed light on the mechanism of sulfur transfer in *Wolinella succinogenes*.

Section 5.2 Materials and Methods

Molecular Cloning. Standard methods were used for DNA manipulations (Sambrook *et al.*, 1989, Ausubel & Brent, 1987). Plasmid DNA was purified with the

Fermentas GeneJet miniprep kit and DNA fragments were purified from agarose gel with the Zymoclean Gel DNA Recovery kit from Zymo Research. The *E. coli* strain MachI (Invitrogen) was used as a recipient for transformations during plasmid construction and for plasmid propagation and storage. Phusion DNA polymerase (New England Biolabs) was used for PCR following the manufacturer's instructions. The plasmid THT is a derivative of pET-28, which was obtained from Novagen. Genomic DNA from *Wolinella succinogenes* FDC 602W (ATCC catalog # 29543D-5) was used as a template for PCR. All PCR-derived DNA was sequenced to ensure that the PCR process had not introduced mutations.

Cloning of W. succinogenes hcyS. The *hcyS* gene was amplified from genomic DNA using standard PCR conditions and the following primer pair: 5'-ggg tag cat atg aat ctc atc atc aac gga gag aat aag-3' and 5'-ccc tac tcg agt tat gcg ccc cct ccc atg aaa tat aaa aag-3'. The amplified product containing the gene was excised from the gel, purified and digested with *Nde*I and *Xho*I, repurified, and ligated into similarly digested and purified pTHT.

Cloning of W. succinogenes hcyF. The *hcyF* gene was amplified from genomic DNA using standard PCR conditions and the following primer pair: 5'-ggg tag cat atg aga gag ttt agc gaa gag gag cta g-3' and 5'-ccc tac tcg agt tag gga ttt tga gca tga ttc acc tcg cag atg ggt tgt tc-3'. The amplified product containing the gene was excised from the gel, purified and digested with *Nde*I and *Xho*I, repurified, and ligated into similarly digested and purified pTHT.

Cloning of W. succinogenes metY. The *metY* gene was amplified from genomic DNA using standard PCR conditions and the following primer pair: 5'-ggg tag cat atg agg gga ttc acc acg agg gcg c-3' and 5'-ccc tac tcg agt taa cat agc gct tgc aaa

ata tcc tc-3'. The amplified product containing the gene was excised from the gel, purified and digested with *Nde*I and *Xho*I, repurified and ligated into similarly digested and purified pTHT.

Cloning of W. succinogenes hcyD. The *hcyD* gene was amplified from genomic DNA using standard PCR conditions and the following primer pair: 5'-ggg tag cat atg ctc aaa atc cct aaa gcg ctc ttt g-3' and 5'-ccc tac tcg agt tag atc acc tcg ata ttt tcg gga g-3'. The amplified product containing the gene was excised from the gel, purified and digested with *Nde*I and *Xho*I, repurified, and ligated into similarly digested and purified pTHT.

Cloning of W. succinogenes fsr. The *fsr* gene was amplified from genomic DNA using standard PCR conditions and the following primer pair: 5'-ggg tag cat atg agc cac tac acc ctac ccc cct ccg tcg-3' and 5'-ccc tac tcg agt tat cgt tttt gaa tcc tca cgc gcc-3'. The amplified product containing the gene was excised from the gel, purified and digested with *Nde*I and *Xho*I, repurified, and ligated into similarly digested and purified pTHT.

Section 5.3 Expression and Purification

Each of the plasmids cloned above was transformed into an expression strain BL21 (DE3) *E. coli* cells. An overnight culture of 10 mL was grown in lysogeny broth media at 37°C supplemented with 50 μg/mL kanamycin, and then introduced into 1 L volume containing 50 μg/mL kanamycin. The 1L cell culture was shaken at 37°C and 200 rpm until the OD_{600} reached 0.6, at which point the temperature was reduced to 15°C and 1 mM of isopropyl-1-β-D galactopyranoside was added.

Approximately 16 hours later, the cells were harvested by centrifugation at 7459 *g* for 20 minutes. The pellet was stored at -80 °C.

For each of the corresponding proteins, the same procedures were applied as follows. The frozen cell pellet was thawed overnight at 4°C. Approximately 30 mL of lysis buffer (20 mM Tris pH 8, 10 mM imidazole, and 333 mM NaCl) were added to the pellet to solubilize it. The cell suspension was sonicated and then centrifuged at 47, 488 *g* for 1 hour at 4°C to remove the cell debris. The clarified lysate was loaded onto a pre-equilibrated Ni-NTA gravity column, after which the column was rinsed with 20 column volumes of the same lysis buffer. The protein was eluted with an elution buffer (20 mM Tris pH 8, 250 mM imidazole, and 333 mM NaCl). The eluted protein was loaded directly onto a size exclusion chromatography (Pharmacia G200 superdex, hiload) column for further purification. The protein fractions from the column were pooled together and concentrated to 15-25 mg/mL using an Amicon Ultra centrifugal filter. Its concentration was determined by the Bradford assay (Sambrook *et al.*, 1989). The protein was confirmed by SDS-PAGE showing at least 95% purity. The pure protein was buffer exchanged into 20 mM Tris, pH 8 and 50 mM NaCl, flash frozen with liquid nitrogen and stored at –80°C.

Section 5.4 Crystallization Screening

For the apo-enzymes, the frozen protein was thawed at 4°C. For the holo-enzymes, the thawed protein was incubated overnight with 5 mM of appropriate ligands made in 10 mM Tris buffer pH 8 at a stock-concentration of 1 M. Crystallization experiments were carried out by the vapor diffusion hanging drop method at 22°C, using the sparse matrix screening solutions (Hampton Research,

Emerald Biosystems), as well as PEG 3350/ion screen, PEG 400/pH screen, sodium malonate screen, ammonium sulfate/pH screen, NaCl grid screen, PEG/LiCl screen, PEG 8000/ion screen, and PEG 6000/pH screen. 1.5 μL of the protein solution was combined with an equal volume of well solution. No hits were obtained from these screens. Thus, efforts to co-crystallize the enzymes with their potential ligands were attempted. Only potential ligands for HcyF and MetY are putatively known based on multiple sequence alignments. Rescreening these two proteins with their corresponding ligands (ATP for HcyF and homocysteine/O-acetylhomoserine for MetY) using Wizard I, Wizard II, Hampton Research Crystal Screens 1 and 2 gave two hits with similar conditions, namely Wizard II conditions 39 and 43. Condition 43 of the Wizard II, which contains 10% PEG 8000, 0.1 M Tris pH 7, and 0.2 M $MgCl_2$, gave needles that were 350 Å long and 10 Å in diameter. They were optimized to bigger needle clusters using 12% PEG 8000, 0.1 M Tris pH 7.6, 0.25 M $MgCl_2$, and 5% n-octyl-ß-D-maltoside. Condition 39 of the Wizard II screen, which contains 10% PEG 8000 and 0.1 M imidazole pH 8.0, gave very thin but long plates that looked similar to needles (Figure 5.2).

Another attempt to crystallize these proteins using *in situ* proteolysis gave showers of microcrystals of HcyF from condition 39 of Crystal Screen 2 (0.2 M magnesium chloride hexahydrate, 0.1 M Tris pH 8.5, and 3.4 M 1.6-hexanediol) (Figure 5.3). Trypsin with a concentration of 10 mg/mL was used as the protease in a 1:100 (v/v) ratio (the volume of trypsin was 1 μL and the protein mixture 100 μl). However, the crystals were not reproducible and microseeding did not generate bigger crystals.

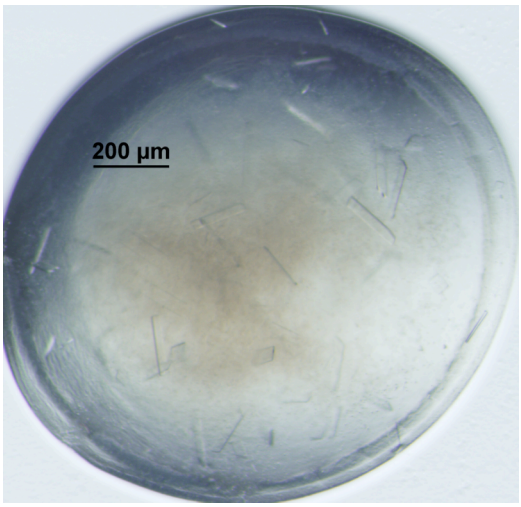


Figure 5.2. Optimized crystals for the MetY/homocysteine complex.

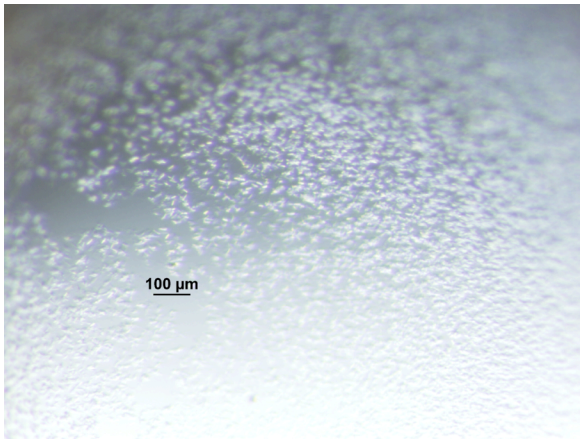


Figure 5.3. In situ proteolysis of HcyF formed microcrystals from condition 39 of Crystal Screen 2.

At the same time attempts to crystallize the apo-enzymes were continued using detergents. However, zero hits from the long list of commercial screens mentioned above indicate that the normal detergent screen method, in which a hit is screened against 24 different detergents, was not possible. To get around this problem, the detergent N, N-dimethyl dodecylamine-N-oxide (LDAO) was chosen because it is one

of the common detergents known to promote crystallization. LDAO was used as an additive against the four commercial screens: Wizard I, Wizard II, Hampton Research Crystal Screen 1, and Crystal Screen 2. One hit was obtained from condition 3 (25% ethylene glycol) of the Hampton Research Crystal Screen 2 for MetY. 3.0 µl of the protein solution was combined with 3.0 µl of well solution. The crystals were very small, ranging from 25 to 50 Å in two dimensions and 50 to 75 Å in the third dimension. These crystals generally have blocky shape and took approximately 90 days to grow (Figure 5.4).

