CRYSTALLOGRAPHIC STUDIES ON ENZYMES INVOLVED IN VITAMIN B_1 BIOSYNTHESIS AND THE DEGRADATION OF VITAMIN B_6

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

Kathryn M. McCulloch

May 2010

CRYSTALLOGRAPHIC STUDIES ON ENZYMES INVOLVED IN VITAMIN B_1 BIOSYNTHESIS AND THE DEGRADATION OF VITAMIN B_6

Kathryn M. McCulloch, Ph. D.

Cornell University 2010

Macromolecular crystallography allows for the structures of biologically interesting molecules to be determined to atomic resolutions. The enzymes involved in metabolic pathways often catalyze interesting and novel chemical reactions, and understanding the structure of these enzymes can offer insight into the function and mechanism of the enzyme. In the thiamin biosynthetic pathway, the two moieties of thiamin are biosynthesized separately and then stitched together by thiamin phosphate synthase. Recent work has shown that one of the heterocyclic moieties is carboxylated; the crystal structure and kinetic characterization of *Bacillus subtilis* thiamin phosphate synthase with carboxylated thiazole phosphate and the pyrimidine moiety demonstrate the activity of this enzyme with this substrate. Additionally, thiamin phosphate synthase is an ideal drug target in some pathogens as thiamin is an essential cofactor and humans do not have this enzyme. The crystal structure of thiamin phosphate synthase from *Mycobacterium tuberculosis* was determined and used for identifying potential drug compounds. The final step in thiamin biosynthesis is the phosphorylation of thiamin phosphate to produce the biologically active thiamin pyrophosphate and the crystal structure of *Aquifex aeolicus* thiamin monophosphate kinase has shown that this reaction proceeds via an in-line phosphate transfer.

While much work has been devoted to the study of the biosynthesis of cofactors, the degradation pathways can also contain interesting chemistry. The

degradation of PLP occurs over seven steps in *Mesorhizobium loti* MAFF303099 and the final three steps have been structurally characterized. 2-Methryl-3-hydroxypyridine-4,5-dicarboxylate decarboxylase shows significant similarity to class II aldolases and uses a retro-Aldol type reaction to catalyze its decarboxylation reaction. The following enzyme, 2-methyl-3-hydroxypyridine-5-carboxylic acid oxygenase, catalyzes an oxidative ring opening and incorporates two atoms of oxygen to produce *E*-2-(acetamidomethylene)succinate. The structure of this enzyme demonstrates its similarity to flavin monooxygenases and offers clues to catalytically important residues. The final enzyme, *E*-2-(acetamidomethylene)succinate hydrolase, is an α/β hydrolase and catalyzes the production of succinic semialdehyde, ammonia, acetate, and carbon dioxide using a modified catalytic triad consisting of a serine, histidine, and aspartic acid.

BIOGRAPHICAL SKETCH

Kathryn was born in Texas and raised in the greater Chicagoland area. She is the eldest daughter of two to Jim and Linda McCulloch. During her high school years in Waukegan, IL she was most fascinated by her science classes, particularly the chemistry and biology courses. Kathryn attended Ball State University in Muncie, IN, graduated *summa cum laude*, and received her Bachelor of Science in Chemistry in the spring of 2004. While at Ball State, Kathryn had the opportunity to participate in the undergraduate research program in the Chemistry department. Her mentor was Dr. Karlett Parra-Belky and Kathryn began studying vacuolar ATPases (V-ATPase) from *Saccharomyces cerevisiae*. Her three years working with Dr. Parra-Belky culminated in a publication in the *Journal of Biological Chemistry* and an award-winning poster at the annual American Society for Biochemistry and Molecular Biology meeting in Boston, MA in June of 2004.

After graduating from Ball State University, Kathryn began her Ph.D. studies in the Chemistry and Chemical Biology department at Cornell University in the fall of 2004. After completing her coursework, she began studying enzyme structure using X-ray crystallography in Prof. Steven Ealick's group in the summer of 2005. Her work has focused on enzymes involved in thiamin biosynthesis and the pyridoxal 5′-phosphate catabolism. Kathryn's future plans include broadening her scientific knowledge through postdoctoral studies and pursuing a career in academia.

For my family, whom I love with all my heart, and for everyone who knows that scientists can have a sense of humor.

ACKNOWLEDGMENTS

I would first like to thank Prof. Steven E. Ealick for mentoring me over the last six years. He has provided me with interesting and challenging projects that have encouraged independent thought and scientific curiosity. The skills I have learned from his group are invaluable. I would also like to thank Prof. Tadhg P. Begley for providing opportunities for collaboration, allowing me to be a member of his group, and being on my committee. Attending his meetings has taught me to critically analyze mechanistic enzymology problems and given me the ability to examine these questions from many different and complementary perspectives. I would also like to thank Prof. Begley for collaborating on the thiamin biosynthesis and pyridoxal 5′-phosphate degradation projects. I owe his former student, Dr. Tathagata Mukherjee, deep thanks. Tathagata has been my close collaborator on the vitamin B_6 catabolism projects and performed all of the kinetic experiments reported in those chapters. I would also like to thank Prof. Brian Crane for agreeing to be a committee member and offering advice on my research. Many thanks to Ms. Leslie Kinsland for the help she's given me in so many ways. I owe Dr. Cynthia Kinsland great thanks for the cloning work and many helpful suggestions during difficult protein purifications. I thank the beamline staff at NE-CAT for all of their help during data collection trips and much practical advice. I thank the members of the Begley lab for their cooperation and always offering their insights into mechanistic querier. Lastly, I would like to thank current and former Ealick group members for the camaraderie, support, practical advice, and many enjoyable work days.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

PLP, Pyridoxal 5′-phosphate

HMP-P, 4-amino-2-methyl-5-(hydroxymethyl)pyrimidine phosphate

HMP-PP, 4-amino-2-methyl-5-(hydroxymethyl)pyrimidine pyrophosphate

ThzP, 4-methyl-5-(β-hydroxyethyl)thiazole phosphate

ATP, adenosine 5′-triphosphate

AMP-PCP, β,γ-methylene adenosine 5′-diphosphate

ADP, adenosine 5′-diphosphate

HMPDdc, 3-hydroxy-2-methylpyridine-4,5-dicarboxylate decarboxylase

MHPCO, 2-methyl-3-hydroxypyridine-5-carboxylic acid oxygenase

FAD, flavin adenine dinucleotide

MHPC, 2-methyl-3-hydroxypyridine-5-carboxylate

E-2AMS, (*E*)-2-(acetamidomethylene)succinate

TPP, thiamin pyrophosphate

TP synthase, thiamin phosphate synthase

PPi, pyrophosphate

CF_3HMP-PP, 4-amino-2-(trifluoromethyl)-5-(hydroxymethyl)pyrimidine pyrophosphate

CO_2-ThzP, 2-carboxy-4-methyl-5-(hydroxyethyl)thiazole phosphate

TbpA, thiamin binding protein

MDCC, N-[2-(1-maleimidyl)ethyl]-7-(diethylamino)coumarin-3-carboxamide

CO_2-EtThz, 3-ethyl-4,5-dimethylthiazole-2-carboxylate

NDMT, 2′-nitrobenzyl-4,5-dimethylthiazole-2-carboxylate

DCM, dichloromethane

NEDMT, 2′-nitrobenzyl-3-ethyl-4,5-dimethylthiazole-2-carboxylate

DTT, dithiothreitol

PEG, polyethylene glycol

CNS, crystallography and NMR systems

TPS, thiamin phosphate synthase

NCI, National Cancer Institute

TEV, tobacco etch virus

X-gal, 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

TMP, thiamin monophosphate

ThiL, thiamin monophosphate kinase

PurM, aminoimidazole ribonucleotide synthetase

PurL, formylglycinamide ribonucleotide amidotransferase

SelD, selenophosphate synthetase

HypE, NiFe hydrogenase maturation protein

SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis

MWCO, molecular weight cutoff

MPD, 2-methyl-2,4-pentanediol

CHESS, Cornell High Energy Synchrotron Source

VAST, Vector Alignment Search Tool

FGAR, formylglycinamide ribonucleotide

LB, Luria-Bertani

IPTG, isopropyl-β-D-thiogalactopyranoside

SelMet, selenomethionyl

NCS, noncrystallographic symmetry

DHPdc, 3,4-dihydroxyphthalate 2-decarboxylase

PHBH, *p*-hydroxybenzoate hydroxylase

MHBH, *m*-hydroxybenzoate hydroxylase

DHPH, 2,6-dihydroxypyridine hydroxylase

p-OHB, *p*-hydroxybenzoate

NAD, nicotinamide adenine dinucleotide

CHAPTER 1

INTRODUCTION

Section 1.1. Structural Biology

Gaining an understanding of the structural basis of molecular processes is the ultimate goal of structural biology. Several methods have evolved allowing the structures of biological molecules to varying resolutions and include electron microscopy, NMR, small angle X-ray scattering, and X-ray crystallography (*1, 2*). Electron microscopy can offer a composite view of many molecules that could be in differing conformations and NMR provides a means of looking at dynamic structures (*1*). Small angle X-ray scattering can offer a view of the molecular envelope (*3*). Of these techniques, X-ray crystallography is among the most powerful. The resolution available by macromolecular crystallography can extend beyond 1 Å and allow for detailed analysis of a static structure.

X-ray crystallographic studies of enzymes, proteins capable of catalyzing a reaction, have proven very useful for advancing mechanistic enzymology. Often, it is possible to trap ligands in a crystal structure, through either cocrystallization or soaking experiments, and this can allow for reaction coordinates to be mapped out. The three dimensional coordinates for nearly every (non-hydrogen) atom at atomic resolution can allow for the identification of key catalytic residues, the proposal for mechanistic action, and the design of additional experiments to provide further evidence of means of action. The 2009 Nobel Prize in Chemistry was awarded to three scientists for the determination of the crystal structure of the ribosome, which is responsible for catalyzing the formation of the peptide bonds found in proteins in the specific order as encoded by the mRNA (*4*). Knowledge of the structure of the ribosome is leading to the development of novel antibiotics (*5*). X-ray crystallography is complemented well with mutagenesis studies, kinetic characterization, labeling

studies, circular dichroism, gel filtration, and other techniques to provide a complete picture of how an enzyme functions.

Section 1.2. Importance of Cofactors

Cofactors are small molecules that can assist an enzyme in catalyzing chemistry that would ordinarily be beyond the scope of chemistry available through amino acid side chains alone. Cofactors may be inorganic, such as metal ions (Fe^{2+}, Mg^{2+}, Mn^{2+}), small organic molecules (thiamin pyrophosphate, pyridoxal 5′-phosphate, flavin), or a combination of the two (heme). Many of the organic cofactors are known as vitamins, and humans must obtain vitamins through the diet, as they are unable to biosynthesize these small molecules. Vitamins are typically biosynthesized by bacteria, fungi, and plants, although to ensure adequate vitamin levels, many foods are now fortified with vitamins, such as cereals and breads.

The B vitamins are a group of eight water soluble vitamins that play important cellular roles and include thiamin, niacin, riboflavin, pantothenate, pyridoxine, biotin, folate, and cobalamin. Vitamin B_1, thiamin, acts to stabilize carbocation intermediates and is often involved in decarboxylation reactions of α-keto acids and is important in carbohydrate metabolism. Deficiency of vitamin B_1 can lead to beri-beri and in alcoholics, Wernicke-Korsakoff syndrome (*6*). Another B vitamin, vitamin B_6, is very important for cellular function. The Enzyme Commission has estimated that up to 4% of all enzymatic activities are dependent on the active form of vitamin B_6, PLP. PLP is typically covalently bound to the enzyme through a Schiff's base at an active site lysine residue and is involved in amino acid metabolism, gluconeogenesis, and lipid metabolism. Deficiency in vitamin B_6 can lead to anemia, as the generation of heme is PLP dependent, and also cause skin lesions similar to those seen with niacin deficiency (*7*). In addition to the critical roles these B vitamins play within the cell, the metabolic pathways of cofactors often contain novel and exciting chemistry (*8*).

Section 1.3. Vitamin B_1 Biosynthesis

The mechanistic role of vitamin B_1 in the stabilization of carbanion intermediates was determined in the 1950's (*9, 10*). However, the biosynthesis of thiamin has gained more attention over the last twenty years. The pathway found in bacteria is fairly well understood and studies continue as the novel and interesting chemical reactions are probed further (*6*). Figure 1.1 shows the biosynthetic pathway for the active form of vitamin B_1, thiamin pyrophosphate, as found in *Bacillus subtilis*. The two heterocyclic moieties are biosynthesized separately, and then joined together by thiamin phosphate synthase. Briefly, the pyrimidine moiety, 4-amino-2-methyl-5-(hydroxymethyl)-pyrimidine phosphate (HMP-P), is formed by a radical SAM dependent rearrangement of 5-aminoimidazole ribotide by ThiC and then phosphorylated by ThiD to form HMP-PP (*11, 12*). The thiazole moiety, 4-methyl-5-(β-hydroxyethyl)thiazole phosphate (ThzP), is formed from deoxy-D-xylulose phosphate, glycine, and a sulfur source by seven enzymes (*6*). After coupling, the final reaction is the phosphorylation of thiamin phosphate to produce thiamin pyrophosphate, the active form of vitamin B_1 within the cell.

Recent advances in the reconstitution of the thiazole moiety in *B. subtilis* have allowed direct characterization of the product of the ThiG-ThiS catalyzed reaction (*13*). The characterized product did not co-migrate with the known standard of thiazole phosphate when analyzed by HPLC and was identified as the carboxylated thiazole phosphate. This molecule is then aromatized by TenI to form carboxylated thiazole phosphate (*6*). Chapter Two reports on the kinetic characterization of *B. subtilis* thiamin phosphate synthase using carboxylated thiamin phosphate and presents the structural complex of thiamin phosphate synthase with the carboxylated thiazole phosphate and a trifluorinated HMP-PP analogue. The active site contains three distinct species, representing an enzyme-intermediate complex containing the

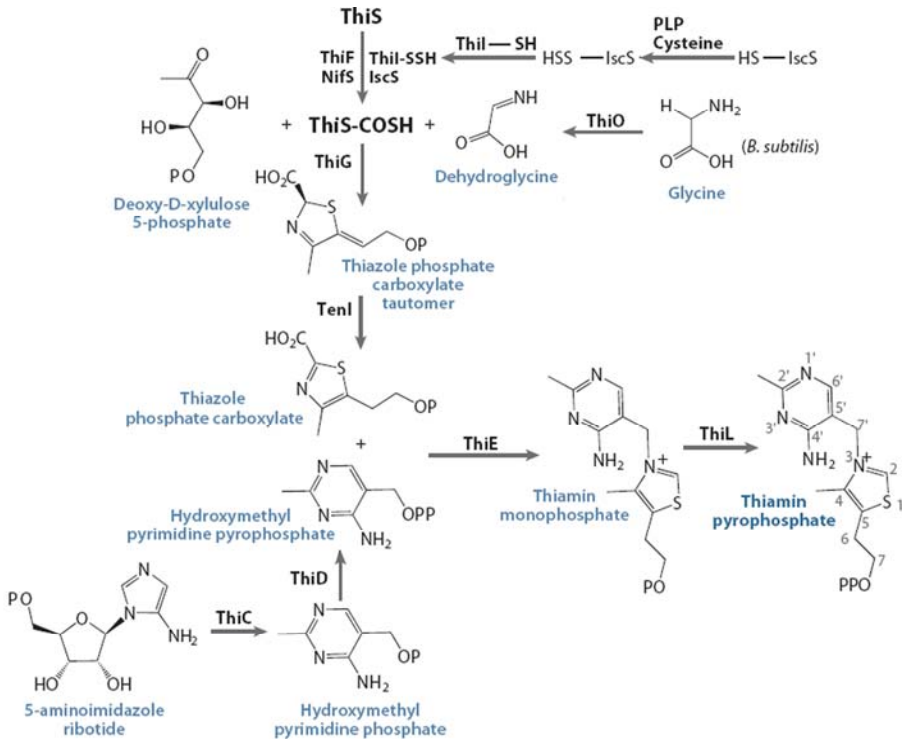


Figure 1.1. Thiamin biosynthesis in *B. subtilis*. This figure is modified from (*3*).

dissociated pyrophosphate, the pyrimidine iminemethide, and the carboxylated thiazole phosphate. This structure, in conjunction with the kinetic data, demonstrates that the carboxylated thiazole phosphate binds to thiamin phosphate synthase and is catalytically active.

Chapter Three reports the structure of thiamin phosphate synthase from *Mycobacterium tuberculosis*. As tuberculosis continues to be a global threat and continues to develop resistance to current drug therapies, new targets must be pursued and new drug treatments developed. Because humans do not synthesize thiamin but instead obtain it through diet, the thiamin biosynthetic pathway is a potential target, and thiamin phosphate synthase has no orthologue in humans. This chapter reports the crystal structure of *M. tuberculosis* thiamin phosphate synthase. This enzyme is very similar to thiamin phosphate synthase from other organisms and is an $(\beta/\alpha)_8$ barrel fold enzyme where the active site is found at the C-terminal end of the β-barrel. Two

loop regions that form interactions with the active site are disordered and the active site is exposed to solvent in this structure. The *M. tuberculosis* thiamin phosphate synthase structure was then used for simulated docking studies and the initial results of this virtual drug screening to identify potential inhibitors of this enzyme are being characterized.

The final enzymatic step in biosynthesizing the active form of thiamin is the phosphorylation of thiamin monophosphate to form thiamin pyrophosphate using ATP. ThiL is a member of a small ATP-binding superfamily known as the PurM ATP-binding superfamily (*14*). Chapter Four reports several structures of complexes of this enzyme, ThiL, from *Aquifex aeolicus* with various substrates and products (*15*). The four complexes determined were AaThiL with ATP, AaThiL with AMP-PCP (a non-hydrolyzable ATP analogue), AaThiL with AMP-PCP and thiamin phosphate, and AaThiL with ADP and thiamin pyrophosphate. These four structures allowed for the reaction to be carefully scrutinized and the phosphate transfer was shown to occur via a direct, in-line transfer from ATP to thiamin phosphate.

Section 1.4. Vitamin B_6 Catabolism

The biosynthesis of cofactors has been fairly well studied; in contrast, the degradation and catabolism of these small molecules is not well characterized. However, the identification of several *Pseudomonas* species capable of growing solely on various vitamin B_6 vitamers indicated that some organisms are able to break down this essential cofactor for use as a source of both carbon and nitrogen (*16, 17*). Two pathways for the degradation of vitamin B_6 were characterized and although the intermediates were isolated and identified, and the enzymatic activities could be tied to specific enzymes, the lack of genomic information meant that the genes responsible went unknown (*16, 18-26*).

More recently, the genome for *Mesorhizobium loti* MAFF303099, a symbiotic bacterium found on the roots of legumes, became available (*27*). The gene for 4-pyridoxine oxidase was found in this bacteria and the gene product was shown to catalyze the first step in vitamin B_6 degradation (*28*). Since that time, the genes responsible for each of the seven enzymatic steps for the conversion of pyridoxal 5′-phosphate to succinic semialdehyde, ammonia, acetate, and carbon dioxide have been identified and the gene products characterized (*28-36*). The degradation pathway of PLP is shown in Figure 1.2. Several of these enzymes are responsible for very interesting chemical conversions, such as an oxidative ring opening of an aromatic ring without the use of a non-heme iron cofactor (2-methyl-3-hydroxypyridine-5-carboxylic acid oxygenase) and the decarboxylation of an aromatic ring without the use of an organic cofactor (2-methyl-3-hydroxypyridine-4,5-dicarboxylate decarboxylase).

Chapter Five reports the gene identification of 2-methyl-3-hydroxypyridine-4,5-dicarboxylate decarboxylase (HMPDdc) and the kinetic and structural characterization of this enzyme from *M. loti* (*33*). Initial sequence analysis showed this gene product has significant similarity to class II aldolases and the gene was initially assigned as a putative aldolase. The gene product was characterized by HPLC analysis and catalytically converted 2-methyl-3-hydroxypyridine-4,5-dicarboxylate to 2-methyl-3-hydroxypyridine-5-carboxylate as determined by NMR characterization. HMPDdc was structurally characterized and shown to be a tetramer with the active site located at the interface between two protomers. A manganese ion was found coordinated to three histidine residues in the active site and comparison of this structure to those found in the Protein Data Bank revealed strong similarity to the class II aldolases. A mechanism has been proposed for the decarboxylation reaction that is similar to the reaction of the class II aldolases.

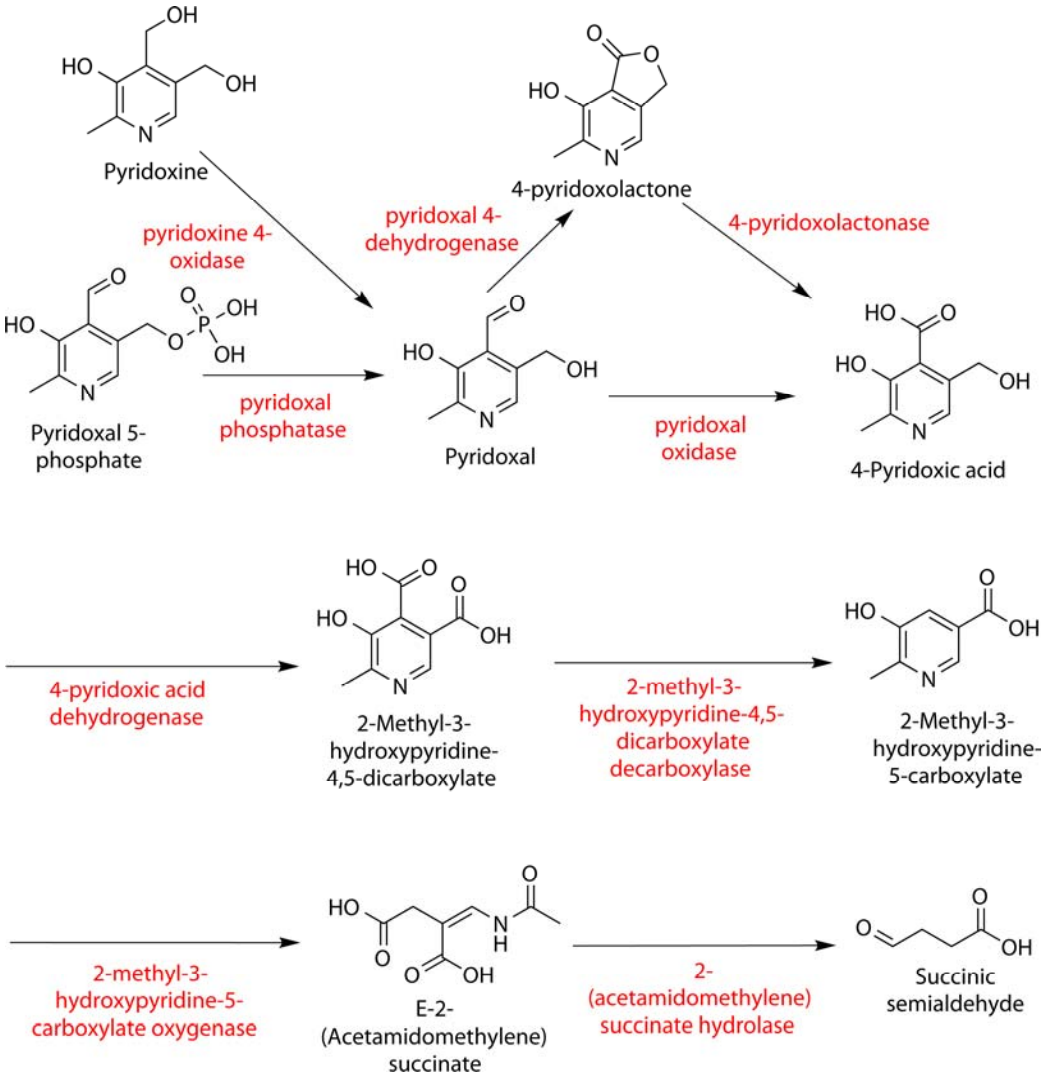


Figure 1.2. The Vitamin B_6 degradative pathway as found in *M. loti* MAFF303099.

Chapter Six reports two crystal structures of 2-methyl-3-hydroxypyridine-5-carboxylic acid oxygenase (MHPCO) with FAD or FAD and MHPC bound (*37*). MHPCO is unique in that after the flavin-dependent hydroxylation of MHPC, a second atom of oxygen (from water) and produces a ring-opened product, *E*-2-(acetamidomethylene)succinate without utilizing a non-heme iron cofactor. Although significant effort has been put forth to probe the mechanism of MHPCO, the exact means for the incorporation of the second oxygen atom remained unclear (*24*, *25*, *38*-

43). The crystal structure confirmed that MHPCO is a member of the flavin monooxygenase superfamily and the ternary complex with FAD and MHPC provided additional insight. MHPC binds in a very large active site pocket and forms primarily water-mediated interactions with the protein, raising additional questions about how MHCPO imparts specificity. However, the position of MHPC in the active site allowed for the identification of several potentially important residues in the active site that could help stabilize or activate MHPC during the addition of the second oxygen atom.

The final enzyme in the vitamin B6 degradative pathway is *E*-2-(acetamidomethylene)succinate hydrolase (*E*-2AMS hydrolase) to produce succinic semialdehyde, ammonia, acetate, and carbon dioxide, and its structure is reported in Chapter Seven (*44*). *E*-2AMS hydrolase is an α/β hydrolase and utilizes a serine/histidine/aspartic acid catalytic triad. This catalytic triad is common in α/β hydrolases, although in *E*-2AMS hydrolase, the structural orientation of the aspartic acid is not in the conserved position. The importance of the serine and aspartic acid were confirmed through mutagenesis; the enzyme was inactive when these residues were mutated. Although attempts at obtaining a complex with *E*-2AMS were unsuccessful, manual positioning of the substrate in the active site identified several additional residues that could be involved in the coordination of *E*-2AMS for proper catalysis.

Section 1.5. Acknowledgements

The work presented in this thesis represents projects that rely heavily on collaborative efforts. Chapter Two's NMR experiment and kinetic characterization of TP synthase with carboxylated thiazole phosphate was performed by Dr. Jeremiah Hanes in collaboration with Dr. Begley's group. The synthesis and kinetic characterization of the thiazolium carboxylate compound presented in Chapter Two

was carried out by Dr. Sameh Abdelwahed from Dr. Begley's research group. Chapter Three involved collaboration with both Divya Ramamoorthy from Dr. Guida's laboratory, who performed virtual drug screening based on the structure of thiamin phosphate synthase from *M. tuberculosis* and docking studies, and Dr. Tomoshige Hiratsuka of the Begley group, who performed the activity assay on the top hits of the virtual drug screen. The gene identification and kinetic characterization of HMPDdc presented in Chapter Five was the work of Dr. Tathagata Mukherjee of the Begley group, as was the activity measurements of *E*-2AMS hydrolase mutants described in Chapter Seven. These collaborations have provided context and these collaborators deserve recognition for their work.

REFERENCES

1. Campbell, I. D. (2002) Timeline: the march of structural biology, *Nat Rev Mol Cell Biol 3*, 377-381.
2. Liljas, A. (2006) On the complementarity of methods in structural biology, *Acta Crystallogr D Biol Crystallogr 62*, 941-945.
3. Rambo, R. P., and Tainer, J. A. (2010) Bridging the solution divide: comprehensive structural analyses of dynamic RNA, DNA, and protein assemblies by small-angle X-ray scattering, *Curr Opin Struct Biol 20*, 128-137.
4. Van Noorden, R. (2009) Structural biology bags chemistry prize, *Nature 461*, 860.
5. Schmeing, T. M., and Ramakrishnan, V. (2009) What recent ribosome structures have revealed about the mechanism of translation, *Nature 461*, 1234-1242.
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. Hisano, M., Suzuki, R., Sago, H., Murashima, A., and Yamaguchi, K. (2010) Vitamin B6 deficiency and anemia in pregnancy, *Eur. J. Clin. Nutr. 64*, 221-223.
8. Begley, T. P., Chatterjee, A., Hanes, J. W., Hazra, A., and Ealick, S. E. (2008) Cofactor biosynthesis--still yielding fascinating new biological chemistry, *Curr Opin Chem Biol 12*, 118-125.
9. Jordan, F. (2003) Current mechanistic understanding of thiamin diphosphate-dependent enzymatic reactions, *Nat. Prod. Rep. 20*, 184-201.

10. Breslow, R. (1958) On the Mechanism of Thiamine Action. IV. Evidence from Studies on Model Systems, *J. Am. Chem. Soc. 80*, 3719-3726.
11. Chatterjee, A., Li, Y., Zhang, Y., Grove, T. L., Lee, M., Krebs, C., Booker, S. J., Begley, T. P., and Ealick, S. E. (2008) Reconstitution of ThiC in thiamine pyrimidine biosynthesis expands the radical SAM superfamily, *Nat. Chem. Biol. 4*, 758-765.
12. Cheng, G., Bennett, E. M., Begley, T. P., and Ealick, S. E. (2002) Crystal structure of 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate kinase from Salmonella typhimurium at 2.3 A resolution, *Structure 10*, 225-235.
13. Hazra, A., Chatterjee, A., and Begley, T. P. (2009) Biosynthesis of the Thiamin Thiazole in Bacillus subtilis: Identification of the Product of the Thiazole Synthase-Catalyzed Reaction, *J. Am. Chem. Soc. 131*, 3225-3229.
14. Li, C., Kappock, T. J., Stubbe, J., Weaver, T. M., and Ealick, S. E. (1999) X-ray crystal structure of aminoimidazole ribonucleotide synthetase (PurM), from the *Escherichia coli* purine biosynthetic pathway at 2.5 Å resolution, *Structure 7*, 1155-1166.
15. McCulloch, K. M., Kinsland, C., Begley, T. P., and Ealick, S. E. (2008) Structural studies of thiamin monophosphate kinase in complex with substrates and products, *Biochemistry 47*, 3810-3821.
16. Rodwell, V. W., Volcani, B. E., Ikawa, M., and Snell, E. E. (1958) Bacterial oxidation of vitamin B6. I. Isopyridoxal and 5-pyridoxic acid, *J Biol Chem 233*, 1548-1554.
17. Snell, E. E., and Haskell, B. E. (1971) *The Metabolism of Vitamin B6, In: Comprehensive Biochemistry*, Vol. 21, Elsevier/North Holland, New York.
18. Ikawa, M., Rodwell, V. W., and Snell, E. E. (1958) Bacterial oxidation of vitamin B6. II. Structure of 260 compound, *J Biol Chem 233*, 1555-1559.

19. Burg, R. W., Rodwell, V. W., and Snell, E. E. (1960) Bacterial oxidation of vitamin B6. II. Metabolites of pyridoxamine, *J Biol Chem 235*, 1164-1169.
20. Burg, R. W., and Snell, E. E. (1969) The bacterial oxidation of vitamin B6. VI. Pyridoxal dehydrogenase and 4-pyridoxolactonase, *J Biol Chem 244*, 2585-2589.
21. Nyns, E. J., Zach, D., and Snell, E. E. (1969) The bacterial oxidation of vitamin B6. 8. Enzymatic breakdown of alpha-(N-acetylaminomethylene) succinic acid, *J Biol Chem 244*, 2601-2605.
22. Sparrow, L. G., Ho, P. P., Sundaram, T. K., Zach, D., Nyns, E. J., and Snell, E. E. (1969) The bacterial oxidation of vitamin B6. VII. Purification, properties, and mechanism of action of an oxygenase which cleaves the 3-hydroxypyridine ring, *J Biol Chem 244*, 2590-2600.
23. Sundaram, T. K., and Snell, E. E. (1969) The bacterial oxidation of vitamin B6. V. The enzymatic formation of pyridoxal and isopyridoxal from pyridoxine, *J Biol Chem 244*, 2577-2584.
24. Kishore, G. M., and Snell, E. E. (1981) Interaction of 2-methyl-3-hydroxypyridine-5-carboxylic acid oxygenase with FAD, substrates, and analogues. Spectral and fluorescence investigations, *J Biol Chem 256*, 4234-4240.
25. Kishore, G. M., and Snell, E. E. (1981) Kinetic investigations on a flavoprotein oxygenase, 2-methyl-3-hydroxypyridine-5-carboxylic acid oxygenase, *J Biol Chem 256*, 4228-4233.
26. Huynh, M. S., and Snell, E. E. (1985) Enzymes of vitamin B6 degradation. Purification and properties of two N-acetylamidohydrolases, *J. Biol. Chem. 260*, 2379-2383.

27. Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S., Watanabe, A., Idesawa, K., Ishikawa, A., Kawashima, K., Kimura, T., Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A., Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimpo, S., Sugimoto, M., Takeuchi, C., Yamada, M., and Tabata, S. (2000) Complete genome structure of the nitrogen-fixing symbiotic bacterium Mesorhizobium loti, *DNA Res 7*, 331-338.

28. Yuan, B., Yoshikane, Y., Yokochi, N., Ohnishi, K., and Yagi, T. (2004) The nitrogen-fixing symbiotic bacterium Mesorhizobium loti has and expresses the gene encoding pyridoxine 4-oxidase involved in the degradation of vitamin B6, *FEMS Microbiol. Lett. 234*, 225-230.

29. Funami, J., Yoshikane, Y., Kobayashi, H., Yokochi, N., Yuan, B., Iwasaki, K., Ohnishi, K., and Yagi, T. (2005) 4-Pyridoxolactonase from a symbiotic nitrogen-fixing bacterium Mesorhizobium loti: cloning, expression, and characterization, *Biochim Biophys Acta 1753*, 234-239.

30. Yokochi, N., Nishimura, S., Yoshikane, Y., Ohnishi, K., and Yagi, T. (2006) Identification of a new tetrameric pyridoxal 4-dehydrogenase as the second enzyme in the degradation pathway for pyridoxine in a nitrogen-fixing symbiotic bacterium, Mesorhizobium loti, *Arch Biochem Biophys 452*, 1-8.

31. Yuan, B., Yokochi, N., Yoshikane, Y., Ohnishi, K., and Yagi, T. (2006) Molecular cloning, identification and characterization of 2-methyl-3-hydroxypyridine-5-carboxylic-acid-dioxygenase-coding gene from the nitrogen-fixing symbiotic bacterium Mesorhizobium loti, *J. Biosci. Bioeng. 102*, 504-510.

32. Mukherjee, T., Kinsland, C., and Begley, T. P. (2007) PLP catabolism: identification of the 4-pyridoxic acid dehydrogenase gene in Mesorhizobium loti MAFF303099, *Bioorg Chem 35*, 458-464.

33. Mukherjee, T., McCulloch, K. M., Ealick, S. E., and Begley, T. P. (2007) Gene identification and structural characterization of the pyridoxal 5'-phosphate degradative protein 3-hydroxy-2-methylpyridine-4,5-dicarboxylate decarboxylase from mesorhizobium loti MAFF303099, *Biochemistry 46*, 13606-13615.

34. Ge, F., Yokochi, N., Yoshikane, Y., Ohnishi, K., and Yagi, T. (2008) Gene identification and characterization of the pyridoxine degradative enzyme 4-pyridoxic acid dehydrogenase from the nitrogen-fixing symbiotic bacterium Mesorhizobium loti MAFF303099, *J Biochem 143*, 603-609.

35. Mukherjee, T., Hilmey, D. G., and Begley, T. P. (2008) PLP catabolism: identification of the 2-(Acetamidomethylene)succinate hydrolase gene in Mesorhizobium loti MAFF303099, *Biochemistry 47*, 6233-6241.

36. Yuan, B., Yokochi, N., Yoshikane, Y., Ohnishi, K., Ge, F., and Yagi, T. (2008) Gene identification and characterization of the pyridoxine degradative enzyme alpha-(N-acetylaminomethylene)succinic acid amidohydrolase from Mesorhizobium loti MAFF303099, *J. Nutr. Sci. Vitaminol. (Tokyo) 54*, 185-190.

37. McCulloch, K. M., Mukherjee, T., Begley, T. P., and Ealick, S. E. (2009) Structure of the PLP degradative enzyme 2-methyl-3-hydroxypyridine-5-carboxylic acid oxygenase from Mesorhizobium loti MAFF303099 and its mechanistic implications, *Biochemistry 48*, 4139-4149.

38. Chaiyen, P., Ballou, D. P., and Massey, V. (1997) Gene cloning, sequence analysis, and expression of 2-methyl-3-hydroxypyridine-5-carboxylic acid oxygenase, *Proc Natl Acad Sci U S A 94*, 7233-7238.

39. Chaiyen, P., Brissette, P., Ballou, D. P., and Massey, V. (1997) Reaction of 2-methyl-3-hydroxypyridine-5-carboxylic acid (MHPC) oxygenase with N-

methyl-5-hydroxynicotinic acid: studies on the mode of binding, and protonation status of the substrate, *Biochemistry 36*, 13856-13864.

40. Chaiyen, P., Brissette, P., Ballou, D. P., and Massey, V. (1997) Unusual mechanism of oxygen atom transfer and product rearrangement in the catalytic reaction of 2-methyl-3-hydroxypyridine-5-carboxylic acid oxygenase, *Biochemistry 36*, 8060-8070.
41. Chaiyen, P., Brissette, P., Ballou, D. P., and Massey, V. (1997) Thermodynamics and reduction kinetics properties of 2-methyl-3-hydroxypyridine-5-carboxylic acid oxygenase, *Biochemistry 36*, 2612-2621.
42. Chaiyen, P., Sucharitakul, J., Svasti, J., Entsch, B., Massey, V., and Ballou, D. P. (2004) Use of 8-substituted-FAD analogues to investigate the hydroxylation mechanism of the flavoprotein 2-methyl-3-hydroxypyridine-5-carboxylic acid oxygenase, *Biochemistry 43*, 3933-3943.
43. Oonanant, W., Sucharitakul, J., Yuvaniyama, J., and Chaiyen, P. (2005) Crystallization and preliminary X-ray crystallographic analysis of 2-methyl-3-hydroxypyridine-5-carboxylic acid (MHPC) oxygenase from Pseudomonas sp. MA-1, *Acta Crystallogr Sect F Struct Biol Cryst Commun 61*, 312-314.
44. McCulloch, K. M., Mukherjee, T., Begley, T. P., and Ealick, S. E. (2010) Structure determination and characterization of the vitamin B(6) degradative enzyme (E)-2-(acetamidomethylene)succinate hydrolase, *Biochemistry 49*, 1226-1235.