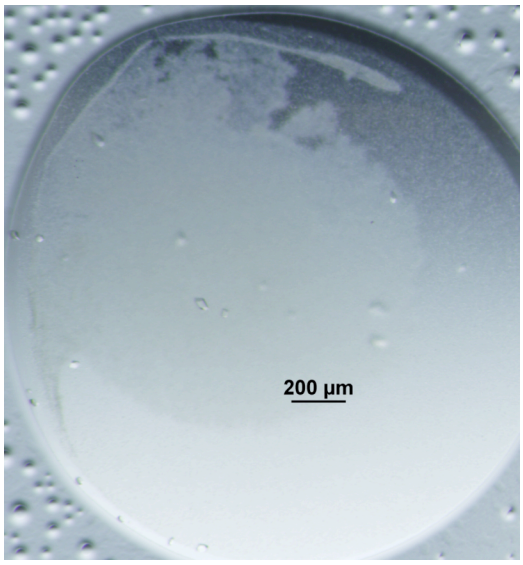


Figure 5.4. Small crystals from the detergent screen after approximately 45 days.

Since the C-terminus of the sulfur carrier protein HcyS must be cleaved at the last residue to expose the gly-gly motif before adenylation can occur, it was hypothesized that HcyS is probably cleaved by the putative metaloprotease HcyD. Attempts to crystallize the HcyS-HcyD complex were carried out. An initial hit was obtained from condition 31 of Wizard I (0.1 M phosphate-citrate pH 4.2, 0.2 M NaCl, and 20% PEG 8000), having long thin needles (Figure 5.5). Crystals took

approximately 16 days to grow and was found to be reproducible. Current efforts are aimed at optimzation.

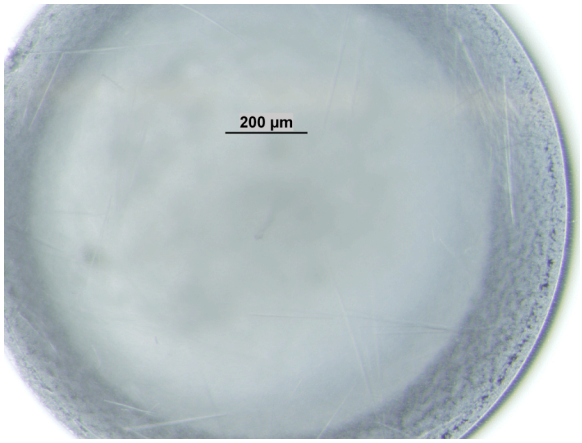


Figure 5.5. Initial hit for the HcyS-HcyD complex.

FSR underwent basic sparse matrix screening as an apo-enzyme, in complex with AMP, and as a protein-protein complex with HcyS. So far only one hit shows needle clusters from condition 35 of Wizard II. Although this condition contains 0.8 M NaH_2PO_4, 1.2 M of K_2HPO_4, and 0.2 M sodium acetate, these needle clusters appear brown suggestive of iron sulfur cluster (Figure 5.6).

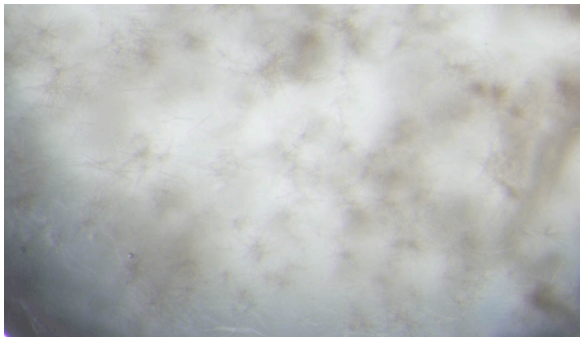


Figure 5.6. Initial hit for apo-FSR.

Section 5.5 Preliminary X-ray Diffraction Analysis and Discussion

The optimized condition 43 of the Wizard II screen (12% PEG 8000, 0.1 M Tris pH 7.6, 0.25 M $MgCl_2$, and 5% n-octyl-ß-D-maltoside) for the homocysteine/MetY complex diffracted to only 8 Å resolution. The crystals from condition 39 of Wizard II of the homocysteine/MetY complex also diffracted to 8 Å resolution. In order to greatly improve the resolution, these crystals were screened against cryoreagents in addition to the commercially available additives from Hampton Research. The only tested cryoreagent that significantly improved the crystal morphology was 2-methylpentane-2,4-diol (MPD) in a narrow range of 5-6%. The resulting crystals (Figure 5.2) were single plates with some thickness. They improved the resolution to about 4 Å and scales in the $P2_1$ space group. Further optimization is under way. The wispy needles of the HcyS-HcyD complex in Figure 5.5 diffracted poorly, and thus they are being optimized as well.

Only the detergent condition of the apo-MetY crystals (Figure 5.4) that diffracted to a relatively high resolution was collected for structure determination. Due to the high concentration of ethylene glycol, no additional cryoprotection was required prior to data collection. X-ray diffraction data from crystals of the apo-enzyme was collected at the F1 beamline outfitted with capillary focusing optics at the Cornell High Energy Synchrotron Source (CHESS), using an Area Detector System Corp. (ADSC) Quantum 270 CCD detector. The data were collected using 0.916 Å radiation over 360° with 1° oscillation range. The data set was indexed, integrated, and scaled using the HKL2000 program suite (Otwinowski & Minor, 1997). The diffraction pattern is shown in Figure 5.7.

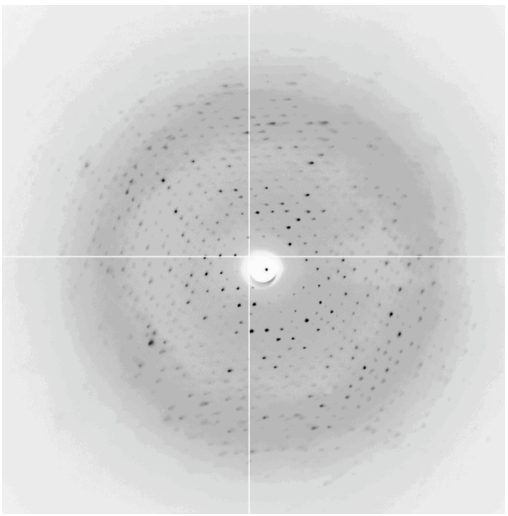


Figure 5.7. Diffraction pattern that diffracted at 2.2 Å but scaled to only 4 Å at best

The resolution went to 2.2 Å but the spots in the higher shells appeared very fuzzy (Figure 5.7). These reflections suggest a dehydration pathology during the crystal transfer process. To overcome this problem, the crystal was soaked into immersion oil before transfer to the goniometer. As seen in Figure 5.8, the blurry spots disappeared and the reflections were well defined. The data were collected and indexed to 2.2 Å. The data collection statistics are shown in Table 5.1.

Table 5.1. Data Collection Statistics for MetY.

	MetY
beamline	F1 CHESS
wavelength (Å)	0.916
resolution (Å)	50-2.2
space group	C2
Molecules / a.s.u.	2
a (Å)	161.8
b (Å)	62.5
c (Å)	91.6
α (°)	90.0
β (°)	120.5
γ (°)	90.0
measured reflections	89132
unique reflections[a]	37506 (1862)
average $\langle I \rangle / \langle \sigma \rangle$[a]	11.0 (2.9)
redundancy[a]	2.4 (2.3)
completeness (%)[a]	96.6 (98.2)
Rsym (%)[a,b]	12.3 (40.1)

[a]Values for the highest-resolution shell are given in parentheses.
[b]Rsym = $\Sigma\Sigma_i \mid I_i - \langle I \rangle \mid / \Sigma \langle I \rangle$, where $\langle I \rangle$ is the mean intensity of the N reflections with intensities I_i and common indices h,k,l.

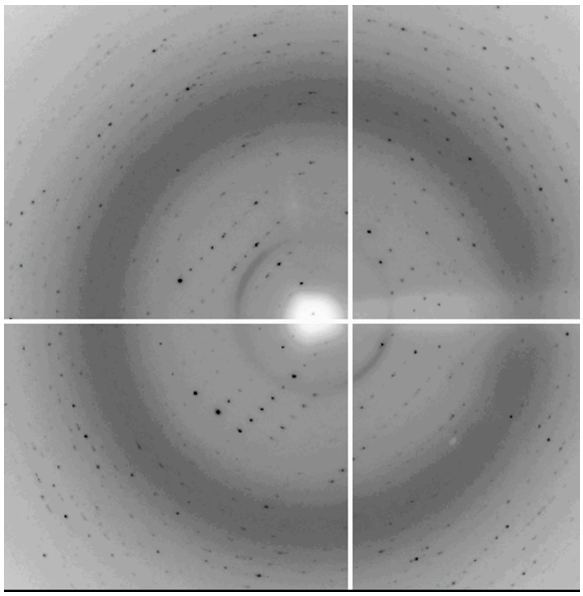


Figure 5.8. The same diffraction pattern after soaking the crystal in immersion oil indexed to 2.2 Å.

The data were indexed to space group C2 with two molecules per asymmetric unit. Size exclusion chromatography indicated a tetrametric biological unit. This experimental observation was supported by applying a two-fold crystallographic symmetry operation to generate a tetramer. Further analysis using the PISA server (Krissinel & Henrick, 2005) confirmed that the most probable oligomeric state for MetY is a tetramer. MetY is being refined using molecular replacement. The current R and R_{free} parameters are 23.0 and 25.7, respectively. Water addition to the electron density map and ligand modeling are under way.

REFERENCES

1. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Nucleic Acids Res. 25, 3389-3402.
2. Ausubel, F. M. & Brent, F. (1987). Current Protocols in Molecular Biology. New York: John Wiley and Sons.
3. Burns, K. E., Baumgart, S., Dorrestein, P. C., Zhai, H., McLafferty, F. W. & Begley, T. P. (2005). J. Am. Chem. Soc. 127, 11602-11603.
4. Dorrestein, P. C., Zhai, H., McLafferty, F. W. & Begley, T. P. (2004). Chem. Biol. 11, 1373-1381.
5. Duda, D. M., Walden, H., Sfondouris, J. & Schulman, B. A. J. Mol. Biol.349, 774-786.
6. Godert, A. M., Jin, M., McLafferty, F. W. & Begley, T. P. J. Bacteriol. 189, 2941-2944.
7. Krissinel, E. & Henrick, K. (2005). CompLife 163-174.
8. Otwinowski, Z. & Minor, W. (1997). Methods Enzymol. 276, 307-326.
9. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual. Plainview, New York.: Cold Spring Harbor Laboratory Press.
10. Settembre, E. C., Dorrestein, P. C., Zhai, H., Chatterjee, A., McLafferty, F. W., Begley, T. P. & Ealick, S. E. Biochemistry 43, 11647-11657.