CHAPTER 2

CRYSTAL STRUCTURE AND KINETIC CHARACTERIZATION OF *BACILLUS SUBTILIS* THIAMIN PHOSPHATE SYNTHASE WITH A CARBOXYLATED THIAZOLE PHOSPHATE

Section 2.1. Introduction

Thiamin pyrophosphate (TPP), the active form of vitamin B_1 in the cell, is an essential cofactor used in the stabilization of carbanion intermediates (*1*). While humans are unable to biosynthesize vitamin B_1 and must obtain it through diet, many microorganisms have developed means to produce TPP (*2*). While both eukaryotic and prokaryotic organisms are known to form TPP, the pathways for its production are different. The eukaryotic pathway, found in plants and fungi, utilizes a single enzyme to form the thiazole moiety from glycine and nicotinamide dinucleotide and two enzymes to generate the pyrimidine portion of TPP (*3-5*). The well-studied bacterial pathway relies on two enzymes to form 4-amino-2-methyl-5-(hydroxymethyl)pyrimidine pyrophosphate (HMP-PP) and 4-methyl-5-(hydroxyethyl)thiazole phosphate (ThzP) is formed using up to six enzymes from deoxy-D-xylulose, a sulfur source, and either glycine or tyrosine (*6, 7*). Both the prokaryotic and eukaryotic pathways, however, after forming the two heterocycles of TPP separately, then couple the pyrimidine and thiazole moieties together to form thiamin monophosphate (*8-10*). This coupling reaction in prokaryotes is catalyzed by thiamin phosphate synthase (TP synthase) and releases pyrophosphate (PP_i) and thiamin monophosphate.

*Bs*TP synthase has been very well studied. Previous work using positional isotope exchange studies has demonstrated that TP synthase first forms a carbocation intermediate on the methyl group at C5 of the pyrimidine ring after release of PP_i in a magnesium dependent reaction shown in Figure 2.1 (*11*). Additionally, the rate of

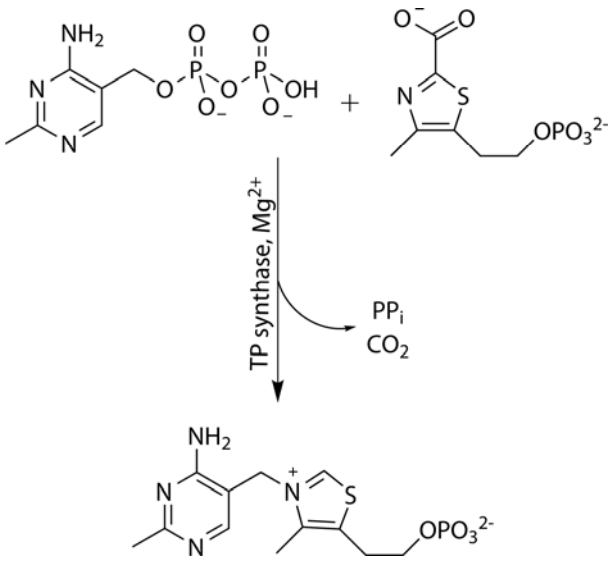


Figure 2.1. Reaction catalyzed by TP synthase.

formation for the pyrimidine carbocation has been determined using transient steady state kinetics and the possibility of an associative S_N2 mechanism has been ruled out (*12*). The crystal structure of TP synthase in complex with thiamin monophosphate and PP_i allowed the active site to be mapped (*8*). Ser130 was identified as a potentially important residue for activating PP_i as a leaving group while bonded to the pyrimidine moiety and mutagenesis of Ser130 to alanine resulted in an inactive enzyme. Subsequent structural work on the S130A mutant and the trifluoromethyl-substituted substrate analogue 4-amino-2-(trifluoromethyl)-5-(hydroxymethyl)pyrimidine pyrophosphate (CF_3HMP-PP), which is a poor substrate for TP synthase due to the electron withdrawing properties of the trifluoromethyl group, allowed intermediates to be trapped crystallographically (*13*). Two intermediate complexes show three discrete species in the active site: ThzP, CF_3HMP, and PP_i. These structures support the dissociative mechanism for TP synthase and allowed the reaction pathway to be mapped.

Recent work on the THI4p from the eukaryotic thiamin biosynthetic pathway has shown that a carboxylated thiazole tautomer intermediate is formed (*14*).

Advances in the reconstitution of TP synthase activity has also allowed for direct characterization of the product of *Bs*TP synthase (*15*). This work has shown that ThzP is not the product of TP synthase, but that a similar thiazole tautomer intermediate to that seen in the eukaryotic system leads to the carboxylated thiazole moiety 2-carboxy-4-methyl-5-(hydroxyethyl)thiazole phosphate (CO_2-ThzP). CO_2-ThzP has now been shown to be a substrate for TP synthase and crystallization experiments in which CO_2-ThzP and CF_3HMP-PP were soaked into TP synthase crystals produce an intermediate complex in which PP_i has dissociated from the pyrimidine moiety and the carboxylate group is still covalently bonded to the thiazole moiety.

Section 2.2. Materials and Methods

Protein Overexpression and Purification. The molecular cloning, protein overexpression, and protein purification has been previously described for TP synthase (*8, 10*). Briefly, the plasmid pYZC6927 was transformed into the *E. coli* cell strain SG13009 (Qiagen) and grown in 5 mL LB supplemented with 50 mg/L kanamycin and 150 mg/L ampicillin overnight. This starter culture was used to inoculate 1 L cultures of LB with the same antibiotic supplements to an OD_{600} of 0.8 at 37 °C with shaking. Overexpression of TP synthase was induced with 1.5 mM isopropyl β-D-thiogalactopyranoside (IPTG) and cells were allowed to continue shaking at 37 °C for six hours. Cells were harvested by centrifugation at 6,000 *g* at 4 °C and then stored at -20 °C until purification.

Pelleted cells from 2 L of culture were resuspended in sonication buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH 7.8) and lysed by sonication. The lysate was then centrifuged at 40,000 *g* for one hour at 4 °C. The supernatant was passed two times over a 2 mL Ni-NTA column (Qiagen) pre-equilibrated with sonication buffer. The column was then washed with 75 mL of sonication buffer and 40 mL of wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10% glycerol, 5 mM imidazole, pH 6.0). TP

synthase was eluted from the column using 10 mL of wash buffer containing 80 mM imidazole. TP synthase was judged to be 95% pure by SDS-PAGE analysis (results not shown) and dialyzed into 50 mM Tris (pH 7.5) 1 mM dithiothreitol (DTT). The sample was concentrated to 20 mg/mL as measured by Bradford assay and stored at -80 °C until use in crystallization trials (*16*).

The overexpression and purifications of recombinant thiaminase I and the thiamin binding protein (TbpA) were performed according to the methods described elsewhere (*17, 18*).

NMR Characterization of TP Synthase Reaction. The reaction catalyzed by TP synthase was monitored using 600 MHz ^{1}H-NMR over 80 minutes. A concentration of 50 μM TP synthase was added to start the reaction in which 3 mM HMP-PP and 3 mM ThzP was present in a buffer composed of 25 mM sodium phosphate (pH 7.5), 50 mM NaCl, 16 mM $MgCl_2$, and 70% D_2O. Scans were taken every 90 seconds.

Steady State Kinetic Parameters of TP Synthase. In order to measure the steady state kinetic parameters of HMP-PP, a coupled enzyme assay was used based upon a previously published method employing thiaminase I and its highly chromophoric substrate 4-nitrothiophenolate (*17*). The conditions for the assay were as follows: 100 nM TP synthase, variable HMP-PP, 1 mM ThzP, 1 μM Thiaminase I, 80 μM 4-nitrothiophenolate, 2 mM TCEP (tris(2-carboxyethyl)phosphine), 50 mM Tris-Cl pH 7.6 at room temperature, 100 mM NaCl, and 8 mM $MgCl_2$. The reactions were started by the addition of HMP-PP and monitored for up to 5 minutes at room temperature in a UV/vis spectrophotometer set measure absorbance at 411 nm. The time dependent decrease in the concentration of 4-nitrothiophenolate was used to quantify the rate of synthesis of thiamin phosphate.

Measurement of Steady State Kinetic Parameters Using Fluorescently Labeled TbpA. The kinetic parameters for the thiazole substrates were estimated using a newly

developed fluorescently labeled mutant *E. coli* thiamin binding protein (TbpA). The N-[2-(1-maleimidyl)ethyl]-7-(diethylamino)coumarin-3-carboxamide (MDCC)-labeled TbpA is analogous to the inorganic phosphate probe developed by Webb and coworkers (*19*). An approximate 50% reduction in florescence occurs upon addition of thiamin phosphate to the labeled protein and the dissociation constant (K_d) was found to be 202 ± 1 nM at a concentration of 436 ± 3 nM MDCC-labeled TbpA. The K_d for thiamin depends on the TbpA concentration, possibly due to dimer formation as was speculated based upon the crystal structure where dimer formation may block the ligand binding site (*18*). A full description of the engineered thiamin probe mentioned here will be published separately, but briefly, the assay conditions used to measure TP synthase kinetics were as follows: 4 nM TP synthase, variable ThzP or CO_2-ThzP, 50 μM HMP-PP, 500 nM MDCC-labeled TbpA, 50 mM Tris-Cl pH 7.6 at room temperature, 100 mM NaCl, 2 mM TCEP, and 8 mM $MgCl_2$. The reaction was started with the addition of TP synthase and monitored at room temperature in a 96-well plate (Greiner plate model 655801) purchased from Omega Scientific (Tarzana, CA, USA) in a Synergy HT Multi-Detection Microplate Reader (BioTek Instruments, Winooski, VT, USA) exciting at a wavelength of 425 nm and recording emission at a wavelength of 462 nm. The reactions were monitored for up to 25 minutes and the fluorescence data were fit by linear regression.

Synthesis and Kinetic Characterization of Thiazolium-2-Carboxylate as a Model Compound. 3-Ethyl-4,5-dimethylthiazole-2-carboxylate (CO_2-EtThz) can be used as a model compound for carboxylated thiamin phosphate as this compound contains a positive charge on the nitrogen atom of the thiazole ring adjacent to the carboxylate group at C2. 4,5-Dimethylthiazole-2-carboxylate was synthesized using a previously reported method (*20*). This thiazole moiety was then used to prepare 2′-nitrobenzyl-4,5-dimethylthiazole-2-carboxylate (NDMT) (1.0 g) by addition of K_2CO_3

(in water and acetonitrile) and 1-(bromomethyl)-2-nitrobenzene. The reaction mixture was stirred overnight at room temperature. 100 mL of water was added and the aqueous layer washed three times with 50 mL of dichloromethane (DCM). The combined organic extracts were dried over magnesium sulfate and evaporated under reduced pressure and purified by column chromatography on silica gel using a 70:30 mixture of hexanes and ethyl acetate as eluent. Yield: 0.998 g (54%); ^{1}H-NMR ($CDCl_3$) δ: 2.44 (3H, s), 2.45 (3H, s), 7.49 (1H, t), 7.66 (2H, m), 8.15 (1H, d). 2′-Nitrobenzyl-3-ethyl-4,5-dimethylthiazole-2-carboxylate (NEDMT) was derived from NDMT using a procedure based upon similar chemistry (*20, 21*). To a solution of NDMT (200 mg, 0.7 mmol) dissolved in dichloromethane (1 mL), Et_3O BF_4 (700 μL, 0.7 mmol, 1 M solution in DCM) was added and allowed to stir overnight at room temperature in darkness. The salt was precipitated through the addition of Et_2O (5 mL) and filtered. The salt was purified by redissolving in DCM (2 mL) and re-precipitated in Et_2O (5 mL) to give a quantitative yield of NEDMT (270 mg, 97%); ^{1}H-NMR ($CDCl_3$) δ: 1.51 (3H, t), 2.59 (3H, s), 2.62 (3H, s), 4.87 (2H, q), 5.85 (2H, s), 7.57 (1H, t), 7.76 (2H, m), 8.15 (1H, d).

The *o*-nitrobenzyl group can be removed through photorelease to generate CO_2-EtThz and the decarboxylation of CO_2-EtThz can be kinetically characterized using ^{1}H-NMR spectroscopy (*22*). To an NMR tube (10 mM, 0.5 mL), NEDMT was prepared using 45 mM phosphate buffer (H2O:D2O 90:10) at pH 7.0. The NMR tube was stored in a larger glass tube to hold the temperature steady at 25 °C and to eliminate light with a wavelength above 300 nm. The sample was then exposed to UV light with a wavelength longer than 300 nm for 20 minutes and then returned to darkness, in which the NMR spectra were taken periodically.

Crystallization of TP Synthase Complex. TP synthase was crystallized using slight modifications to the methods previously described (*8, 13*). Crystals were grown

using the hanging drop vapor diffusion method over six months at 4 °C with a reservoir solution of 50 mM Tris buffer (pH 7.7), 75 mM $MgCl_2$, 26% PEG 4000, and 1 mM DTT. Drops consisted of 3 µl protein solution mixed with 3 µl reservoir solution. Crystals belong to the space group $P4_22_12$ with a = 96.0 Å and c = 59.3 Å with a Matthew's coefficient of 2.74 $Å^3$/Da and 55% solvent with one molecule in the asymmetric unit (*23*). A single TP synthase crystal was allowed to soak for 30 minutes in the crystallization solution supplemented with 10 mM CO_2-ThzP and 10 mM CF_3HMP-PP. The crystal was quickly transferred to a cryoprotectant of the soak solution with 10% glycerol before being flash frozen by plunging the crystal into liquid nitrogen.

X-Ray Data Collection and Processing. A single wavelength dataset was collected at 0.97920 Å at 100 K on the NE-CAT beamline 24-ID-C at the Advanced Photon Source. An oscillation range of 1° was used over 75 frames with 1 second exposure on a Quantum315 detector (Area Detector Systems Corp.). A single native crystal diffracted to 1.95 Å. The data were indexed, integrated, and scaled using the HKL2000 suite of programs (*24*). The data collection statistics are given in Table 2.1.

Structure Determination, Model Building, and Refinement. The monomeric coordinates of *B. subtilis* TP synthase (PDB ID: 1G4T) with all water and ligand atoms removed was used as a search model for molecular replacement in CNS (*25*). Rounds of refinement in CNS consisted of rigid body refinement, simulated temperature annealing, restrained individual B-factor refinement, and minimization refinement, followed by manual model adjustment in COOT (*25*, *26*). Water molecules were added during later rounds of refinement using CNS and ligands were then manually placed in COOT (*27*). The structure geometry was verified using PROCHECK; one residue (Asp93) was found in the disallowed region, although the

density for this residue is clear (*28*). The refinement statistics for the final model are given in Table 2.2. PyMOL and ChemDraw were used to generate figures (*29*).

Table 2.1. Summary of Data Collection Statistics

	BsTPS CO_2-ThzP
Source	APS 24-ID-C
Resolution (Å)	1.95
Energy (eV)	12662
Space Group	$P4_22_12$
a (Å)	96.0
b (Å)	96.0
c (Å)	59.3
Matthew's Coef.	2.7
% solvent	55
Mol/a.s.u.	1
Measured Reflections	69987
Unique reflections	20400 (2006) [a]
Average I/σ	17.3 (4.3)
Redundancy	3.4 (3.4)
Completeness (%)	98.3 (99.6)
Rsym [b] (%)	6.2 (33.0)

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

Analytical Ultracentrifugation. Sedimentation velocity analytical ultracentrifugation was performed to determine the oligomeric state of TP synthase in solution using a ProteomeLab XL-1 protein characterization system (Beckman Coulter). Samples were buffer exchanged to 35 mM Tris, 150 mM NaCl, and 2 μM β-mercaptoethanol (pH 7.5) and then diluted to Abs_{280} of 0.35 to 1.2 prior to centrifugation. SEDNTERP was used to calculate the partial specific volume of TP synthase and the buffer density and viscosity while the program SEDFIT was used to analyze the data (*30, 31*).

Section 2.3. Results

NMR Characterization of TP Synthase with CO_2-ThzP and HMP-PP. The formation of TPP was monitored over time using 600 MHz ^{1}H-NMR and observing

the disappearance and emergence of proton peaks within the spectra, shown in Figure 2.2. The control reaction (Figure 2.2A), utilizing HMP-PP and ThzP as substrates, illustrates the shift of the proton on C5 of HMP-PP upfield in the product TP. The proton at C2 of ThzP is shifted downfield in TP. Figures 2.2B and 2.2C each demonstrate TP synthase will utilize CO_2-ThzP in the formation of TP, as indicated by the appearance of the protons at 7.96 ppm and 9.26 ppm and the disappearance of the proton at C5 of HMP-PP. Performing this reaction in 22% D_2O with HMP-PP and CO_2-ThzP showed the gradual appearance and disappearance of the proton at C2 of the thiazole moiety (Figure 2.2C) in addition to the proton shifts seen in Figure 2.2B.

Table 2.2. Summary of Data Refinement Statistics

	BsTPS CO_2-ThzP
Resolution (Å)	50.00 – 1.95
No. of protein atoms	1635
No. of ligand atoms	37
No. of water atoms	212
Reflections in working set	18991
Reflections in test set	970
R factor[a] (%)	19.8
R_{free}[b] (%)	23.2
Rmsd from ideal	
Bonds (Å)	0.0052
Angles (º)	1.2
Average B-factor (Å^3)	28.9
Ramachandran Plot	
Most favored (%)	91.0
Additionally allowed (%)	7.9
Generously allowed (%)	0.5
Disallowed (%)	0.5

[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.

Steady State Kinetic Characterization of TP Synthase with HMP-PP and CO_2-ThzP. A coupled enzyme assay was used to determine the steady state kinetic parameters of HMP-PP at a saturating concentration of ThzP by plotting the rate of production of TP (Figure 2.3A). This previously described assay monitored the

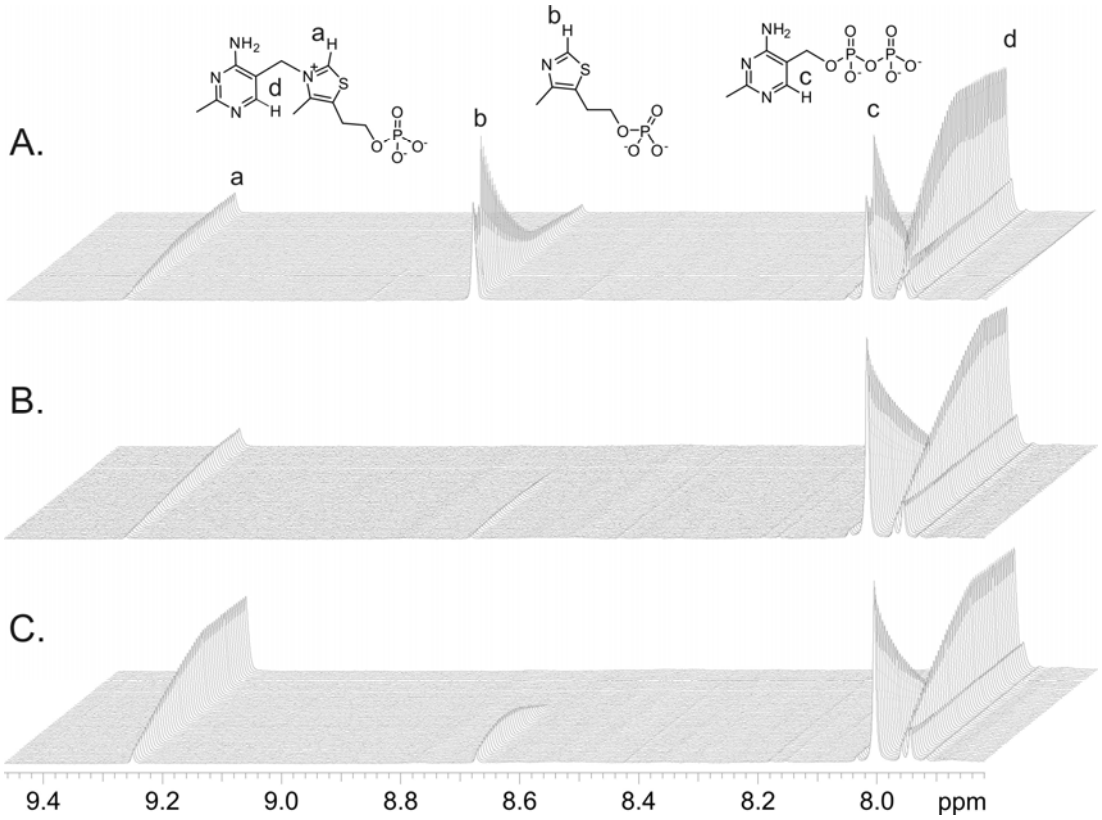


Figure 2.2. 600 MHz ^{1}H-NMR of the TP synthase catalyzed formation of TP using ThzP or CO_2-ThzP as substrates. (A) Control reaction showing the formation of TP monitored as a function of time (~0-80 min, with the longest time in back). The appearance of the peak corresponding to the C2 proton of TP (proton **a**) appears to be substoichiometric to the aromatic proton **d** due to its exchange with deuterium in the buffer composed of 70% D_2O. (B) An identical reaction was performed as in the control, except that ThzP was substituted with CO_2-ThzP. Again TP appeared, albeit at a slightly reduced steady state rate. (C) An identical reaction as to that described in B was performed except that the amount of D_2O was reduced to 22% to more clearly inspect the formation of a peak corresponding to the C2 proton of ThzP.

absorbance at 411 nm and coupled the activity of TP synthase to thiaminase I and 4-nitrothiophenolate (*17*). The data were fit to the Michaelis-Menton equation and defined a k_{cat} of 0.021 ± 0.002 s^{-1} and a K_m of 3.5 ± 1.4 μM for HMP-PP. The steady state parameters of ThzP and CO_2-ThzP were determined by plotting the rate of TP formation as a function of substrate concentration under saturating HMP-PP conditions (Figure 2.3B). Due to the relatively tight binding of the ThzP substrates to TP synthase, a fluorescent assay based upon a fluorescently tagged mutant *E. coli*

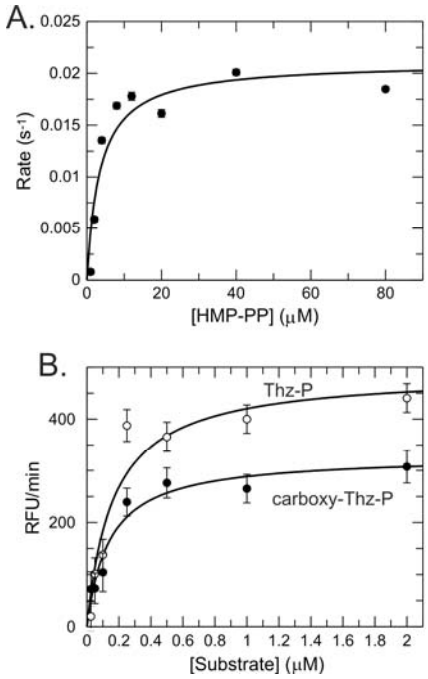


Figure 2.3. Steady state kinetics of TP formation using HMP-PP, ThzP, and CO_2-ThzP as substrates. (A) Plot of the rate of TP production as a function of HMP-PP concentration (at saturating ThzP concentration) using a coupled assay employing thiaminase I and a chromophoric co-substrate, 4-nitrothiophenolate. A fit of the data to the Michaelis-Menten equation defined a k_{cat} of 0.021 ± 0.002 s^{-1} and a K_m of 3.5 ± 1.4 μM for HMP-PP. (B) The rate of formation of thiamin-P is plotted as a function of either Thz-P or carboxy-Thz-P using a fluorescent assay based upon the use of a fluorescently labeled mutant *E. coli* TbpA. The data were fit using the Michaelis-Menten equation to extract a k_{cat} of 490 ± 60 RFU/min^{-1} (~0.021 s^{-1}) and a K_m of 165 ± 65 nM for ThzP and a k_{cat} of 330 ± 30 RFU/min^{-1} (~0.014 s^{-1}) and a K_m of 140 ± 40 nM for CO_2-ThzP.

TbpA was used. The fluorescence of TbpA decreases by roughly 50% upon the binding of TP. The data were again fit to the Michaelis-Menton equation. The k_{cat} and K_m for ThzP are 490 ± 60 RFU/min^{-1} (~0.021 s^{-1}) and 165 ± 65 nM, respectively. A k_{cat} of 330 ± 30 RFU/min^{-1} (~0.014 s^{-1}) and a K_m of 140 ± 40 nM were calculated for CO_2-ThzP.

Kinetic Characterization of the Non-enzymatic Decarboxylation of CO_2-EtThz. NEDMT was prepared using modified synthetic protocols as a precursor to CO_2-EtThz with a protecting group on the carboxylate group that is removable through exposure to UV light with a wavelength longer than 300 nm as shown in Figure 2.4 (*20, 21*).

The rate of decarboxylation of CO_2-EtThz to form EtThz can then be determined by monitoring the peak shift corresponding to the protons on the ethyl group at the 3 position of the thiazole ring (Figure 2.5). The rate of decarboxylation was determined using a first order equation and the relative areas beneath the peaks at 1.40 ppm and 1.51 ppm, and *k* was found to be 4.6×10^{-6} s^{-1}. The half-life for CO_2-EtThz is 41.5 hours.

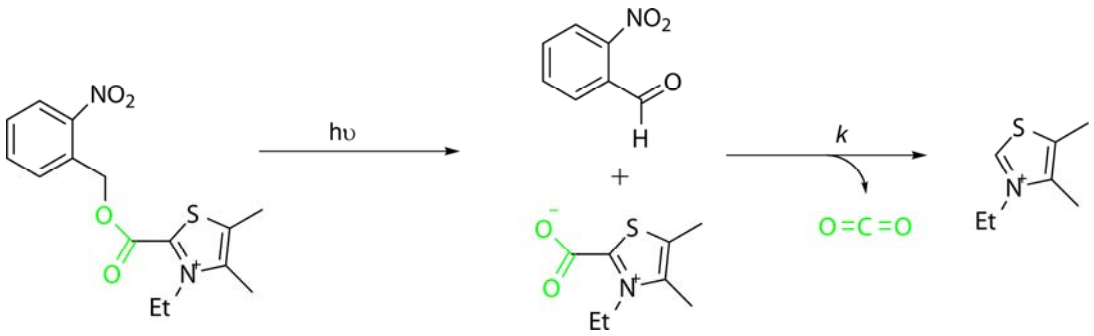


Figure 2.4. Photolytic deprotection of CO2-EtThz and non-enzymatic decarboxylation to form EtThz.

Overall Structure of TP Synthase Complex. The structure of TP synthase with CF_3-HMP-PP and CO_2-ThzP bound was determined to 1.95 Å using the structure of *Bs*TP synthase (PDB ID: 1G4T) as the search model for molecular replacement. TP synthase crystallized in the space group $P4_22_12$ with a = 96.0 Å and c = 59.3 Å and a corresponding Matthew's coefficient of 2.74 $Å^3$/Da for one molecule in the asymmetric unit (*23*). The final *Bs*TP synthase model consists of one polypeptide chain with 219 residues, 212 water molecules, and one molecule each of CO_2-ThzP, CF_3HMP, and PP_i. The N-terminal 14 residues and the C-terminal two residues are disordered and are not included in the model. TP synthase is an $(\beta/\alpha)_8$ barrel with an additional α-helix covering the N-terminal end of the β-barrel and the active site is located at the C-terminal end of the β-barrel (Figure 2.6). Analysis of crystal packing suggests *Bs*TP synthase adopts a monomeric assembly and was confirmed by analytical ultracentrifugation, as seen in Figure 2.7 (*32*).

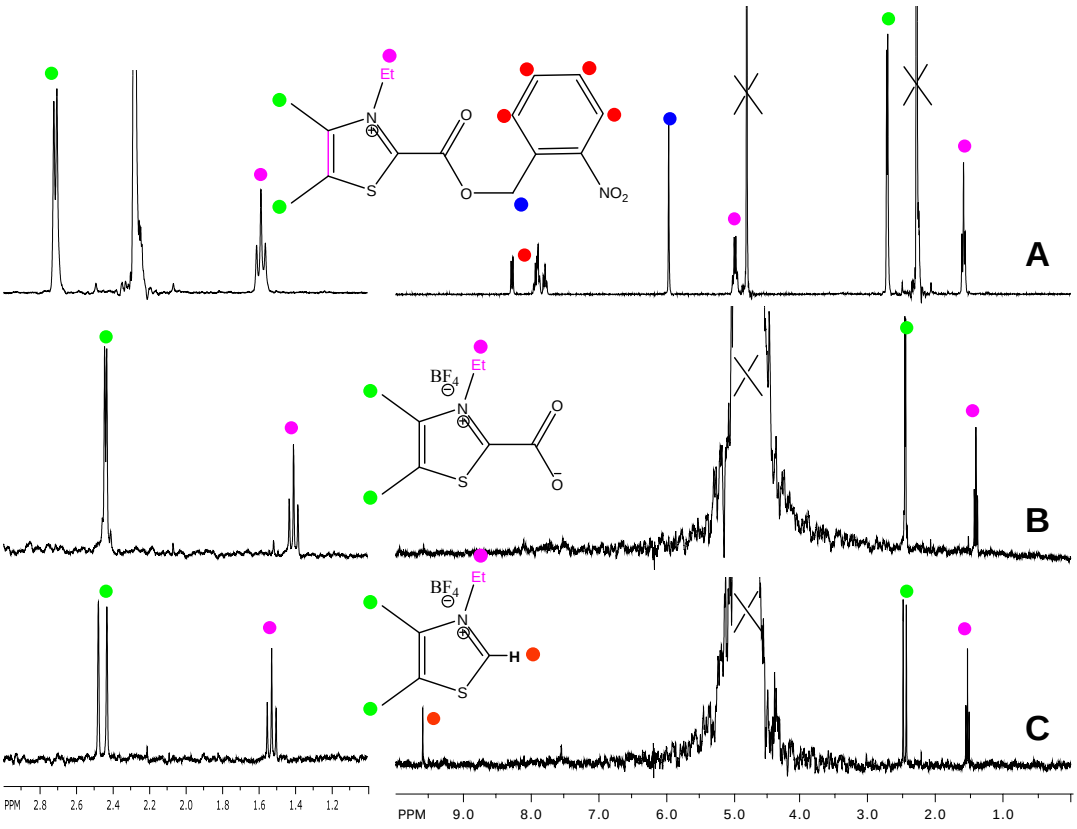


Figure 2.5. ^{1}H-NMR spectra of the deprotection and decarboxylation of CO_2-EtThz. C2-Proton of the thiazolium ring produces a peak at 9.6 ppm. (A) NEDMT before photolysis. (B) NEDMT after photolysis. (C) Decarboxylation of NEDMT after photolysis.

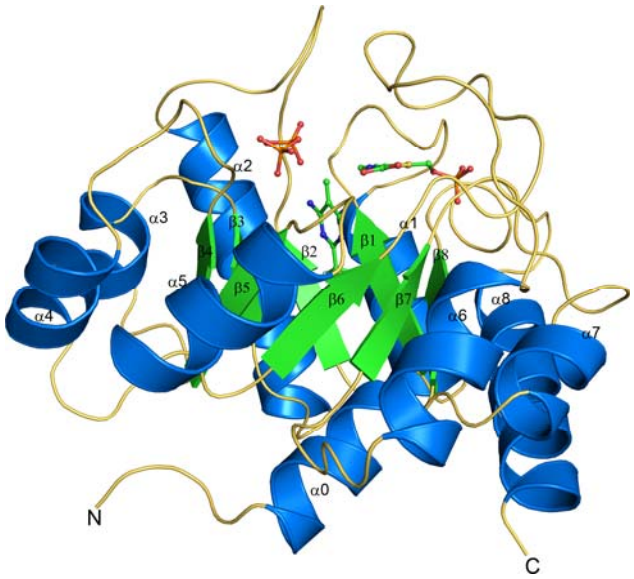


Figure 2.6. Ribbon diagram of the monomeric TP synthase structure with ligands shown in ball and stick representation. The secondary structural elements are labeled, with α-helices shown in blue and the β-strands colored green.

Binding of Substrate in TP Synthase. The active site of TP synthase is located at the C-terminal end of the β-barrel and several loop regions contribute to interactions between TP synthase and the ligands. These loop regions, particularly the loop connecting β6 to α6, close over the active site and largely shield the active site from solvent. The TP synthase crystal was soaked for thirty minutes with 10 mM CF_3HMP-PP and 10 mM CO_2-ThzP. The resulting density is clear for CF_3HMP, CO_2-ThzP, and PP_i (Figure 2.8A). CF_3HMP forms π-stacking interactions with His107 and binds in a mostly hydrophobic environment, surrounded by Ile31, Ile186, Val184, Tyr29, Tyr147, and Ser206. Gln37, from β2, forms two hydrogen bonds of 3.1 and 3.3 Å with the N3 and N4 atoms of the pyrimidine ring, shown in Figure 2.8B. The density of the PP_i group is discrete and separate from that of the CF_3HMP moiety. The negative charge of PP_i is stabilized through interactions with Arg59, Lys61, and Lys159. Additionally, a magnesium ion interacts with both phosphate groups and is

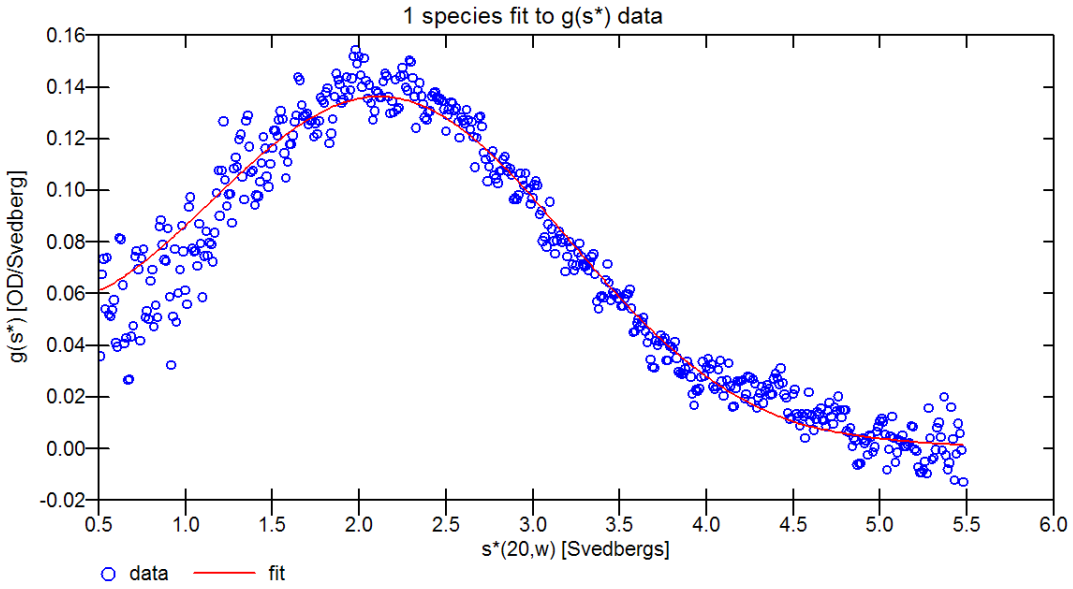


Figure 2.7. Analytical ultracentrifugation results showing the major species in solution is a monomer.

coordinated by Asp93 and Asp112. Asn92 is oriented towards one phosphate group, and the other phosphate group can interact with His132 and the amide nitrogen atom of the backbone at Gln110 via two water molecules (Figure 2.8). CO_2-ThzP binds roughly 90° rotated from CF_3HMP. The thiazole ring of CO_2-ThzP stacks between Lys159 and Ile136 and the carboxylate group at the C2 atom forms a strong water-mediated hydrogen bond to His132. The carboxylate group also forms two additional hydrogen bonds with water molecules in the active site. The phosphate tail of CO_2-ThzP is well coordinated, interacting with Thr156, Thr158, Ser209, and the amide nitrogen atom of the protein backbone at Gly188 (Figure 2.8B). The oxygen atoms of the phosphate group of CO_2-ThzP also form hydrogen bonds with three water molecules. One of these water molecules is positioned by the carbonyl oxygen atom of the main chain at Ser206 and amide nitrogen atoms at Gly187 and Ile189; the second water molecule coordinated to the phosphate group forms hydrogen bonds with the backbone carbonyl oxygen atom of Ile189 and the backbone amide nitrogen atom of Ala210.

Section 2.4. Discussion

CO_2-ThzP As A Substrate for TP Synthase. Monitoring the ^{1}H-NMR time course of TP synthase with HMP-PP and either ThzP or CO_2-ThzP shows that both thiazole moieties lead to the production of thiamin phosphate. The control experiment, using HMP-PP and ThzP, demonstrated the slight shift of the proton at C5 of the pyrimidine ring and the dramatic relocation of the C2 proton peak of the thiazole moiety upon the formation of the carbon-nitrogen bond to generate thiamin phosphate over 80 minutes. The experiment was then repeated using CO_2-ThzP and HMP-PP in 70% D_2O and proton peaks **a** and **d** grew in, demonstrating that the product of TP synthase when incubated with CO_2-ThzP is thiamin phosphate, not CO_2-thiamin phosphate. Complete conversion to thiamin phosphate on this time scale

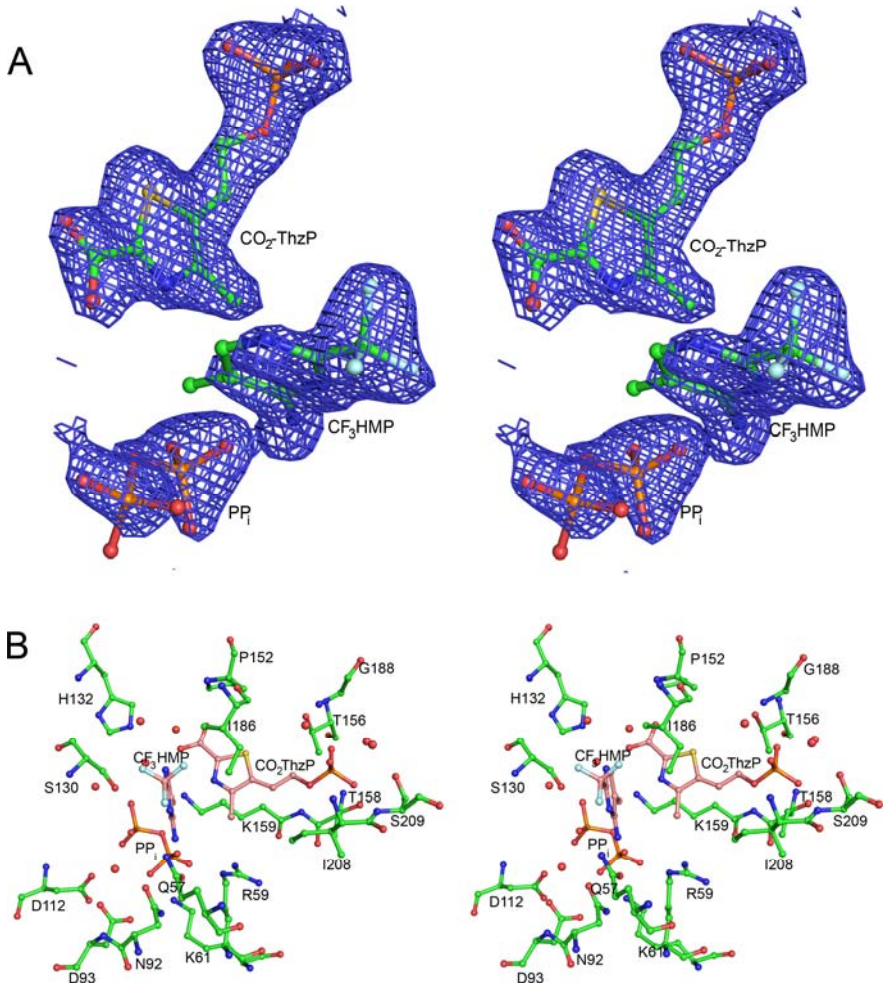


Figure 2.8. Stereoview diagrams of the TP synthase active site with ligands bound. Three species are observed in the active site. Residues and ligands are shown in ball and stick representation and colored according to atom type. Red nonbonded spheres are water molecules. (A) Ligand orientation with composite omit electron density shown as a blue mesh at a contour level of 2.0 σ. (B) Substrate binding interactions with TP synthase. Substrate carbon atoms are colored salmon, while carbon atoms belonging to TP synthase are shown in green.

suggests the decarboxylation of CO_2-thiamin phosphate occurs quickly. A slight increase of the proton at C2 of the thiazole moiety (proton peak **b**), which then disappeared, was surprising and to verify that this was not an artifact, the reaction was repeated at a lower percentage of D_2O to enhance the deuterium-proton exchange. This effect was increased when 20% D_2O was used and suggests that after CO_2-

thiamin phosphate is formed in the active site of TP synthase, a fraction of this species reacts in the reverse reaction with PP_i and ThzP is then dissociated from the active site. No accumulation of ThzP has been observed when TP synthase is incubated with only CO_2-ThzP, or when TP synthase is incubated with both CO_2-ThzP and CF_3HMP-PP, indicating that the formation of CO_2-thiamin phosphate must occur before the detection of ThzP is possible.

Kinetic Characterization of TP Synthase With CO_2-ThzP. Previous characterization of the steady state kinetic parameters for TP synthase relied upon a well characterized indirect assay that converts thiamin phosphate via oxidation to thiochrome phosphate, a highly fluorescent molecule (*11, 15, 33*). The steady state kinetic parameters for HMP-PP reported here were determined using a coupled assay with thiaminase I and the results agreed fairly well with those determined for HMP-PP previously using the linear portion of burst kinetics, demonstrating that the coupled assay is fairly accurate for determination of k_{cat} (*12, 17*). TP synthase binds the thiazole moiety fairly well, so the steady state kinetic parameters for CO_2-ThzP and ThzP were determined using a fluorescently tagged mutant of *E. coli* thiamin binding protein (TbpA) (Hanes, J.W., Begley, T.P., unpublished work). The k_{cat} is similar for both thiazole moieties, but not identical. The k_{cat} for CO_2-ThzP is slightly lower (~67% that of ThzP), but the k_{cat}/K_m are very close in value, ~0.1 $\mu M^{-1}s^{-1}$ for CO_2-ThzP and ~0.13 $\mu M^{-1}s^{-1}$ for ThzP. This data indicates that these two molecules are approximately equally good substrates for TP synthase.

Non-enzymatic Decarboxylation of CO_2-EtThz. The non-enzymatic decarboxylation of CO_2-EtThz can be directly characterized through the removal of a photolyzable protecting group on the carboxylate group and monitored using ^{1}H-NMR (Figure 2.5). The ethyl group on the nitrogen atom of the thiazole ring generates a positive charge on the nitrogen atom and allows for formation of the ylide after

decarboxylation and is thus a fair model for CO_2-thiamin phosphate, which also has a positive charge on the thiazole nitrogen atom. Removal of the photoprotecting group *o*-nitrobenzyl is rapid upon exposure to UV light with a wavelength greater than 300 nm, followed by the slow decarboxylation of CO_2-EtThz, as indicated by the 41.5 hour half-life of this species. The inability to detect CO_2-thiamin phosphate and the relatively quick complete conversion of CO_2-ThzP and HMP-PP to thiamin phosphate within 80 minutes suggest that the decarboxylation of the product of TP synthase must be enzyme-catalyzed.

Oligomeric Assembly of BsTP Synthase. The TP synthase complex with CO_2-ThzP, CF_3HMP, and PP_i crystallized in the space group $P4_22_12$ with unit cell parameters a = 96 Å and c = 59 Å and one molecule in the asymmetric unit. The previous structures of *Bs*TP synthase crystallized in a different space group, $P4_32_12$, with unit cell parameters a = 76 Å and c = 139 Å with two molecules in the asymmetric unit (*8, 13*). The dimer reported previously orients α2 running antiparallel to α2 of the second chain and forms additional interactions between the N-terminal residues and the N-terminus of α2. Several water molecules also mediate the dimer interactions (*8*). The β-barrels of the two protomers face in opposite directions so that the active sites are on opposing sides of the dimer. Other structurally characterized thiamin binding proteins belonging to the $(\beta/\alpha)_8$ barrel family, including TP synthase (PDB ID: 1XI3) and TenI, form dimers in which the β-barrels run parallel to each other and consequently have the active sites on the same face of the dimer (*34, 35*). Additionally, the structure of a putative TP synthase from *Bacteroides thetaiotamicron* is a monomer (PDB ID: 3CEU). The structure reported here of *Bs*TP synthase is also a monomer and the oligomeric state in solution was confirmed using analytical ultracentrifugation with an apparent molecular weight of 29 kDa (Figure 2.7). The various oligomeric states observed for TP synthase and TenI proteins indicate that

while the tertiary structure of these enzymes is well conserved, the oligomeric assembly is not an important feature for function.

Comparison of CO_2-ThzP and ThzP Binding in TP Synthase. The crystal structure of TP synthase with CO_2-ThzP, CF_3HMP, and PP_i allows for comparisons to the structures available of TP synthase with ThzP and the reaction pathway mapped out previously (*13*). The current structure most closely resembles the intermediate complex obtained by soaking overnight TP synthase S130A crystals with CF_3HMP-PP and ThzP. While the CF_3HMP-PP and CO_2-ThzP soak was only 30 minutes, the use of wild type TP synthase resulted in a similar intermediate, the pyrimidine iminemethide, being trapped. Nearly all the protein side chains of active site residues are found in the same conformations; the most variability is seen in the loop connecting β6 to α6 and the shift in loop conformation is less than 1.5 Å. A comparison of the current structure with the product complex of TP synthase with thiamin phosphate bound (PDB ID: 1G4S) shows that the thiazole moiety has moved roughly 0.75 Å further from the pyrimidine moiety, and the pyrimidine moiety is tilted at approximately 15° with respect to the pyrimidine ring of thiamin phosphate (*13*). The pyrimidine moiety has been shifted less than 1 Å closer to CO_2-ThzP when compared with the substrate complex (PDB ID: 1G69) but this could be due to the trifluoromethyl substitution as the ternary complex of CF_3HMP, ThzP, and PP_i has the pyrimidine moiety positioned in the same manner (PDB ID: 1G6C).

Comparison of BsTP Synthase Active Site to TenI Active Site. Recent work has shown that the true product of the ThiS-ThiG thiazole synthase reaction is not ThzP, but a tautomer of CO_2-ThzP (*15*). TenI, a highly homologous protein to TP synthase, was originally thought to catalyze the formation of thiamin phosphate but structural characterization of TenI showed that the presence of a bulky residue in the active site precluded binding of the pyrimidine moiety (*8, 34, 36*). TenI has now been shown to

catalyze the aromatization of the thiazole tautomer by deprotonation at the C2 of the thiazole ring by an active site histidine residue (A. Hazra, A. Chatterjee, T.P. Begley, unpublished work). The comparison of TenI to TP synthase showed that HMP-PP binding would lead to significant steric clashes with Leu119 (*34*). Additionally, the magnesium binding site was not conserved, so the pyrophosphate group would not be stabilized after dissociation from HMP. ThzP was modeled in the active site of TenI and the model suggested TenI would be capable of binding the thiazole portion of thiamin phosphate. Figure 2.9 compares the structures of TP synthase with TenI through the superposition of TenI upon TP synthase. While the pyrimidine moiety does form steric clashes, CO_2-ThzP superimposes well into the active site of TenI and the phosphate group of the thiazole moiety is marked by a sulfate group in the TenI structure. The loop region connecting β6 to α6 is disordered in TenI and would be expected to become ordered upon substrate binding. Several residues involved in the coordination of the thiazole moiety are conserved (Figure 2.9). The thiazole ring is

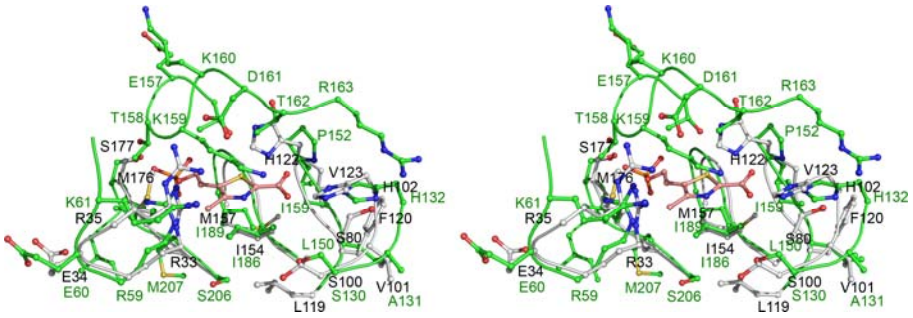


Figure 2.9. Stereoview diagram of a superposition of CO_2-ThzP from TP synthase with TenI from *B. subtilis*. TP synthase is shown with green carbon atoms, while carbon atoms from TenI are colored gray. Water molecules and secondary structural elements have been omitted for clarity and CO_2-ThzP from the TP synthase structure is shown in salmon.

stacked against an isoleucine residue and the phosphate tail of the thiazole moiety is coordinated by a backbone amide nitrogen atom 2.8 Å away and the side chain of Ser209. His132 is oriented toward the carboxylate group on CO_2-ThzP, although the distance (3.9 Å) is too long to be a hydrogen bond. Another positively charged

residue common to both TP synthase and TenI is Arg59, found near the methyl group at C4. The putative catalytic residue of TenI, His122, is not conserved in TP synthase (Pro152).

The disordered loop of TenI could contain additional residues important for stabilization or aromatization of the thiazole tautomer. A sequence alignment of five TenI and TP synthase proteins was prepared to identify potentially important residues for either the decarboxylation reaction or the aromatization reaction and is shown in Figure 2.10 (*37-39*). Several of the strictly conserved residues are found at the C-terminal end of β-strands, corresponding to residues in or near the active site. Arg59, Ser209, and several glycines are all conserved. Interestingly, Ser130, a key residue for the dissociation of PP_i in the TP synthase reaction, is also absolutely conserved. His122 of TenI, which could potentially be required for aromatization, is conserved only among other TenI proteins and is often a proline residue in TP synthase. Lys159, a potentially important residue for the decarboxylation of CO_2-ThzP, is found on the loop between β7 and α7. When a thiazole moiety is bound in the active site, this loop is ordered in TP synthase and Lys159 adopts an extended conformation over the thiazole ring with the amino terminus directly above the carboxylate group (Figure 2.8B). However, this residue is also conserved among TenI proteins.

Mechanistic Implications. Previous work has demonstrated that the rate limiting step of this reaction is the dissociation of product from the active site of TP synthase and that the coupling proceeds through a dissociative mechanism that produces a pyrimidine carbocation (*11, 12*). Steady state kinetic characterization of TP synthase has revealed that ThzP and CO_2-ThzP are roughly equivalent substrates. The decarboxylation of CO_2-ThzP appears to occur after the coupling of the thiazole moiety to the pyrimidine moiety, as this is the only way for a small amount of ThzP to be transiently detected by the NMR monitoring. Additionally, the rapid formation of

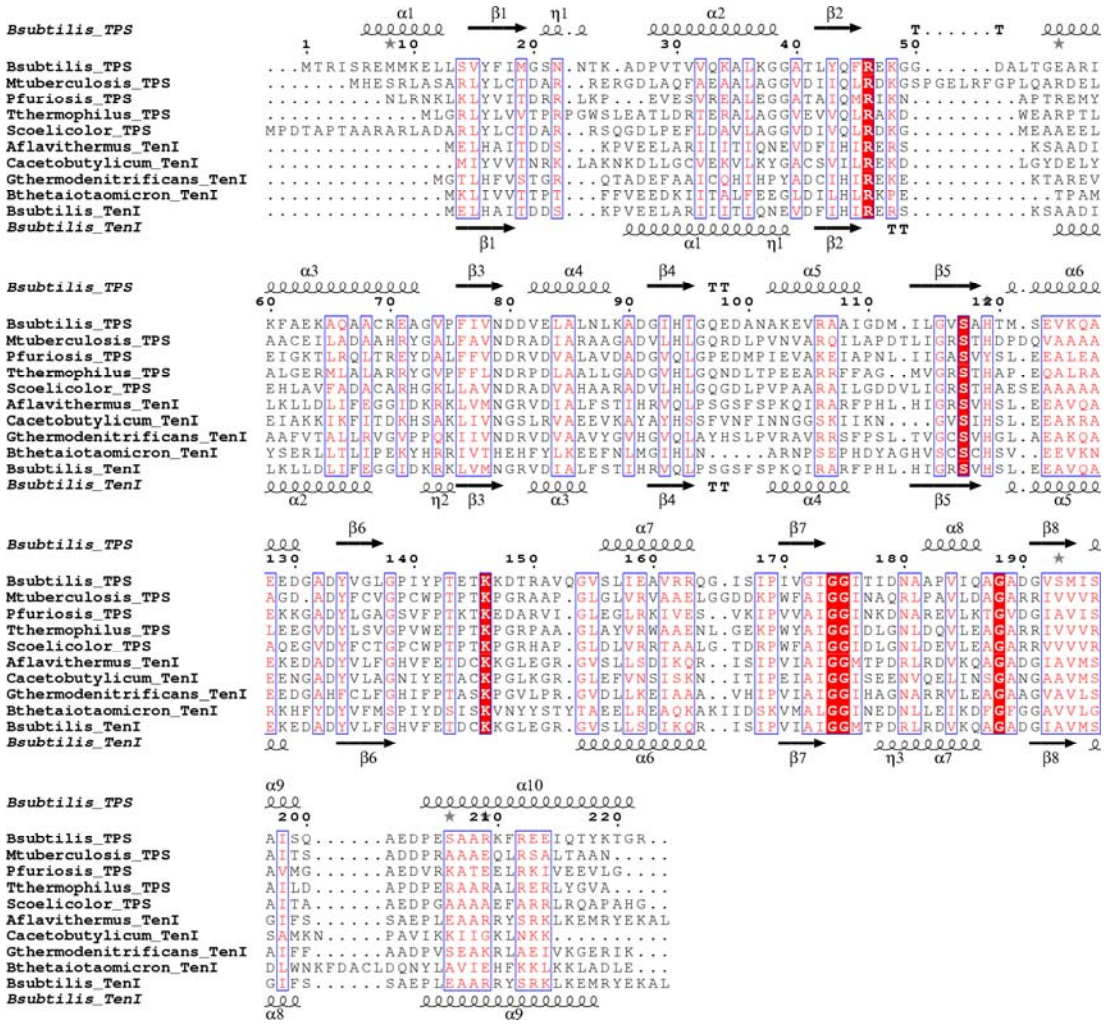


Figure 2.10. Sequence alignment of various TP synthases and TenI proteins. Secondary structural elements for TP synthase are shown above the alignment and structural elements from TenI are shown below the alignment. Absolutely conserved residues are highlighted in red while less strictly conserved residues are colored red.

thiamin phosphate indicates that either CO_2-thiamin phosphate is unstable and spontaneously decarboxylates nonenzymatically or TP synthase enhances the decarboxylation reaction. Characterization of the model compound CO_2-EtThz, however, demonstrates that this compound is very stable, with a half-life of more than a day and a half. CO_2-thiamin phosphate, which also contains a positively charged nitrogen atom in the thiazole ring, therefore most likely undergoes an enzyme-catalyzed decarboxylation. Ser130 is too far from the carboxylate group to participate

(5.0 Å) and His132 is nearly 4 Å, making this an unlikely residue for the catalysis. Lys159, with the amide group in close proximity of the carboxylate group of CO_2-ThzP, could be involved in this reaction. However, the presence of Lys159 in the TenI proteins, which do not catalyze the decarboxylation of the thiazole tautomer, could indicate that Lys159 provides a hydrogen bond only to the carboxylate group to stabilize negative charge. Alternatively, the lack of positive charge on the nitrogen atom of the thiazole ring in CO_2-ThzP, the product of TenI, could preclude decarboxylation with no means for stabilization of the decarboxylated product. Future studies probing the role of this absolutely conserved lysine residue will provide more information on the mechanism of the decarboxylation of CO_2-thiamin phosphate.

REFERENCES

1. Jordan, F. (2003) Current mechanistic understanding of thiamin diphosphate-dependent enzymatic reactions, *Nat. Prod. Rep. 20*, 184-201.
2. Spenser, I. D., White, R. L. (1996) Biosynthesis of Vitamin B1 (Thiamin): An Instance of Biochemical Diversity, *Angew. Chem., Int. Ed. 36*, 1032-1046.
3. Haas, A. L., Laun, N. P., and Begley, T. P. (2005) Thi20, a remarkable enzyme from Saccharomyces cerevisiae with dual thiamin biosynthetic and degradation activities, *Bioorg. Chem. 33*, 338-344.
4. White, R. L., Spenser, I.D. (1982) Thiamin biosynthesis in yeast. Origin of the thiazole moiety., *J. Am. Chem. Soc. 104*, 4934-4943.
5. Wightman, R., and Meacock, P. A. (2003) The THI5 gene family of Saccharomyces cerevisiae: distribution of homologues among the hemiascomycetes and functional redundancy in the aerobic biosynthesis of thiamin from pyridoxine, *Microbiology 149*, 1447-1460.
6. Begley, T. P., Downs, D. M., Ealick, S. E., McLafferty, F. W., Van Loon, A. P., Taylor, S., Campobasso, N., Chiu, H. J., Kinsland, C., Reddick, J. J., and Xi, J. (1999) Thiamin biosynthesis in prokaryotes, *Arch. Microbiol. 171*, 293-300.
7. Settembre, E., Begley, T. P., and Ealick, S. E. (2003) Structural biology of enzymes of the thiamin biosynthesis pathway, *Curr. Opin. Struct. Biol. 13*, 739-747.
8. Chiu, H. J., Reddick, J. J., Begley, T. P., and Ealick, S. E. (1999) Crystal structure of thiamin phosphate synthase from Bacillus subtilis at 1.25 A resolution, *Biochemistry 38*, 6460-6470.
9. Nosaka, K., Nishimura, H., Kawasaki, Y., Tsujihara, T., and Iwashima, A. (1994) Isolation and characterization of the THI6 gene encoding a bifunctional

thiamin-phosphate pyrophosphorylase/hydroxyethylthiazole kinase from Saccharomyces cerevisiae, *J. Biol. Chem. 269*, 30510-30516.

10. Zhang, Y., Taylor, S. V., Chiu, H. J., and Begley, T. P. (1997) Characterization of the Bacillus subtilis thiC operon involved in thiamine biosynthesis, *J. Bacteriol. 179*, 3030-3035.
11. Reddick, J. J., Nicewonger, R., and Begley, T. P. (2001) Mechanistic studies on thiamin phosphate synthase: evidence for a dissociative mechanism, *Biochemistry 40*, 10095-10102.
12. Hanes, J. W., Ealick, S. E., and Begley, T. P. (2007) Thiamin phosphate synthase: the rate of pyrimidine carbocation formation, *J. Am. Chem. Soc. 129*, 4860-4861.
13. Peapus, D. H., Chiu, H. J., Campobasso, N., Reddick, J. J., Begley, T. P., and Ealick, S. E. (2001) Structural characterization of the enzyme-substrate, enzyme-intermediate, and enzyme-product complexes of thiamin phosphate synthase, *Biochemistry 40*, 10103-10114.
14. Chatterjee, A., Schroeder, F. C., Jurgenson, C. T., Ealick, S. E., and Begley, T. P. (2008) Biosynthesis of the thiamin-thiazole in eukaryotes: identification of a thiazole tautomer intermediate, *J. Am. Chem. Soc. 130*, 11394-11398.
15. Hazra, A., Chatterjee, A., and Begley, T. P. (2009) Biosynthesis of the Thiamin Thiazole in Bacillus subtilis: Identification of the Product of the Thiazole Synthase-Catalyzed Reaction, *J. Am. Chem. Soc. 131*, 3225-3229.
16. 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.
17. Hanes, J. W., Kraft, C. E., and Begley, T. P. (2007) An assay for thiaminase I in complex biological samples, *Anal. Biochem. 368*, 33-38.

18. Soriano, E. V., Rajashankar, K. R., Hanes, J. W., Bale, S., Begley, T. P., and Ealick, S. E. (2008) Structural similarities between thiamin-binding protein and thiaminase-I suggest a common ancestor, *Biochemistry 47*, 1346-1357.

19. Brune, M., Hunter, J. L., Corrie, J. E., and Webb, M. R. (1994) Direct, real-time measurement of rapid inorganic phosphate release using a novel fluorescent probe and its application to actomyosin subfragment 1 ATPase, *Biochemistry 33*, 8262-8271.

20. Castells, J., Calahorra, F.L., and Domingo, L. (1985) Use of Thiazolium-2-Carboxylates to Induce Benzoin Condensations, *Tet. Lett. 26*, 5457-5458.

21. Schlegel, J., and Maas, G. (1999) Propyne Iminium Salts by N-Alkylation of Alkynyl Imines, *Synthesis 1*, 100-106.

22. Murray, C. J., and Duffin, K.L. (1991) Determination of Rates of Proton Exchange of Thiamine Hydrochloride by ^{1}H-NMR Spectroscopy, *J. Chem. Ed. 68*, 683-684.

23. Matthews, B. W. (1968) Solvent content of protein crystals, *J. Mol. Biol. 33*, 491-497.

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

25. 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 Cryst. D54*, 905-921.

26. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Cryst. D60*, 2126-2132.

27. Schuttelkopf, A. W., and van Aalten, D. M. (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes, *Acta Cryst. D60*, 1355-1363.

28. 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. Cryst. 26*, 283-291.

29. DeLano, W. L. (2002) The PyMOL Molecular Graphics Systems, DeLano Scientific, San Carlos, CA.

30. Brown, P. H., and Schuck, P. (2006) Macromolecular size-and-shape distributions by sedimentation velocity analytical ultracentrifugation, *Biophys. J. 90*, 4651-4661.

31. Philo, J. S. (2006) Improved methods for fitting sedimentation coefficient distributions derived by time-derivative techniques, *Anal. Biochem. 354*, 238-246.

32. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state, *J. Mol. Biol. 372*, 774-797.

33. Park, J. H., Dorrestein, P. C., Zhai, H., Kinsland, C., McLafferty, F. W., and Begley, T. P. (2003) Biosynthesis of the thiazole moiety of thiamin pyrophosphate (vitamin B1), *Biochemistry 42*, 12430-12438.

34. Toms, A. V., Haas, A. L., Park, J. H., Begley, T. P., and Ealick, S. E. (2005) Structural characterization of the regulatory proteins TenA and TenI from Bacillus subtilis and identification of TenA as a thiaminase II, *Biochemistry 44*, 2319-2329.

35. 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.

36. Rodionov, D. A., Vitreschak, A. G., Mironov, A. A., and Gelfand, M. S. (2002) Comparative genomics of thiamin biosynthesis in procaryotes. New genes and regulatory mechanisms, *J Biol Chem 277*, 48949-48959.

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

38. Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H. Y., Cohoon, M., de Crecy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E. D., Gerdes, S., Glass, E. M., Goesmann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R., Jamshidi, N., Krause, L., Kubal, M., Larsen, N., Linke, B., McHardy, A. C., Meyer, F., Neuweger, H., Olsen, G., Olson, R., Osterman, A., Portnoy, V., Pusch, G. D., Rodionov, D. A., Ruckert, C., Steiner, J., Stevens, R., Thiele, I., Vassieva, O., Ye, Y., Zagnitko, O., and Vonstein, V. (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes, *Nucleic Acids Res 33*, 5691-5702.

39. 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.

CHAPTER 3

CRYSTAL STRUCTURE AND IDENTIFICATION OF POTENTIAL INHIBITOR COMPOUNDS FOR *MYCOBACTERIUM TUBERCULOSIS* THIAMIN PHOSPHATE SYNTHASE

Section 3.1. Introduction

Mycobacterium tuberculosis presents a significant threat to human health (*1*). *M. tuberculosis* is an opportunistic disease which has also recently begun to develop worrying resistance to the drug therapies that are currently available (*2*). *M. tuberculosis* requires a long course of six months of antibiotic treatment, which can lead to patients stopping treatment early before the infection has been completed (*1*). The drugs available have been on the market for up to half a century with no new treatments released. Several of these available drugs, such as isoniazid and pyrazinamide, two first line drugs, target cell wall biosynthesis or membrane energy metabolism (*3*). In extreme drug resistant *M. tuberculosis*, it is common for mutations to render these treatments inaffective (*3*). The development of new drugs targeting different pathways within *M. tuberculosis* would expand the arsenal of treatment options.

Thiamin biosynthesis in *M. tuberculosis* presents an ideal pathway for inhibiting the growth of the bacterium. Thiamin pyrophosphate, the biologically active form of vitamin B_1, consists of two heterocyclic moieties, a pyrimidine and a thiazole ring. Thiamin pyrophosphate is required by all cells for the stabilization of carbanion intermediates and plays an important role in carbohydrate metabolism (*4*). Thiamin biosynthesis is an attractive target because humans do not biosynthesize thiamin but instead obtain the essential vitamin through diet, making the enzymes of this pathway less likely than other critical metabolic pathways to inadvertently harm the patient. The *de novo* thiamin biosynthetic pathway in bacteria biosynthesizes the

two moieties separately and then couples them together in the penultimate step to form thiamin phosphate (*5*). 4-amino-2-methyl-5-(hydroxymethyl)-pyrimidine pyrophosphate (HMP-PP) is formed by two enzymes from 5-aminoimidazole ribotide and 4-methyl-5-(hydroxyethyl)thiazole phosphate (ThzP) is formed by seven enzymes from a sulfur source, glycine (or tyrosine), and deoxy-D-xylulose 5-phosphate (*5*). While this biosynthetic pathway is fairly well studied and understood, recent work has confirmed that some bacteria, such as *Bacillus subtilis*, can salvage thiamin and degraded forms of thiamin from the environment (*6, 7*). Implicated genes for thiamin salvage encode for an ABC transporter and are known as ThiX-ThiY-ThiZ or YkoE-YkoD-YkoC (*8*). Examination of the genome of *M. tuberculosis* shows that this bacterium does not contain any of the genes responsible for thiamin salvage, indicating that *M. tuberculosis* is reliant up on the *de novo* biosynthetic pathway to generate its thiamin (*8*).

Thiamin phosphate synthase (TPS) catalyzes the coupling of HMP-PP with ThzP in a magnesium dependent reaction to produce thiamin phosphate and pyrophosphate (PP_i), shown in Figure 3.1 (*9*). *Bacillus subtilis* TPS has been extensively studied and is mechanistically well understood (*10-13*). BsTPS has been shown to catalyze a dissociative reaction, releasing PP_i and generating a carbocation on the methyl group at C2 on pyrimidine moiety which can then react with Thz to form thiamin phosphate. The rate of carbocation formation has been determined, as have several crystal structures of BsTPS with different intermediates mapping the reaction pathway (*11, 12*). The work presented here describes the crystal structure of MtTPS, the results of a virtual drug screening to identify potential inhibitors, and the effectiveness of these compounds based on a novel *in vivo* assay.

Section 3.2. Materials and Methods

Materials. The NCI Diversity Set compounds were obtained from the NCI.

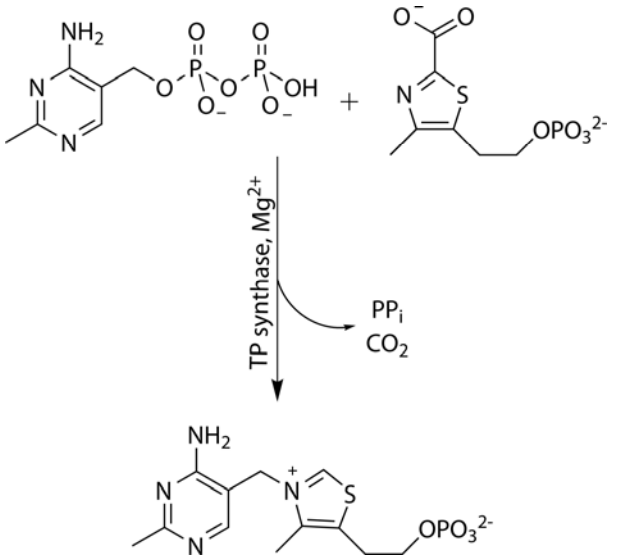


Figure 3.1. Reaction catalyzed by TPS.

Cloning of MtTPS. Standard methods were used for DNA manipulations (*14, 15*). Plasmid DNA was purified with the Qiagen Miniprep kit and DNA fragments were purified from agarose gel with the Nucleospin Purification kit (Macherey-Nagel). *E. coli* strain MachI (Invitrogen) was used as a recipient for transformations during plasmid construction and for plasmid propagation. Phusion DNA polymerase (New England Biolabs) was used for PCR following the manufacturer's recommendations. The *M. tuberculosis thiE* gene was amplified from genomic DNA by PCR with the following primer pair: 5'-TAG CAT ATG CAC GAA TCC CGT CTG GCA TCG G-3' and 5'-CTA CTC GAG TTA GTT CGC TGC TGT AAG CGC C-3'. The PCR product was verified by sequencing and then subcloned into the plasmid pET-28nTEV using the NdeI/XhoI cut sites. The overexpression vector pET-28nTEV is based on the pET-system from Novagen and encodes for kanamycin resistance and an N-terminal 6xHis tag cleavable by TEV protease and is under the control of the T7*lac* promoter. Again, clones were screened by restriction digest. A correct clone was named pMtThiE.28nTEV.

Overexpression and Purification of MtTPS. The plasmid MtThiE.28nTEV was transformed into the *E. coli* B834(DE3) cell line for overexpression. Overnight cultures were grown in 5 mL LB supplemented with 50 mg/L kanamycin at 37°C with shaking and used for the inoculation of 1 L cultures of LB containing 50 mg/L kanamycin. The large cultures were grown at 37°C to an OD600 of 0.7, when the temperature was lowered to 15.5°C. Once the cultures were cooled, overexpression of MtTPS was induced by the addition of 0.5 mM isopropyl β-D-thiogalactopyranoside and allowed to shake overnight. Cells were harvested the next morning by centrifugation at 4°C at 6000×*g* for 15 minutes and the supernatant discarded. The cell pellet resulting from 4 L of culture was stored at -20° until purification.

To purify MtTPS, the cell pellet was thawed and resuspended in purification buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10% (v/v) glycerol, pH 8.0) containing 10 mM imidazole. The cells were then lysed by three cycles of sonication. All steps after lysis were carried out at 4°C. After lysis the cell lysate was clarified by centrifugation for 1 hour at 40000×g and the pelleted debris discarded. The supernatant was twice passed over a 2 mL Ni-NTA column (Qiagen) pre-equilibrated with purification buffer containing 10 mM imidazole. The column was then washed with 60 mL of this buffer, followed by washing with 40 mL purification buffer containing 20 mM imidazole to remove any contaminants binding nonspecifically to the column. MtTPS was eluted from the column using 8 mL of purification buffer spiked with 250 mM imidazole. The sample was judged to be ~70% pure by SDS-PAGE analysis (results not shown). Further purification was carried out using size exclusion chromatography (HiLoad 26/60 Superdex 200 pg, GE Healthcare) to increase purity to ~85% as determined by SDS-PAGE. The storage buffer for MtTPS was 20 mM Tris (pH 8.0), 50 mM NaCl, 3% glycerol. MtTPS was concentrated to ~9 mg/mL as determined by Bradford assay and stored at -80°C until used for

crystallization experiments (*16*). The overall yield of pure protein was low at only 2 mg/L of culture.

Crystallization of MtTPS. Initial crystallization leads were determined using the hanging drop vapor diffusion method at room temperature temperature by mixing an equal volume of reservoir solution with protein sample (1 μL of each) over 450 μL reservoir solution using commercially available sparse matrix screens (Emerald BioSystems). After three months an initial hit was found. Optimized crystal rods grew to 150 μm long and 30 μm thick in 1.6 M NaH_2PO_4, 0.4 M K_2HPO_4, and 100 mM phosphate-citrate buffer (pH 4.4) over three weeks. MtTPS crystals formed in the space group $C222_1$ with a = 84.5 Å, b = 90.9 Å, and c = 124.8 Å. Two molecules are located in the asymmetric unit, corresponding to a Matthew's coefficient of 2.54 and 51% solvent (*17*). Crystals were prepared for data collection by transferring briefly to a cryoprotectant consisting of the crystallization solution supplemented with 16% glycerol, followed by flash freezing through plunging the crystal in liquid nitrogen.

X-Ray Data Collection and Processing. A dataset for MtTPS was collected at the Advanced Photon Source at Argonne National Laboratory on the NE-CAT 24-ID-X beamline. The oscillation method was used, collecting 100° of data with a rotation range of 1° for each frame. The MtTPS crystal was exposed for 1 second per frame and diffraction was recorded using a Quantum315 detector (Area Detector Systems Corp.). The dataset diffracted to 2.32 Å resolution and the data were indexed, integrated, and scaled using the HKL2000 suite of programs (*18*). The data collection statistics are presented in Table 3.1.

Structure Determination, Model Building, and Refinement. The structure of MtTPS was determined using molecular replacement. Thiamin phosphate synthase from *Pyrococcus furiosus* (PDB ID: 1XI3) with all ligands and water molecules removed, was chosen as the search model (*19*). The search model was then prepared

using CHAINSAW in the CCP4 suite of programs to generate a search model as similar as possible to MtTPS (*20, 21*). The rotation and translation search functions were carried out using MOLREP and a solution was found (*22*). The molecular replacement solution was then manually rebuilt in COOT, followed by rounds of refinement in CNS (*23, 24*). In later stages of refinement water molecules were added to the model. The final model contains 427 protein residues, 118 water molecules, and 2 phosphate ions with one residue on each chain, Asp82, adopting a disallowed conformation, although the density is clear (*25*). The final data refinement statistics are presented in Table 3.2. Figures were prepared using ChemDraw and Pymol (*26*).

Table 3.1. Summary of Data Collection Statistics

Source	APS 24-ID-C
Resolution (Å)	2.32
Energy (eV)	12662
Space Group	$C222_1$
a (Å)	84.5
b (Å)	90.9
c (Å)	124.8
Matthew's Coef.	2.5
% solvent	51
Mol/a.s.u.	2
Measured Reflections	60007
Unique reflections	16709 (1119) [a]
Average I/σ	14.7 (4.4)
Redundancy	3.6 (2.8)
Completeness (%)	83.1 (56.5)
Rsym [b] (%)	7.6 (17.4)

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

Virtual Drug Screening. The MtTPS structure was used to screen the National Cancer Institute's (NCI) diversity set of potential inhibitor compounds. The Schrödinger program Maestro was used to manipulate the 2.3 Å structure of MtTPS during screening. The small molecules were then docked *in silico* using the GLIDE program. The docking scores for the top hits from the NCI diversity set are presented

in Table 3.3. The more negative scores indicate a more negative binding free energy between the enzyme and the small molecule.

Table 3.2. Summary of Data Refinement Statistics

Resolution (Å)	50.00 – 2.32
No. of protein atoms	427
No. of ligand atoms	10
No. of water atoms	118
Reflections in working set	18498
Reflections in test set	912
R factor[a] (%)	22.1
R_{free}[b] (%)	25.9
Rmsd from ideal	
Bonds (Å)	0.0058
Angles (º)	1.2
Average B-factor (Å^3)	31.4
Ramachandran Plot	
Most favored (%)	92.6
Additionally allowed (%)	6.0
Generously allowed (%)	0.9
Disallowed (%)	0.6

[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.

In Vivo Thiamin Riboswitch Inhibition Assay. The gene for β-galactosidase was inserted into the ThiC gene of *B. subtilis*, which is under the control of the thiamin riboswitch. The cells are supplied low levels of thiamin for survival. Cells grown in the presence of thiamin and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) are white, but when given a lower amount of thiamin or thiamin antimetabolites the cells turn blue. Successful inhibition of TPS by the molecules identified in the *in silico* screen will turn the cells blue.

Section 3.3. Results

Overall Structure of MtTPS. The final model of MtTPS contains two protomers in the asymmetric unit. Each chain consists of residues 0-148, 154-166, and 170-221. The loop regions at both the N and C termini of α6 are disordered. The

final model also has 115 water molecules two phosphate ions. MtTPS monomer adopts the $(\beta/\alpha)_8$ barrel fold as seen in other TPS structures (Figure 3.2). One residue from the histidine purification tag is visible, His0, and extends away from the core of

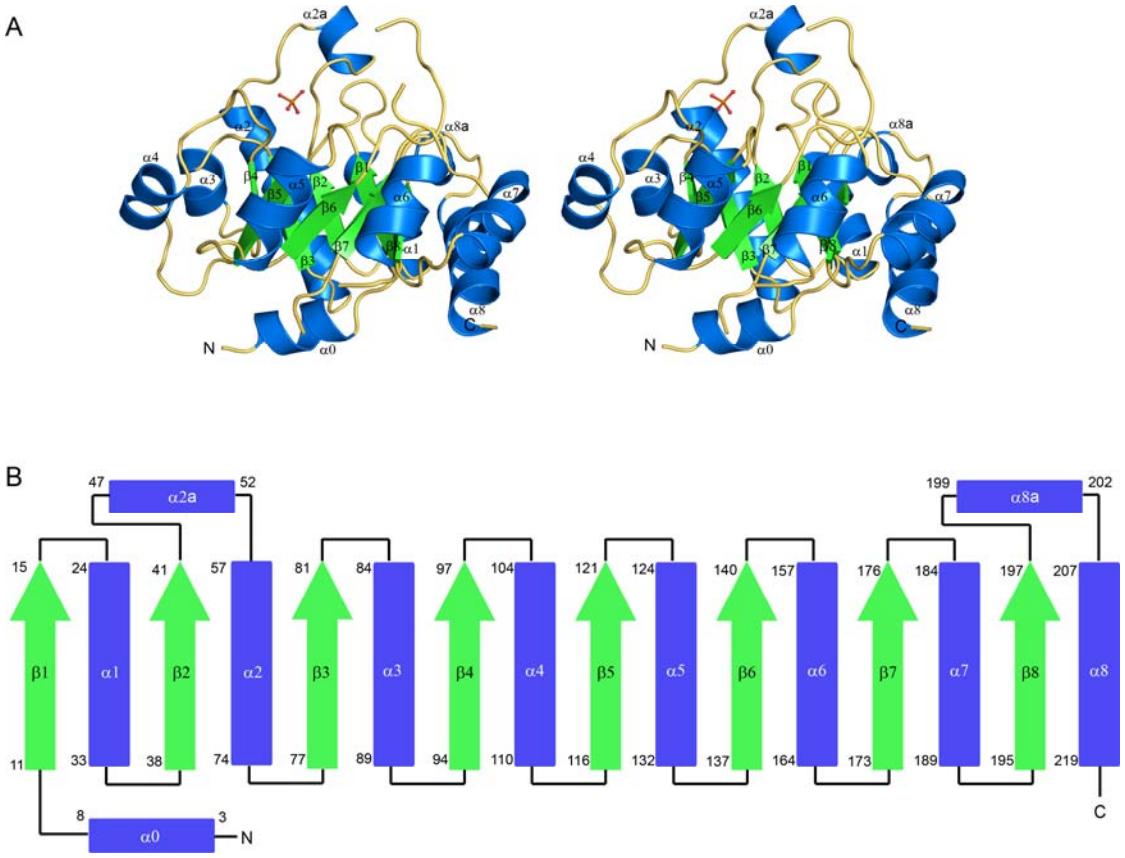


Figure 3.2. Monomeric structure of MtTPS. (A) Stereoview diagram of MtTPS depicted in cartoon representation. The secondary structural elements are labeled α-helices are colored blue, β-strands colored green, and loop regions shown in yellow. The phosphate ion is shown in ball and stick representation and colored according to atom type. (B) Topology diagram of MtTPS using the same color scheme as described for A.