CHAPTER 6

MECHANISM OF SULFUR TRANSFER OF PROTEIN THIOCARBOXYLATE CATALYZED BY O-ACETYLHOMOSERINE SULFHYDRYLASE IN METHIONINE BIOSYNTHESIS OF *WOLINELLA SUCCINOGENES*

Section 6.1 Introduction

Wolinella succinogenes, originally found in cattle rumen, is a gram-positive bacterium. Being is a member of *Helicobacteraceae hepaticus*, it is a close relative of *Helicobacter pylori* and *Campylobacter jejuni*, which are known to be harmful pathogens in humans and animals. Specifically, *H. pylori* has been linked to ulcers and gastric cancer and *C. jejuni* has been known to cause Guillain-Barre syndrome (*1*). O-acetylhomoserine sulfhydrylase (EC 4.2.99.10) is a sulfide-utilizing enzyme in L-cysteine and L-methionine biosynthetic pathways employed by various enteric bacteria and fungi. OAHS is known to catalyze the conversion of O-acetylhomoserine to homocysteine using hydrogen sulfide from an unknown source (*2*). While the source of sulfur is of great interest, the identification of this source has eluded researchers. Nevertheless, recent biochemical studies in our laboratory (unpublished data) have revealed the likely sulfur source and led to a mechanistic proposal of the new pathway in *Wolinella succinogenes* (Figure 6.1). A gene cluster containing genes for a protein thiocarboxylate, a putative metalloprotease, an adenylating enzyme, sulfur assimilating proteins, and proteins for methionine biosynthesis was found in this species. In this proposed pathway, the sulfur of the sulfate is incorporated into

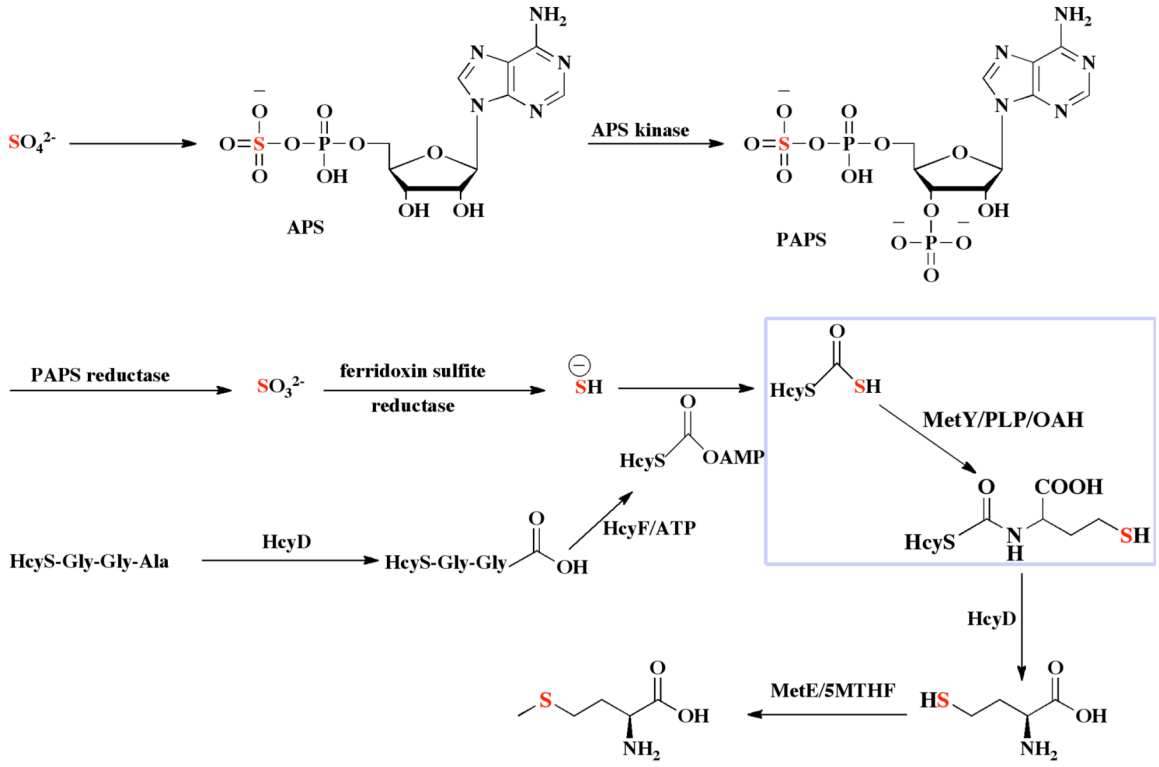


Figure 6.1. Proposed pathway for sulfur assimilation in *Wolinella succinogenes*. The blue box indicates the specific reaction discussed in this study. The transferred sulfur atom is highlighted in red.

adenosine-5'-phosphosulfate (APS) by an unknown enzyme. The APS is phosphorylated at the 3' hydroxyl group by APS kinase to generate 3'-phosphoadenosine-5'-phosphosulfate (PAPS), which is acted upon by PAPS reductase to generate sulfite. Ferridoxin sulfite reductase (FSR) converts the sulfite to sulfide and transfers the sulfide to the activated, truncated HcyS protein, which has been truncated at the C-terminus by HcyD and adenylated by HcyF using ATP. The resulting HcyS thiocarboxylate reacts with O-acetylhomoserine (OAH) catalyzed by MetY to form the HcyS-homocysteine adduct. HcyD cleaves the homocysteine from the adduct, which is then methylated by MetE using 5-methyltetrahydrofolate as the cofactor to produce methionine. As illustrated in Figure 6.1, the sulfur for

homocysteine in this species originated from sulfate. In particular, the sulfur in the reaction catalyzed by MetY comes from a protein thiocarboxylate (HcyS-COSH) rather than from sulfide as previously proposed (*2*).

In this study, we report the first crystal structure of OAHS at 2.2 Å resolution. By comparing its structure with other homologs through DALI (*3*), it was discovered that MetY has the same α/β fold as cystathionine gamma lyase (CGL) and methionine gamma lyase (MGL). In addition, MetY has similar residues in the PLP binding site as the two lyases. Thus, MetY belongs to the so-called Cys/Met metabolism PLP-dependent family of enzymes, which include at least 200 enzymes from various species (*4*). Although these enzymes are found in yeast, bacteria, and plants, none of them are known to exist in humans. Therefore, this enzyme might be a potential target for antibiotic development. This family of enzymes catalyzes a variety of reactions including γ-replacement, γ-elimination, and β-elimination (*5*). The structure of MetY in this study, together with biochemical data, led to sub-classification of MetY as a γ-elimination enzyme and to a mechanistic proposal of sulfur transfer *via* a protein thiocarboxylate catalyzed by MetY.

Section 6.2 Materials and Methods

Molecular Cloning of Native MetY. Standard methods were used for DNA manipulations (*6, 7*). Plasmid DNA was purified with the Fermentas GeneJet miniprep kit and DNA fragments were purified from agarose gel with the Zymoclean Gel DNA Recovery kit from Zymo Research. *E. coli* strain MachI (Invitrogen) was used as a recipient for transformations during plasmid construction and for plasmid propagation and storage. Phusion DNA polymerase (New England Biolabs) was used

for PCR following the manufacturer's instructions. The plasmid THT is a derivative of pET-28, which was obtained from Novagen. Genomic DNA from *Wolinella succinogenes* FDC 602W (ATCC catalog # 29543D-5) was used as a template for PCR. All PCR-derived DNA was sequenced to insure that the PCR process had not introduced mutations.

The *mety* gene was amplified from genomic DNA using standard PCR conditions and the following primer pair: 5'-ggg tag cat atg agg gga ttc acc acg agg gcg c-3' and 5'-ccc tac tcg agt taa cat agc gct tgc aaa ata tcc tc-3'. The amplified product containing the gene was excised from the gel, purified and digested with *Nde*I and *Xho*I, repurified and ligated into similarly digested and purified pTHT.

Expression and Purification of Native MetY. The plasmid described above was transformed into expression strain BL21(DE3) *E. coli* cells. An overnight culture of 10 mL was grown in lysogeny broth media at 37°C supplemented with 50 μg/mL kanamycin, and then introduced into 1 L volume containing 50 μg/mL kanamycin. The 1L cell culture was shaken at 37°C and 200 rpm until the OD_{600} reached 0.6, at which point the temperature was reduced to 15°C and 1 mM of isopropyl-1-β-D galactopyranoside (IPTG) was added. Approximately 16 hours later, the cells were harvested by centrifugation at 7459 *g* for 20 minutes. The pellet was stored at –80 °C.

The frozen cell pellet was thawed overnight at 4°C. Approximately 30 mL of lysis buffer (20 mM Tris pH 8, 10 mM imidazole, and 333 mM NaCl) were added to the pellet to solubilize it. The cell suspension was sonicated and then centrifuged at 47, 488 *g* for 1 hour at 4°C to remove the cell debris. The clarified lysate was loaded onto a pre-equilibrated Ni-NTA gravity column, after which the column was rinsed with 20 column volumes of the same lysis buffer. MetY was then eluted with 10 mL

of elution buffer (20 mM Tris pH 8, 250 mM imidazole, and 333 mM NaCl). The eluted protein was loaded directly onto a size exclusion column (Pharmacia G200 superdex, hiload) for further purification. The protein fractions from the column were pooled together and concentrated to 25-30 mg/mL using an Amicon Ultra centrifugal filter. The MetY concentration was determined by the method of Bradford (*8*). The protein was confirmed to be at least 95% pure by SDS-PAGE. The pure protein was buffer exchanged into 20 mM Tris pH 8 and 50 mM NaCl, flash frozen with liquid nitrogen and stored at -80°C.