MtTPS into the solvent region. An N-terminal α-helix (α0) blocks the bottom of the β-barrel. MtTPS also contains two additional insertions to the $(\beta/\alpha)_8$ fold. The monomer has an average surface area of 9800 $Å^2$ and is approximately 45 Å wide, 45 Å long, and 35 Å tall. The β-strands of the $(\beta/\alpha)_8$ barrel are rotated 45° in respect to the axis of the β-barrel. The core of the β-barrel has a ring of aromatic residues,

Tyr12, Tyr137, and Phe174 and the N-terminal end of the β-barrel is composed of hydrophobic residues: Leu6, Ala9, Ile38, Val60, Leu77, Val94, and Leu116.

Oligomeric Assembly of MtTPS. MtTPS crystallizes with two chains in the asymmetric unit where the chains are related by a noncrystallographic twofold axis (Figure 3.3A). The C-terminal end of α1, α2, and the following loops from one protomer interacts with α0 from the second protomer, and the α8 helices run antiparallel to each other. The association between the two chains is weak with eleven water molecules found at the interface and no hydrogen bonds or hydrophobic patches

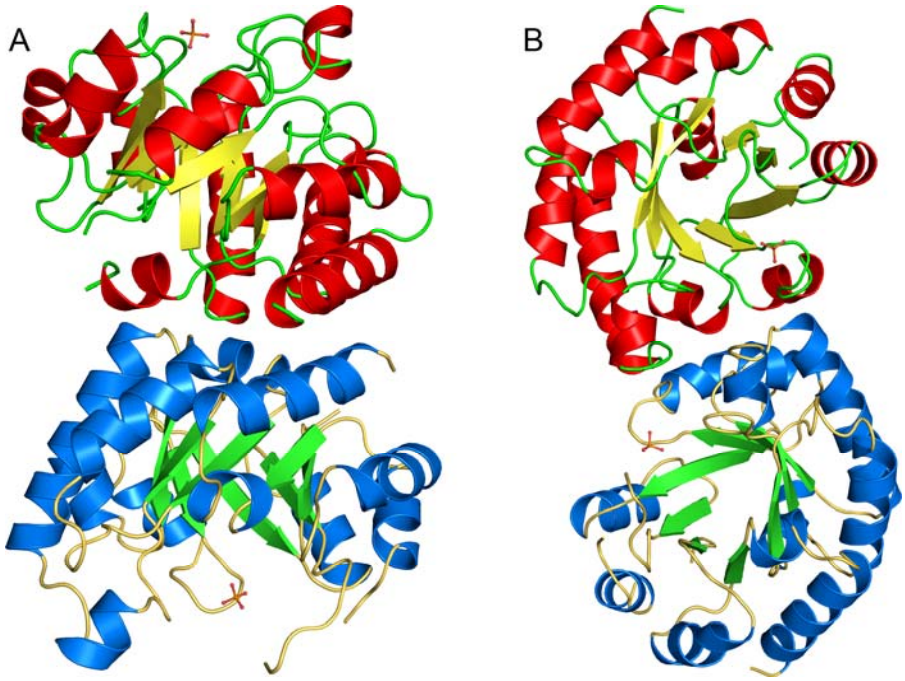


Figure 3.3. Packing Interactions of MtTPS. (A) Packing of monomer A and monomer B in the asymmetric unit. Secondary structural elements of the A chain are colored according to the scheme described in Figure 3.2. (B) Cartoon representation of the dimeric assembly of MtTPS chain A using the symmetry related chain to complete the dimer. The phosphate group is shown in ball and stick representation.

packing together. The symmetry-related molecules, however, show that MtTPS is a dimer with the A chain forming a dimer with a symmetry-related A protomer and the same occurring with the B chain (Figure 3.3B). The $(\beta/\alpha)_8$ barrels of the dimer run roughly parallel to each other so that the C-terminal ends of the barrels lie on the same

face of the dimer. The A_2 dimer buries a total of 2440 Å^2 of surface area at the dimer interface and the B_2 dimer buries 2390 Å^2 total surface area (*27*). Interactions between the protomers occur primarily in loops 2, 3, and 4. Additionally, α3 and α4 pack approximately perpendicularly against α3* and α4* (where * indicates the second chain) and the N-terminus of α2 interacts with loop 2. The short α2a runs antiparallel to α2a*. Two salt bridges are found at the dimer interface between Arg83 and Asp82*, and Asp82 and Arg83*. Two symmetry-related hydrophobic patches are also found at the dimer interface. This area packs Ile86, Leu61, Ala57, and Ala89 together with Leu102*, Val106*, and Ile 110*.

MtTPS Phosphate Binding Site. The putative active site of MtTPS is located at the C-terminal end of the $(\beta/\alpha)_8$ barrel and is marked by the presence of a phosphate ion from the crystallization conditions. In contrast to other residues of the β-barrel, residues at the C-terminal end of the β-barrel many charged amino acid residues occupy the active site. The phosphate ion is bound in a highly solvent accessible position adjacent to loops 2 and 4 and forms interactions with residues from β4, β3, and β5, along with two water molecules (Figure 3.4). The environment around the phosphate ion is positively charged and stabilizes the negative charge of the phosphate group. The phosphate is coordinated directly with His122 and the amide nitrogen atom of the protein backbone at Gln99. The backbone amide nitrogen atom of Arg100 is also oriented toward the phosphate binding pocket, although the distance is too great to interact directly with the phosphate group. Two arginine residues, Arg42 and Arg100, are each pointed toward the phosphate group and one nitrogen atom of the guanidinium side chain from Arg100 forms a water-mediated interaction with the phosphate group. A water molecule is hydrogen bonded to an oxygen atom of the phosphate group and is also positioned near Arg42 and Asn81, although the distances are long, 3.3 Å to Asn81 and 3.4 Å to Arg42.

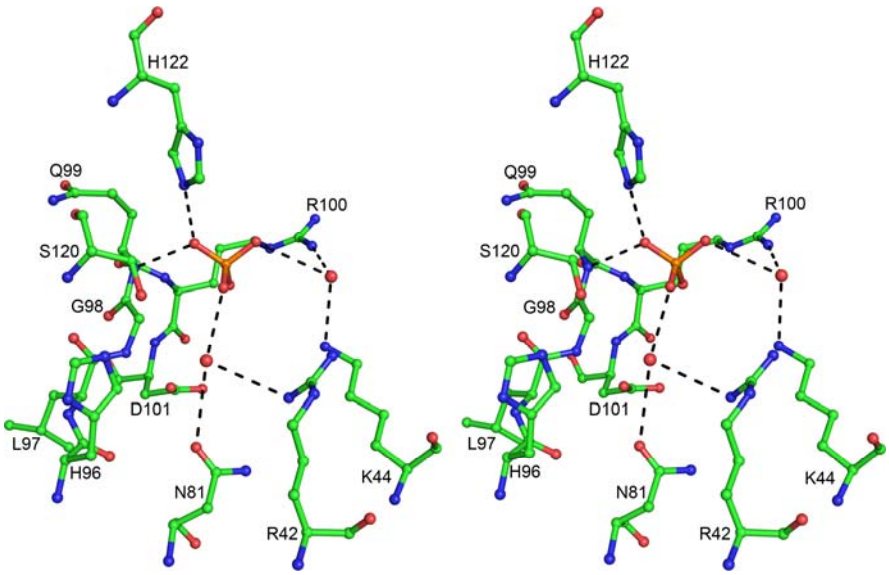


Figure 3.4. Stereoview diagram of MtTPS phosphate binding site shown in ball and stick representation. Atoms are colored depending on type. Nonbonded red spheres represent water molecules. Hydrogen bonds are shown with dashed lines.

Identification of Potential TPS Inhibitor Compounds. The 2.3 Å structure of MtTPS was used for the *in silico* screening of the NCI diversity set of compounds and then simulated docking was performed to determine the theoretical free energy of binding. The top compounds are presented in Table 3.3. The program GLIDE provides different scoring algorithms, and the scoring results from the free binding energy are provided in Table 3.3. Also provided is the amount of relative β-galactosidase activity compared to cells exposed to no inhibitor compounds and the proposed antibiotic type for each of the effective compounds. Figure 3.5 presents the structures for the three most active compounds.

In Vivo Thiamin Riboswitch Assay Activity. B. subtilis cells with the thiC gene interrupted with the gene for β-galactosidase are sensitive to the demands of the cell for increased thiamin production. Treatment of these cells with inhibitors of MtTPS indirectly reduces the concentration of thiamin available to the cell and the cells then turn blue when supplied with X-gal when the riboswitch attempts to upregulate thiamin through thiC and instead upregulates the β-galactosidase gene product. The

Table 3.3. *In silico* NCI Diversity Set screening and *in vivo* screening results

Rank	NCS Title	CAS	Dock Score	β-Gal Activity	Antibiotic Type
1	19803	17912-87-7	-9.4798	++	CF_3HMP type
2	408734		-8.9731	±	
3	60423	6940-93-8	-8.5456	n/a	
4	42014	6308-29-8	-8.4926	±	
5	114436		-8.4356	±	
6	16722	5463-22-9	-8.2503	±	
7	159686	13898-58-3	-8.1733	±	
8	25368		-7.9443	±	
9	30930		-7.6710	±	
10	227383	69746-62-9	-7.6315	±	
11	16736	530-07-4	-7.5299	±	
12	279895		-7.4347	±	
13	78623	64567-62-0	-7.2554	±	
14	93945	40367-32-6	-7.2508	±	
15	60239	6627-85-6	-7.1149	±	
16	107022	568-80-9	-6.9718	+++	CF_3HMP type
17	78999	64679-65-8	-6.9625	±	
18	51683	6943-27-7	-6.9553	±	
19	39938		-6.9049	±	
20	319471		-6.8223	+	CF_3HMP type
21	122819	29767-20-2	-6.6852	±	
22	36586	446-72-0	-6.5941	±	
23	362639	72177-30-1	-6.4667	±	
24	142277		-6.4546	±	
25	111847	894-93-9	-6.3576	(+)	Typical antibiotic type?

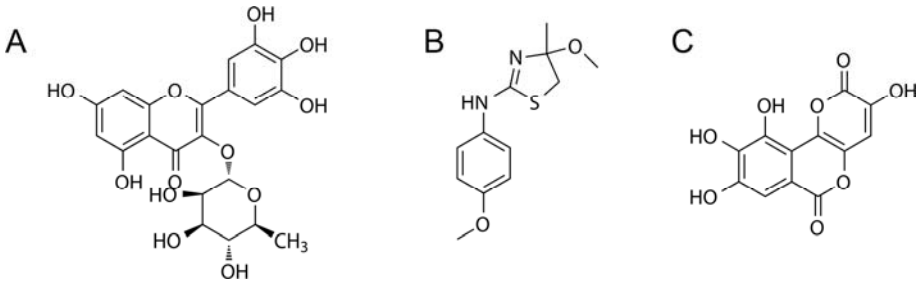


Figure 3.5. Chemical structures of identified compounds showing MtTPS inhibition. (A) Myricitrin (NSC compound 19803). (B) NSC compound 319471. (C) Galloflavin (NSC compound 107022).

top 25 hits from the *in silico* docking study were screened for increased β-galactosidase activity through monitoring the cell color and ranked. These results are presented in Table 3.3. Three compounds, NSC 107022, NSC19803, and NSC 319471, turned the ***B. subtilis*** cells blue and thus showed increased β-galactosidase activity. The chemical structures of these molecules are shown in Figure 3.5.

Section 3.4. Discussion

Comparison of MtTPS to Other TP Synthases. MtTPS adopts the $(\beta/\alpha)_8$ barrel fold seen among the other structurally characterized TPS proteins. The sequence identity between MtTPS and the other TPS proteins is low, ranging from 36% identical for *Pyrococcus furiosis* TPS (PDB ID: 1XI3) to 18% identical when compared to TPS from *Bacteroides thetaiotamicron* (PDB ID: 3CEU). To determine how structurally similar these proteins are, a DALI search was performed using the MtTPS structure and a representative selection of enzymes is shown in Table 3.4 (*28*). As expected several TPS were identified, as well as BsTenI, which has been compared to BsTPS previously (*12, 29*).

MtTPS was then compared specifically to other TPS proteins. The monomeric structures superimpose well, with RMSD values below 2.5 Å (Table 3.4). The differences between BsTPS with thiamin phosphate bound and the unliganded MtTPS

are slight (*10*). The α0 helix at the N-terminal end of the β-barrel is rotated approximately 60° with respect to α0 in BsTPS and completely blocks the entrance to the β-barrel from the N-terminal face of the enzyme. BsTPS has an inserted α-helix between β8 and α8; the same inserted α-helix is present in MtTPS. However, MtTPS has an additional α-helix inserted after β2 and before α2 (Figure 3.2). The helix α2a appears to contribute to the formation of the dimer. The presence of α2a in MtTPS has also forced the loop connecting β1 to α1 to adopt a more open conformation away from the active site. The corresponding region in BsTPS is found closer to the active site and also forms contacts with α8a.

Another difference between BsTPS and MtTPS is the oligomeric assembly of the molecules. BsTPS has previously been identified as a dimer based on the crystal structure, but the assignment was not definite as the interactions were weak and fairly nonspecific. The BsTPS dimer has been described as adopting a *trans* conformation with the active sites on opposite sides of the dimer (*10, 12*). In contrast to BsTPS, BsTenI and PfTPS form a much tighter dimer with hydrophobic patches on the surface of the protein packing together. BsTenI and PfTPS are both *cis* dimers where the $(\beta/\alpha)_8$ barrels run roughly parallel to each other and the active sites lie on the same face of the dimer. The asymmetric unit of MtTPS does not contain either of these dimers (Figure 3.3A), but instead contains two half-dimers where the *cis* dimer is

Table 3.4. Enzymes Identified as Structurally Similar Through DALI

Protein	**PDB ID**	**Z Score**	**RMSD**	**% Identical**	**# aligned residues**
PfTPS	1XI3	28.7	1.7	36	195
BsTPS	1G69	26.6	2.2	28	206
BsTenI	1YAD	22.6	2.0	28	184
Tryptophan synthase	1UJP	18.1	2.8	18	194
Hexulose 6-Phosphate synthase	3F4W	18.1	2.6	17	211
Triosephosphate isomerase	1W0M	17.2	2.7	17	181

formed using each chain's symmetry-related molecule to produce two slightly different copies of the *cis* dimer.

Active Site Comparison. The structure of BsTPS with thiamin phosphate bound (PDB ID: 2TPS) was compared with BsTPS to closely examine the putative active site of MtTPS (*10*). The structural alignment reveals that most of the active site residues are strictly conserved with very little variation (Figure 3.6). The residues surrounding the pyrimidine moiety, Ser120, His122, Tyr137, Tyr12, Gln40, and Asn81 are conserved between the two structures and adopt the same conformations, even in the absence of ligand in MtTPS. The amino acids forming interactions with the thiazole moiety are also mostly conserved, but the conformations seen in MtTPS are slightly different than those seen in BsTPS (Figure 3.6). Two arginine residues of MtTPS form clashing interactions with the phosphate groups of the BsTPS structure. Arg42 is in a slightly more extended conformation than Arg59 of BsTPS and is thus

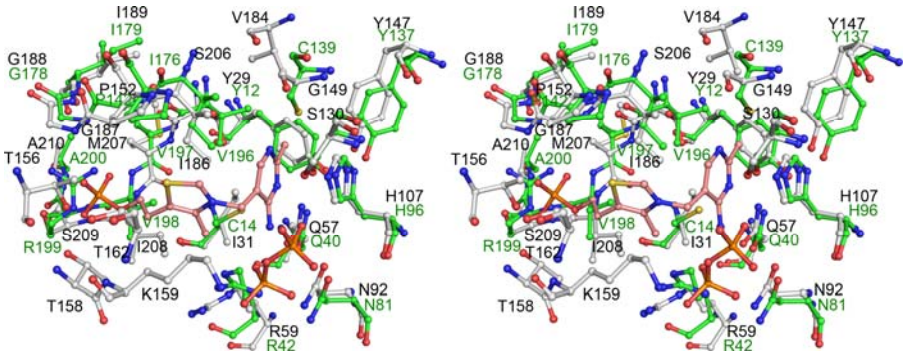


Figure 3.6. Stereoview diagram of MtTPS superimposed with BsTPS with thiamin phosphate bound. The carbon atoms of MtTPS are colored green while the carbon atoms from BsTPS are shown in grey. The carbon atoms of thiamin phosphate are salmon colored. Water molecules have been excluded for clarity.

too close to two oxygen atoms of PP_i. This residue could easily adopt a slightly different orientation, however, and instead of clashing with PP_i, contribute stabilizing interactions. The second arginine residue, Arg199, is not conserved between MtTPS and BsTPS, where a serine residue occupies this position. Reorientation of Arg199 would allow this residue to offer a positively charged residue to favorably interact

with the negatively charged phosphate tail of the thiazole moiety. One striking difference is that the presence of the thiazole moiety in the active site has triggered the loop connecting β6 to α6 to become ordered in the BsTPS structure and Lys159 adopts an extended conformation over the thiazole ring. While this region is disordered in MtTPS, the loop between β6 and α6 is mostly conserved and Lys159 from BsTPS corresponds to Lys149 in MtTPS. Additionally, this lysine residue is absolutely conserved among all TPS and TenI proteins.

An interesting feature of the MtTPS active site when compared to the structure of BsTPS is the presence of two active site cysteine residues, Cys14 and Cys139, seen in Figure 3.6. Cys14 replaces the large-sized, hydrophobic Ile31 in BsTPS. The second cysteine residue superimposes upon Gly149 of BsTPS. The residue at this position is one of the primary means for distinguishing TPS from TenI; TenI proteins have a bulky hydrophobic residue at this position, often a leucine residue. This bulk prevents HMP-PP from fitting in the active site and TenI can instead only bind to thiazole moieties. TPS molecules most commonly have much smaller residues at this position, such as the glycine residue observed in both BsTPS and PfTPS. The smaller side chain allows HMP-PP to bind deeply in the active site without forming any steric clashes between the methyl group at C4 of the pyrimidine with the protein. The superposition shown in Figure 3.6 illustrates that the sulfur of cysteine is very close to HMP-PP (2.2 Å to the methyl carbon at C4). This close proximity may indicate that Cys139 adopts a different conformation upon HMP-PP binding or that HMP-PP binds slightly less deeply within the active site.

Inhibition of MtTPS. The virtual library screening of the relatively small NCI Diversity set of compounds identified the top 50 inhibitors based on the free binding energy and the scoring algorithm used by the software program GLIDE (Table 3.3). To further characterize these compounds, a novel assay capable of indicating thiamin

depletion was developed. The gene for β-galactosidase was inserted into the thiC gene of *B. subtilis*, which is under the direct control of the thiamin riboswitch. As the thiC gene has been interrupted, the cells are now unable to biosynthesize HMP-PP, so a low level of thiamin must be supplied for the cells to survive. When the cells are provided with X-gal and sufficient thiamin, the riboswitch does not upregulate the expression of β-galactosidase and the colonies remain white. However, when the cells are deprived of enough thiamin, or when thiamin antimetabolites are present along with X-gal, the riboswitch upregulates the production of β-galactosidase and the colonies become blue.

The top 25 virtual library hits were assayed by this new technique and three compounds resulted in an upregulation of β-galactosidase (Table 3.3 and Figure 3.5) and therefore were inhibiting the production of thiamin phosphate and ideally causing this inhibition by binding to MtTPS. Two of the small molecules, NSC 107022 and myricitrin, are highly hydroxylated and could potentially form many hydrogen bonds with MtTPS. Myricitrin, shown in Figure 3.5A, also had the lowest free binding energy in the *in silico* study (Table 3.3). The third compound, NSC 319471 (Figure 3.5B), has several structural characteristics in common with thiamin phosphate. Both compounds contain an aromatic ring with a hydrophobic substituent para to the bridge to the second heterocycle. The adjoining rings are each five-member rings containing one nitrogen atom and one sulfur atom, although the positions are not conserved. NSC319471 could be acting as a substrate mimic, as could myricitrin. Structural studies will be conducted to confirm the mode of binding for these compounds and the K_d values will be determined as well.

REFERENCES

1. Cohen, K., and Meintjes, G. (2010) Management of individuals requiring antiretroviral therapy and TB treatment, *Curr. Opin. HIV AIDS 5*, 61-69.
2. Zignol, M., Hosseini, M. S., Wright, A., Weezenbeek, C. L., Nunn, P., Watt, C. J., Williams, B. G., and Dye, C. (2006) Global incidence of multidrug-resistant tuberculosis, *J. Infect. Dis. 194*, 479-485.
3. Zhang, Y., and Yew, W. W. (2009) Mechanisms of drug resistance in Mycobacterium tuberculosis, *Int. J. Tuberc. Lung Dis. 13*, 1320-1330.
4. Jordan, F. (2003) Current mechanistic understanding of thiamin diphosphate-dependent enzymatic reactions, *Nat. Prod. Rep. 20*, 184-201.
5. 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.
6. Jenkins, A. H., Schyns, G., Potot, S., Sun, G., and Begley, T. P. (2007) A new thiamin salvage pathway, *Nat. Chem. Biol. 3*, 492-497.
7. Jenkins, A. L., Zhang, Y., Ealick, S. E., and Begley, T. P. (2008) Mutagenesis studies on TenA: a thiamin salvage enzyme from Bacillus subtilis, *Bioorg. Chem. 36*, 29-32.
8. Rodionov, D. A., Vitreschak, A. G., Mironov, A. A., and Gelfand, M. S. (2002) Comparative genomics of thiamin biosynthesis in procaryotes. New genes and regulatory mechanisms, *J. Biol. Chem. 277*, 48949-48959.
9. Zhang, Y., Taylor, S. V., Chiu, H. J., and Begley, T. P. (1997) Characterization of the Bacillus subtilis thiC operon involved in thiamine biosynthesis, *J. Bacteriol. 179*, 3030-3035.

10. Chiu, H. J., Reddick, J. J., Begley, T. P., and Ealick, S. E. (1999) Crystal structure of thiamin phosphate synthase from Bacillus subtilis at 1.25 A resolution, *Biochemistry 38*, 6460-6470.

11. Hanes, J. W., Ealick, S. E., and Begley, T. P. (2007) Thiamin phosphate synthase: the rate of pyrimidine carbocation formation, *J. Am. Chem. Soc. 129*, 4860-4861.

12. Peapus, D. H., Chiu, H. J., Campobasso, N., Reddick, J. J., Begley, T. P., and Ealick, S. E. (2001) Structural characterization of the enzyme-substrate, enzyme-intermediate, and enzyme-product complexes of thiamin phosphate synthase, *Biochemistry 40*, 10103-10114.

13. Reddick, J. J., Nicewonger, R., and Begley, T. P. (2001) Mechanistic studies on thiamin phosphate synthase: evidence for a dissociative mechanism, *Biochemistry 40*, 10095-10102.

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

15. Sambrook, J., Fritsch, G. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Guide*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

16. 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.

17. Matthews, B. W. (1968) Solvent content of protein crystals, *J. Mol. Biol. 33*, 491-497.

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. Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003) A graphical user interface to the CCP4 program suite, *Acta Cryst. D59*, 1131-1137.

21. Stein, N. (2008) CHAINSAW: a program for mutating pdb files used as templates in molecular replacement, *J. Appl. Cryst. 41*, 641-643.

22. Vagin, A., Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement, *J. Appl. Cryst. 30*, 1022-1025.

23. 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 Cryst. D54*, 905-921.

24. Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010) Features and Development of Coot, *Acta Cryst. D66*, 486-501.

25. 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. Cryst. 26*, 283-291.

26. DeLano, W. L. (2002) The PyMOL Molecular Graphics Systems, DeLano Scientific, San Carlos, CA.

27. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state, *J. Mol. Biol. 372*, 774-797.

28. Holm, L., and Sander, C. (1998) Touring protein fold space with Dali/FSSP, *Nucleic Acids Res. 26*, 316-319.

29. Toms, A. V., Haas, A. L., Park, J. H., Begley, T. P., and Ealick, S. E. (2005) Structural characterization of the regulatory proteins TenA and TenI from Bacillus subtilis and identification of TenA as a thiaminase II, *Biochemistry 44*, 2319-2329.

CHAPTER 4

STRUCTURAL STUDIES OF THIAMIN MONOPHOSPHATE KINASE IN COMPLEX WITH SUBSTRATES AND PRODUCTS[1]

Section 4.1. Introduction

Thiamin, also known as vitamin B_1, is required by all living organisms and is an essential vitamin in the human diet. The active form of thiamin, thiamin pyrophosphate (TPP), stabilizes acyl carbanion intermediates and plays important roles in carbohydrate metabolism and in the pentose phosphate pathway (*1*). Often, the reactions catalyzed by thiamin pyrophosphate involve decarboxylation reactions of α-keto acid. In these reactions, the thiamin ylide adds to the ketone. Loss of carbon dioxide from the resulting adduct gives the thiamin-stabilized acyl carbanion, which can then be protonated by an active site acid. Release of the cofactor generates the aldehyde product. (*2*).

The biosynthesis of thiamin pyrophosphate in prokaryotic systems has been studied both structurally and mechanistically (*1, 3*). The thiazole and pyrimidine moieties are biosynthesized separately. Six enzymes act upon deoxy-D-xylulose 5-phosphate, glycine, and cysteine to form the thiazole moiety. The pyrimidine pyrophosphate is formed by a rearrangement of 5-aminoimidazole ribonucleotide catalyzed by ThiC, followed by phosphorylation by HMP-P kinase. The thiazole and pyrimidine moieties are then joined together to form thiamin monophosphate (TMP) in a reaction catalyzed by thiamin phosphate synthase. Thiamin monophosphate kinase (ThiL) is the final enzyme in the thiamin biosynthetic pathway and is responsible for the phosphorylation of TMP to form TPP (Figure 4.1), the biologically relevant form of the cofactor (*4*).

---

[1] Reproduced with permission from McCulloch, K.M., Kinsland, C., Begley, T.P., and Ealick, S.E. (2008) *Biochemistry* 47: 3810-3821. Copyright 2008 American Chemical Society

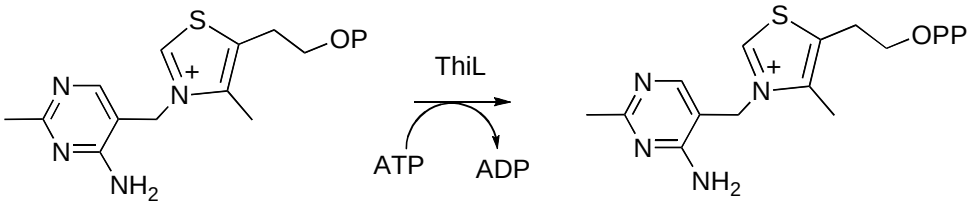


Figure 4.1. Reaction catalyzed by ThiL.

In addition to ThiL′s role in thiamin biosynthesis, ThiL has been postulated to be a member of the PurM ATP binding superfamily (*5*). To date, only five proteins have been identified as members of this superfamily: aminoimidazole ribonucleotide synthetase (PurM), formylglycinamide ribonucleotide amidotransferase (PurL), selenophosphate synthetase (SelD), the NiFe hydrogenase maturation protein HypE, and ThiL (*5*). Interestingly, PurL and PurM catalyze sequential reactions in the purine biosynthetic pathway and the product of the reaction catalyzed by PurM is 5-aminoimidazole ribonucleotide, the small molecule transformed into the pyridine moiety of thiamin pyrophosphate. With the exception of ThiL these enzymes are believed to proceed via a phosphoimidate intermediate (*6*). It is unclear why ThiL belongs to this superfamily because the TMP phosphorylation reaction is more likely to proceed by an inline phosphate transfer in which a phosphorylated enzyme intermediate would be unnecessary.

The structure of *Aquifex aeolicus* ThiL (AaThiL) was previously determined by the New York Structural Genomics Group (PDB ID: 1VQV) but no substrate or product complexes were available. We initiated additional structural studies to investigate the place of ThiL within the PurM ATP binding superfamily. AaThiL was co-crystallized with β,γ-methylene adenosine 5′-triphosphate (AMP-PCP) as a binary complex and these crystals were then soaked with TMP to yield the ternary complex AMP-PCP/TMP. Additionally, AaThiL was co-crystallized with the reaction products, TPP and adenosine 5′-diphosphate (ADP) to form the ADP/TPP complex.

These structures allow for comparison to the structures of the other members of the PurM ATP binding superfamily, especially at the active sites. Examination of the active site of ThiL, with substrates and products bound, has eliminated the possibility of a phosphoimidate intermediate on ThiL and strongly supports a direct, inline transfer of the γ-phosphate of ATP to TMP.

Section 4.2. Materials and Methods

Gene Synthesis. Primers for gene synthesis of *Aquifex aeolicus thiL* were designed using the GeMS website (*7*) (http://software.kosan.com/GeMS/). Gene synthesis was carried out using the method described by Kodumal (*8, 9*). The 5′ synthon was cloned into pSTBlue1 (Novagen) and the 3′ synthon was cloned into pCR4-Blunt (Invitrogen). After sequencing, the two synthons were merged by Single Overlap Extension PCR. The resulting PCR fragment was cloned into pENTR-TEV-D-TOPO (Invitrogen). Sequencing revealed a missing base in the overlap region. This was repaired by site-directed mutagenesis using KOD DNA polymerase (New England Biolabs) and a standard PCR protocol. The following primer and its reverse complement were used: 5′-GTT AAA CGC GCG TGT GAA TTC TAC AAG TGC GAA GTG GTC GG-3′. Parental DNA was digested with *Dpn*I prior to transformation. Clones were screened for the introduction of an *Eco*RI site and verified by sequencing. A clone with the correct sequence was named pAaThiL.ET and used as the entry vector in an LR reaction using the Gateway system (Invitrogen) with a destination vector based on the Novagen pET system. The destination vector encodes an N-terminal 6xHis tag and the pENTR-TEV-D-TOPO encodes for a TEV protease site between the fusion tag and the AaThiL gene. A correct clone was named pAaThiL.XF1 and used for protein production.

Protein Overexpression. pAaThiL.XF1 was transformed into the B*R2 *Escherichia coli* cell line. Cells were grown in LB media at 37 ºC with shaking (200

rpm) to an $O.D._{600}$ of 0.6. Overexpression of ThiL was induced by 1 mM isopropyl-β-D-thiogalactopyranoside and cells were incubated at 37 ºC for three hours. Cells were harvested by centrifugation, resuspended in lysis buffer (300 mM NaCl, 50 mM Na_2HP0_4, and 5 mM imidazole, pH 8.0) and lysed by sonication. The crude lysate was then centrifuged at 40000 *g* for 25 minutes at 4 ºC. The cleared lysate was applied to a Ni-NTA column (Qiagen) and washed with the above buffer for 75 X the column volume. The column was then washed with wash buffer (300 mM NaCl, 50 mM Na_2HP0_4, and 30 mM imidazole, pH 8.0) to remove any nonspecific binding proteins for roughly 20 column volumes. AaThiL was then eluted from the column using elution buffer (300 mM NaCl, 50 mM Na_2HP0_4, and 250 mM imidazole, pH 8.0) in five column volumes. Protein was buffer exchanged into 10 mM Tris and 20 mM NaCl, pH 7.7 by overnight dialysis and was then concentrated using an Amicon Ultra centrifugal filter with a molecular weight cutoff of 10 kDa until a concentration of 13 mg/mL was reached as determined by the Bradford assay (*10*). Protein purity was verified by SDS-PAGE to be 95% pure (results not shown).

Protein Crystallization, Data Collection, and Structure Determination. Crystallization experiments were conducted using the hanging drop vapor diffusion method at 22 ºC by combining equal volumes of protein and reservoir solution. Initial conditions were found using sparse matrix screens Crystal Screen 1 and 2 (Hampton Research) and Wizard Screen 1 and 2 (Emerald Biosystems). To obtain binary complexes, AaThiL was incubated with 4 mM AMP-PCP or ATP and 4 mM $MgCl_2$ on ice for one hour prior to crystallization experiments. Optimized conditions for the AaThiL binary complexes were 100 mM HEPES pH 7.4, 6-11% isopropyl alcohol, 200 mM $(NH_4)_2SO_4$, and 4-6 % 2-methyl-2,4-pentanediol (MPD). 100 mM $CaCl_2$ was used as an additive in a 1:9 ratio with the reservoir solution described above. Crystals reached their maximum size of 200 x 100 x 100 μm^3 within three days.

Cryoprotection of these crystals was done by increasing the MPD concentration to 15% and adding 2% ethylene glycol to the crystallization conditions. To form a ternary substrate complex, TMP was soaked into the cocrystallized AaThiL and AMP-PCP crystals by transferring the crystals to a solution containing the mother liquor and 20 mM TMP and allowing the crystals to soak for 5 minutes before freezing. To obtain AaThiL ADP/TPP crystals, AaThiL was incubated on ice with 4 mM ADP, 4 mM TPP and 4 mM $MgCl_2$. Optimized conditions for the product complex were 100 mM imidazole pH 8.0, 7-10% PEG 8000, and 100 to 250 mM $Ca(C_2H_3O_2)_2$. These crystals reached their maximum size of 200 x 80 x 80 μm^3 within one week and were cryoprotected using 15% ethylene glycol in addition to the crystallization conditions.

Datasets were collected for AaThiL with AMP-PCP, AMP-PCP/TMP, and ADP/TPP at the Advanced Photon Source 24-ID-C beamline using a Quantum315 detector (Area Detector Systems Corp) at the wavelength 0.9795 Å. The dataset for AaThiL complexed with ATP was collected at the A1 station at the Cornell High Energy Synchrotron Source (CHESS) using Quantum210 detector (Area Detector Systems Corp) and a wavelength of 0.9771 Å. All datasets were collected using an oscillation range of 0.5° to resolve the diffraction patterns. Crystals were indexed in the $P2_12_12_1$ space group with cell dimensions a = 61.2 Å, b = 67.1 Å, and c = 203.5 Å for the binary complexes, giving a Matthews coefficient of 2.82 and a solvent content of 56%. The binary structure with AMP-PCP bound was determined to 2.20 Å resolution and the structure with ATP bound was determined to 1.98 Å resolution. Cell dimensions were a = 60.8 Å, b = 66.8 Å, and c = 197.0 Å for the ternary complexes containing either TMP or TPP with a smaller Matthews coefficient of 2.67 and 54% solvent (*11*). The substrate analogue ternary complex, AMP-PCP/TMP, was solved to 2.60 Å resolution and the product ternary complex ADP/TPP was

determined to 1.48 Å resolution. The data collection and refinement statistics are given in Tables 4.1 and 4.2.

Table 4.1. Summary of Data Collection Statistics

	AaThiL ATP	AaThiL AMP-PCP	AaThiL AMP-PCP/TMP	AaThiL ADP/TPP
Beamline	APS 24 ID-C	APS 24 ID-C	APS 24 ID-C	APS 24 ID-C
Resolution (Å)	2.30	2.20	2.60	1.48
Wavelength (Å)	0.9795	0.9795	0.9795	0.9795
Space Group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
a (Å)	61.3	61.2	60.8	60.6
b (Å)	67.1	67.1	66.8	66.2
c (Å)	200.7	203.5	196.9	197.0
Reflections	226383	309648	71312	412005
Unique reflections	34823 (1980)	43410 (4269)	23068 (2062)	118539 (11700)
Average I/σ	31.5 (1.9)	31.6 (6.0)	23.9 (6.1)	20.3 (3.4)
Redundancy	6.5 (3.6)	7.1 (6.6)	3.2 (3.2)	3.6 (2.5)
Completeness (%)	92.4 (53.4)	99.8 (99.9)	90.0 (82.6)	89.1 (88.9)
Rsym (%)	5.2 (36.5)	5.1 (32.5)	3.8 (14.3)	6.6 (28.6)

$R_{sym} = \Sigma\Sigma_i|I_i - \langle I\rangle|/\Sigma\langle I\rangle$, where $\langle I\rangle$ is the mean intensity of the N reflections with intensities I_i and common indices *h,k,l*.

Data were indexed, integrated, and scaled using the HKL2000 suite of programs (*12*). Structures were solved using molecular replacement with thiamin monophosphate kinase structure (PDB ID: 1VQV) as the search model (*13*). Ligands and missing residues were manually added using COOT (*14*). Structures were then refined using rigid body, simulated annealing, B factor refinement and minimization with CNS (*15*). Figures were generated using ChemDraw and PYMOL (*16*).

Section 4.3. Results

Structure of the AaThiL Protomer. The ternary complex models of AaThiL contain all 306 possible residues, as well as two to six residues of the His tag. However, the binary complexes lacking TMP or TPP have no density for the final six to eleven C-terminal residues and they are not included in the model. The AaThiL protomer is composed of two domains, A and B, as seen in Figure 4.2A. The A

Table 4.2. Summary of Data Refinement Statistics

	ThiL ATP	ThiL AMP-PCP	ThiL AMP-PCP/TMP	ThiL ADP/TPP
Resolution (Å)	50.00 – 1.98	50.00 – 2.20	50.00 – 2.60	50.00 – 1.48
# of protein atoms	4778	4750	4881	4906
# of ligand atoms	62	62	116	116
# of water atoms	101	90	14	547
Reflections in working set	32928	38005	19849	102350
Reflections in test set	2053	4312	2260	11388
R factor[a] (%)	22.21	21.7	21.3	20.1
Rfree[b] (%)	24.84	25.7	27.1	21.9
Rmsd from ideals				
Bonds (Å)	0.006	0.006	0.007	0.004
Angles (º)	1.3	1.3	1.4	1.2
Avg B factor (Å^2)	50.9	45.5	56.1	23.9
Ramachandran Plot				
Most favored (%)	91.2	92.2	87.8	93.7
Additionally allowed (%)	8.4	7.4	11.2	5.6
Generously allowed (%)	0.4	0.4	0.4	0.7
Disallowed (%)	0.0	0.0	0.6	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 (10%) excluded from all stages of refinement.

domain consists of the first 140 residues and adopts an α/β fold. The mixed β-sheet has a strand order of β1↑β2↓β5↑β3↓β4↓, shown in Figure 4.2B. The β-strands are unusually long, with β2 and β3 each having eleven residues, β5 ten residues, and β9 nine residues. The fifth strand, β1, is shorter with only five residues. The β-sheet is flanked on one side by four α-helices. Two of the α-helices, α3 and α4, are also longer than average, each with eighteen residues. The B domain, consisting of the C-terminal 160 residues, has an α/β fold as well. The mixed β-sheet has a strand order of β7↓β9↑β6↓β10↑β8↓β11↓. The β-strands are much shorter than the β-strands found in the A domain. The longest β-strand, β10, has eight residues, and the two shortest strands, β8 and β11, each have four residues. β10 has a sharp twist, seen in Figure 4.2A, that results in a bend in the β-sheet. The β-sheet is flanked on both sides

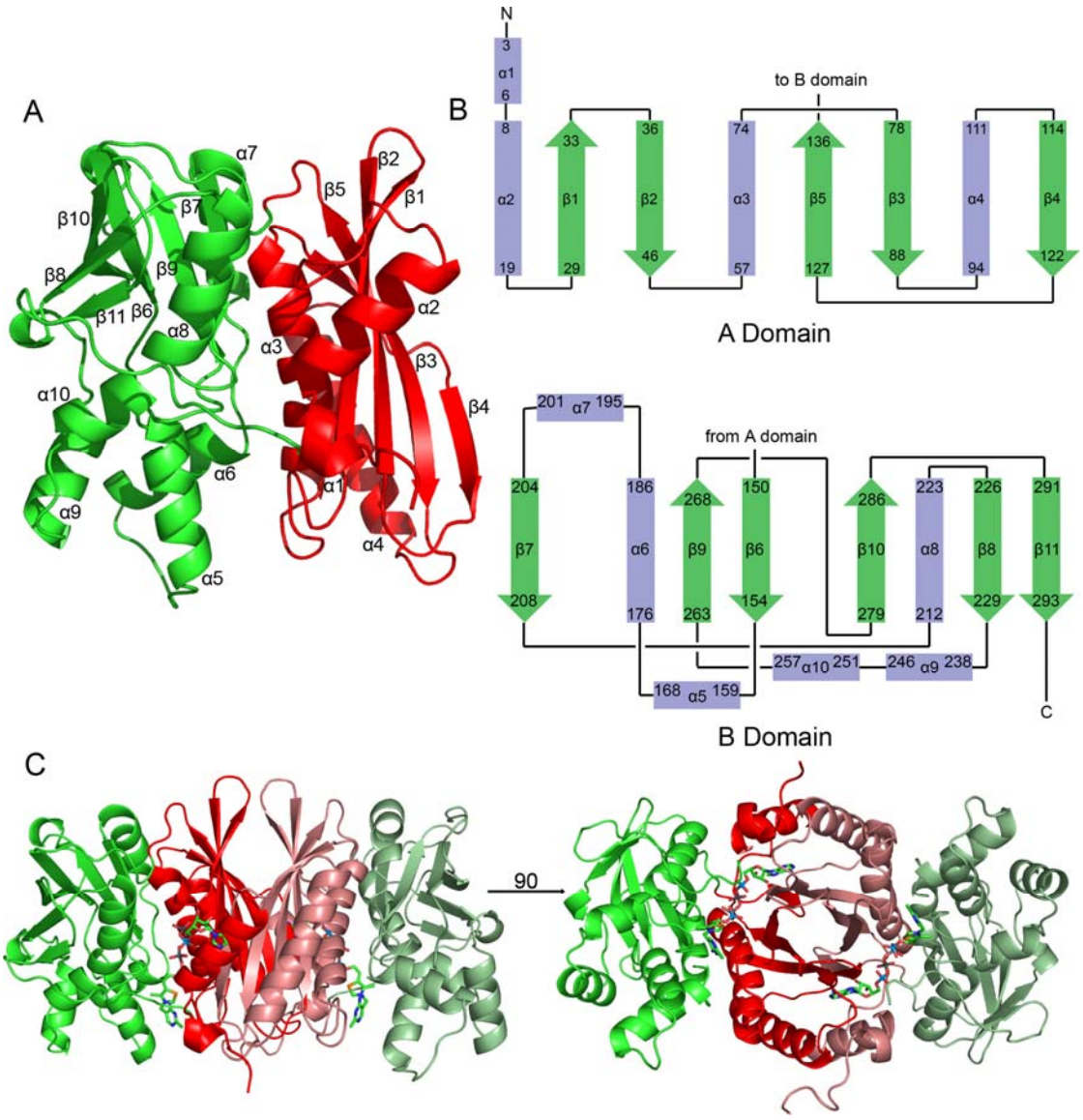


Figure 4.2. Overall structure of AaThiL. (A) Ribbon diagram of the monomer of AaThiL. The A domain is shown in red and the B domain is shown in green. All figures were generated using PyMol (*16*). (B) Topology diagram of AaThiL. The A domain and the B domain are shown separately. (C) Ribbon diagram of the dimer of AaThiL. The A domain is colored in red and the B domain has been colored green in one monomer, and to emphasize domain interactions, the A and B domains of the second monomer are shown in muted tones. The products, TPP and ADP, are shown as stick representations. A 90° rotation provides both a side view of the dimer as well as the view down the β-barrel.

by α-helices. Five α-helices, α5, α6, α8, α9, and α10, are found on one side of the β-

sheet, while α7 is the only α-helix on the other side of the β-sheet. The A domain and B domain are linked by a short connective loop between β5 and β6.

Structure of the AaThiL Dimer. The unit cell contains one dimer per asymmetric unit (Figure 4.2C). The surface area of the dimer is approximately 25500 $Å^2$, with 6300 $Å^2$ buried at the interface between protomers (*17*). The primary interaction between protomers occurs in the A domain in which the twofold related β-sheets come together to form an eight stranded β-barrel. The short β1 strand stacks on top of β4 to form one long strand that is of a length comparable to the other β-strands in the β-barrel. This β-barrel is flanked by α-helices from the A domains. α2 of one protomer and α4 of the second protomer run parallel to each other and hydrophobic residues along each α-helix interact with the opposite α-helix. A noncrystallographic twofold axis runs through the β-barrel. The interface between protomers is stabilized primarily through hydrophobic interactions and the packing of β-strands against each other. A disulfide bond forms between Cys34 and Cys34*, where ″*″ indicates the twofold related protomer. Ten hydrogen bonds form between protomers. The carboxylate group of Glu20 forms a hydrogen bond to the amine of Lys105*, the amide group Asn46 is hydrogen bonded to both the hydroxyl group of Ser121 and the oxygen atom of the amide group of the side chain of Asn87, and the carboxylate of Asp91 is hydrogen bonded to the amine of Lys4. The hydroxyl group of Tyr33 is hydrogen bonded to the carboxylate group of Glu135, and the amine of Lys125 forms a hydrogen bond to the carboxylate of Glu124. In addition to these hydrogen bonds, four hydrogen bonds are formed between the backbone atoms of the two protomers. The carbonyl oxygen atoms of Met1, Asp91, Leu92, and Glu90 form hydrogen bonds to the nitrogen atoms of the amide bonds in Val94, Arg2, Leu3, and Lys4, respectively. The ends of the β-strands are composed of mostly hydrophilic residues that are solvent exposed. Additionally, several highly conserved aspartate residues

that contribute to the formation of the active site are found within the β-barrel facing away from the β-barrel core.

ATP Binding Site. The AaThiL dimer has two equivalent active sites, located mostly within a single protomer but with some interactions from the opposite protomer. The active site is located in a cleft between the A and B domains and also involves β-strands β3 and β4 of the A domain of the twofold related protomer. The interactions between the residues of the active site and ATP are shown in Figure 4.3A and Figure 4.4. ATP binds deeply within the active site cleft and has no exposure to the solvent. The adenine base of ATP binds in a highly hydrophobic pocket, rich with highly conserved isoleucine and valine residues, formed on one side by the β4 strand

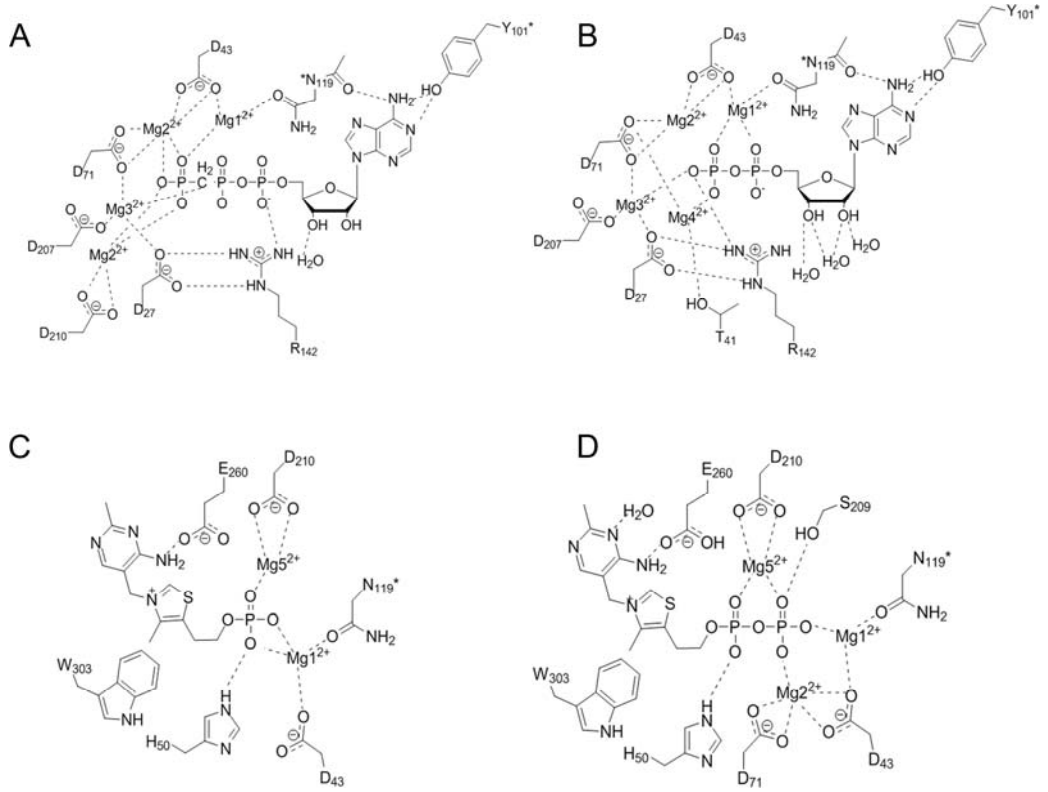


Figure 4.3. Schematic diagrams for the active sites of the substrate and product complexes. Interactions between ligand and residues, ions, and water molecules are shown with dashed lines. (A) Binding of AMP-PCP within the active site of AaThiL. (B) Binding of ADP within the active site of AaThiL. (C) Binding of TMP within the active site of AaThiL. (D) Binding of TPP within the active site of AaThiL.

from the opposing protomer. The hydroxyl group of Tyr101* forms hydrogen bonds to N6 and N1 of the adenine base. The amino group of the adenine base is also hydrogen bonded to the backbone carbonyl group of Asn119*. The hydroxyl groups of the ribose sugar form hydrogen bonds to three ordered water molecules that are involved in water-mediated hydrogen bonding to the protein backbone at Gly25, Asp26, Asp27, Thr28, and Ala29; these residues form a turn between strands β1 and β2. The phosphate tail in both binary complex structures is precisely positioned by the presence of three magnesium ions. These magnesium ions are coordinated to several strictly conserved aspartate residues, Asp27, Asp43, Asp71, and Asp207. Mg1 is coordinated to both the α-phosphate and β-phosphate groups of AMP-PCP or ATP and is also coordinated by the carboxylate group of Asp43, the carbonyl groups of the backbone at Thr42, Asn119*, and two water molecules. Mg2 coordinates the β-phosphate and γ-phosphate groups and is positioned by Asp43, Asp71, and one water molecule. The third magnesium ion is coordinated to the same phosphate groups as Mg2 and is held in position by Asp27, Asp71, Asp207, and one water molecule. The β-phosphate group is also hydrogen bonded to Arg142, the only positively charged residue located within the active site. In the AMP-PCP/TMP complex, Mg1 moves 2.9 Å to coordinate to both ligands while maintaining its coordination to the protein. The other magnesium ions remain coordinated to the same residues and phosphate groups as observed in the binary complexes. In addition to the three magnesium ions seen in the binary complexes, a fourth magnesium ion is found coordinated to the oxygen atom of the β-phosphate group, which is not coordinated to any of the other magnesium ions. Mg4 is also coordinated to Thr41, Asp71, the carbonyl group of Asp27, and three water molecules. A fifth magnesium ion, Mg5, is found coordinated to the α-phosphate and β-phosphate groups of TPP, as well as three water molecules and Asp210.

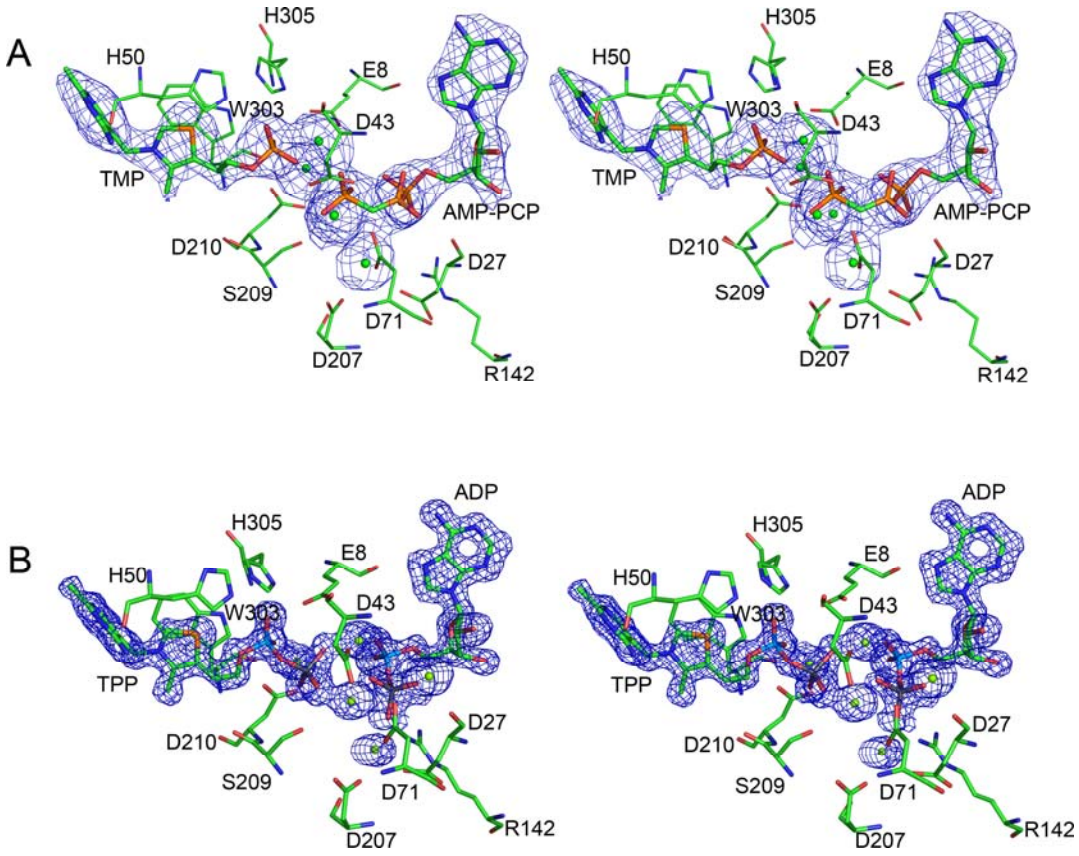


Figure 4.4. Stereoview diagram of the active site of the AaThiL ternary complexes. (A) Stereoview of the active site in the AMP-PCP and TMP structure. The $1F_O$-F_C map is contoured at 2.5 σ and is shown in blue. (B) Stereoview of the active site in the ADP and TPP structure. The magnesium ions, shown as green spheres, have moved to adjust to the shift of the transferred phosphate group. The $1F_O$-F_C map is again contoured at 2.5 σ and shown in blue.

The conversion of ATP to ADP results in movement of the magnesium ions. Mg1 remains coordinated to the α-phosphate and β-phosphate groups but loses its coordination to Thr42 and is instead bound by only the two water molecules, Asp43, and Asn119*. Mg2 coordinates to the same oxygen atom as Mg1 as well as Asp71, Asp43, and two water molecules. Asp27, Asp71, Asp207, a water molecule and the β-phosphate group of ADP are coordinated to Mg3.

Thiamin Binding Site. ATP binds deeply within the active site and in both binary complexes the C-terminal end (six to ten residues) is disordered. Binding of TMP leads to ordering of the C-terminus. In the structure with AMP-PCP, the

electron density of the A chain can be seen through Pro300 and for the B chain through Lys297. Ordering of the C-terminus in the ternary complex results in a lid forming over the active site, shielding the substrates from the solvent. Additionally, the ordering of the C-terminus causes the c axis of the unit cell to shorten by approximately 6 Å (203 Å to 197 Å).

Several residues of the lid interact directly with TMP. Trp303 stabilizes the thiazole ring through π stacking and this residue, while not strictly conserved, is always an aromatic residue, as seen in the sequence alignment (Figure 4.5). Strictly conserved His305 is coordinated to an oxygen atom of the α-phosphate group of TMP, one of the few interactions between a positively charged residue and the phosphate tail. The pyrimidine portion of TMP is partially exposed to the solvent and with few interactions between the pyrimidine ring and the active site residues. The amino group of the pyrimidine ring forms a hydrogen bond to Glu260, which is conserved among ThiL's. N3 of the pyrimidine ring is hydrogen bonded to an ordered water molecule, which provides water-mediated hydrogen bonding to the backbone carbonyl oxygen atom of His50 and the hydroxyl group of Tyr55. The phosphate group of TMP is positioned by a magnesium ion. Mg1 is coordinated to the phosphate group as well as two phosphate groups of AMP-PCP, Asp43, and Asn119*.

Conversion of TMP to TPP results in some movement of the magnesium ions. These interactions are shown in Figures 4.3B and 4.4. The γ-phosphate group of ATP moves approximately 2.3 Å to form the products ADP and TPP. Three magnesium ions, Mg1, Mg2 and Mg5, and Ser209 coordinate the β-phosphate group of TPP. Mg1 is also coordinated to Asn119* and Asp43, Mg2 is positioned between the β-phosphate group of TPP and the β-phosphate group of ADP, and Mg5 interacts with both the α-phosphate and β-phosphate groups of TPP. Attempts to prepare a binary

complex with TMP were unsuccessful suggesting that ATP binding precedes TMP binding.

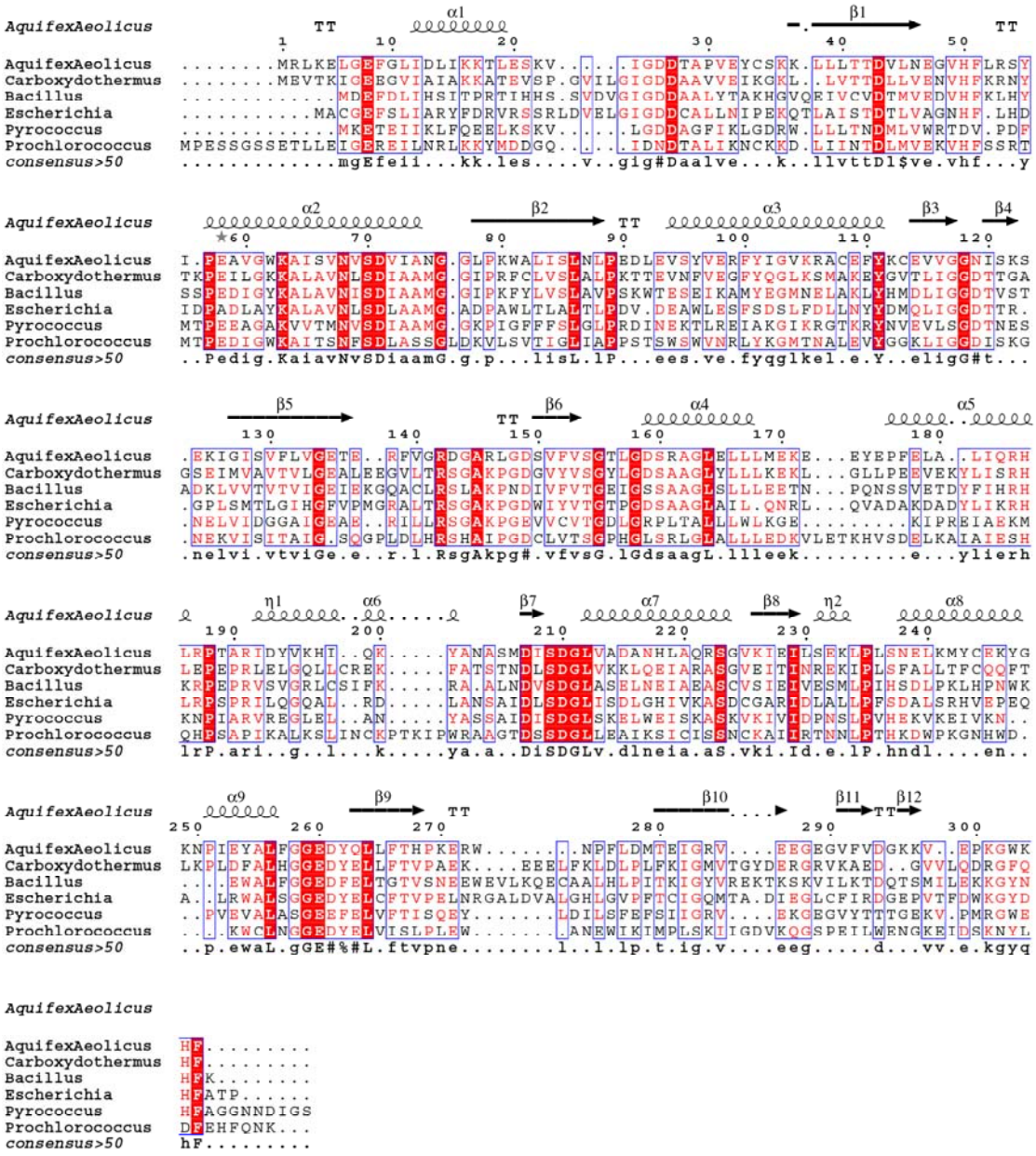


Figure 4.5. Sequence alignment of randomly selected ThiL proteins. Strictly conserved residues are blocked in red, while less conserved residues are in red font. Secondary structure from AaThiL is shown above the alignment. Alignments were performed using ClustalW (*34*) and ESPript (*35*).

Section 4.4. Discussion

Conformational Changes in ThiL upon Substrate Binding. The lack of substrate or product bound to ThiL in the deposited structure (PDB ID: 1VQV) presented several questions about the enzyme's mechanism of action. Our results showed that upon binding ATP or AMP-PCP in the binary complexes, several disordered regions in the unliganded structure become ordered, although the C terminus remains disordered. Residues 7-9, 25, 32, 33, 289, and 301-306 from the A chain and residues 25, 33-36, and 301-306 from the B chain are disordered in unliganded AaThiL. Gly7, Glu8, and Phe9 make up a loop between α1 and α2. Binding of ATP appears to order this loop and the side chain of Glu8 points into the active site. The loop near the ATP molecule between α2 and β1, composed of residues 20-28, lacks density for Gly25 and the conformation of this loop changes upon binding of ATP. The unliganded AaThiL structure includes two phosphate ions bound to each protomer; however, these phosphate groups are located on the surface of the protein and do not correspond to phosphate group locations in the AaThiL active site.

Binding of TMP results in some additional changes in the structure of AaThiL. The rmsd is 0.6 Å between the AMP-PCP and AMP-PCP/TMP structures. The most noticeable and important structural change is the ordering of the C-terminal residues upon the binding of TMP. The last five residues are mostly conserved among ThiL's, as seen in Figure 4.5. Trp303 is consistently an aromatic residue and stabilizes the bound thiamin group by providing π stacking with the thiazole ring. His305 is strongly conserved, as is Phe306. His305 directly interacts with the α-phosphate group of TMP, providing stabilization. Phe306 fits neatly into a pocket of the hydrophobic residues Leu3, Leu11, Leu88, Val97, and Ile120. This pocket shows no movement or conformational changes of residues resulting from TMP binding.

Within the active site, two additional magnesium ions, Mg4 and Mg5, are bound in the ternary substrate complex. These magnesium ions play a key role in positioning the phosphate group of TMP. The highly conserved aspartate residues, Asp27, Asp43, Asp71, and Asp207, retain the side chain conformations seen in the binary complexes upon the binding of TMP.

Little change occurs with the conversion of substrates to products, with an rmsd between these two structures of 0.4 Å. The γ-phosphate group moves 2.4 Å toward TMP to become the β-phosphate group of TPP, and in turn Mg1 moves 2.5 Å to maintain its coordination to this phosphate group and help stabilize the transition state. The β-phosphate and α-phosphate groups of ADP and ATP do not move and the α-phosphate of TMP and TPP also remains in place. The side chains of all active site residues, including the residues of the C-terminal tail, are in the same conformation in the substrate complex and the product complex.

Comparison of ThiL to Other Proteins. The structure determination of *E. coli* PurM (EcPurM) revealed that this enzyme adopted a novel fold that binds ATP (*5*). BLAST searching using this enzyme identified four other enzymes, PurL, SelD, HypE, and ThiL as likely members of this ATP binding superfamily. These five enzymes each contain a short signature amino acid sequence, Dx_4GA/GxP, which characterizes members of the superfamily. Thus far, no other sequence motif has been identified for this superfamily.

Structures have been determined for every member of the superfamily except SelD. There are two forms of PurL, designated large and small. The structure of *Salmonella typhimurium* PurL (StPurL) (*6*) represents the large PurL's, which are found in eukaryotes and Gram negative bacteria. StPurL is 140 kDa in mass and is a multifunctional enzyme in which the formylglycinamide ribonucleotide synthetase domain shows twofold pseudosymmetry and adopts the fold of a PurM dimer. Small

PurL's are found in Gram positive bacteria and are represented by the structure of *Thermatoga maritima* PurL (TmPurL) (*18*). TmPurL is 80 kDa in mass and corresponds to the formylglycinamide ribonucleotide synthetase domain of large PurL. TmPurL and other small PurL's require two additional enzymes, PurS and PurQ, to be active. Recently the structure of *Thermococcus kodakaraensis* KOD1 HypE (TkHypE), involved in the maturation of NiFe hydrogenase, was determined (*19*). Like ThiL and PurM, HypE is a homodimer in the crystal structure.

A DALI search was performed using AaThiL as the search structure, and the results are summarized in Table 4.3 (*20*). The most structurally similar proteins to ThiL are as expected the other members of the PurM ATP binding superfamily. The top DALI score corresponded to HypE. HypE requires carbamoylation at a conserved cysteine residue at the C-terminus; however, this modification was not present in the crystal structure (*19*). An ATP-dependent dehydration transfers the resulting cyanide ligand to a HypCD complex. Although TkHypE forms a dimer crystallographically, the biologically active form is uncertain because *E. coli* HypE has also been reported as a monomer (*21*).

Table 4.3. Summary of DALI search results.

Protein	PDB ID	Z Score	RMSD	# aligned residues
ThiL	1yaw	42.2	1.1	287
TkHypE	2z1f	28.2	2.6	263
TmPurL (AMPPCP/FGAR)	2hs4	19.6	3.1	249
EcPurM	1cli	19.6	3.1	242
StPurL	1t3t	14.2	3.0	240
TmPurL	1vk3	13.3	3.5	210
YjgF	1qu9	8.4	2.4	94
YjgH	1pf5	6.6	2.9	91
YjgF	2ewc	5.4	3.2	92

AaThiL and TkHypE were structurally aligned using the DaliLite Pairwise comparison (*22*). The Z score for these two structures was 28.2 and the rmsd was 2.6

Å, reflecting the similarity of these two structures (Figure 4.6A). No density for ATP is observed in the HypE structure; however, the C-terminal tail shows different conformations depending on whether or not ATP was added during crystallization. Inclusion of ATP in the crystallization conditions resulted in an inward conformation where the C-terminal tail folded back into the active site, suggestive that after carbamoylation at Cys338 the C-terminal end could enter the active site for dehydration. The position of Cys338 corresponds to the thiazole ring in the AaThiL structures. For both AaThiL and TkHypE the C-terminal tail appears to form a lid over the active site when both substrates are present.

The superpositions of EcPurM and TmPurL upon AaThiL are shown in Figures 4.6B and 4.6C, respectively. In both superpositions, the A domains are more structurally conserved than the B domains. The rmsd for the entire EcPurM structure when compared to AaThiL is 3.2 Å and drops to 2.1 Å when only the A domains are compared. The TmPurL structure has an overall rmsd of 3.2 Å, and the rmsd of the A domain is 2.0 Å. Comparisons of the B domains yield rmsd values similar to that of the overall structure, 3.1 Å for EcPurM and 3.0 Å for TmPurL. The higher degree of similarity in the A domains for the members of the PurM ATP binding superfamily is not surprising as the A domains are responsible for forming the dimer interface and the binding of ATP occurs almost entirely within the A domain. The B domain interacts with the second substrate and is expected to show greater variation because each enzyme has evolved to catalyze a different chemical reaction.

Despite this strong structural similarity, the sequence identity is low between AaThiL and other PurM superfamily members, ranging from 20% identity with TkHypE to 13% identity with StPurL. The only conserved sequence based on multiple sequence alignments is the signature sequence motif of Dx_4GA/GxP. To identify any other conserved sequence motifs, a structural alignment was carried out

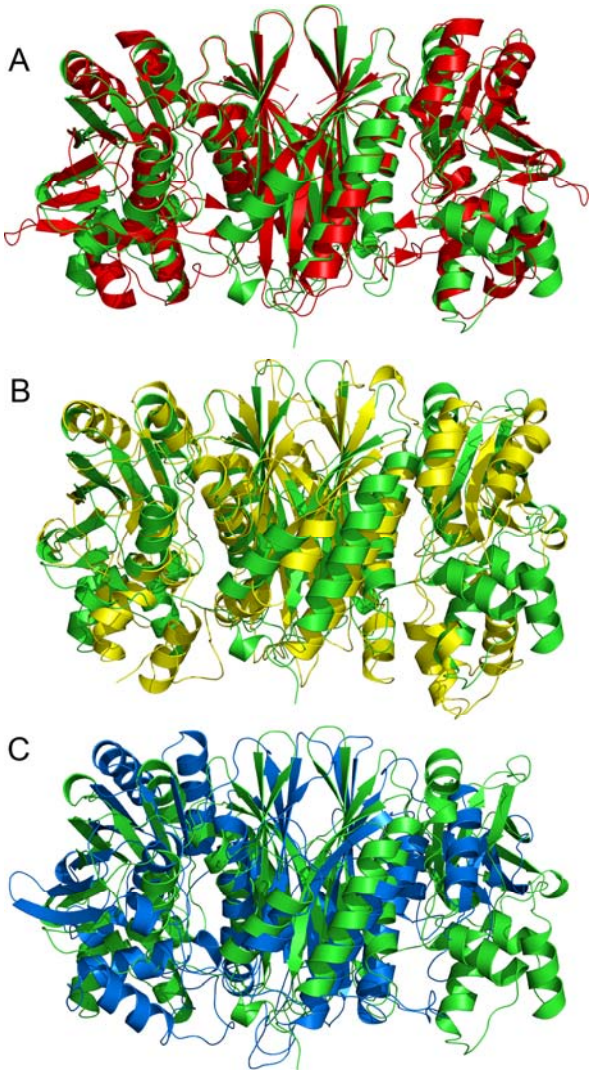


Figure 4.6. Comparison of AaThiL to other members of the PurM ATP binding superfamily. (A) Ribbon diagram of TkHypE superimposed on AaThiL. TkHypE is colored in red and AaThiL is green. (B) Ribbon diagram of EcPurM superimposed on AaThiL. EcPurM is colored in yellow and AaThiL is colored in green. (C) Ribbon diagram of TmPurL superimposed on AaThiL. TmPurL is colored in blue and AaThiL is colored in green.

using the program VAST (*23*). As seen in Figure 4.7, the signature sequence motif is structurally conserved, as well as five glycine residues, but no other residue is conserved within the superfamily. The aspartate residue from the signature motif is responsible for coordinating a magnesium ion in the active sites of TmPurL and AaThiL and by inference, the active sites of StPurL, EcPurM and HypE for which no

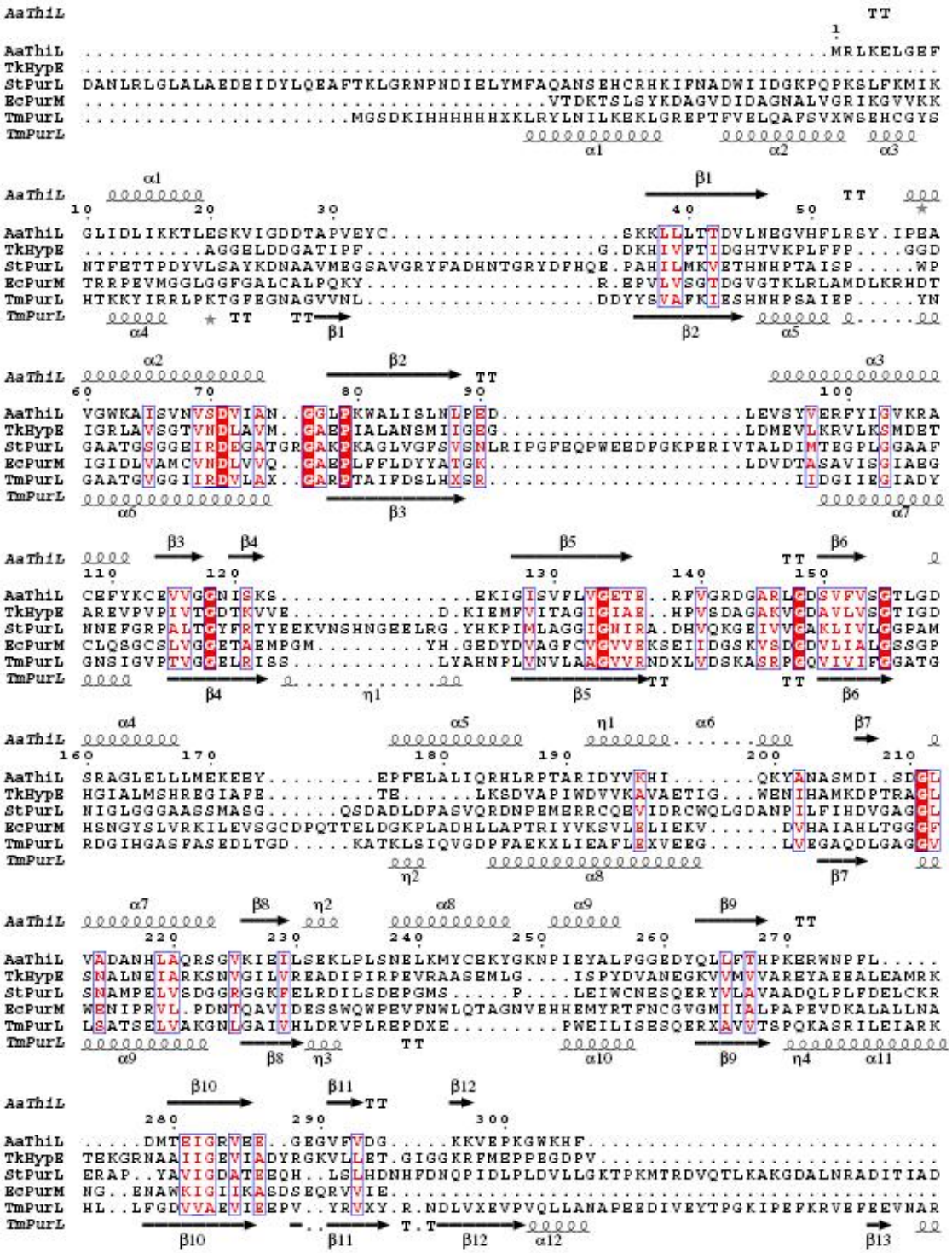


Figure 4.7. Structural sequence alignment of all available structures of the PurM ATP binding superfamily. The conserved residues are shown blocked in red, while less conserved residues are in red font. The only conserved residues are the signature sequence for the superfamily and glycine residues important for folding. The structural alignment was done using VAST.

complex structure are available. This magnesium ion coordinates the β-phosphate and γ-phosphate groups of ATP. The conserved glycine residues are found at the ends of β-strands or α-helices and most likely play structural roles within the PurM ATP binding superfamily.

Also identified through the DALI search with Z scores ranging from 8.4 to 5.4 and rmsd values between 2.4 Å and 3.1 Å for 90 to 95 residues aligned were a series of proteins belonging to the YjgF family. The function of the YjgF family of proteins is largely unassigned, although one family member, YabJ, has been implicated in the regulation of purines (*24*). The YjgF family of proteins (PDB ID: 1QU9) adopt a fold that is similar to that of chorismate mutase with a six-stranded mixed β-sheet β1↑β2↓β3↑β6↓β4↑β5↑ flanked on one side by two α-helices (*25*). Like the β-strands of the A domain of the PurM superfamily, these β-strands are unusually long and contain eight to ten residues. ThiL has a strand order of β1↑β2↓β5↑β3↓β4↓ and in the DALI alignment strands β1, β2, β5, β3 and β4 of AaThiL superimpose on strands β2, β3, β6, β4 and β5, respectively, of YjgF. YjgF is a trimer and forms a 12-stranded central β-barrel that is reminiscent of the eight-stranded central β-barrel of the AaThiL dimer. Clefts that are postulated to be the active site are formed between the subunits of the YjgF trimer. After the DALI alignment the cleft of YjgF superimposes on the active site of ThiL (Figure 4.8B). Examination of the YjgF cleft and the active site of ThiL failed to reveal the conserved aspartate residue from the signature motif. Therefore, the shared structural motif is unlikely to bind ATP in the YjgF family.