Intein Chemistry to Make HcyS-COSH. All chemical reagents were purchased from Sigma-Aldrich (St. Louis, MO). HcyS in pTYB1 was over-expressed in *E. coli* BL21(DE3). Lysogeny broth cultures containing 100 μg/mL of ampicillin were grown at 37°C until an OD_{600} of 0.6-0.8, when the temperature was reduced to 15°C and the cultures were induced with a final concentration of 0.5 mM IPTG. Further growth was carried out at 15°C for 12-16 h with constant agitation. The cultures were harvested by centrifugation and lysed by sonication on ice in 20 mM Tris, 500 mM NaCl, 1 mM EDTA, 0.1% (v/v) Triton X-100, pH 7.8. The samples were then loaded onto a 20 mL bed volume of chitin bead column at a rate of 0.5 mL/min and washed with 300 mL of 20 mM Tris, 500 mM NaCl, 1 mM EDTA, pH 7.8 at rate of 2 mL/min. Cleavage of the protein was carried out at 4°C for 12-16 h with 30 mL of 50 mM Na_2S to give truncated HcyS-COSH. The protein was desalted into 100 mM Tris, 150 mM NaCl, 2 mM TCEP, 30% glycerol, pH 8.0 by dialysis using Novagen D-tube dialyzer Maxi (MWCO 3.5 kDa) and stored as frozen aliquots at -80°C.

Enzymatic Reaction of HcyS-COSH with MetY and OAH. 95 μL of 159 μM truncated HcyS-COSH and 11 μL of 1.4 mM MetY were incubated with 1.25 μL of 10

mM O-acetyl-L-homoserine at room-temperature for 1 h. They were then desalted into 50 mM ammonium acetate, pH 6.0 and analyzed by MALDI-MS. The time-period of incubation was reduced to 2 min when it was observed that the higher molecular-weight 7907 Da adduct was formed.

Crystallization Experiments. The frozen protein was thawed at 4°C and incubated overnight with 5 mM homocysteine dissolved in 10 mM of Tris buffer pH 8 at a stock-concentration of 1 M. Co-crystallization with 5 mM homocysteine and OAH ligands were initially carried out by the hanging drop vapor diffusion method at 22°C using sparse matrix screening solutions (Hampton Research, Emerald Biosystems). 1.5 µl of the protein solution was combined with an equal volume of well solution. The homocysteine/MetY complex initially crystallized in 12% (w/v) PEG 8000, 0.1 M Tris pH 7.6, and 0.25 M $MgCl_2$ and generated crystals that diffracted to 8 Å resolution. Optimization of this condition against different precipitating agents led to a condition consisting of 10% (w/v) PEG 8000, 0.1 M imidazole pH 7.5, 0.2 M $Mg(NO_3)_2$, and 5% MPD. This condition produced a new crystal form that diffracted to 4 Å resolution and scales in the $P2_1$ space group. Further optimization is underway. For the apo-MetY, extensive screenings led to one condition containing 25-30% (w/v) ethylene glycol and 5% (v/v) N-dodecyl-N, N-dimethylamine-N-oxide detergent as an additive. The crystals were small blocks, ranging from 25 to 50 Å in two dimensions and 50 to 75 Å in the third dimension. These crystals took approximately three months to grow and were used for X-ray diffraction experiments.

Data Collection and Processing. Due to the high concentration of ethylene glycol, no additional cryoprotection was required prior to data collection. X-ray

diffraction data from crystals of the apo-enzyme was collected at the F1 beamline outfitted with capillary focusing optics at the Cornell High Energy Synchrotron Source (CHESS), using an Area Detector System Corp. (ADSC) Quantum 270 CCD detector. The data were collected using 0.916 Å radiation over 360° with 1° oscillation range. The data set was indexed, integrated, and scaled using the HKL2000 program suite (*9*). Data collection and statistics are shown in Table 6.1.

Structure Determination. The MetY structure was solved by molecular replacement using MOLREP (*10*) and the TtOAHS structure (2CTZ) as the search model (*11*). CHAINSAW (*12*) was used to remove all side chains of the search model that were not conserved in the sequence alignment with MetY. Refinement was performed using REFMAC (*13*), and alternated with successive manual model building using COOT (*14*) guided by mF_o-DF_c, map, $2f_o$-f_c map, and composite omit electron density maps generated using CNS (*15*) . Water molecules were added after the R_{factor} and R_{free} converged. The geometry of MetY was validated using PROCHECK (*16*). All figures were produced using Pymol and ChemBioDraw (*17*). Refinement statistics are tabulated in Table 6.2.

Table 6.1. Data Collection Statistics for MetY.

	MetY
beamline	F1 CHESS
wavelength (Å)	0.916
resolution (Å)	50-2.2
space group	C2
Molecules / a.s.u.	2
a (Å)	161.8
b (Å)	62.5
c (Å)	91.6
α (°)	90.0
β (°)	120.5
γ (°)	90.0
measured reflections	89132
unique reflections[a]	37506 (1862)
average <I> / <σ(I)>[a]	11.0 (2.9)
redundancy[a]	2.4 (2.3)
completeness (%)[a]	96.6 (98.2)
Rsym (%)[a,b]	12.3 (40.1)

[a]Values for the highest-resolution shell are given in parentheses.

[b]Rsym = $\Sigma\Sigma_i \mid I_i - \langle I \rangle \mid / \Sigma \langle I \rangle$, where $\langle I \rangle$ is the mean intensity of the N reflections with intensities I_i and common indices h,k,l.

Table 6.2. Data Refinement Statistics for MetY.

	MetY
resolution (Å)	50-2.2
no. of protein atoms	5355
no. of water atoms	163
reflections in working set	38264
reflections in test set	2565
R-factor[a] (%)	23.0
R_{free}[b] (%)	25.7
rmsd from ideals	
bonds (Å)	0.017
angles (°)	1.831
avg B factor (Å^2)	23.1
Ramachandran Plot	
most favored (%)	92.7
additionally allowed (%)	7.1
generously allowed (%)	0.0
disallowed (%)	0.2

[a]R-factor = Σ_{hkl}| |F_{obs}| − k |F_{cal}| | / Σ_{hkl} |F_{obs}| where F_{obs} and F_{cal} are observed and calculated structure factors, respectively.

[b]For R_{free}, the sum is extended over a subset of reflections (5%) excluded from all stages of refinement.

Section 6.3 Results

Enzymatic Reaction of HcyS-COSH with MetY/PLP and OAH. MetY was tested as a possible enzyme for homocysteine biosynthesis. Figure 6.2A illustrates this reaction, in which HcyS-COSH was synthesized using intein chemistry. Analysis by MALDI-TOF after incubation of HcyS-COSH with MetY and OAH for 2 minutes, revealed a new peak (Figure 6.2B). A single dominant peak was observed with a molecular weight of 7810 Da in the graph of normalized intensity versus mass per unit charge.

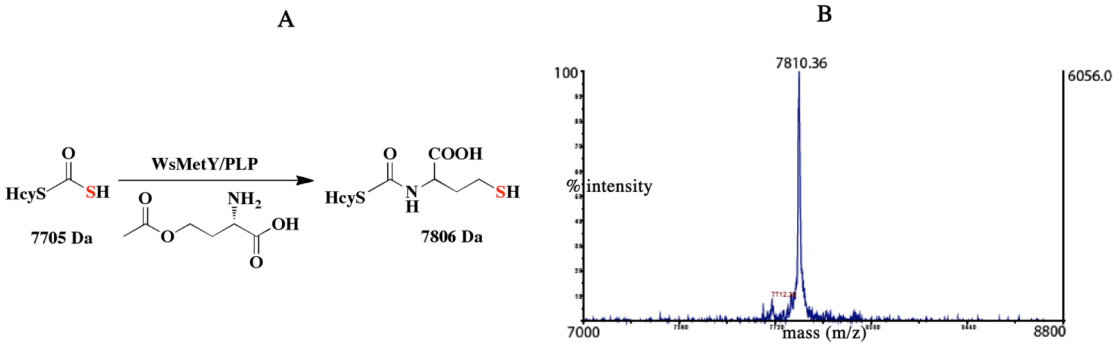


Figure 6.2. (A) The proposed reaction and the expected molecular weights catalyzed by MetY. (B) the MALDI mass spectrum for HcyS-homocysteine adduct.

Crystal Structure of MetY. The unit cell was calculated to have a Matthews coefficient of 2.23 for two molecules in the asymmetric unit, which corresponds to a solvent content of 44.8%. Although only two subunits are found in the asymmetric unit (Figure 6.3A) the biological unit of MetY appears to be a homotetramer (Figure 6.3D), based on size exclusion chromatography data. Each subunit contains an active site and thus there are four active sites per tetramer. These are arranged such that the active site of each monomer faces one side and the two active sites of adjacent

monomers face the opposite side. The dimension of the monomer is approximately 50 x 50 x 70 Å and the tetramer 95 x 85 x 75 Å. The average interface surface area between any two adjacent pair of monomers is about 1018 $Å^2$ for a given average monomeric surface area of 15127 $Å^2$ as calculated by the PISA server (*18*).