Active Site Comparison. While structures of EcPurM, StPurL, and TkHypE have nothing bound, the structures of both AaThiL and TmPurL have been determined with substrates or products bound to the active site (*18*). The superposition of the AaThiL and TmPurL active sites is shown in Figure 4.9. The two AMP-PCP

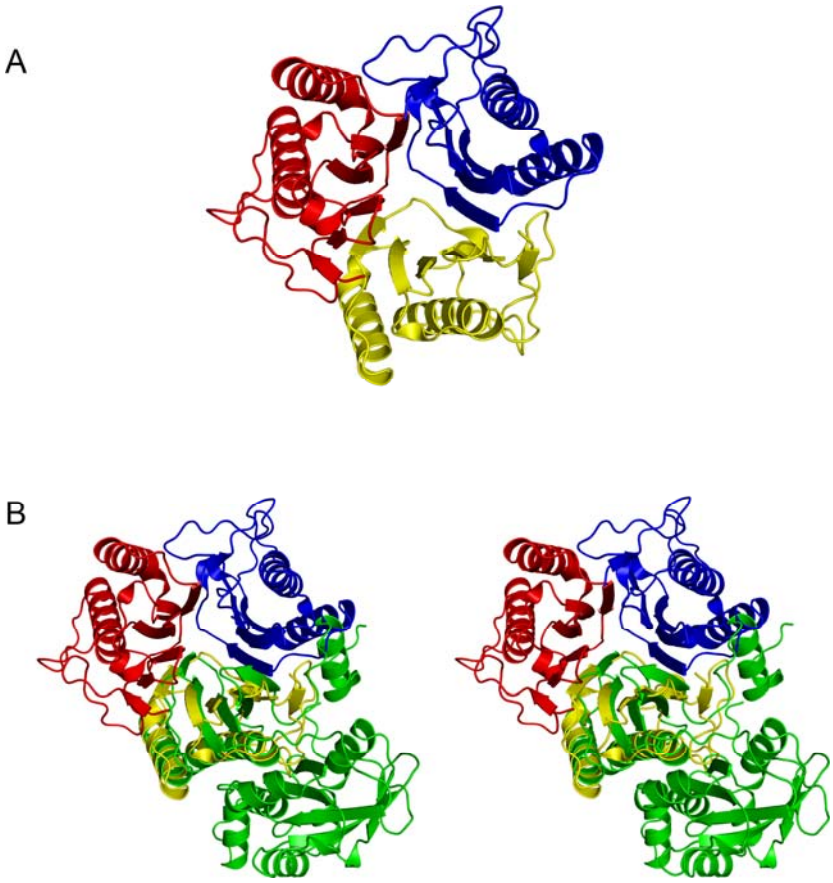


Figure 4.8. Comparison of the YjgF gene product from *E. coli* and AaThiL. (A) The structure of the trimer of YjgF. (B) Stereoview diagram of the superpositioning of the YjgF trimer on the monomer of AaThiL. AaThiL is colored in green, while the monomers of YjgF are colored in yellow, blue, and red.

molecules bind similarly although the ribose moiety is slightly rotated in TmPurL. This results in a slight shift in the phosphate tail placing the γ-phosphate groups 2.3 Å apart. Overall, the secondary structural elements surrounding AMP-PCP superimpose well. Two magnesium ions, Mg1 and Mg2, are present in the active sites of both structures and precisely position ATP for phosphate transfer. A third magnesium ion in the AaThiL structure, Mg3, has no structural equivalent in the TmPurL structure and coordinates the α-phosphate and β-phosphate groups of AMP-PCP. In other ATP-dependent enzymes this role of positioning the phosphate tail is often played by positively charged residues, which are lacking in the PurM ATP binding superfamily.

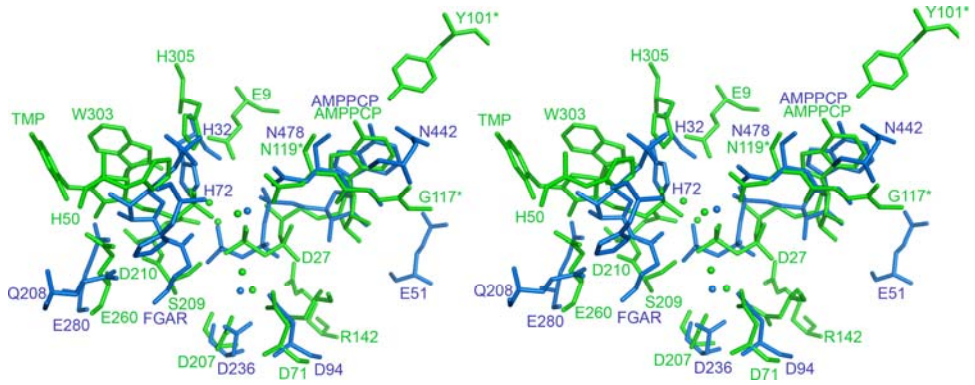


Figure 4.9. Superposition of the active sites of AaThiL and TmPurL. The ThiL structure is shown in green and the PurL structure is shown in blue.

The TMP and formylglycinamide ribonucleotide (FGAR) binding sites are in similar locations but show fewer similarities than the ATP binding sites. Both substrates have a phosphate group; however, the phosphate group plays a different role in the two enzymes. The phosphate group of TMP is directly involved in the reaction as the acceptor of the γ-phosphate group of ATP. The TMP phosphate group is pointed into the active site and is aligned with the γ-phosphate group of AMP-PCP. The transfer is facilitated by Mg1 and Mg2 in AaThiL. In contrast, the phosphate group of FGAR is not directly involved in the reaction catalyzed by PurL; rather it is the carbonyl group of FGAR that is phosphorylated. The oxygen atom of the carbonyl group of FGAR is 2.1 Å from the oxygen atom of the phosphate group of TMP and no magnesium ions are located near FGAR in the TmPurL structure. Each structure has two conserved histidine residues that form hydrogen bonds to TMP or FGAR; no other positively charged residues hydrogen bond to TMP, and FGAR has only one additional hydrogen bond to a positively charged residue.

Comparison to Proteins Binding Thiamin. TPP in cells usually acts as a cofactor where it participates in a variety of enzymatic reactions. TPP dependent enzymes require either a magnesium ion or calcium ion and the TPP cofactors exist in similar conformations when bound to TPP-dependent enzymes (*26*). For example, in

yeast pyruvate decarboxylase (PDB ID: 1PYD) (*27*) and yeast transketolase (PDB ID: 1AY0) (*28*), N4 is positioned to deprotonate C2 of the thiazole ring. In contrast, TPP when not serving as a cofactor, as well as thiamin and TMP, are usually found in different conformations when bound to proteins. For example, in thiamin phosphate synthase (PDB ID: 1G4S), TMP is bound such that N4 is pointed away from C2 (*29*). In this conformation, N4 is unable to deprotonate C2, preventing the carbanion from forming (*26*). In AaThiL the pyrimidine ring of TPP is rotated such that N4 is pointed toward C2 of the thiazole ring; however, the conformation TPP adopts in ThiL is different from both described above. While N4 is pointed toward C2, the pyrimidine ring is rotated about 100° compared to TPP as a cofactor, which may prevent TPP activation in the ThiL binding site.

Mechanistic Implications. The presence of a phosphoimidate intermediate (Figure 4.10) was proposed as a common theme among the members of the PurM superfamily (*6*). Positional isotope exchange studies performed on SelD suggest a phosphorylated enzyme intermediate, although the phosphorylated residue has not been definitively identified (*30, 31*). Studies on PurL in which the amide carbonyl oxygen atom from FGAR is labeled with O^{18} show that this oxygen atom is found in the inorganic phosphate product, also supporting a phosphoimidate intermediate (*32*). These observations suggest that ThiL might transfer the phosphate to TMP indirectly through a phosphoenzyme intermediate rather than a directly through an inline mechanism (*33*). Inspection of the AaThiL complexes revealed no candidate (e.g., backbone amide or amino acid side chain) with suitable geometry for formation of a phosphoenzyme intermediate. The binary complexes showed that Ser209 is possibly positioned for phosphorylation; however, the ternary complexes with AMP-PCP/TMP or ADP/TPP showed that Ser209 is better positioned to provide a hydrogen bond to the γ-phosphate of ATP or the β-phosphate of TPP, and is poorly positioned to take

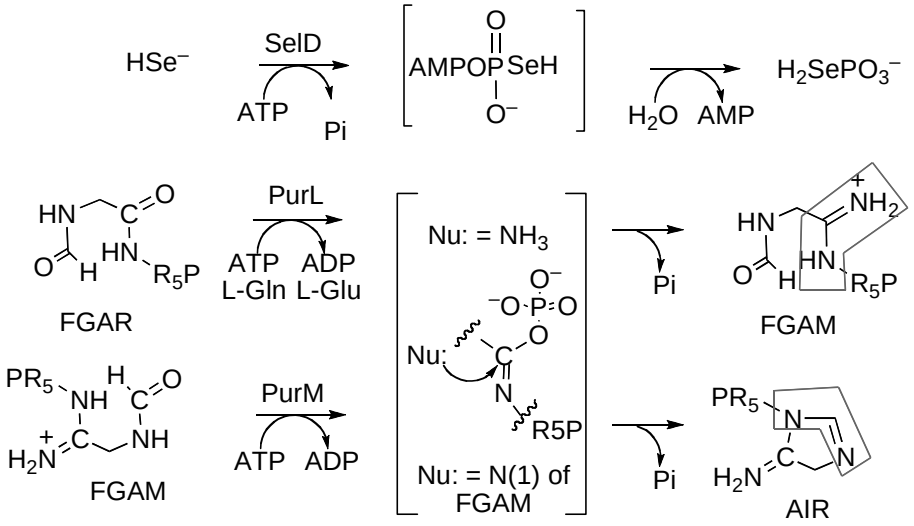


Figure 4.10. Proposed phosphoimidate intermediates for SelD, PurL, and PurM.

part in a double displacement mechanism. Additionally, the γ-phosphate of ATP is positioned only 2.7 Å away from the β-phosphate of TPP with an ideal geometry for inline transfer. The active site of ThiL is relatively simple; only two of the protein side chains participate directly in catalysis (Arg142 and Ser209). The five magnesium ions and several water molecules present in the ternary substrate and product complexes orient the substrates, stabilize the trigonal bipyramidal intermediate and activate ADP as a leaving group (Figure 4.3). These observations support a direct inline transfer rather than an indirect double transfer via a phosphoenzyme intermediate.

REFERENCES

1. Settembre, E., Begley, T. P., and Ealick, S. E. (2003) Structural biology of enzymes of the thiamin biosynthesis pathway, *Curr Opin Struct Biol 13*, 739-747.
2. Jordan, F. (2003) Current mechanistic understanding of thiamin diphosphate-dependent enzymatic reactions, *Nat Prod Rep 20*, 184-201.
3. Begley, T. P., Downs, D. M., Ealick, S. E., McLafferty, F. W., Van Loon, A. P., Taylor, S., Campobasso, N., Chiu, H. J., Kinsland, C., Reddick, J. J., and Xi, J. (1999) Thiamin biosynthesis in prokaryotes, *Arch. Microbiol. 171*, 293-300.
4. Webb, E., and Downs, D. (1997) Characterization of thiL, encoding thiamin-monophosphate kinase, in Salmonella typhimurium, *J Biol Chem 272*, 15702-15707.
5. Li, C., Kappock, T. J., Stubbe, J., Weaver, T. M., and Ealick, S. E. (1999) X-ray crystal structure of aminoimidazole ribonucleotide synthetase (PurM), from the *Escherichia coli* purine biosynthetic pathway at 2.5 Å resolution, *Structure 7*, 1155-1166.
6. Anand, R., Hoskins, A. A., Stubbe, J., and Ealick, S. E. (2004) Domain Organization of *Salmonella typhimurium* Formylglycinamide Ribonucleotide Amidotransferase Revealed by X-ray Crystallography, *Biochemistry 43*, 10328-10342.
7. Jayaraj, S., Reid, R., and Santi, D. V. (2005) GeMS: an advanced software package for designing synthetic genes, *Nucleic Acids Res 33*, 3011-3016.
8. Kodumal, S. J., Patel, K. G., Reid, R., Menzella, H. G., Welch, M., and Santi, D. V. (2004) Total synthesis of long DNA sequences: synthesis of a

contiguous 32-kb polyketide synthase gene cluster, *Proc Natl Acad Sci U S A 101*, 15573-15578.

9. Kodumal, S. J., and Santi, D. V. (2004) DNA ligation by selection, *Biotechniques 37*, 34, 36, 38 passim.
10. Sambrook, J., Fritsch, G. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Guide*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
11. Matthews, B. W. (1968) Solvent content of protein crystals, *J. Mol. Biol. 33*, 491-497.
12. Otwinowski, Z., and Minor, W. (1997) Processing of x-ray diffraction data collected in oscillation mode, *Methods Enzymol. 276*, 307-326.
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. Cowtan, P. E. a. K. (2004) Coot: Model-Building Tools fo Molecular Graphics, *Acta Crystallographica Section D 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 Cryst. D54*, 905-921.
16. DeLano, W. L. (2002) The PyMOL Molecular Graphics System.
17. Krissinel, E., and Henrick, K. (2005) Detection of Protein Assemblies in Crystals, In *CompLife 2005* (al., B. e., Ed.), pp 163-174, Springer-Verlag, Berlin Heidelberg.

18. Morar, M., Anand, R., Hoskins, A. A., Stubbe, J., and Ealick, S. E. (2006) Complexed structures of formylglycinamide ribonucleotide amidotransferase from Thermotoga maritima describe a novel ATP binding protein superfamily, *Biochemistry 45*, 14880-14895.

19. Watanabe, S., Matsumi, R., Arai, T., Atomi, H., Imanaka, T., and Miki, K. (2007) Crystal structures of [NiFe] hydrogenase maturation proteins HypC, HypD, and HypE: insights into cyanation reaction by thiol redox signaling, *Mol Cell 27*, 29-40.

20. Holm, L., and Sander, C. (1998) Touring protein fold space with Dali/FSSP, *Nucleic Acids Res. 26*, 316-319.

21. Blokesch, M., Paschos, A., Bauer, A., Reissmann, S., Drapal, N., and Bock, A. (2004) Analysis of the transcarbamoylation-dehydration reaction catalyzed by the hydrogenase maturation proteins HypF and HypE, *Eur J Biochem 271*, 3428-3436.

22. Holm, L., and Park, J. (2000) DaliLite workbench for protein structure comparison, *Bioinformatics 16*, 566-567.

23. Madej, T., Gibrat, J. F., and Bryant, S. H. (1995) Threading a database of protein cores, *Proteins 23*, 356-369.

24. Sinha, S., Rappu, P., Lange, S. C., Mantsala, P., Zalkin, H., and Smith, J. L. (1999) Crystal structure of Bacillus subtilis YabJ, a purine regulatory protein and member of the highly conserved YjgF family, *Proc Natl Acad Sci U S A 96*, 13074-13079.

25. Volz, K. (1999) A test case for structure-based functional assignment: the 1.2 A crystal structure of the yjgF gene product from Escherichia coli, *Protein Sci 8*, 2428-2437.

26. Malandrinos, G., Louloudi, M., and Hadjiliadis, N. (2006) Thiamine models and perspectives on the mechanism of action of thiamine-dependent enzymes, *Chem Soc Rev 35*, 684-692.

27. Dyda, F., Furey, W., Swaminathan, S., Sax, M., Farrenkopf, B., and Jordan, F. (1993) Catalytic centers in the thiamin diphosphate dependent enzyme pyruvate decarboxylase at 2.4-A resolution, *Biochemistry 32*, 6165-6170.

28. Wikner, C., Nilsson, U., Meshalkina, L., Udekwu, C., Lindqvist, Y., and Schneider, G. (1997) Identification of catalytically important residues in yeast transketolase, *Biochemistry 36*, 15643-15649.

29. Peapus, D. H., Chiu, H. J., Campobasso, N., Reddick, J. J., Begley, T. P., and Ealick, S. E. (2001) Structural characterization of the enzyme-substrate, enzyme-intermediate, and enzyme-product complexes of thiamin phosphate synthase, *Biochemistry 40*, 10103-10114.

30. Mullins, L. S., Hong, S.-B., Gibson, G. E., Walker, H., Stadtman, T. C., and Raushel, F. M. (1997) Identification of a Phosphorylated Enzyme Intermediate in the Catalytic Mechanism for Selenophosphate Synthetase, *J. Am. Chem. Soc. 119*, 6684-6685.

31. Walker, H., Ferretti, J. A., and Stadtman, T. C. (1998) Isotope exchange studies on the Escherichia coli selenophosphate synthetase mechanism, *Proc Natl Acad Sci U S A 95*, 2180-2185.

32. Schendel, F. J., Mueller, E., Stubbe, J., Shiau, A., and Smith, J. M. (1989) Formylglycinamide ribonucleotide synthetase from Escherichia coli: cloning, sequencing, overproduction, isolation, and characterization, *Biochemistry 28*, 2459-2471.

33. Prasher, D. C., Carr, M. C., Ives, D. H., Tsai, T. C., and Frey, P. A. (1982) Nucleoside phosphotransferase from barley. Characterization and evidence for

ping pong kinetics involving phosphoryl enzyme, *J Biol Chem 257*, 4931-4939.

CHAPTER 5

GENE IDENTIFICATION AND STRUCTURAL CHARACTERIZATION OF THE PYRIDOXAL 5′-PHOSPHATE DEGRADATIVE PROTEIN 3-HYDROXY-2-METHYLPYRIDINE-4,5-DICARBOXYLATE DECARBOXYLASE FROM *MESORHIZOBIUM LOTI* MAFF303099[1]

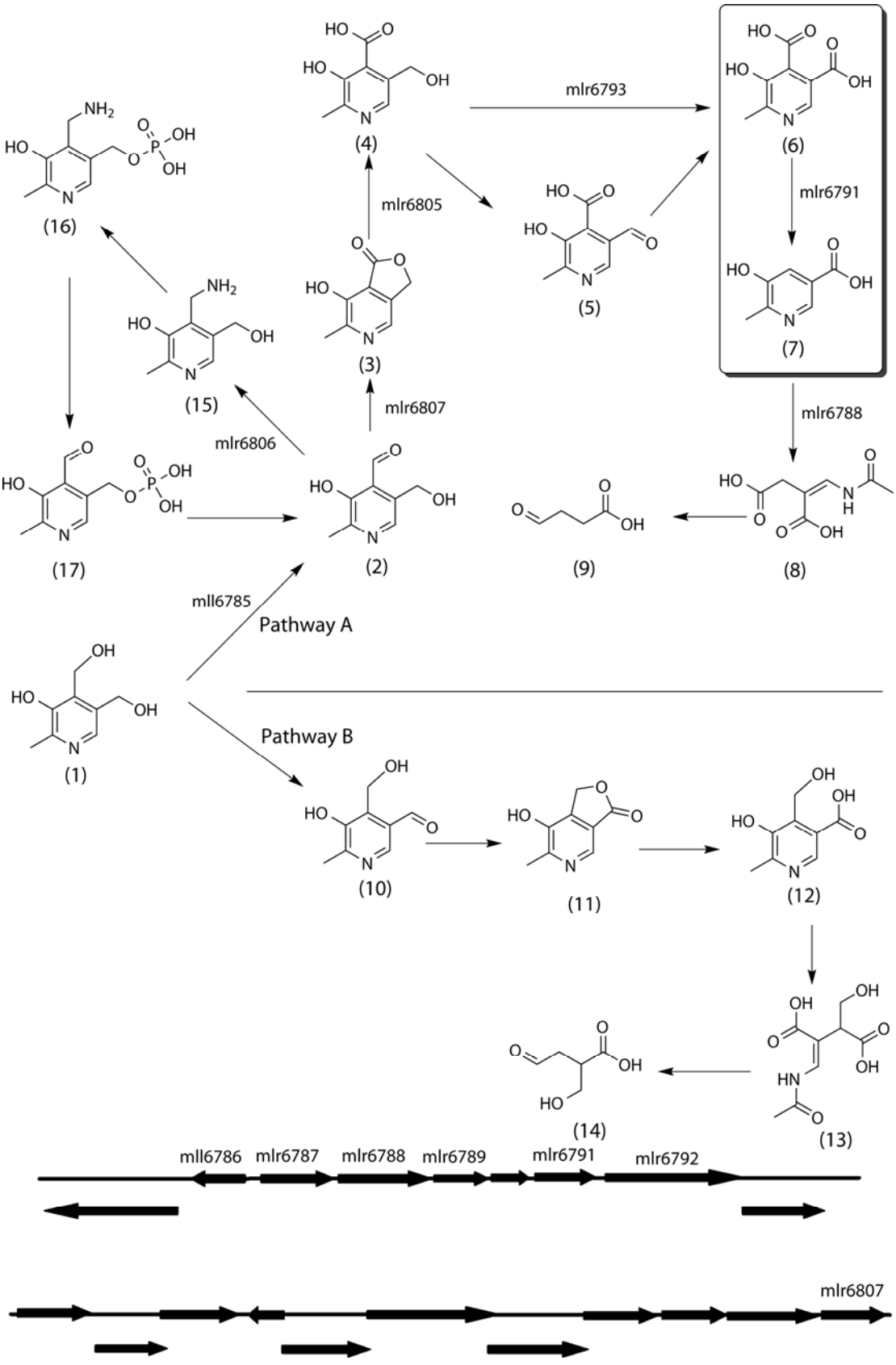
Section 5.1. Introduction

In contrast to our understanding of cofactor biosynthetic pathways, very little is known about cofactor catabolism. Cofactor catabolism is likely to be rare because cofactors are trace metabolites and therefore not good food sources for bacteria. Pyridoxine, **1** (vitamin B_6) catabolism is the best understood cofactor catabolic pathway and a small number of bacteria that can grow on vitamin B_6 as the sole source of carbon and nitrogen have been identified (*1*). Two catabolic pathways have been proposed (*2*). In the first pathway of Figure 5.1, (pathway A), found in *Pseudomonas sp.* MA-1, vitamin B_6 is degraded in eight steps to form succinic semialdehyde, **9**, while in the second pathway (pathway B) observed in *Pseudomonas* IA and in *Arthrobacter* Cr-7, vitamin B_6 is catabolized in seven steps to 2-(hydroxymethyl)-4-oxobutanoate, **14**. A related catabolic pathway in *Mesorhizobium loti* MAFF303099 which is very similar to the degradative pathway A has recently been discovered. 3-hydroxy-2-methylpyridine-5-carboxylate, **7** is an intermediate in this catabolic pathway. The gene coding for the enzyme producing this metabolite has not previously been discovered and is the subject of this paper.

Previously, the genes encoding pyridoxine-4-oxidase (mlr6785), 4-pyridoxolactonase (mlr6805), pyridoxal-4-dehydrogenase (mlr6807) and the 2-methyl-3-hydroxypyridine-5-carboxylic acid oxygenase (mlr6788) were identified in

[1] Reproduced with permission from Mukherjee, T., McCulloch, K.M., Ealick, S.E., and Begley, T.P. (2007) *Biochemistry* 46: 13606-13615. Copyright 2007 American Chemical Society

Figure 5.1. The vitamin B_6 catabolic pathways. The pathway A is observed in Pseudomonas sp. MA-1 (*3*) while the pathway B is observed in *Pseudomonas* IA and in *Arthrobacter* Cr-7. A minor variation of pathway A is seen in *Mesorhizobium loti* MAFF303099 where the formation of **5** is not seen. The reaction shown in the box is catalyzed by HMPDdc. The bottom figure shows the gene organization in *M. loti*. The genes that have been identified to participate in PLP degradation are identified in catabolic pathway A.


mlr6793
mlr6805
mlr6791
mlr6788
mlr6807
mlr6806
mll6785
Pathway A
Pathway B
(1)
(2)
(3)
(4)
(5)
(6)
(7)
(8)
(9)
(10)
(11)
(12)
(13)
(14)
(15)
(16)
(17)
mll6786
mlr6787
mlr6788
mlr6789
mlr6791
mlr6792
mlr6807

Mesorhizobium loti MAFF303099 (*4-7*). Recently we have reported the identification of a fifth gene (mlr6793), encoding 4-pyridoxic acid dehydrogenase (*8*). These genes are not part of an operon, but are all close to each other on the *M. loti* chromosome (Figure 5.1). This suggested that other PLP catabolic genes might also be found in this region. In particular, we felt that the mlr6791gene, annotated as ribulose-5-phosphate 4-epimerase (*9*), was a likely candidate for the decarboxylase gene because the epimerase and the decarboxylase catalyzed reactions both proceed via enzyme stabilized enolate intermediates. In this paper, we report the cloning and overexpression of mlr6791, demonstrate that the purified gene product catalyzes the decarboxylation of 3-hydroxy-2-methylpyridine-4,5-dicarboxylate **6** and describe the structure of this enzyme. While the decarboxylation of hydroxy-substituted benzene rings is a common motif in biosynthesis, the mechanism of this reaction is still poorly characterized. The structural studies described here suggest that catalysis of such decarboxylations proceeds by an aldolase-like mechanism.

Section 5.2. Experimental Procedures

Materials. A dehydrated form of Luria-Bertani (LB) broth was purchased from EMB Chemicals, (Gibbstown, NJ). Ampicillin and isopropyl-β-D-thiogalactopyranoside (IPTG) were obtained from Lab Scientific Inc. (Livingston, N.J.). 4-pyridoxic acid, NAD, TRIS hydrochloride, DTT and M9 minimal salts were from Sigma (St. Louis, MO.). Triethylamine was from Fisher (Fairlawn, NJ). Trifluoroacetic acid (TFA), methanol (HPLC grade), sodium chloride, imidazole, 2-mercaptoethanol and L-(+)-selenomethionine were from Acros Organics (Morris Plains, NJ). Sodium dihydrogen phosphate monohydrate, calcium chloride, ferrous sulfate heptahydrate, manganese chloride and magnesium sulfate were from Mallinckrodt Baker Inc. (Phillipsburg, NJ). Deuterium oxide (D_2O) was purchased from Cambridge Isotope Laboratories Inc. (Andover, MA). Microcon YM-10

centrifugal filter devices (10,000 MWCO) and the Amicon Ultra centrifugal filter device (10,000 MWCO) were obtained from Millipore (Billerica, MA). The Supelcosil LC-18-T column for HPLC was from Supelco, (Bellefonte, PA). MEM vitamin solution, *E. coli* strain MachI and the Gateway system were from Invitrogen (Carlsbad, CA). Nucleospin Purification kit, Phusion DNA polymerase, *E. coli* BL21(DE3) and the Ni-NTA superflow resin were obtained from Macherey-Nagel (Easton, PA), New England Biolabs (Ipswich, MA), Novagen (San Diego, CA) and Qiagen (Valencia, CA) respectively.

Molecular Cloning. Standard methods were used for DNA manipulations (*10*), (*11*). Plasmid DNA was purified with the Qiagen Miniprep kit and DNA fragments were purified from agarose gel with the Nucleospin Purification kit. *Escherichia coli* strain Mach1 was used as a recipient for transformations during plasmid construction and for plasmid propagation. Phusion DNA polymerase was used for PCR following the manufacturer's recommendations. The pENTR-TEV-D-TOPO and the Gateway system were used following the manufacturer's instructions with slight modifications.

Cloning of M. loti mlr6791. The *M. loti* mlr6791 gene was amplified from genomic DNA by PCR with the following primer pair: 5'-CAC CAT GCG TCG GAA GGT CTT CGA AGA G-3' AND 5'-TCA GGC GAG GCC TGC TTG CCT GAG G-3'. The PCR product was purified and used in a topoisomerase mediated reaction with pENTR-TEV-D-TOPO essentially following the manufacturer's instructions. Clones were screened by PCR and verified by sequencing. In initial sequencing, no clones were found with a completely correct sequence. One clone was subjected to standard site-directed mutagenesis with the following complementary primer pair: 5'-GGA TAC GTT CGG GCA CAT ATC TGC CCG TGA CCC CGA G-3' and 5'-CTC GGG GTC ACG GGC ACA TAT GTG CCC GAA CGT ATC C-3'. Colonies from the mutagenesis were screened with the M13-reverse primer and the mutant specific

primer 5'-GAT ACG TTC GGG CAC ATA TC-3'. PCR positive clones were sequenced and a correct clone was used in an LR recombination reaction with the plasmid pDESTF1, which is a Gateway adapted vector based on the pET-system. The plasmid pDESTF1 encodes an N-terminal 6xHis tag and is under the control of the T7/lac /promoter. Clones were screened by restriction digestion. A correct clone was named pMl5335.XF1.

Overexpression and Purification. The plasmid pMl5335.XF1 was used to transform *E. coli* BL21(DE3). The cells were grown in 1L LB medium, containing 100 μg/mL of ampicillin, at 37 °C with agitation until the culture reached an OD_{590} of 0.6, at which point overexpression was induced by adding IPTG to a final concentration of 0.5 mM, the temperature was lowered to 15 °C and the cells were allowed to grow for a further 12 h before being harvested. The yield of the purified protein was 20 mg/L.

For selenomethionine (SelMet) protein, the plasmid pMl5335.XF1 was transformed into *E. coli* B834(DE3) cells, a strain auxotrophic for methionine. Cells were grown at 37 °C with shaking in minimal M9 media which was supplemented with 20 mg/L of all amino acids except methionine, 1X MEM vitamin mix, 0.4% glucose, 50 mg/L L-selenomethionine, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 25 mg/L $FeSO_4$, and 100 mg/L ampicillin. When cells reached an $O.D._{600}$ of 0.6, the temperature was lowered to 15 °C and protein overexpression was induced with 0.5 mM IPTG. Cells were harvested after 18 h by centrifugation at 8500 *g* for 15 min. Cells were lysed by sonication and the cell lysate was cleared by centrifugation at 40000 *g* for 45 min at 4 °C. SelMet-HMPDdc was purified using a Ni-NTA affinity chromatography column. The sample was loaded in buffer A spiked with 10 mM imidazole, which contains 50 mM NaH_2PO_4, 300 mM NaCl, and 3 mM β-mercaptoethanol to prevent oxidation of the selenomethionine. The column was then washed with 10 column volumes of

buffer A with 30 mM imidazole and the sample was eluted from the column with buffer A containing 250 mM imidazole. Protein was buffer exchanged into 10 mM Tris-HCl pH 7.7, 20 mM NaCl, and 1 mM DTT using an Econo-Pac 10 DG desalting column (BioRad). SelMet-HMPDdc was judged to be greater than 95% pure by SDS-PAGE gel analysis (results not shown). The yield of SelMet-HMPDdc was comparable to native protein at 20 mg per L. The protein sample was concentrated to 10 mg/mL as determined by Bradford assay using bovine serum albumin as a standard (*11*).

HPLC analysis. HPLC analysis of the enzymatic reaction mixture was performed on a Hewlett-Packard 1100 instrument using a Supelcosil LC-18-T (15 cm X 4.6 mm, 3.0 μM) column. Solution A contained water, solution B contained 100 mM sodium phosphate buffer at pH 6.6 and solution C contained methanol. The following linear gradient was used: 0% to 10% solution A and 100% to 90% solution B for 0 to 5 min, 10% to 48% solution A, 90% to 40% of solution B and 0% to 12% of solution C from 5-12 min, 48% to 50% solution A, 40% to 30% of solution B and 12% to 20% of solution C in 12-14 min, 50% to 30% solution A, 30% to 10% of solution B and 20% to 60% of solution C in 14-18 min, 30% to 0% solution A, 10% to 100% of solution B and 60% to 0% of solution C in 18-20 min and 0% of solution A, 100% of solution B and 0% of solution C in 20-25 minutes. The flow rate was 1 mL/min and the absorbance was measured at 254 nm (characteristic for NAD, 3-hydroxy-2-methylpyridine-5-carboxylate **7**) and 320 nm (characteristic of 4-pyridoxic acid **4,** 3-hydroxy-2-methylpyridine-4,5-dicarboxylate **6** and 3-hydroxy-2-methylpyridine-5-carboxylate,**7**). Under these conditions the following compounds were readily separated (retention time in parenthesis): 4-pyridoxic acid, **4**, (17.8 min), 3-hydroxy-2-methylpyridine-4,5-dicarboxylate **6** (2.8 min), 3-hydroxy-2-methylpyridine-5-carboxylate **7 (**5.5 min) and NAD (12.9 min).

Enzymatic synthesis of the substrate. 3-hydroxy-2-methylpyridine-4,5-dicarboxylate **6** was enzymatically synthesized from 4-pyridoxic acid, **4**. The 5 mL enzymatic reaction mixture containing 12 mM NAD, 6 mM 4- pyridoxic acid, **4** and 100 μM freshly purified 4-pyridoxic acid dehydrogenase (*8*) in 100 mM sodium phosphate buffer at pH 8.0 was incubated overnight at room temperature. It was subsequently concentrated by lyophilizing and redissolving in a minimum volume of water. It was then filtered through YM-10 Microcon centrifugal filter at 14,000 *g* for 30 min to remove the protein and the filtrate was purified by HPLC. 100 mM ammonium acetate, pH 6.6, was used instead of 100 mM sodium phosphate buffer at pH 6.6 as solution B, to facilitate removal of the buffer salts from the isolated 3-hydroxy-2-methylpyridine-4,5-dicarboxylate **6** during lyophilization. Compound **6** is a stable white solid. ^{1}H NMR (300 MHz, D_2O) δ 2.69 (s, 3H, CH_3) and 8.42 (s, 1H, C_6-H).

Reaction time course. A time course was determined with a reaction mixture (1 mL) containing 1.58 mM 3-hydroxy-2-methylpyridine-4,5-dicarboxylate **6**, 2 μM pure enzyme and 10 μM $MnCl_2$ in 100 mM Tris HCl at pH 8.0 containing 100 mM NaCl and 2 mM DTT. At various time points, 100 μL of the reaction mixture was quenched by addition to 100 μL of 10% TFA. This mixture was filtered through Microcon YM-10 and 100 μL of the filtrate was analyzed by HPLC (Figure 5.2A).

Product purification and characterization. A reaction mixture (10.0 ml) containing 12 mM NAD, 6 mM 4-pyridoxic acid, **4**, 100 μM 4-pyridoxic acid dehydrogenase, 100 μM of HMPDdc and 1mM $MnCl_2$ in 100 mM Tris HCl at pH 8.0 containing 100 mM NaCl and 2 mM DTT was incubated overnight at room temperature. It was then filtered through an YM-10 Microcon centrifugal filter at 14,000 g for 30 min. The desired enzymatic product (retention time 5.5 min) was purified by HPLC over multiple injections. Methanol was removed by rotary

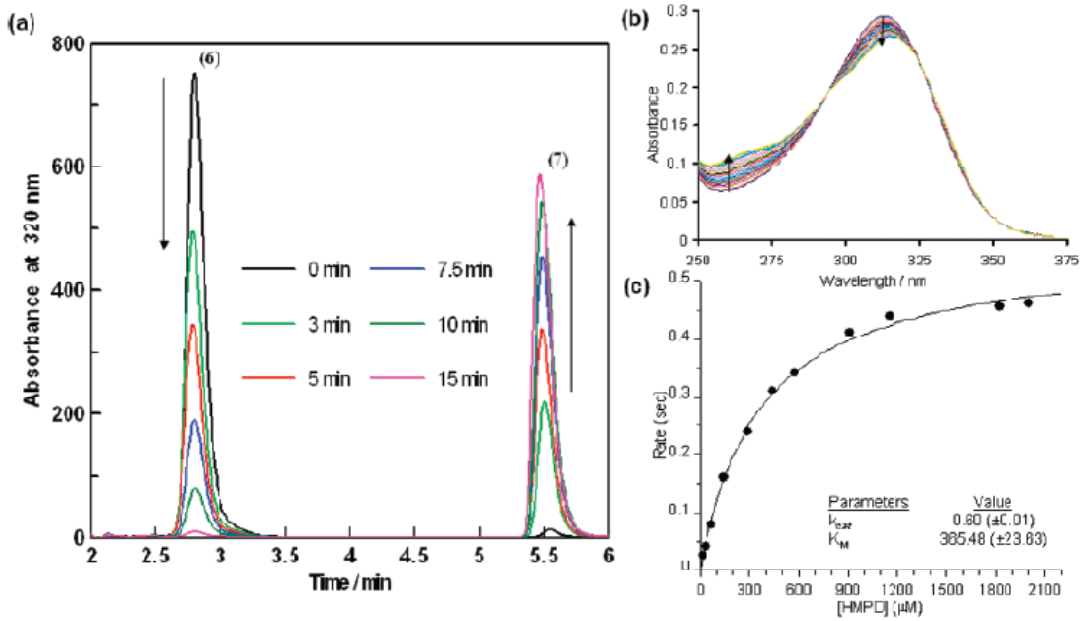


Figure 5.2. (A) HPLC trace showing the disappearance of 3-hydroxy-2-methylpyridine-4,5-dicarboxylate **6** (retention time of 2.8 minutes) and the appearance of 3-hydroxy-2-methylpyridine-5-carboxylate **7** (retention time of 5.5 min). (B) UV visible spectra of the enzymatic reaction mixture (500 □L) containing 1.58 mM of 3-hydroxy-2-methylpyridine-4,5-dicarboxylate **6**, 4 □M $MnCl_2$ and 780 nM of the HMPDdc in 100 mM Tris HCl at pH 8.0 containing 100 mM NaCl and 2 mM DTT were taken at 1 minute intervals over 20 minutes. (C) The steady state kinetic parameters for HMPDdc were determined by monitoring the absorbance at 265 nm over time.

evaporation; TFA and water were removed under high vacuum overnight. The resulting white powder was characterized by NMR and ESI-MS. ^{1}H NMR (300 MHz, D_2O) δ 2.46 (s, 3H, CH_3), 7.68 (s, 1H, C4-H) and 8.12 (s, 1H, C6-H). ESI-MS (Esquire-LC_00146 instrument, Bruker, Negative ion mode) m/z = 152, (mono anionic 3-hydroxy-2-methylpyridine-5-carboxylate **7)**. Fragmentation analysis resulted in the formation of a species with m/z = 108 (M-44, decarboxylation of **7**).