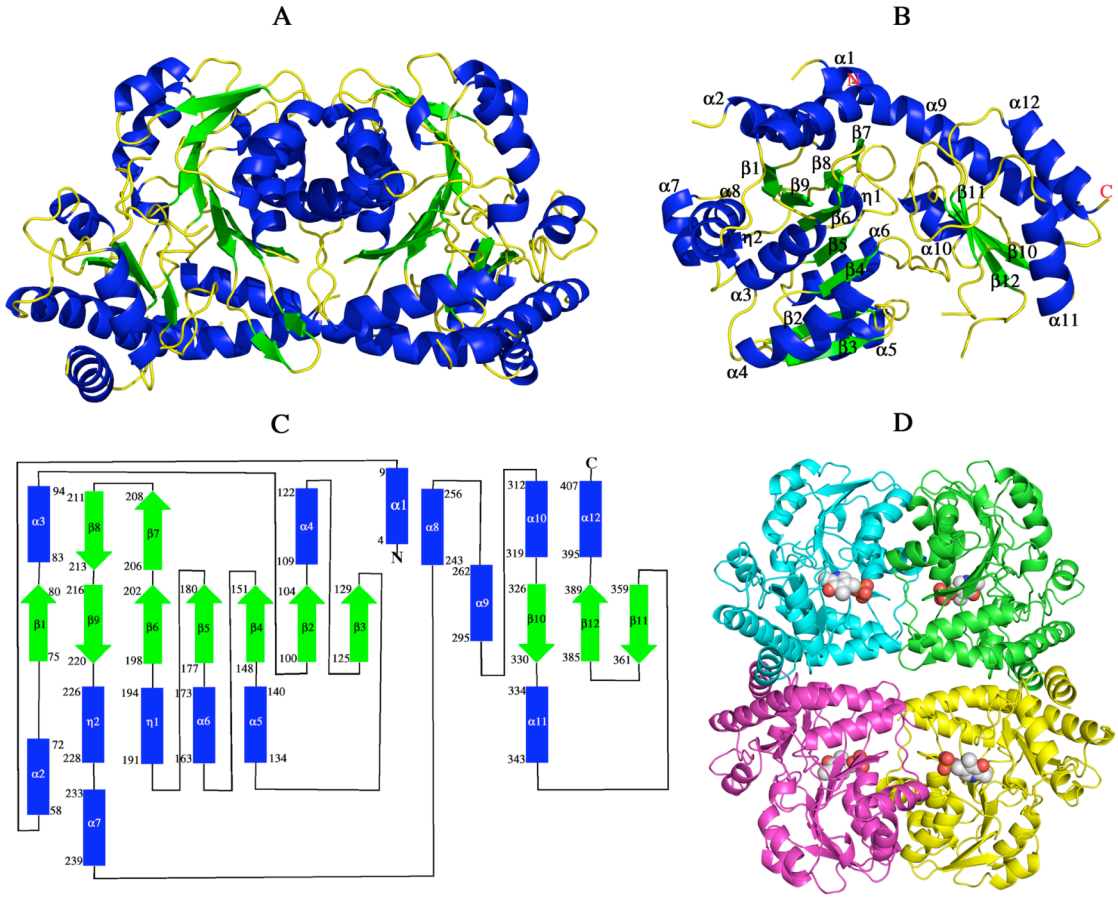


Figure 6.3. Structures of MetY. (**A**) Dimer observed in the asymmetric unit. (**B**) Monomer with labeled and numbered secondary structures. (**C**) Topology diagram containing the same labels as in (**B**. (**D**) The biological tetramer. PLP is represented as spheres. Letters N and C in the topology diagram denote N- and C-termini.

Monomeric Subunit of MetY. The monomer has the α/β fold and consists of a large central seven-stranded beta sheet, a two-stranded sheet, and a three-stranded sheet. All of these sheets are flanked by a total of ten 10 alpha helices and two 3_{10}

helices (Figures 6.3B, 6.3C). Each monomer contains one PLP cofactor and 407 amino acids. The electron density for two segments was absent in the structure of MetY. The longer segment is located at the N-terminus and is comprised of about 50 residues while the smaller one is at the C-terminus and consists of about 10 residues. The monomer is composed of at least two domains. The large domain contains the PLP binding site and is comprised of the seven-stranded beta sheet, which is flanked on both sides by α2, α3, α4, α5, α6, α7, α8, η1, η2, and part of α9. Six of the seven strands, β1-β6, in the large sheet are parallel and one is anti-parallel (β9). The smaller domain resides near the C-terminal end and contains a slightly twisted anti-parallel three-stranded sheet, which is flanked on both sides by α9, α10, α11, and α12. β7 and β8 are continuous with β6 and β9 and extend to form another small sheet. α9 is an extremely long helix, consisting of 34 residues and spanning the two domains. The compact monomer is achieved by the tight packing between the two domains.

Active Site of MetY. While the enzyme was crystallized without any substrates bound, the active site was easily located based on structural homology to *Pseudomonas putida* MGL (PpMGL), and yeast CGL (yCGL) (*5, 19*). The search for the active site was facilitated by the fact that the space occupied by the active site is relatively large and open to the solvent. The stereo drawing and schematic illustration of the active site are shown in Figures 6.4A and 6.4B, respectively. The PLP cofactor was manually positioned in the active site in order to identify important residues. The active site is located near the N-terminus of α3. In the absence of PLP, the active site is filled with water molecule. Without PLP in the active site, Thr204, Lys205, and F108 shift significantly from their assumed positions based on structural alignment of

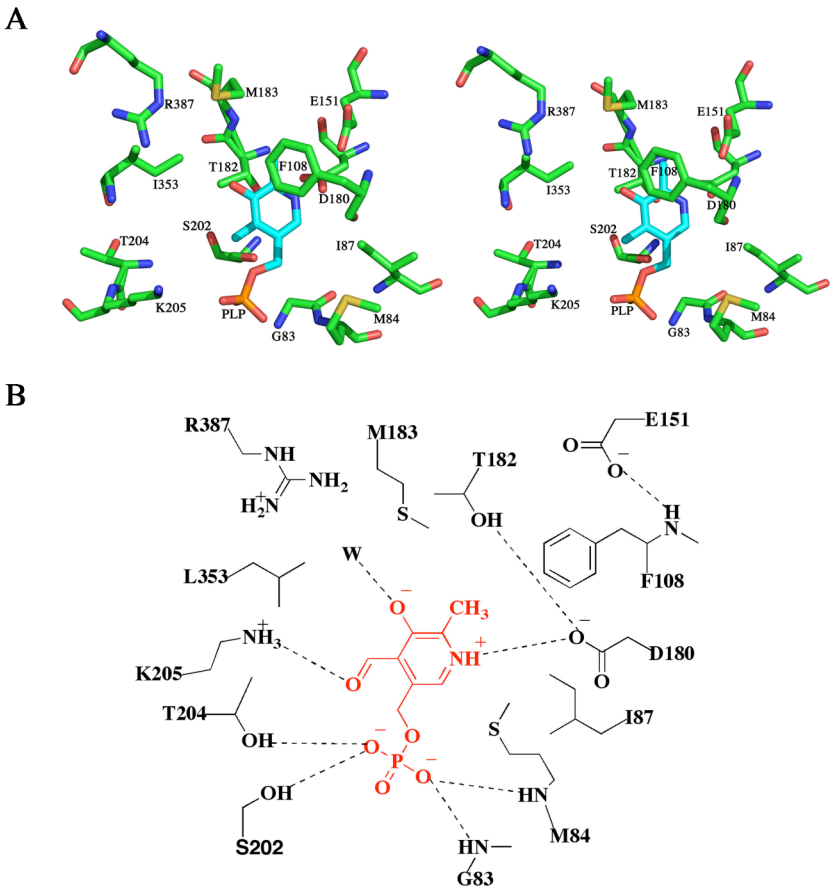


Figure 6.4. The active site of MetY in stereo (**A**) and schematic (**B**). The PLP cofactor shown in cyan and red, respectively, in both panels is manually positioned in the active site based on superposition of structural homologs. The dashed lines indicate possible hydrogen bonds and letter 'W' water molecule.

the homologs. The phosphate group of PLP potentially makes hydrogen bonds with Thr204, Ser202, and the amide hydrogen atoms of Gly83 and Met84. Asp180 is within salt bridge distance to the N1H of PLP. Leu353, Met183, I87, and Met84 could interact with the hydrophobic region of PLP. Glu151 does not appear to interact directly with the PLP but makes a hydrogen bond to Phe108, which could stabilize the pyridine ring of PLP by π-stacking interaction. Thr182 and Ser202 are about 3 Å directly below the plane of the PLP's pyridine ring according to the orientation shown

in Figure 6.4. These two residues hydrogen bond with the π-electron ring of the pyridine ring of PLP. In addition, Thr182 also hydrogen bonds to Asp180. No residues in the active site appear to interact with the methyl group of PLP. The hydroxyl group of PLP hydrogen bonds to an ordered water molecule in the structures of the homologs. Hence, the same might be the case for the PLP of MetY. With exception of the residues missing from the structure, all of the residues that bind PLP originate from the large PLP binding domain.

Section 6.4 Discussion

Enzymatic Reaction of HcyS-COSH with MetY/PLP and OAH. Recent literature (*20-22*) has demonstrated that CysO thiocarboxylate in cysteine biosynthetic pathway of *Mycobacterium tuberculosis* is converted to CysO-cysteine adduct by PLP-dependent O-phospho-L-serine sulfhydrylase or CysM and O-phospho-L-serine. Thus, if MetY catalyzes a similar reaction as that of CysM using PLP as a cofactor, OAH, and HcyS-COSH substrates, then the expected product is the HcyS-homocysteine adduct with a molecular weight of 7806 Da. A single peak of the MALDI spectrum on Figure 6.2B shows the expected mass of 7810 Da, which is well within experimental error. This confirms that MetY utilizes a similar catalytic mechanism as CysM. Furthermore, the HcyS-homocysteine adduct was shown to be cleaved by HcyD to release homocysteine, which was characterized by NMR and HPLC (unpublished data).

Structure of MetY. Although the asymmetric unit contains a dimer, the tetramer can be generated by a two-fold symmetry operation on the dimer. This structure was confirmed by the PISA server (*18*) (Figure 6.3). Furthermore, the

OAHS from *Schizosaccharomyces pombe* was also shown biochemically to be tetrameric (*23*). The arrangement of the subunits of MetY is not uniform and the diagonal subunits of the tetramer do not interact. Each monomer of the tetramer packs differently against two adjacent monomers. Specifically, each monomer of the tetramer interacts intimately with one adjacent monomer primarily via a long loop, α3, α4, and α8. The two active sites of these two monomers are about 20 Å apart and they are equivalent to the so-called active dimer in PpMGL, and yCGL (*5, 19*). The interaction of the same monomer with the other adjacent monomer involves primarily short loops, α1 and α9. The active sites of these latter two monomers are about 37 Å apart.

With the exception of the missing residues in the N-terminus, the monomer contains two domains: the large PLP binding domain and the small C-terminal domain. A search using the DALI server (*3*) identified several structural homologs. The top representative hits are tabulated in Table 6.3. The top hit with a Z-score of 45.4 and aligned sequence identity of 41% comes from TtOAHS, the molecular replacement search model. The next two representative hits with a Z-score of 41.9 belong to the Cys/Met metabolism PLP-dependent family of enzymes, specifically CGL and MGL. This family of enzymes utilizes PLP to carry out β-elimination, γ-elimination, and γ-replacement reactions (*5*). MGL and CGL also function biologically as a tetramer (*5, 19, 24, 25*). The high degree of similarity suggests that MetY most likely performs the same type of γ-elimination reaction as those enzymes. In addition, the missing N-terminal residues in MetY most likely constitute the third domain that was observed in structures of γ-elimination enzymes. These residues are

known to stabilize the active dimer-dimer interface (*19*), which may explain why only one active dimer is found in the asymmetric unit of MetY.

Table 6.3. Enzymes Structurally Similar to MetY as Evidenced by DALI.

protein	**PDB ID**	**Z score**	**rmsd**	**% identical**	**no. aligned residues**	**no. residues**
OAHS	2CTZ	45.4	1.7	41	344	421
CGL	3E6G	41.9	1.5	33	322	368
MGL	1GC2	41.9	1.8	38	325	370

One residue, Tyr187, is in the disallowed region of the Ramachandran plot. Tyr187 has clear electron density at 2.2 Å resolution allowing for its unambiguous placement. This bulky side chain is part of the helix turn and this probably explains why it was slightly twisted into an unfavorable conformation.

Comparison of MetY with its Structural Homologs. Primary sequence alignment (*26, 27*) of MetY against TtOAHS, PpMGL, and yCGL (Figure 6.5) shows that these sequences have relatively high sequence identity. The first 58 residues vary the most among the four species. This N-terminal domain was proposed to play a role in substrate recognition among the MGL family of enzymes (*19*). An insertion of 19 and 26 amino acids for MetY and TtOAHS, respectively, starting at residue 241 of MGL may be unique for OAHSs. The result of this insertion is very obvious when viewing superimposed structures of these enzymes (Figure 6.6). The insertion in MetY covers $\eta 2$, $\alpha 7$, and the loop between them and might differentiate OAHSs from other Cys/Met metabolism PLP-dependent enzymes. The conserved residues marked

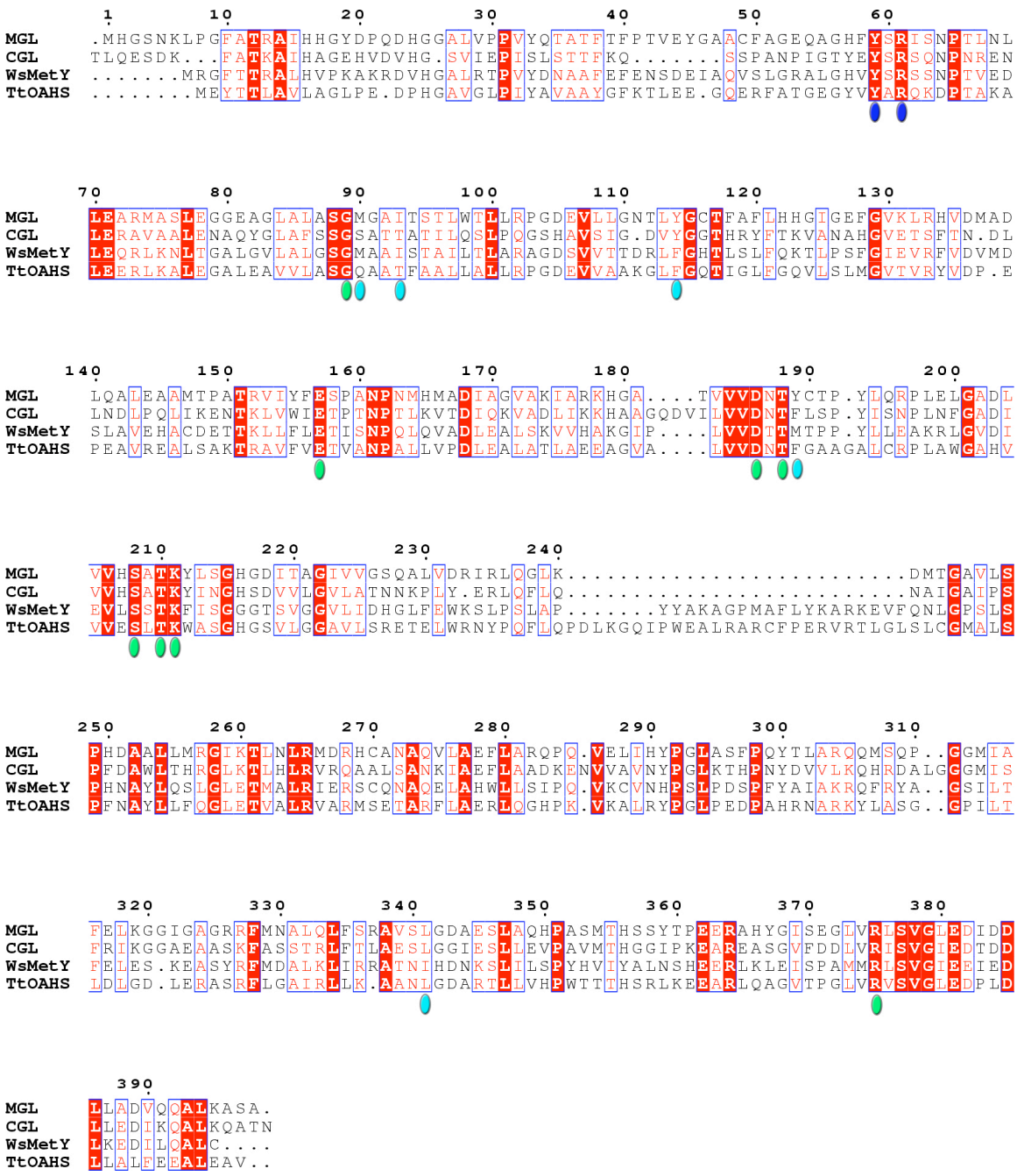


Figure 6.5. Sequence alignment of MetY with TtOAHS (**2CB1**), yeast CGL (**1N8P**), and PpMGL (**2O7C**). The conserved active sites residues are marked with green ellipses, and other active site residues of MetY with cyan ellipses. The two conserved residues that bind to the phosphate group of adjacent monomer's PLP are marked with blue ellipses. The numbering scheme is based on the sequence of PpMGL.

with blue ellipses are only observed in the MGL and CGL structures (*5, 19*) but not in TtOAHS and MetY. These residues of PpMGL and yCGL bind to the phosphate group of PLP of an adjacent monomer. Hence, the equivalent residues from each monomer of TtOAHS and MetY probably bind to the phosphate group of PLP of an adjacent monomer as well. This implies that each subunit requires two conserved residues from the adjacent monomer of the active dimer. Thus, the minimal functional unit is the active dimer. Figure 6.6 demonstrates that these enzymes have nearly identical structures despite sharing only 33-41 percent identity. As illustrated in Table 6.3**,** the r.m.s. distance ranges from 1.5 to 1.8 Å for 322 to 344 aligned residues. These results suggest that MetY, MGL, TtOAHS, and CGL probably evolved from a common ancestral enzyme.

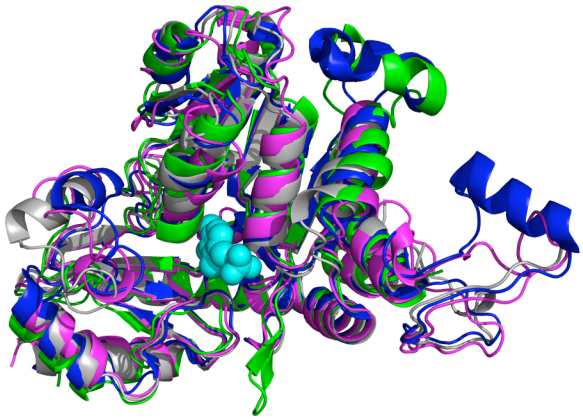


Figure 6.6. Structural superposition of the monomer of MetY (green) with that of TtOAHS (2CB1, blue), yeast CGL (1N8P, magenta), and PpMGL (2O7C, grey). The PLP molecule is shown using spheres.

Comparison of the MetY Active Site with Structural Homologs. Since MetY, TtOAHS, PpMGL, yCGL share the same fold it is of interest to compare their active sites. In fact, superposition of the active sites (Figure 6.7) of these homologs and

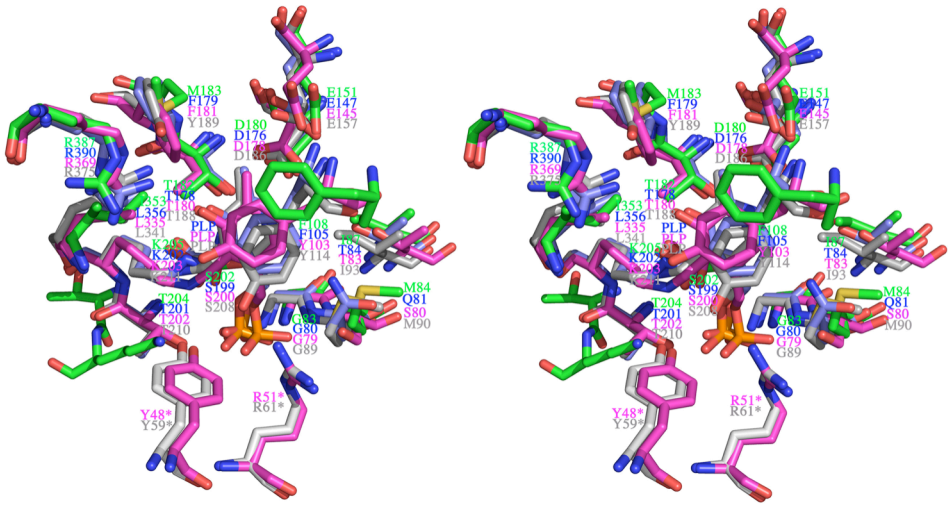


Figure 6.7. Superposition of the active sites of MetY with the same set of enzymes in Figure 6.6 using the same coloring scheme, except that the color of PLP is the same as its corresponding enzyme. Residues marked with an asterisk come from a neighboring monomer.