Steady state kinetic parameters. The steady state kinetic parameters for HMPDdc were determined by monitoring the change in absorbance at 265 nm over time. To a reaction mixture (500 μL) containing 1 μM enzyme and 5 μM $MnCl_2$, varying concentrations of 3-hydroxy-2-methylpyridine-4,5-dicarboxylate **6** were

added. The rate of formation of 3-hydroxy-2-methylpyridine-5-carboxylate **7** was monitored over 3 min at 265 nm for each concentration of the substrate. The K_M and k_{cat} for the enzyme were determined by fitting the rate of product formation as a function of substrate concentration using non-linear regression to the Michaelis-Menten equation using Grafit 5.0.11 (Erithacus Software Ltd., Surray, UK). All solutions were made in 100 mM Tris HCl at pH 8.0 containing 100 mM NaCl and 2 mM DTT.

Protein Crystallization. The hanging drop vapor diffusion method was used with 1 μL of SelMet protein solution and 1 μL of reservoir solution at 22 °C. Sparse matrix screens were used to determine initial hits (Crystal Screen and Crystal Screen 2, Hampton Research). Optimized crystallization conditions consisted of 6-9% polyethylene glycol 8000 and 100 mM Tris buffer with a pH ranging from 7.0 to 7.5. Crystals grew in a conical shape to a size of approximately 0.3 mm x 0.1 mm in roughly one week and were cryoprotected by a quick transfer into crystallization conditions with an additional 17% glycerol. SelMet-HMPDdc crystals were then flash frozen by plunging in liquid nitrogen and stored frozen until data collection. Crystals belong to the space group I4 with unit cell parameters a = 72.0 Å and c = 90.4 Å. The unit cell contains one monomer per asymmetric unit with a solvent content of 45% and a Matthews coefficient of 2.24 $Å^3$/Da (*12*).

X-Ray Data Collection and Processing. A single wavelength anomalous diffraction dataset was collected on a SelMet-HMPDdc crystal at the NE-CAT 24-ID-C beamline at the Advanced Photon Source using a Quantum315 detector (Area Detector Systems Corp). The dataset was collected at the maximum f'' for selenium as determined using a fluorescence scan of the SelMet-HMPDdc crystal. The crystal diffracted to 1.9 Å resolution and data was collected over 360° using a 1° oscillation

range. The HKL2000 suite of programs was used to index, integrate, and scale the data (*13*). The data collection statistics are given in Table 5.1.

Table 5.1. Summary of Data Collection Statistics

	SelMet HMPDdc
Beamline	APS NE-CAT 24-IDC
Resolution (Å)	1.90
Wavelength (Å)	0.97918
Space Group	*I*4
a (Å)	72.0
c (Å)	90.4
Reflections	133301
Unique reflections	34563 (2646)
Average I/σ	27.0 (4.0)
Redundancy	3.9 (2.0)
Completeness (%)	96.9 (75.1)
Rsym (%)	5.6 (14.8)

$R_{sym} = \Sigma\Sigma_i \left| I_i - \langle I \rangle \right| / \Sigma \langle I \rangle$, where <I> is the mean intensity of the N reflections with intensities I_i and common indices *h,k,l*.

Structure Determination, Model Building, and Refinement. The computer program hkl2map was used to determine the positions of the Se atoms using data cut off at 2.2 Å resolution to maximize the anomalous signal (*14*). Three of the possible five selenium atoms were located. The program autoSHARP was used for refinement of the heavy atom positions, phasing, calculation of residual maps, density modification, and automated model building (*15*). Automated model building built 222 of 234 residues with the correct side chain. Examination of the density modified maps allowed the manual building of residues 139-141, 209-216, and 234 using COOT (*16*). Refinement of the HMPDdc model was then performed using CNS (*17*). A metal binding site was identified and modeled as a manganese ion, based on the magnitude of the electron density, coordination geometry, and the results of an EXAFS scan performed (results not shown). Water molecules were added as refinement continued also using CNS. The data refinement statistics are given in Table 5.2. The HMPDdc structure was verified using PROCHECK and no residues

were located in disallowed regions of the Ramachandran plot (*18*). Figures were generated using Pymol (*19*).

Table 5.2. Summary of Data Refinement Statistics

	SelMet HMPDdc
Resolution (Å)	50.00 – 1.90
# of protein atoms	1852
# of metal atoms	1
# of water atoms	205
Reflections in working set	16904
Reflections in test set	906
R factor[a] (%)	19.6
Rfree[b] (%)	24.1
Rmsd from ideals	
Bonds (Å)	0.0058
Angles (º)	1.1
Avg B factor (Å^2)	30.3
Ramachandran Plot	
Most favored (%)	88.8
Additionally allowed (%)	10.7
Generously allowed (%)	0.5
Disallowed (%)	0.0

[a] R factor = $\sum_{hkl} \left| \left| F_{obs} \right| - k \left| F_{cal} \right| \right| / \sum_{hkl} \left| F_{obs} \right|$ 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 (10%) excluded from all stages of refinement.

Section 5.3. Results

Product purification and characterization. The product of the enzymatic reaction, was purified as a stable white powder. It was identified as 3-hydroxy-2-methylpyridine-5-carboxylate **7** by NMR and ESI-MS analysis. The presence of a singlet at 7.68 ppm in the aromatic region, corresponding to the C-4 hydrogen of **7**, indicated a clean conversion of substrate **6** to product **7**. The time course for the HMPDdc catalyzed decarboxylation of 3-hydroxy-2-methylpyridine-4,5-dicarboxylate **6** to 3-hydroxy-2-methylpyridine-5-carboxylate **7**, as analyzed by HPLC is shown in Figure 5.2A.

Steady state kinetic parameters. UV visible spectra of the enzymatic reaction mixture (500 μL), containing 1.58 mM of 3-hydroxy-2-methylpyridine-4,5-dicarboxylate **6**, 4 μM $MnCl_2$ and 780 nM of the HMPDdc in 100 mM Tris HCl at pH 8.0 containing 100 mM NaCl and 2 mM DTT, taken at various time points, Figure 5.2B, showed an increase in absorbance at 265 nm. Both 3-hydroxy-2-methylpyridine-4,5-dicarboxylate **6** and 3-hydroxy-2-methylpyridine-5-carboxylate **7** absorb at this wavelength. The extinction coefficients of 3-hydroxy-2-methylpyridine-4,5-dicarboxylate **6** and 3-hydroxy-2-methylpyridine-5-carboxylate **7** at 265 nm in 100 mM Tris HCl at pH 8.0 containing 100 mM NaCl and 2 mM DTT were determined to be 1017 $M^{-1}cm^{-1}$ and 3623 $M^{-1}cm^{-1}$ respectively. The rate of product formation was determined using the difference in molar extinction coefficients (ΔA_{265} = 2615 $M^{-1}cm^{-1}$). Steady state kinetic parameters were obtained from the concentration dependence of the rate of formation of 3-hydroxy-2-methylpyridine-5-carboxylate **7** at constant concentration of HMPDdc under saturating concentration of $MnCl_2$. The enzymatic reaction exhibited Michealis Menten kinetics with K_M and k_{cat} of 366 μM and 0.6 s^{-1} respectively. The k_{cat}/K_M for HMPDdc was determined to be 1530 $M^{-1}s^{-1}$ (Figure 5.2C).

Monomeric Structure of HMPDdc. The structure of HMPDdc was determined at 1.9 Å resolution using SAD phasing. All 234 residues of the protein and 3 residues from the 6X N-terminal His tag were modeled into the final structure of the monomer, as well as 205 water molecules and a manganese ion. The monomer is composed of a single domain with an α/β/α fold (Figure 5.3). The central 7 stranded mixed β-sheet is mostly antiparallel with a strand order of β5↑β6↑β7↓β4↑β1↓β2↑β3↓ where only β5 and β6 run parallel to each other as seen in Figure 5.3B. The β-sheet forms a half-barrel, with four α-helices flanking one side of the β-sheet and two α-helices flanking

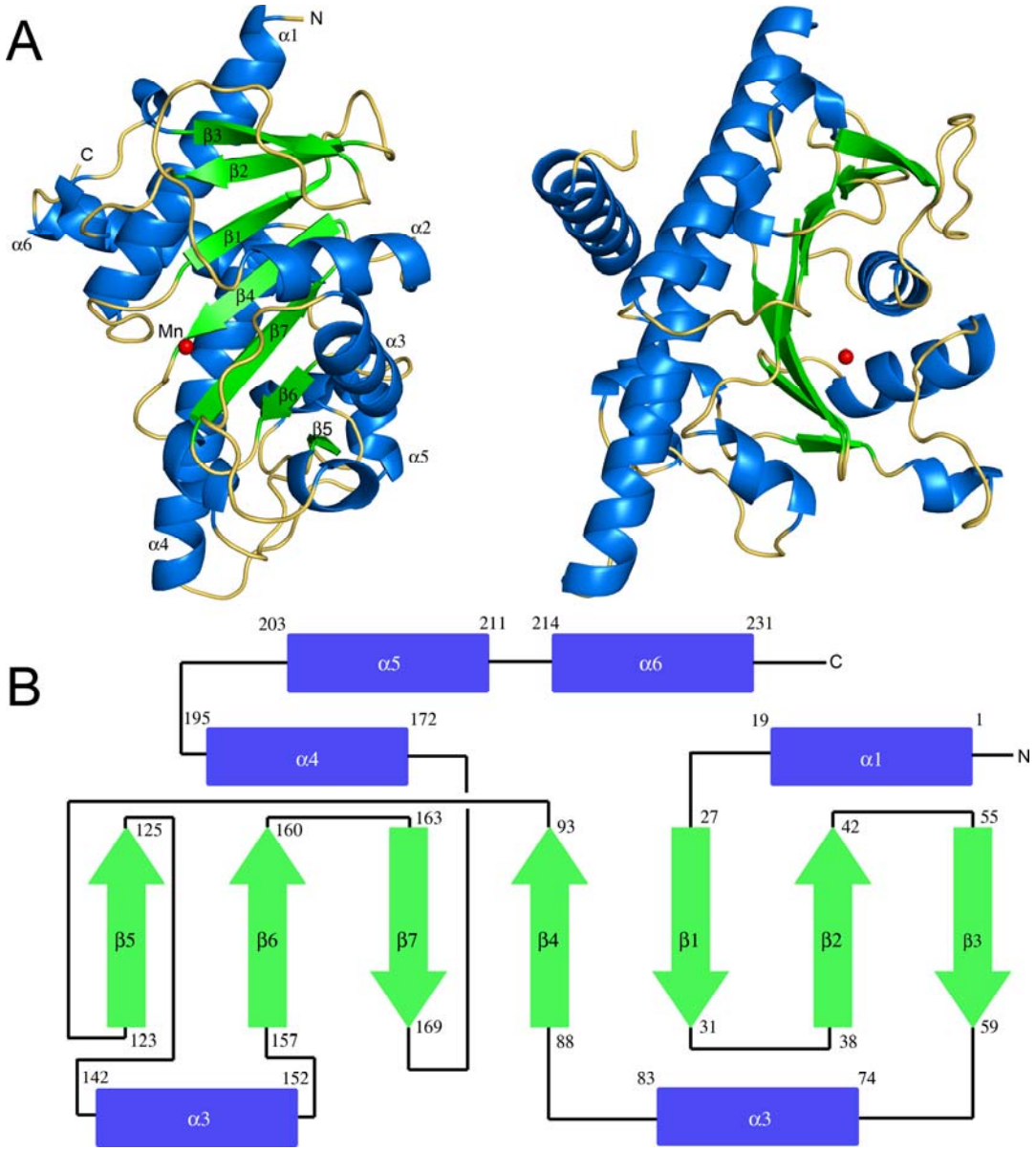


Figure 5.3. Monomeric structure of *Ml*HMPDdc. (A) Ribbon diagram of HMPDdc at two orientations 90° rotated from each other. The secondary structure is labeled and α-helices are colored in blue, β-strands are colored in green, and loop regions are yellow. The manganese ion is shown as a non-bonded red sphere. (B) Topology diagram of HMPDdc.

the opposite side. Three of the α-helices, α1, α4 and α6, are unusually long with 19, 24 and 18 amino acids, respectively. There are also three 3_{10} helices.

Tetrameric Structure of HMPDdc. The quaternary structure of HMPDdc is a tetramer formed by using the fourfold crystallographic axis of the space group I4 and

is shown in Figure 5.4. The tetramer is roughly 65 Å wide, 45 Å tall and 80 Å across the diagonal. A channel with a diameter of 10 Å runs through the tetramer. α-Helices α1, α4, and α6 face the channel and it is composed mostly of hydrophilic side chains and backbone carbonyl groups. The opening is much wider on the top, near the N and C termini, and nearly closed on the bottom by the last turn of α4 and the side chain of Lys194. The subunit interface is formed by interactions between the C-terminal end of one HMPDdc subunit (α4, α5, α6, and the connecting loops) and the loops between β2 and β3, β6 and β7, and the loop between the third 3_{10} helix and α3 from a neighboring subunit. The first 3_{10} helix also contributes to the interface between subunits.

Approximately 3000 $Å^2$ of surface area is buried at the interface between subunits (*20*). The residues lining the subunit interface are mostly hydrophobic, including a patch rich in aromatic residues, including Phe25, Phe177, Phe180, and Tyr181. There are two sets of salt bridges at the interface, one between Arg128 and Glu205 and the second between Arg162 and Glu126. The arginine residues are found close to each other and there also appears to be some stacking between these side chains. The interface is also stabilized by five hydrogen bonds: Asp23 and Thr24 both

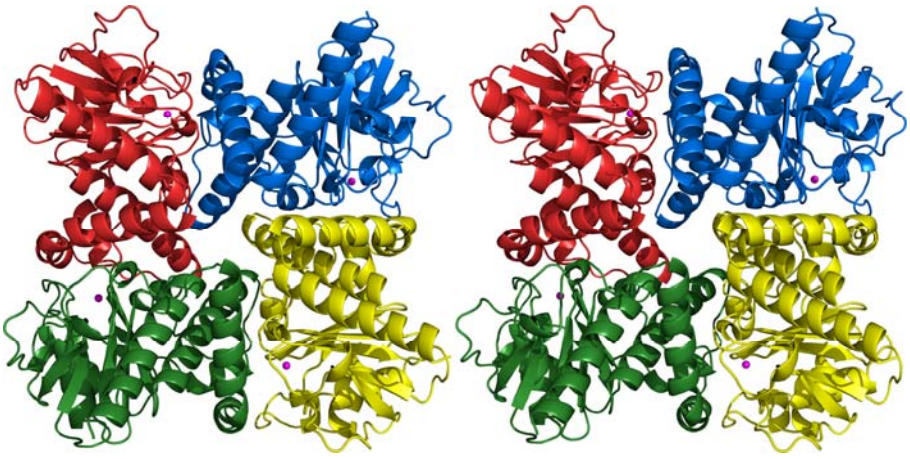


Figure 5.4. Stereoview diagram of the tetrameric structure of HMPDdc. The tetramer has been color coded by subunit. The manganese ion is shown in magenta.

hydrogen bonding to Tyr181, Ser48 to His223, and Asp105 to Asn197. Three backbone carbonyl groups form hydrogen bonds: the carbonyl group from Ile195 to the side chain of Thr106, the carbonyl group of Asp133 to the hydroxyl group of Ser202, and the carbonyl oxygen atom of Glu134 to the oxygen of the Thr208 side chain.

Metal Binding Site. A cleft 13 Å long, 8 Å wide, and 10 Å deep forms near the hydrophobic patch at the interface between two subunits and is shown in Figure 5.5. The cleft is found in the middle of the tetramer and lies closer to the external solvent than to the channel that runs through the tetramer. This cleft contains six histidine residues in close proximity to each other: His27, His92, His94, His163, and His177 from one subunit and His113* from the neighboring subunit.

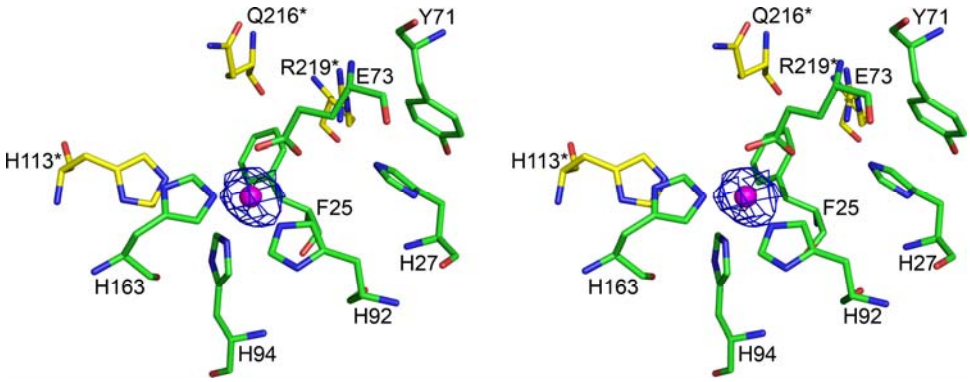


Figure 5.5. Stereoview diagram of the metal binding site. Composite omit density is shown around the manganese ion at a contour level of 1.0 σ and is shown in blue. The manganese ion is magenta. Residues shown in green are from one subunit and the residues shown in yellow are from the neighboring subunit.

In addition to the histidine residues, a large unexpected peak in the electron density was found in this cleft and was modeled as a manganese ion. No metal was added during purification, and treatment of the protein sample with EDTA was unsuccessful in removing the metal, showing that the metal is tightly bound in *Ml*HMPDdc. An EXAFS scan identified the metal as a manganese ion. The metal has a tetrahedral coordination to His92, His94, His163, and Glu73 (Figure 5.5). The

bonding distance between the manganese ion and His92 is 2.2 Å, while the distances to the other residues coordinated to the metal are longer than expected with a distance of 2.7 Å for His94, His163 and Glu73. Adjacent to the bound metal is a pocket 8 Å long and 7 Å wide, which could potentially accommodate 3-hydroxy-2-methylpyridine-4,5-dicarboxylate **6** . Hydrophobic residues that protrude into and line the cleft include Phe25, Phe28, and Tyr71 and hydrophilic residues in the pocket include His113*, Gln216*, and Arg219*.

Section 5.4. Discussion

Characterization of the mlr6791gene. The mlr6791 gene product was shown to catalyze the decarboxylation of 3-hydroxy-2-methylpyridine-4,5-dicarboxylate **6** to 3-hydroxy-2-methylpyridine-5-carboxylate **7.** The product was purified by HPLC and characterized by NMR and ESI-MS.

Comparison of HMPDdc to Other Proteins. An iterative BLAST (*21*) search using the non-redundant protein sequence database revealed a large cluster of proteins with sequence identities to HMPDdc ranging from 20 to 40%. Most of these proteins are annotated as class II aldolases/adducins or hypothetical proteins. Alignment of the top hits identified five strictly conserved residues and several highly conserved residues, as shown in Supplemental Figure 5.1 (*22-23*). Pro47 is strictly conserved and is found at the interface between monomers as the first residue of a 3_{10} helix in HMPDdc. Two of the residues coordinating to the manganese ion, Glu73 and His94, are also absolutely conserved. The two other histidine residues, His92 and His163, are mostly conserved among the proteins, with His92 being replaced by an arginine and His163 being replaced by asparagine in a class II aldolase from *Nocardioides sp. JS614*. His77 is strictly conserved, and is one of the six histidine residues found near the putative active site cleft at the interface. The final conserved residue, Gly164, is

adjacent to His163, which coordinates to the manganese ion and is the first residue of a β-strand.

A BLAST search was then conducted on a class II aldolase/adducin-like protein from *Burkholderia sp. 383*, which had the highest sequence similarity to *Ml*HMPDdc in the initial BLAST search. Most of the sequences with significant alignments were the same class II aldolases/adducins and hypothetical proteins identified using HMPDdc as a search subject; however, at a lower similarity level several 3,4-dihydroxyphthalate 2-decarboxylase (DHPdc) sequences were identified. DHPdc is found in the phthalate catabolism pathway of gram-positive bacteria and catalyzes a decarboxylation reaction of an aromatic ring very similar to the reaction catalyzed by HMPDdc (*24-25*). Alignment of these DHPdc sequences with the sequence of HMPDdc and the nearest class II aldolase identified several residues that are conserved (*23*). The sequence alignment prepared using ESPript (*22*) is shown in Supplemental Figure 5.2. All metal coordinating residues are conserved and His77 and His27, two histidine residues found near the cleft between subunits, are also conserved. The aromatic residues near the active site, Phe25, Tyr71, His113, and Phe138, are not conserved in DHPdc but are replaced by hydrophobic residues. Gln216 in HMPDdc is found to be an asparagine residue in DHPdc and Arg219 is replaced by a threonine residue.

The *Ml*HMPDdc monomer was submitted to DALI to identify structurally related proteins (*26*). Five proteins were identified as being structurally similar to HMPDdc with Z-score greater than 10; all other proteins have Z scores less than 4. Top hits included two L-fuculose 1-phosphate aldolases, one L-ribulose 5-phosphate epimerase, a rhamnulose 1-phosphate aldolase, and the C terminal fragment of phosphomethylpyrimidine kinase from *Pyrococcus furiosus* (*27-29*). The aldolases and epimerase belong to the superfamily of AraD-like proteins and class II aldolases.

L-Fuculose 1-phosphate aldolase from *E. coli* showed the highest structural similarity with a Z score of 24.4 and an r.m.s.d. of 2.4 Å for 205 of a possible 210 residues. The sequence identity between HMPDdc and the aldolase is about 20%. The Z score for *E. coli* L-ribulose 5-phosphate epimerase was 22.5 with an r.m.s.d. of 2.4 Å for 207 of 223 residues.

The structures of the L-fuculose 1-phosphate aldolases, L-ribulose 5-phosphate epimerase, and the rhamnulose 1-phosphate aldolase superimpose well on the structure of HMPDdc, as seen in Figure 5.6. The topology of the long β-sheet is conserved for each of the aldolases and HMPDdc; however, HMPDdc has a C-terminal α-helix not observed in the other structures. Rhamnulose 1-phosphate aldolase is the least similar to HMPDdc and has two extra β-strands and four extra α-helices. All of the enzymes shown in Figure 5.6 adopt a tetrameric oligomeric state.

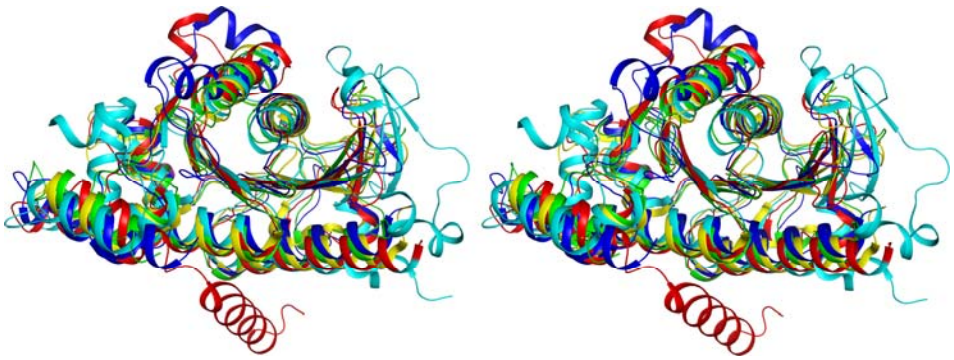


Figure 5.6. Stereoview diagram of the superposition of the top four DALI hits on HMPDdc. HMPDdc is shown in red, 1-ribulose-5-phosphate-4-epimerase is shown is blue, rhamnulose-1-phosphate aldolase is in cyan, 1-fuculose-1-phosphate aldolase from *E. coli* is green, and 1-fuculose-1-phosphate aldolase from *B. thetaiotaomicron* is colored yellow.

Despite the strong structural similarity between HMPDdc and the class II aldolase family, there is low sequence conservation and apart from the coordination of the metal ion few of the putative active site residues are conserved. The class II aldolase superfamily members each bind a zinc ion in the active site, while HMPDdc has a manganese ion bound. L-fuculose 1-phosphate aldolase from *Bacterosides*

thetaiotaomicron has the highest sequence identity at 20% and the sequence identity between HMPDdc and rhamnulose 1-phosphate aldolase from *E. coli* is only 15%. The zinc ion is found in the same cleft and coordinated by three conserved histidine residues. The fourth residue coordinated to the metal ion is a glutamate residue in HMPDdc and the three aldolases, and an aspartate residue in the epimerase. Several glycine and proline residues are conserved, most likely playing roles in positioning structural elements. His77 is conserved, as is Ser29, which is hydrogen bonded to His 77. Arg31, Arg84, Pro85, and Asp86 are all structurally conserved and are found in turns directly exposed to solvent. Glu183 is conserved and is found within α4 and faces the channel that runs through the tetramer.

Active Site Comparison. The structure of L-fuculose 1-phosphate aldolase from *E. coli* with an inhibitor bound (PDB ID: 4FUA) was used to compare the active sites of the class II aldolase and HMPDdc (Figure 5.7) (*27*). The zinc ion is located closer to the coordinating residues in the aldolase structure, only 2.0 Å from His94 and His 155 and 2.1 Å from His 92. The corresponding distances in HMPDdc to the manganese ion are 2.7 Å for His 94 and His163 and 2.2 Å for His92. With the inhibitor phosphoglycolohydroxamate bound Glu73 has been pushed out of position

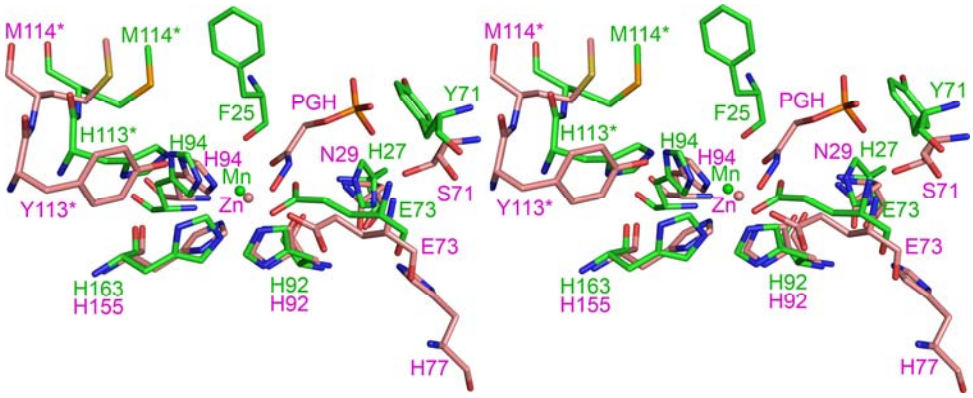


Figure 5.7. Stereoview diagram of the active sites of HMPDdc and 1-fuculose1-phosphate aldolase. HMPDdc is colored in green and 1-fuculose1-phosphate aldolase is colored in pink. The phosphoglycolohydroxamate ligand is abbreviated as PGH.

for coordination to the zinc ion, suggesting the same could occur upon binding of 3-hydroxy-2-methylpyridine-4,5-dicarboxylate **6** . His113* is replaced by Tyr113*, which points toward the zinc ion. The adjacent residue, Met114*, is found in both structures. His27 is replaced by Asn29. Phe138 is conserved, although adopting a different side chain orientation. Gly28 still composes the floor of the binding pocket and the pocket is roughly the same size in both structures. No other residues are conserved within the binding pocket, unsurprising given that the substrates are very different.

Structural Implications for Mechanism. Enzyme-catalyzed decarboxylations constitute a well-studied family of reactions and the role of pyridoxal phosphate, the pyruvoyl cofactor, imine formation with lysine and metal ions in the catalysis of these reactions is now well established. The general rule for the catalysis of such decarboxylations is that the enzyme provides a mechanism for the stabilization of an enolate intermediate by charge delocalization such as shown in Figure 5.8A (*30*). The decarboxylation of orotidine monophopshate that occurs during pyrimidine biosynthesis is a well-known exception (*31*). The decarboxylation of hydroxy-substituted benzene rings is a common motif in polyketide biosynthesis. It is generally assumed that this reaction proceeds by initial tautomerization of the phenol, followed by decarboxylation of the resulting keto-acid. However, this proposal has not been experimentally validated and only one structure of a hydroxybenzoic acid decarboxylase has been reported (PDB ID: 2DVX).

The Protein Data Bank contains three zinc-dependent decarboxylases: 2,6-dihydroxybenzoate decarboxylase (PDB ID: 2DVX), α-amino-β-carboxymuconate-ε-semialdehyde decarboxylase (PDB ID: 2HBV) and a protein of unknown function similar to α-acetolactate decarboxylase (PDB ID: 1XV2) (*32*); however, none of these are structurally similar to HMPDdc. In contrast several members of the class II

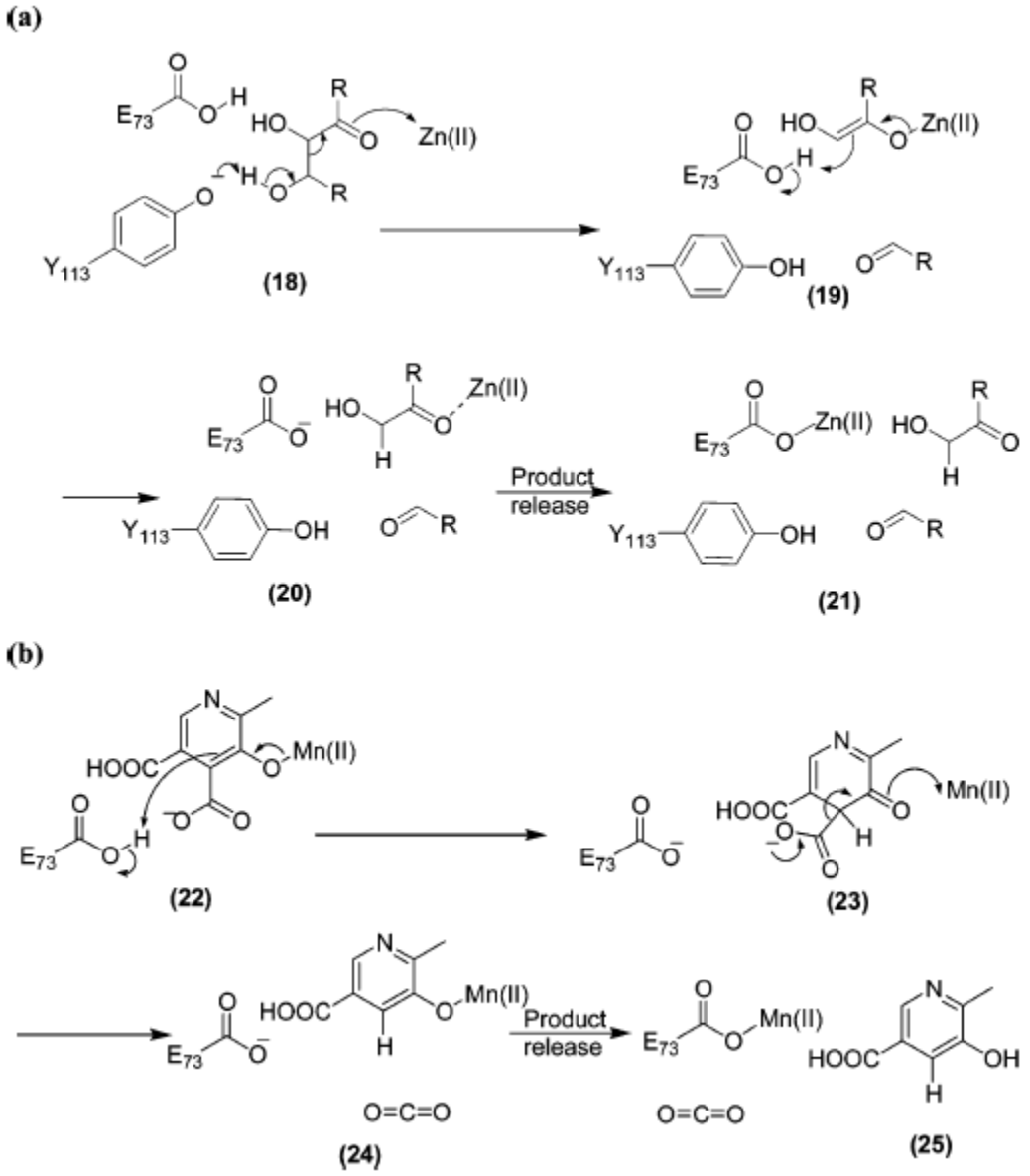
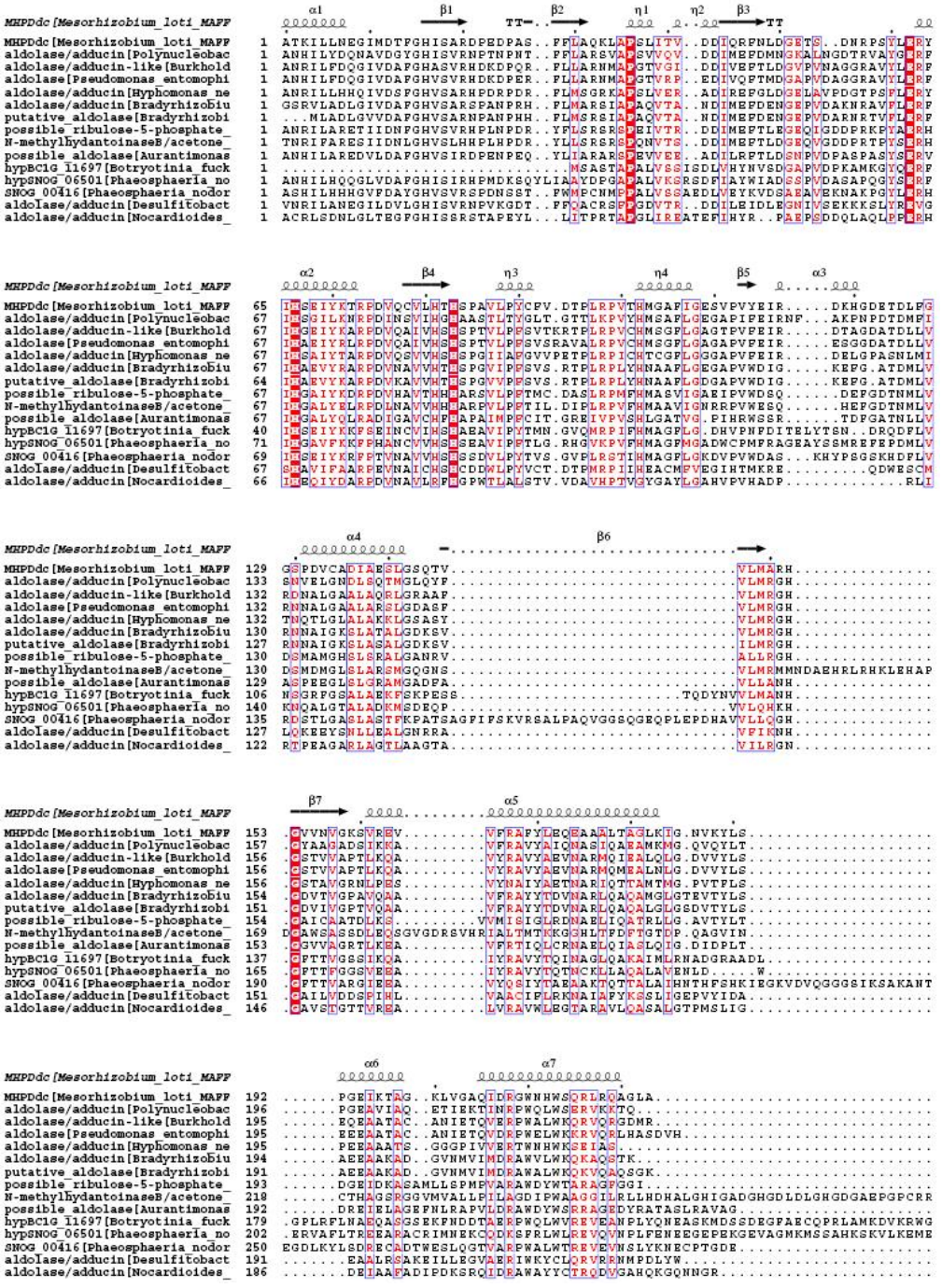


Figure 5.8. (A) The retroaldol condensation reaction catalyzed by fuculose aldolase. (B) The proposed mechanism for the decarboxylation of 3-hydroxy-2-methylpyridine-4,5-dicarboxylate **6**.

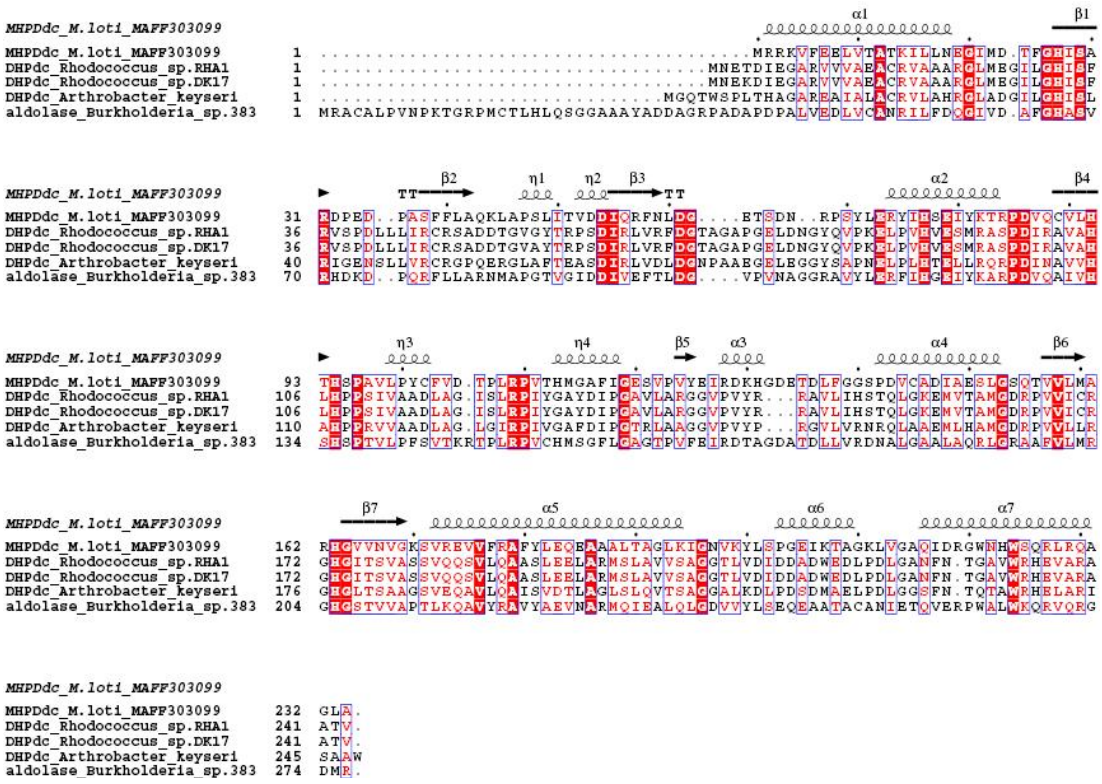
aldolase/adducin family were revealed by a DALI search. The structural similarity between fuculose aldolase and HMPDdc suggests that the two reactions share common mechanistic features (*33*). For the aldolase catalyzed reaction, Tyr113 initiates the retroaldol reaction by alcohol deprotonation, Figure 5.8A. The resulting enolate is stabilized by the active site zinc ion. Protonation of this enolate by Glu73

followed by product release completes the reaction. In the resting state of the enzyme, Glu73 is coordinated to the zinc ion and is released from the metal upon substrate binding. Based on this proposal, we suggest an analogous mechanism for HMPDdc. In this mechanism, binding of the substrate displaces Glu73 from the manganese ion replacing it with the substrate hydroxyl. Glu73 then provides the proton for the keto-enol tautomerization to give **23** (Figure 5.8B). The decarboxylation reaction is analogous to the retroaldol reaction except that is does not need a base as the carboxylate is likely to be deprotonated under the reaction conditions. Displacement of the product from the manganese ion by Glu73 followed by product dissociation completes the reaction. The testing of this mechanistic proposal is in progress and will require additional structural and mechanistic studies.