MetY allowed the identification of the active site of MetY. As expected, the active site residues are relatively conserved. In addition to our biochemical data, which show that MetY needs PLP for activity, MetY appears yellow in the color of PLP throughout the purification procedures. PLP must have dissociated from the active site during crystallization under these conditions. By comparing the active site residues of MetY with TtOAHS (2CB1), PpMGL (*19*), and yCGL (*5*), the active site residues of MetY can be easily identified. The equivalent residues to Phe108 of MetY in the other three structures form π-stacking interactions with the pyrimidine ring of PLP, and thus increase the electron sink capacity of PLP (*28, 29*). Hence, Phe108 is presumed to possess the same role. The conserved Glu151 does not interact directly with the PLP but hydrogen bonds to the amide backbone of Phe108 and positions it in the active site. In the absence of PLP, Phe108 is oriented slightly off from the pyrimidine ring of PLP. The conserved Thr182 and Ser202 form van der Waals

interactions with the PLP's pyridine ring as well as hydrogen bonds to the π-electron ring. Together with Phe108, Thr182 and Ser202, they sandwich or lock PLP in the active site. Thr182 also hydrogen bonds to the amide hydrogen of the conserved Asp180 to presumably position it to make the salt bridge with N1H of PLP. This presumed salt bridge helps increase the electrophilicity of the pyridine ring of PLP. The equivalent residues of the conserved Lys205 covalently link to PLP, and those of the conserved Thr204 and Ser202 residues make hydrogen bonds with the phosphate group of PLP. As for Phe108, both Lys205 and Thr204 move away from the active site in the absence of PLP. In fact, the active site is wide open to the solvent in the absence of PLP as shown by surface representation in Figure 6.8. The backbone amide groups of the Met84 and the conserved Gly83 are in position to hydrogen bond to the phosphate group of PLP. The conserved R387 makes no direct contact with PLP but its equivalents in PpMGL and yCGL are known to bind to the carboxylate group of incoming substrates (*5, 19*). Leu353, Met183, I87, and Met84 are in position to interact with the hydrophobic region of PLP. No residues in the active site appear to interact with the methyl group of PLP.

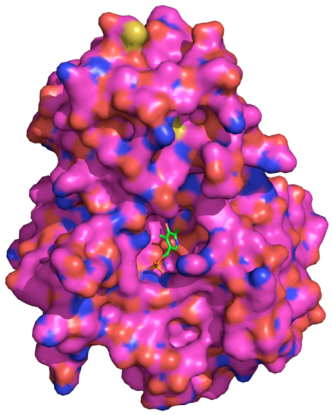


Figure 6.8. Surface representation of a subunit of MetY showing a PLP (sticks) molecule manually positioned in the active site.

Mechanistic Implications. The primary sequence and structural alignments of MetY against its homologs suggests that MetY belongs to the Cys/Met metabolism PLP-dependent family of enzymes, which includes CysO and CysM. Our biochemical and structural data suggest that HcyS and MetY can perform the same catalytic mechanism of sulfur transfer in making homocysteine and subsequently methionine. Furthermore, the active site of MetY of the tetramer is very accessible to the C-terminus of HcyS (Figure 6.8). If MetY and HcyS were to form a complex, the ratio of the two must be 1:1. From the biochemical data and the structure of MetY, the mechanism of sulfur transfer to generate the HcyS-homocysteine adduct is proposed (Figure 6.9**)**. In this proposal, Lys205 of MetY is covalently linked to PLP by an internal aldimine **1** which, as seen in its structural homologs, reacts with OAH to generate **2**. All the nucleophiles are assumed to be deprotonated by a nearby base before adding to an electrophile. Intermediate **2** liberates Lys205 to produce **3**. The external aldimine **3** is tautomerized to generate **4,** and then **5**, which undergoes γ-elimination to make **6**. HcyS thiocarboxylate attacks the activated conjugated double bond to form **7**, which is tautomerized to **8,** and then **9**. Lys205 covalently reattaches to PLP (**10**) to regenerate the internal aldimine **1** and release **11**, which undergoes S, N-acyl shift to form the HcyS-homocysteine adduct.

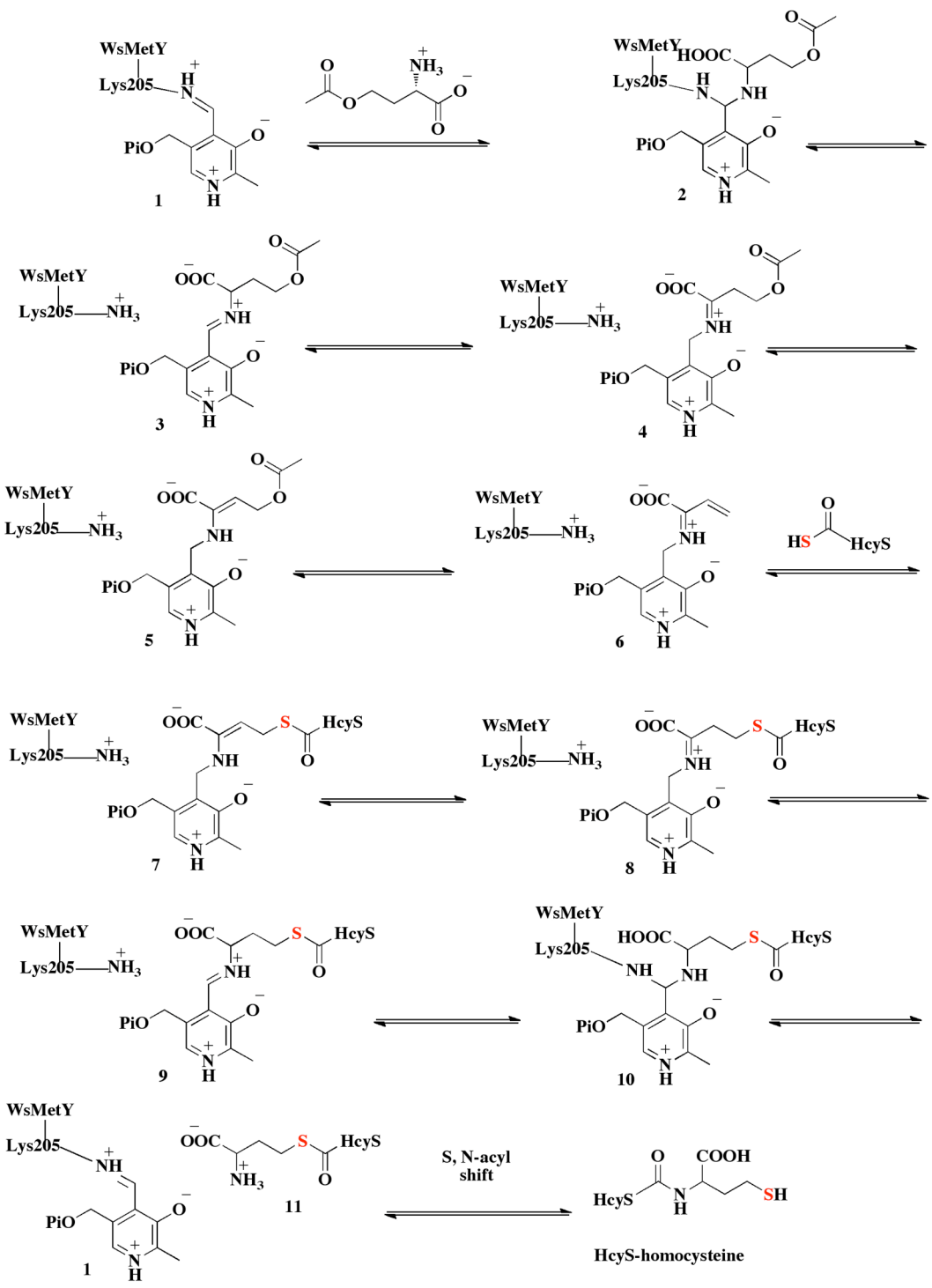


Figure 6.9. Mechanistic proposal for the sulfur transfer from HcyS-COSH to OAH catalyzed by MetY. The sulfur atom is highlighted in red for clarity.

Significance. This mechanistic study of MetY, together with the structure of TtOAHS, suggests that MetY, and perhaps all OAHSs, belong to the Cys/Met metabolism PLP-dependent family of enzymes. The facts that CysM does not have the same fold nor active site geometry as those of MetY (*22, 30*), is about 85 residues shorter than MetY, and catalyzes β-elimination of CysO (*22*) suggest that MetY belongs to the γ-elimination subclass of the Cys/Met metabolism PLP-dependent family of enzymes. In other words, our studies uncover the ability of MetY, and perhaps all OAHSs, to catalyze γ-elimination reactions in addition to the known function of direct sulfide utilization to produce homocysteine *via* the direct sulfhydrylation pathway. This study also broadens our understanding of the mechanism by which the class of protein thiocarboxylate, such as HcyS-COSH, transfers sulfur to small molecules.

REFERENCES

1. Mileni, M., MacMillan, F., Tziatzios, C., Zwicker, K., Haas, A. H., Mantele, W., Simon, J., and Lancaster, C. R. (2006) Heterologous production in *Wolinella succinogenes* and characterization of the quinol: fumarate reductase enzymes from *Helicobacter pylori* and *Campylobacter jejuni, Biochem. J. 395*, 191-201.
2. Yamagata, S. (1989) Roles of O-acetyl-L-homoserine sulfhydrylases in micro-organisms, *Biochimie 71*, 1125-1143.
3. Holm, L., and Sander, C. (1998) Touring protein fold space with Dali/FSSP, *Nucleic Acids Res. 26*, 316-319.
4. Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M., and Sonnhammer, E. L. (2002) The Pfam protein families database, *Nucleic Acids Res. 30*, 276-280.
5. Messerschmidt, A., Worbs, M., Steegborn, C., Wahl, M. C., Huber, R., Laber, B., and Clausen, T. (2003) Determinants of enzymatic specificity in the Cys-Met-metabolism PLP-dependent enzymes family: crystal structure of cystathionine gamma-lyase from yeast and intrafamiliar structure comparison, *Biol. Chem. 384*, 373-386.
6. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Vol. 3, Cold Spring Harbor Laboratory Press, Plainview, New York.

7. Ausubel, F. M., and Brent, F. (1987) *Current Protocols in Molecular Biology*, John Wiley and Sons, New York.

8. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, *Anal. Biochem. 72*, 248-254.

9. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, *Methods Enzymol. 276*, 307-326.

10. Vagin, A., and Teplyakov, A. (2000) An approach to multi-copy search in molecular replacement, *Acta Crystallogr. D Biol. Crystallogr. 56*, 1622-1624.

11. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank, *Nucleic Acids Res. 28*, 235-242.

12. Schwarzenbacher, R., Godzik, A., Grzechnik, S. K., and Jaroszewski, L. (2004) The importance of alignment accuracy for molecular replacement, *Acta Crystallogr. D Biol. Crystallogr. 60*, 1229-1236.

13. Collaborative Computational Project-Number 4. (1994) The CCP-4 suite: programs for protein crystallography, *Acta. Crystallogr. D 50*, 760-763.

14. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr. D Biol. Crystallogr. 60*, 2126-2132.

15. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography &

NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr. D 54*, 905-921.

16. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283-291.

17. DeLano, W. L. (2002) Unraveling hot spots in binding interfaces: progress and challenges, *Curr. Opin. Struct. Biol. 12*, 14-20.

18. Krissinel, E., and Henrick, K. (2005) Detection of Protein Assemblies in Crystals, in *CompLife*, pp 163-174, Springer-Verlag, Berlin.

19. Kudou, D., Misaki, S., Yamashita, M., Tamura, T., Takakura, T., Yoshioka, T., Yagi, S., Hoffman, R. M., Takimoto, A., Esaki, N., and Inagaki, K. (2007) Structure of the antitumour enzyme L-methionine gamma-lyase from *Pseudomonas putida* at 1.8 Å resolution, *J. Biochem. 141*, 535-544.

20. Burns, K. E., Baumgart, S., Dorrestein, P. C., Zhai, H., McLafferty, F. W., and Begley, T. P. (2005) Reconstitution of a new cysteine biosynthetic pathway in *Mycobacterium tuberculosis*, *J. Am. Chem. Soc. 127*, 11602-11603.

21. O'Leary, S. E., Jurgenson, C. T., Ealick, S. E., and Begley, T. P. (2008) O-phospho-L-serine and the thiocarboxylated sulfur carrier protein CysO-COSH are substrates for CysM, a cysteine synthase from *Mycobacterium tuberculosis*, *Biochemistry 47*, 11606-11615.

22. Jurgenson, C. T., Burns, K. E., Begley, T. P., and Ealick, S. E. (2008) Crystal structure of a sulfur carrier protein complex found in the cysteine biosynthetic pathway of *Mycobacterium tuberculosis*, *Biochemistry 47*, 10354-10364.

23. Yamagata, S. (1984) O-Acetylhomoserine sulfhydrylase of the fission yeast *Schizosaccharomyces pombe*: partial purification, characterization, and its probable role in homocysteine biosynthesis, *J. Biochem. 96*, 1511-1523.

24. Motoshima, H., Inagaki, K., Kumasaka, T., Furuichi, M., Inoue, H., Tamura, T., Esaki, N., Soda, K., Tanaka, N., Yamamoto, M., and Tanaka, H. (2000) Crystal structure of the pyridoxal 5'-phosphate dependent L-methionine gamma-lyase from *Pseudomonas putida*, *J. Biochem. 128*, 349-354.

25. Nikulin, A., Revtovich, S., Morozova, E., Nevskaya, N., Nikonov, S., Garber, M., and Demidkina, T. (2008) High-resolution structure of methionine gamma-lyase from *Citrobacter freundii*, *Acta Crystallogr. D Biol. Crystallogr. 64*, 211-218.

26. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res. 22*, 4673-4680.

27. Gouet, P., Courcelle, E., Stuart, D. I., and Metoz, F. (1999) ESPript: analysis of multiple sequence alignments in PostScript, *Bioinformatics 15*, 305-308.

28. Clausen, T., Huber, R., Prade, L., Wahl, M. C., and Messerschmidt, A. (1998) Crystal structure of *Escherichia coli* cystathionine gamma-synthase at 1.5 A resolution, *EMBO J. 17*, 6827-6838.

29. Hayashi, H., Inoue, Y., Kuramitsu, S., Morino, Y., and Kagamiyama, H. (1990) Effects of replacement of tryptophan-140 by phenylalanine or glycine

on the function of *Escherichia coli* aspartate aminotransferase, *Biochem. Biophys Res. Commun. 167*, 407-412.

30. Agren, D., Schnell, R., and Schneider, G. (2009) The C-terminal of CysM from *Mycobacterium tuberculosis* protects the aminoacrylate intermediate and is involved in sulfur donor selectivity, *FEBS Lett. 583*, 330-336.

CHAPTER 7

SUMMARY AND CONCLUSIONS

The work from Chapter 2 reports the crystal structures of SpUP in complex with the products (R1P/Ura) as well as the substrate (Urd) determined at 1.8 Å and 2.2 Å resolution, respectively. The structures reveal that SpUP has the same overall monomeric fold as that of NP-I family of enzymes. Comparing the active site of SpUP with EcUP suggests that the two enzymes have the same transition state. However, SpUP utilizes different residues to stabilize the negative charge on the base. A multiple sequence alignment of SpUP against representative UPs reveals that among all the representative species in the alignments, 22 out of 53 sequences contain the Lys and His residues equivalent to K162 and H169 of SpUP. None of these have the equivalents of Arg168 and Arg223 of EcUP. The other 31 sequences, including EcUP, show the reverse. These 31 enzymes possess the two equivalent arginine residues, but none of them have the equivalent K162 and H169 of SpUP. These results not only suggest a new way of clustering of UP's but also uncovers a subclass of UP's that utilize different active-site residues to stabilize the transition state. The two discriminating residues, Lys162 and His169, together with UP's specificity loop may be used to distinguish between UP and PNP enzymes at the level of primary sequence, and thus may allow proper annotation.

The crystal structures of the bPNP/M_1G-dR/PO_4 and bPNP/M_1G-dR/SO_4 complexes in Chapter 3 were determined at 2.4 Å resolution. Although the electron density for the sugar moiety for both complexes is weak, it clearly shows the presence of M_1G in the active site. One surprising result was that the sulfate ion cleaved M_1G-dR into a glycal intermediate and the M_1G base. The active sites of bPNP/M_1G-dR/PO_4 and bPNP/M_1G-dR/SO_4 complexes were described individually and compared. The comparison reveals that both the phosphate and the sulfate ions can

cleave M_1G-dR. A comparison of the structure of the bPNP/M_1G-dR/SO_4 complex with the EcPNPinosine/SO_4 complex, showed that the Glu201 residue of bPNP, which is nearly coplanar with M_1G in the base binding pocket, could not make an indispensable hydrogen bond to M_1G. This most likely destabilized M_1G-dR and facilitated the unexpected cleavage by the sulfate ion. This unfavorable van der Waals interaction most likely induced the bond cleavage and prevented the bond formation. These comparisons may explain why bPNP can only catalyze the cleave of M_1G-dR but not its synthesis.

Chapter 4 shows step-by-step the entire design and synthesis of a light-activated inhibitor based on the known crystal structure of the ThiS-ThiG complex. The analog was successfully synthesized in relatively good yield measured by Ellman's assay. We also demonstrated the proof-of-concept to covalently link an aminothiol ligand to ThiS using intein chemistry. In particular, we have shown that ThiS can be covalently linked to compounds **37**, **44**, and **46** to give **39**, **45**, and **47**, respectively, and confirmed the resulting covalent adducts using ESI-MS and FT-MS. The objective of future studies is to attach compound **59** to ThiS using intein chemistry to generate **69**. Photolysis of **69** using Hanovia lamp in the presence of ThiG should regenerate methyl ketone **27** and facilitate the Schiff base linkage with Lys96 of ThiG to produce **28**. Sodium borohydride reduction of **28** will trap the proposed analog **4**, which will be confirmed by SDS PAGE, FT-MS, and X-ray crystallography.

In Chapter 5, the crystallization and diffraction screenings of selected enzymes in the methionine biosynthetic and sulfur assimilation pathways of *Wolinella succinogenes* were described. The crystallization conditions for the MetY/homocysteine complex was optimized from 8 Å to 4 Å resolution by screening with different cryoreagents. *In situ* proteolysis, gave showers of microcrystals for

HcyF protein. However, these crystals were not reproducible by this method and microseeding from these microcrystals, so far, has not been not successful. The condition for HcyS-HcyD complex generates reproducible needles but these crystals diffracted poorly. Initial sparse matrix screens gave one hit from condition 35 of Wizard II. Although this condition contains 0.8 M NaH_2PO_4, 1.2 M of K_2HPO_4, and 0.2 M sodium acetate, these needle clusters appear brown suggestive of iron sulfur cluster. Detergent screens for the apo-MetY produced small crystals that diffracted to 2.2 Å resolution. These crystals were used for X-ray diffraction experiments. However, the initial data set appeared very fuzzy at high resolution, suggesting a dehydration pathology. To overcome the problem, the crystals were soaked into immersion oil before transfer to the goniometer. The images became clear and thus the data were collected and indexed to space group C2. Further optimizations of the rest of the hits are in progress.

Chapter 6 reports the first crystal structure of OAHS or MetY at 2.2 Å resolution. By comparing its structure with other homologs, it was discovered that MetY has nearly a identical α/β fold to CGL and MGL. In addition, MetY has similar residues in the PLP binding site as the two lyases. Thus, MetY belongs to the Cys/Met metabolism PLP-dependent family of enzymes. The fact that CysM has meither the same fold nor active site geometry as MetY, is about 85 residues shorter than MetY, and catalyzes β-elimination of CysO suggest that MetY belongs to the γ-elimination subclass of the Cys/Met metabolism PLP-dependent family of enzymes. In other words, our studies uncover the ability of MetY, and perhaps all OAHSs, to catalyze γ-elimination reactions in addition to the known function of direct sulfide utilization to produce homocysteine *via* the direct sulfhydrylation pathway. This study also broadens our understanding of the mechanism by which the class of protein thiocarboxylate, such as HcyS-COSH, transfers sulfur to small molecules.