APPENDIX



Supplemental Figure 5.1. Sequence alignment of proteins identified through BLAST to be similar to *Ml*HMPDdc. Strictly conserved residues are highlighted in read while less strictly conserved residues are shown in red font. The secondary structure for HMPDdc is shown above the alignment.



Supplemental Figure 5.2. Sequence alignment of decarboxylase proteins identified through BLAST with HMPDdc and the aldolase submitted to BLAST. Strictly conserved residues are highlighted in red and less strictly conserved residues are shown in red font. The secondary structure of HMPDdc is shown above the alignment.

REFERENCES

1. Rodwell, V. W., Volcani, B. E., Ikawa, M., and Snell, E. E. (1958) Bacterial oxidation of vitamin B6. I. Isopyridoxal and 5-pyridoxic acid, *J Biol Chem 233*, 1548-1554.
2. Snell, E. E., and Haskell, B. E. (1971) *The Metabolism of Vitamin B6, In: Comprehensive Biochemistry*, Vol. 21, Elsevier/North Holland, New York.
3. Yagi, T., Kishore, G. M., and Snell, E. E. (1983) The bacterial oxidation of vitamin B6. 4-Pyridoxic acid dehydrogenase: a membrane-bound enzyme from Pseudomonas MA-1, *J Biol Chem 258*, 9419-9425.
4. Yuan, B., Yoshikane, Y., Yokochi, N., Ohnishi, K., and Yagi, T. (2004) The nitrogen-fixing symbiotic bacterium Mesorhizobium loti has and expresses the gene encoding pyridoxine 4-oxidase involved in the degradation of vitamin B6, *FEMS Microbiol Lett 234*, 225-230.
5. Funami, J., Yoshikane, Y., Kobayashi, H., Yokochi, N., Yuan, B., Iwasaki, K., Ohnishi, K., and Yagi, T. (2005) 4-Pyridoxolactonase from a symbiotic nitrogen-fixing bacterium Mesorhizobium loti: cloning, expression, and characterization, *Biochim Biophys Acta 1753*, 234-239.
6. Yokochi, N., Nishimura, S., Yoshikane, Y., Ohnishi, K., and Yagi, T. (2006) Identification of a new tetrameric pyridoxal 4-dehydrogenase as the second enzyme in the degradation pathway for pyridoxine in a nitrogen-fixing symbiotic bacterium, Mesorhizobium loti, *Arch Biochem Biophys 452*, 1-8.
7. Yuan, B., Yokochi, N., Yoshikane, Y., Ohnishi, K., and Yagi, T. (2006) Molecular cloning, identification and characterization of 2-methyl-3-hydroxypyridine-5-carboxylic-acid-dioxygenase-coding gene from the nitrogen-fixing symbiotic bacterium Mesorhizobium loti, *J Biosci Bioeng 102*, 504-510.

8. Mukherjee, T., Kinsland, C., and Begley, T. P., . (2007) PLP catabolism: Identification of the 4-Pyridoxic acid Dehydrogenase gene in Mesorhizobium loti MAFF303099, *Biorganic Chemistry in Press*.
9. Schuck, P. (2003) On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation, *Anal. Biochem. 320*, 104-124.
10. Ausubel, F. M., and Brent, F., (Eds.) (1987) *Current Protocols in Molecular Biology*, John Wiley and Sons, New York.
11. Sambrook, J., Fritsch, G. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Guide*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
12. Matthews, B. W. (1968) Solvent content of protein crystals, *J. Mol. Biol. 33*, 491-497.
13. Otwinowski, Z., and Minor, W. (1997) Processing of x-ray diffraction data collected in oscillation mode, *Methods Enzymol. 276*, 307-326.
14. Pape, T., and Schneider, T. R. (2004) HKL2MAP: a graphical user interface for phasing with SHELX programs, *J Appl Cryst 37*, 843-844.
15. Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2006) Automated Structure Solution With autoSHARP, *Methods Mol Biol 364*, 215-230.
16. Cowtan, P. E. a. K. (2004) Coot: Model-Building Tools for Molecular Graphics, *Acta Crystallogr. D 60*, 2126-2132.
17. 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.

18. 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.

19. DeLano, W. L. (2002) The PyMOL Molecular Graphics Systems, DeLano Scientific, San Carlos, CA.

20. Hasel, W., Hendrickson, T. F., and Still, W. C. (1988) A rapid approximation to the solvent-accessible surface areas of atoms, *Tetrahedron Comp. Meth. 1*, 103-116.

21. 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.

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

23. 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.

24. Habe, H., Miyakoshi, M., Chung, J., Kasuga, K., Yoshida, T., Nojiri, H., and Omori, T. (2003) Phthalate catabolic gene cluster is linked to the angular dioxygenase gene in Terrabacter sp. strain DBF63, *Appl Microbiol Biotechnol 61*, 44-54.

25. Eaton, R. W. (2001) Plasmid-encoded phthalate catabolic pathway in Arthrobacter keyseri 12B, *J Bacteriol 183*, 3689-3703.

26. Holm, L., and Sander, C. (1998) Touring protein fold space with Dali/FSSP, *Nucleic Acids Res. 26*, 316-319.

27. Dreyer, M. K., and Schulz, G. E. (1996) Catalytic mechanism of the metal-dependent fuculose aldolase from Escherichia coli as derived from the structure, *J Mol Biol 259*, 458-466.

28. Luo, Y., Samuel, J., Mosimann, S. C., Lee, J. E., Tanner, M. E., and Strynadka, N. C. (2001) The structure of L-ribulose-5-phosphate 4-epimerase: an aldolase-like platform for epimerization, *Biochemistry 40*, 14763-14771.

29. Kroemer, M., and Schulz, G. E. (2002) The structure of L-rhamnulose-1-phosphate aldolase (class II) solved by low-resolution SIR phasing and 20-fold NCS averaging, *Acta Crystallogr D Biol Crystallogr 58*, 824-832.

30. Begley, T. P., and Ealick, S. E. (2004) Enzymatic reactions involving novel mechanisms of carbanion stabilization, *Curr Opin Chem Biol 8*, 508-515.

31. Miller, B. G., and Wolfenden, R. (2002) Catalytic proficiency: the unusual case of OMP decarboxylase, *Annu Rev Biochem 71*, 847-885.

32. 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.

33. Samuel, J., Luo, Y., Morgan, P. M., Strynadka, N. C., and Tanner, M. E. (2001) Catalysis and binding in L-ribulose-5-phosphate 4-epimerase: a comparison with L-fuculose-1-phosphate aldolase, *Biochemistry 40*, 14772-14780.

CHAPTER 6

STRUCTURE OF THE PLP DEGRADATIVE ENZYME 2-METHYL-3-HYDROXYPYRIDINE-5-CARBOXYLIC ACID OXYGENASE FROM *MESORHIZOBIUM LOTI* MAFF303099 AND ITS MECHANISTIC IMPLICATIONS[1]

Section 6.1. Introduction

Two pathways for the degradation of pyridoxal-phosphate vitamers (vitamin B_6), have been identified in soil microbes able to grow on vitamin B_6 as the sole source of carbon and nitrogen (*1*). The first pathway, which converts vitamin B_6 to succinic semialdehyde, occurs in *Pseudomonas sp.* MA-1 (Figure 6.1), while the second pathway has been identified in *Pseudomonas* IA and *Arthrobacter* Cr-7 and converts pyridoxine to 2-(hydroxymethyl)-4-oxobutanoate. The enzymes and intermediates for both pathways have been characterized (*2-4*). However, the corresponding genes remained unidentified until the recent discovery that the first pathway is also present in the symbiotic bacterium *Mesorhizobium loti* MAFF303099 (*5*). All the *M. loti* PLP catabolic genes have now been identified and the corresponding catabolic reactions have been characterized (*6-12*). One of the later enzymes on this pathway, 2-methyl-3-hydroxypyridine-5-carboxylic acid oxygenase (MHPCO), catalyzes a particularly interesting reaction in which 2-methyl-3-hydroxypyridine-5-carboxylic acid (MHPC) undergoes an oxidative ring opening to generate *E*-2-acetamidomethylene succinate, the first acyclic intermediate on the pathway (**7** to **8** in Figure 6.1).

MHPCO (E.C. 1.14.12.4) has been classified as a flavin-dependent monooxygenase based on its sequence and biochemical similarity to other enzymes in

[1] Reproduced with permission from McCulloch, K.M., Mukherjee, T., Begley, T.P., and Ealick, S.E. (2009) *Biochemistry* 48: 4139-4149. Copyright 2009 American Chemical Society

this family, which includes the well-characterized *p*-hydroxybenzoate hydroxylase (PHBH) (*13, 14*). This class of enzymes utilize FAD and molecular oxygen to form a C(4a)-hydroperoxy-FAD intermediate which then donates one oxygen atom to the substrate (*15*). The second reaction catalyzed by MHPCO, the pyridine ring-opening reaction, is an unusual reaction for an enzyme of the monooxygenase family because most oxidative ring-opening reactions of aromatic compounds are catalyzed by iron-dependent dioxygenases. MHPCO was confirmed as a flavin-dependent monooxygenase, incorporating one oxygen atom from molecular oxygen and one oxygen atom from water to produce *E*-2-(acetamidomethylene)succinate (*E*-2AMS) the ring-opened product (*16*). Previous work has shown the *Pseudomonas* MA-1 MHPCO to crystallize and diffract to reasonable resolution, although the structure was

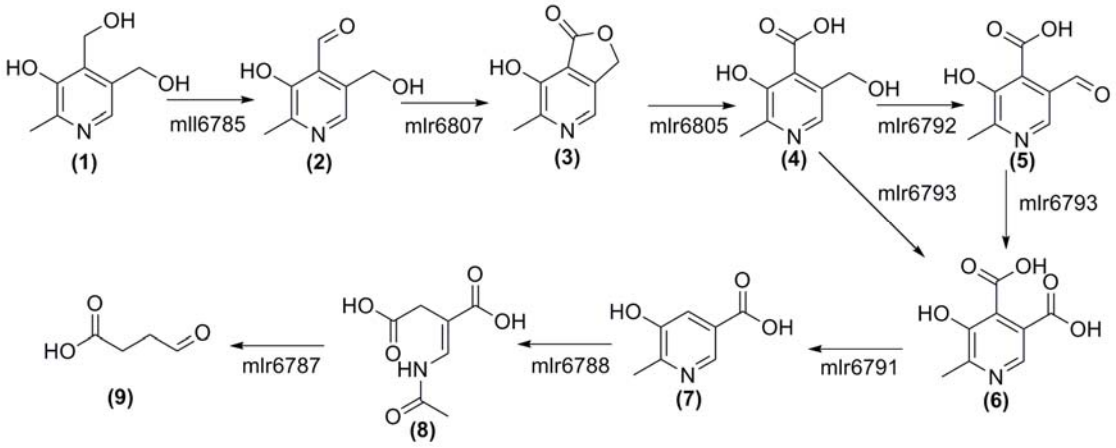


Figure 6.1. The vitamin B_6 degradative pathway found in *M. loti*. The genes are named as using the RhizoBase numbering system. **1**, pyridoxine; **2**, pyridoxal; **3**, 4-pyridoxolactone; **4**, 4-pyridoxic acid; **5**, 5-formyl-2-methyl-3-hydroxypyridine-4-carboxylate; **6**, 2-methyl-3-hydroxypyridine-4,5-dicarboxylate; **7**, 2-methyl-3-hydroxypyridine-5-carboxylate; **8**, 2-(acetamidomethylene)succinate; **9**, succinic semialdehyde

not reported (*17*). We have determined the 2.1 Å resolution crystal structures of *M. loti* MHPCO and its complex with MHPC. These structures confirm the relationship of MHPCO to other members of the flavin monooxygenase family and provide insight into the catalytic mechanism.

Section 6.2. Materials and Methods

Materials. M9 minimal salts and $CaCl_2$ were purchased from Sigma (St. Louis, MO). Ni-NTA resin came from Qiagen (Valencia, CA) and Acros Chemicals (Morris Plains, NJ) supplied the imidazole and L-selenomethionine. Dextrose, $FeSO_4$, $MgSO_4$, NaCl, Tris buffer, β-mercaptoethanol, chloramphenicol, ampicillin, and all natural amino acids were purchased from Fisher Scientific (Fairlawn, NJ). 100X MEM vitamin mix were from Invitrogen (Carlsbad, CA), as was the Gateway system used for DNA manipulations.

Cloning of M. loti MHPCO. Standard methods were used for DNA manipulations (*18, 19*). The *M. loti* MHPCO gene was amplified from genomic DNA (purified from ATCC strain # 700743) by PCR with the following primer pair: 5'-CAC CAT GGC CAA TGT AAA CAA AAC TCC-3' and 5'-CTA CTG CGG CCA CGA GTA GAC ACG GCG CAG C-3'. The PCR product was purified and used in a topoisomerase-mediated reaction with pENTR-TEV-D-TOPO (Invitrogen) following the manufacturer's instructions. Clones were screened by PCR and verified by sequencing. A correct clone was used in an LR recombination reaction with the plasmid pDESTF1, a Gateway-adapted vector based on the pET-system from Novagen that encodes an N-terminal 6xHis tag under the control of the T7*lac* promoter. Clones were again screened by restriction digestion. A correct clone was named pMl5332.XF1.

Overexpression and Purification of Selenomethionyl (SeMet) MHPCO. E. coli B834(DE3) was transformed first using the chaperone-encoding plasmid pGro7 (Takara Bio Inc.) and then using pMl5332.XF1. Cells were grown with shaking at 37 °C in minimal media supplemented with 20 mg/L of all amino acids except methionine, 50 mg/L L-selenomethionine, 1X MEM vitamin mix, 0.4% glucose, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 25 mg/L $FeSO_4$, 35 mg/L chloramphenicol, and 100 mg/L

ampicillin. Chaperone proteins GroES and GroEL were induced at cell culture inoculation by addition of 2.5 g/L L-arabinose. When cells reached an OD_{600} of 0.7 the temperature was lowered to 15 °C and overexpression of MHPCO was induced with 0.5 mM isopropyl β-D-thiogalactopyranoside. After overnight growth, cells were pelleted by centrifugation at 4 °C for 15 min at 6,000 *g* and then stored at -20 °C until purification. Cells were lysed by sonication and the cell lysate was cleared by centrifugation at 40,000 *g* for 45 min at 4 °C. SeMet MHPCO was purified from the clarified cell lysate by nickel affinity chromatography. The clarified lysate was twice passed over the column equilibrated with Buffer A (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, and 3 mM β-mercaptoethanol (β-ME)) and 10 mM imidazole and the column was washed with this buffer for 20 column volumes. The column was then washed with Buffer A containing 35 mM imidazole for 15 column volumes. Buffer A and 250 mM imidazole were used to elute SeMet MHPCO from the nickel column and the protein was 80% pure by SDS-PAGE analysis (results not shown). Size exclusion chromatography (HiLoad 26/60 Superdex 200 pg, GE Healthcare) was used to further purify SeMet MHPCO to >95% homogeneity. One liter of culture produced 0.5 mg pure SeMet MHPCO. The protein was then buffer exchanged into 20 mM Tris (pH 7.7), 50 mM NaCl, and 1 mM DTT and concentrated to 8 mg/mL.

Enzymatic Synthesis of the Substrate. The enzymatic synthesis and purification of 2-methyl-3-hydroxypyridine-4,5-dicarboxylic acid (MHPD) has been previously described (*10*). MHPC was enzymatically synthesized from MHPD in a 10 mL reaction mixture containing 5 mM MHPD and 100 μM of MHPD decarboxylase in 100 mM sodium phosphate buffer (pH 8.0) incubated over night at room temperature. MHPC was purified by the HPLC (*10*) and upon lyophilization yielded a stable white solid. ^{1}H NMR (300 MHz, D_2O) δ 2.46 (s, 3H, CH_3), 7.68 (s, 1H, C4-H),

and 8.12 (s, 1H, C6-H). The purified MHPC was dissolved in 100 mM Tris (pH 7.7) to a final concentration of 100 mM and stored frozen at -20 °C until use.

Activity Assay for MHPCO. The recombinant His-tagged MHPCO was assayed by monitoring the disappearance of NADH absorbance at 340 nm as previously described (*12*).

Analytical Ultracentrifugation. Samples of MHPCO were prepared as described above and analyzed by analytical ultracentrifugation to determine the oligomeric state of the enzyme. To obtain MHPCO with cleaved His tag, the sample was dialyzed overnight into 50 mM Tris (pH 8.0), 150 mM NaCl, and 2 mM DTT at 4 °C. It was then incubated with TEV protease at 4 °C for 20 hours and then passed over a Ni-NTA column to remove uncleaved MHPCO and the TEV protease. The cleavage reaction was monitored by SDS-PAGE analysis. The protein samples, both with and without the His tag, were extensively dialyzed into 20 mM Tris (pH 8.0) 150 mM NaCl and then subjected to sedimentation velocity analytical ultracentrifugation on a ProteomeLab XL-I protein characterization system (Beckman Coulter). Data were analyzed using the programs SEDNTERP and DCDT+ (*20*).

Protein Crystallization. Initial crystallization conditions were identified using the hanging drop vapor diffusion method (Crystal Screens 1 and 2, Hampton Research; Wizard Screens 1 and 2, Emerald Biosystems) at room temperature. Hanging drops were set up by mixing 1 μL of reservoir solution with 1 μL of protein sample. Optimized crystallization conditions were 6-9% PEG8000 and 100 mM Tris (pH 8.1-8.5). Yellow plate-like crystals grew to 100-200 μm × 50-100 μm in five days. To obtain crystals complexed with the substrate MHPC, SeMet MHPCO crystals were transferred to a drop consisting of the reservoir solution and 10 mM MHPC then allowed to soak for 3-5 h. This crystallization condition, supplemented with 17% glycerol, was used as cryoprotectant and for crystals soaked with MHPC, 10

mM MHPC was also present in the cryoprotectant. After briefly soaking the crystals in the cryoprotectant, they were flash frozen by plunging in liquid nitrogen. Both SeMet MHPCO and the substrate complex crystals belong to the space group C2 with unit cell parameters a = 111.3 Å, b = 130.2 Å, c = 89.5 Å, and β = 122.6°, for unliganded MHPCO. The asymmetric unit contains two monomers with a Matthew's number of 3.28 $Å^3$/Da and solvent content of 60% (*21*).

X-Ray Data Collection and Processing. For structure determination, data were collected on a single SeMet MHPCO crystal at the Advanced Photon Source beamline 24-ID-E (Argonne National Laboratory) using a Quantum315 detector (Area Detector Systems Corporation). To maximize the anomalous signal, a single wavelength anomalous diffraction experiment was conducted at the peak *f″* for selenium. Data were collected to 2.1 Å resolution using a 1° oscillation range for 330 frames. Data collection on a crystal containing SeMet MHPCO complexed with MHPC was done on the 24-ID-C beamline using a Quantum315 detector. Data were collected at 0.9795 Å over 180° using a 1° oscillation range. Datasets were indexed, integrated, and scaled using the HKL2000 suite of programs (*22*). The data collection statistics are summarized in Table 6.1.

Structure Determination, Model Building, and Refinement. Se atom positions were located using data cut off at 2.5 Å resolution and the program hkl2map; 18 of a possible 20 Se atoms were located (*23-25*). These heavy atom positions were then used in CNS for phasing, density modification, and calculation of electron density maps (*26*). Further improvement to the maps and automated model building were performed using RESOLVE, which succeeded in modeling 377 residues with correct side chains and 196 residues as alanine out of 758 total residues in the asymmetric unit (*27*). Iterative rounds of model building in COOT and refinement in CNS successfully positioned residues 11-379 for both chains (*28*). An FAD molecule was then modeled

into each chain, followed by addition of water molecules using CNS. Lastly, for the complex MHPC was modeled into MHPCO before a final round of refinement. The refinement statistics are summarized in Table 6.2. The geometry of MHPCO was verified using PROCHECK which identified one residue, Ala51, in each chain in the disallowed region (*29*). Figures were generated using PYMOL and ChemBioDraw (*30*).

Table 6.1. Summary of Data Collection Statistics

	SeMet MHPCO	MHPCO + MHPC
Beamline	APS NE-CAT 24-IDE	APS NE-CAT 24-IDC
Resolution (Å)	2.10	2.10
Wavelength (Å)	0.97918	0.97949
Space Group	*C*2	*C*2
a (Å)	111.3	110.4
b (Å)	130.2	129.5
c (Å)	89.5	89.1
β (°)	122.6	122.6
Reflections	408805	208350
Unique reflections[a]	121606 (11380) [b]	58150 (4574)
Average I/σ	13.9 (3.2)	12.6 (3.5)
Redundancy	3.4 (2.5)	3.6 (3.0)
Completeness (%)	98.6 (91.8)	94.4 (74.2)
Rsym[c] (%)	12.0 (33.5)	9.4 (21.3)

[a]Unique reflections include Bijvoet pairs.
[b]Values in parentheses are for the highest resolution shell.
[c]$R_{sym} = \Sigma\Sigma_i \mid I_i - \langle I \rangle \mid / \Sigma \langle I \rangle$, where <I> is the mean intensity of the N reflections with intensities I_i and common indices *h,k,l*.

Section 6.3. Results

MHPCO Activity. The recombinant His-tagged MHPCO was purified using metal ion chromatography and was found to be active using the assay previously described (*12*). The activity of recombinant MHPCO with the His tag cleaved by digestion with TEV protease was comparable to the activity of the His-tagged MHPCO sample.

Structure of the MHPCO monomer. The structure of MHPCO was determined at 2.1 Å resolution using SAD phasing. The final model of the MHPCO monomer is

Table 6.2. Summary of Data Refinement Statistics

	SeMet MHPCO	MHPCO + MHPC
Resolution (Å)	50.00 – 2.10	50.00 – 2.10
# of protein atoms	5730	5730
# of ligand atoms	108	130
# of water atoms	744	483
Reflections in working set	59384	52927
Reflections in test set	3175	2867
Rfactor[a] (%)	19.3	20.1
Rfree[b] (%)	22.4	22.9
Rmsd from ideals		
Bonds (Å)	0.0055	0.0061
Angles (º)	1.3	1.3
Avg B factor (Å^2)	19.4	22.7
Ramachandran Plot		
Most favored (%)	90.0	90.3
Additionally allowed (%)	9.7	9.1
Generously allowed (%)	0.0	0.3
Disallowed (%)	0.3	0.3

[a] R factor = $\sum_{hkl} \left| \left| F_{obs} \right| - k \left| F_{cal} \right| \right| / \sum_{hkl} \left| F_{obs} \right|$ 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.

shown in Figure 6.2 and contains residues 11-379 of the protein and one molecule of FAD. The final model of the ternary substrate complex also contains one molecule of MHPC. MHPCO shows strong structural similarity to other members of the flavin-dependent hydroxylase family. MHPCO comprises three domains exhibiting mixed α/β folds. The N-terminal domain has a five-stranded parallel β-sheet (β7↑β2↑β1↑β11↑β18↑) flanked on one side by a small, three stranded antiparallel β-sheet (β10↑β9↓β8↑) and seven α-helices arranged in two layers on the opposite side. This domain forms most of the interactions with the FAD molecule. The second domain consists of a six-stranded mostly antiparallel β-sheet with a strand order of β5↓β4↑β14↑β15↓β16↑β13↓ that is flanked on the solvent exposed side by three short α-helices and forms part of the substrate binding pocket on the other face. The third domain connects the first two domains and consists of α11 stretching the length

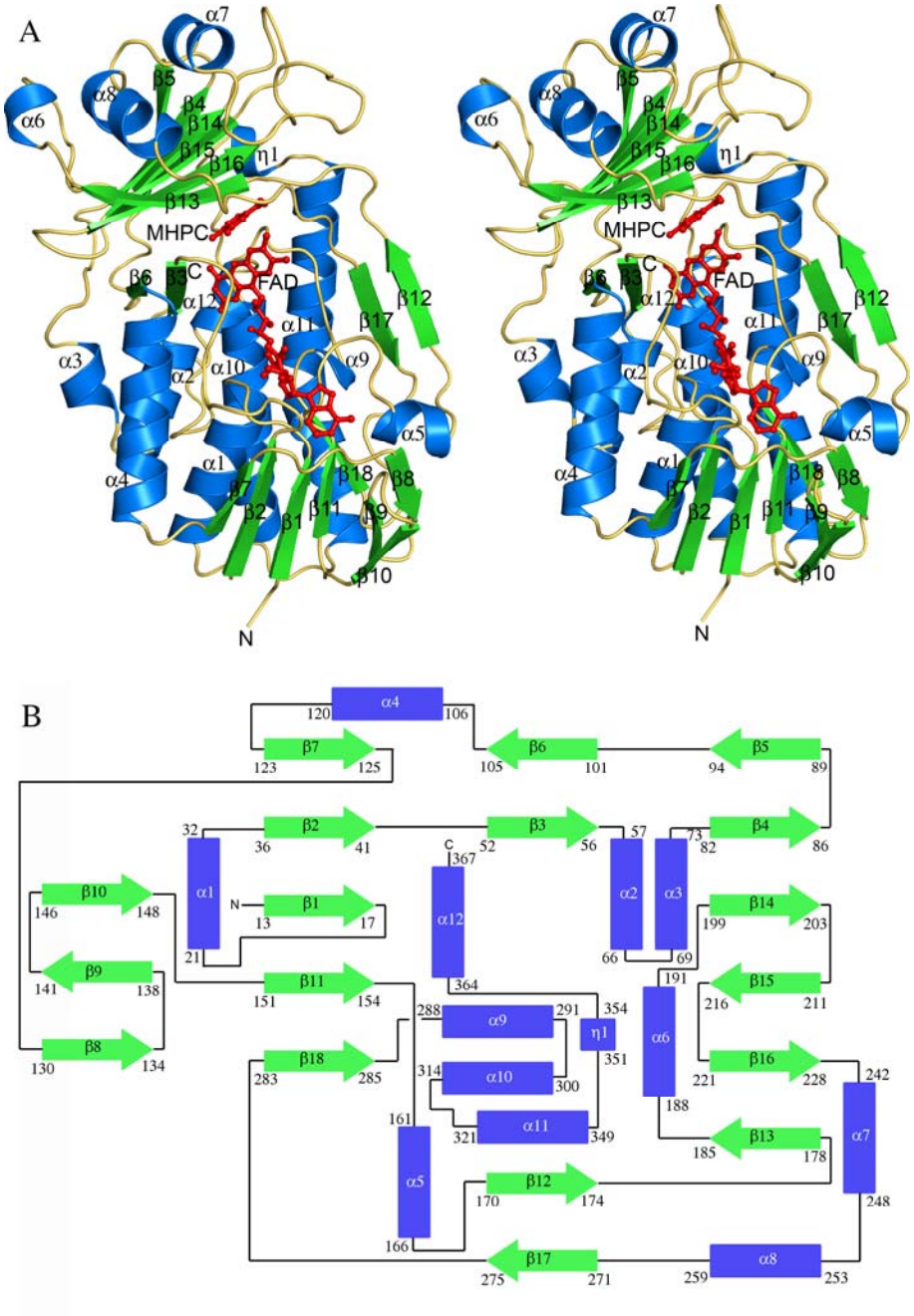


Figure 6.2. The monomeric structure of MHPCO. (A) Stereoview ribbon diagram of MHPCO with FAD and MHPC bound and secondary structural elements labeled. The α-helices are shown in blue and the β-strands are shown in green with loops colored yellow. The bound ligands, FAD and MHPC, are shown in stick representation and colored red. (B) Topology diagram of MHPCO with the same color scheme as 6.2A.

of MHPCO and the two stranded antiparallel β-sheet β3↓β6↑. Domain 3 is largely solvent exposed on the back side of MHPCO.

Oligomeric Structure. Two protomers are found within the asymmetric unit lying antiparallel to each other and the dimer spans approximately 75 Å, (Figure 6.3A). The two active sites are independent of each other, but face in the same direction and are located more than 40 Å apart. The total surface area for the dimer of MHPCO is 28750 $Å^2$ and the total surface area buried is 1375 $Å^2$ at the dimeric interface (*31*). The primary interactions occur through β12 to form an extended β-sheet consisting of β17↑β12↓β12*↑β17*↓ where strands marked with an asterisk are located in the opposing protomer. β12 is five residues long and consists of mostly hydrophilic residues, stretching from Glu170 to Val174. Five hydrogen bonds are formed between carbonyl oxygen atoms and nitrogen atoms from the protein backbone, stabilizing the dimer formation and ranging in length from 2.65 Å and 3.1 Å. A nitrogen atom of Arg172 is involved in an additional hydrogen bond to the carbonyl oxygen atom of the side chain of Asn170*. One salt bridge is also present at this interface, between Arg172 and Asp164*. These interactions are shown in Figure 6.3B. The oligomeric state of MHPCO was further investigated using analytical ultracentrifugation. To determine if the N-terminal 6X-His tag was responsible for dimer formation over tetramer formation, sedimentation velocity analytical ultracentrifugation experiments were performed on MHPCO samples with and without the His tag (*32*). These experiments showed that both sets of samples formed tetramers in solution with a molecular weight of roughly 160 kDa, as shown in Figures 6.3C and 6.3D.

FAD Binding Site. The FAD molecule binds within a long cleft roughly 30 Å long and 15 Å wide. FAD is tightly bound in a non-covalent fashion and copurified with the enzyme. Each protomer within the asymmetric unit contains one bound FAD

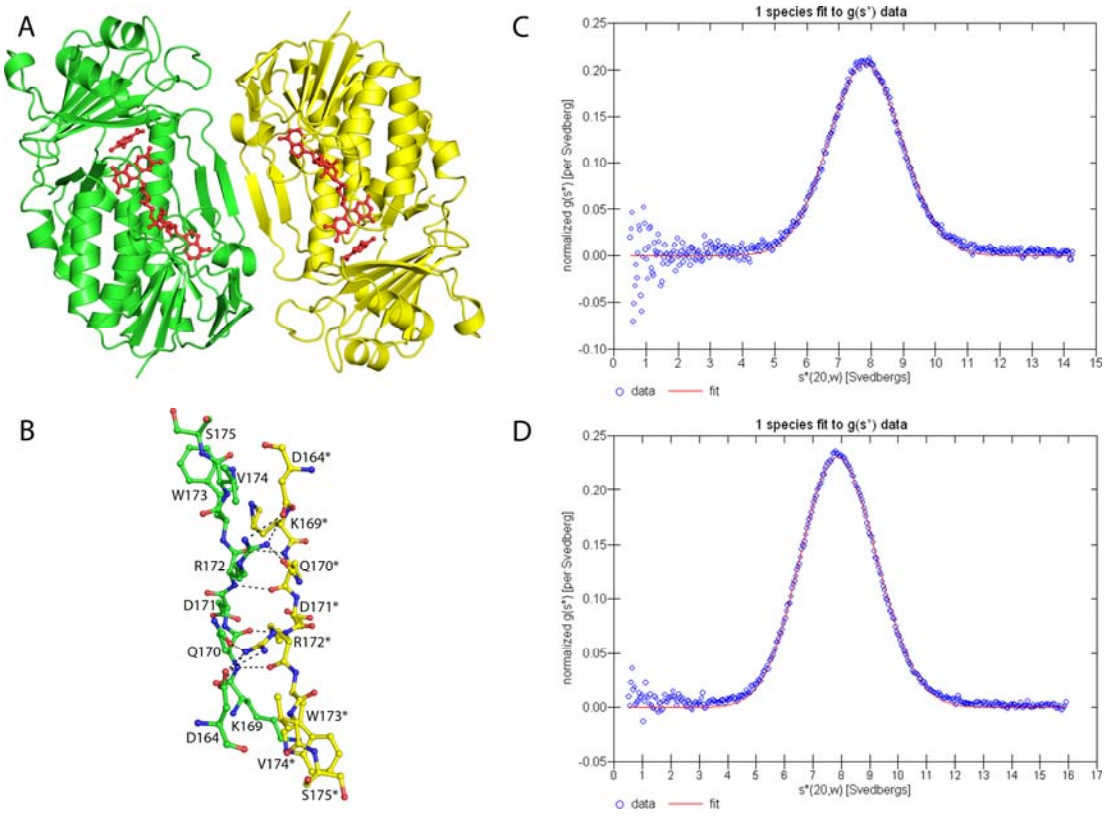


Figure 6.3. The dimeric structure of MHPCO. (A) Ribbon diagram of the MHPCO dimer with FAD and MHPC bound and shown in red. The first protomer is shown in green and the opposing protomer is colored yellow. Hydrogen bonds and salt bridges are shown using dashed lines. (B) Ball and stick representation of the dimeric interface at β12β12*. The color scheme is the same as Figure 6.4A. (C) Overlay of sedimentation velocity analytical ultracentrifugation data with the fitted curve for a single species. The data points are shown as open blue circles and the best fit line is shown in red. The overlay for MHPCO with the N-terminal His tag attached. (D) The overlay for MHPCO with the N-terminal His tag cleaved. The fits were prepared using the program DCDT+.

molecule (Figure 6.4). The adenine base and ribose sugar are partially buried and surrounded by the hydrophilic residues Glu41, Lys42, Glu129, Asp156 and Lys161. N1 of the adenine base forms a hydrogen bond to the amide nitrogen of Ala130 and both N6 and N7 form water-mediated hydrogen bonds to Lys161. The ribose sugar is stabilized through water-mediated hydrogen bonds to the backbone carbonyl oxygen atom of Arg47. The 2′-hydroxyl group also forms a strong hydrogen bond (2.5 Å) with Glu41. The α-phosphate group is largely solvent exposed, while the β-phosphate

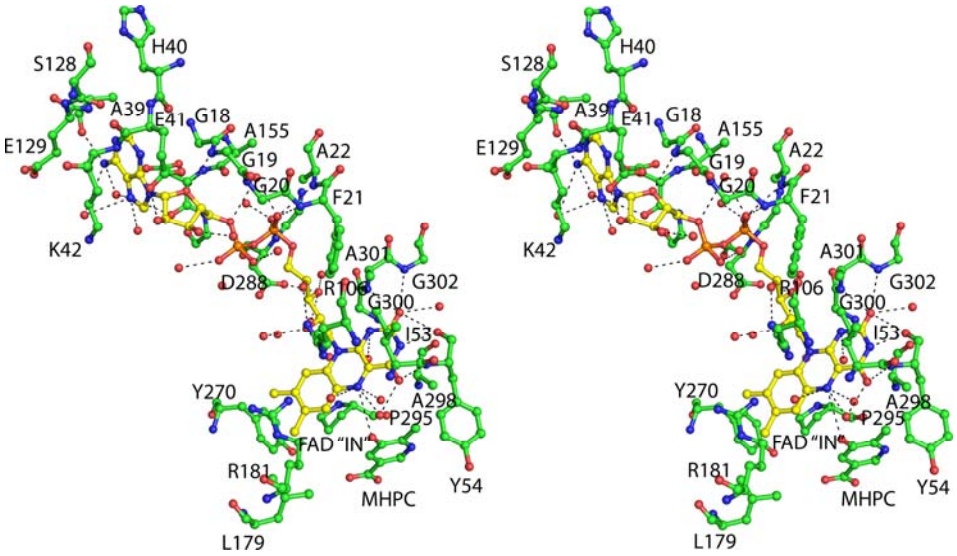


Figure 6.4. Stereoview diagram of FAD conformation found in MHPCO both with and without MHPC bound. Residues from MHPCO are shown in green using the ball and stick representation and the interactions with the protein indicated using black dashed lines. Water molecules are shown as red nonbonded spheres. FAD adopts the "IN" conformation and is shown with carbon atoms colored yellow.

group interacts with two water molecules and two backbone amide nitrogen atoms found at the start of helix α1. The isoalloxazine ring adopts an "IN" conformation for both structures and is positioned in a large pocket lined with hydrophobic and aromatic residues. The N1 atom and the carbonyl oxygen atom of C2 are hydrogen bonded to the amide nitrogen atom of Ala301 and Gly302, the first residues of helix α10. N3 is hydrogen bonded to the main chain carbonyl oxygen atom of Tyr54. N5 and the carbonyl oxygen of C4 form a water-mediated hydrogen bond to Tyr223. There are no interactions with the rest of the isoalloxazine ring.

Substrate Binding Site. Soaking MHPCO crystals in a reservoir solution containing 10 mM MHPC for 3 h resulted in MHPC binding in each active site of the MHPCO dimer (Figure 6.5). The active site forms a pocket 15 Å by 10 Å by 12 Å and is much larger than the MHPC molecule. In addition to MHPC, nine water molecules are found in the binding pocket. A channel 7 Å wide lined with aromatic and hydrophobic residues points away from the FAD binding site leading from the back

side of the dimer interface to the active site (Figure 6.6). This tunnel is roughly 17 Å long. MHPC binds beneath the loop connecting α9 to α10 and is stacked 5 Å above Tyr54. The C2 carbon atom of MHPC is 5.3 Å from the C4a atom of the isoalloxazine ring. MHPC has no direct interactions with the residues lining the active site; instead, all interactions are mediated through hydrogen bonds with water (Figure 6.5). The hydroxyl group is hydrogen bonded to Tyr223 through a water molecule (2.4 Å from the hydroxyl group to the water molecule and 2.7 Å to Tyr223) and N1 is coordinated to Tyr82 through another water molecule 2.6 Å from N1 and 3.0 Å from the hydroxyl group of Tyr82. Each oxygen atom of the carboxylate group is hydrogen bonded to a water molecule that is part of a water molecule network leading down the tunnel to the solvent. Arg211 is oriented 3.2 Å and 90° from the carboxylate group of MHPC and is positioned through a salt bridge with Asp201, but is poorly positioned to contribute a hydrogen bond to MHPC.

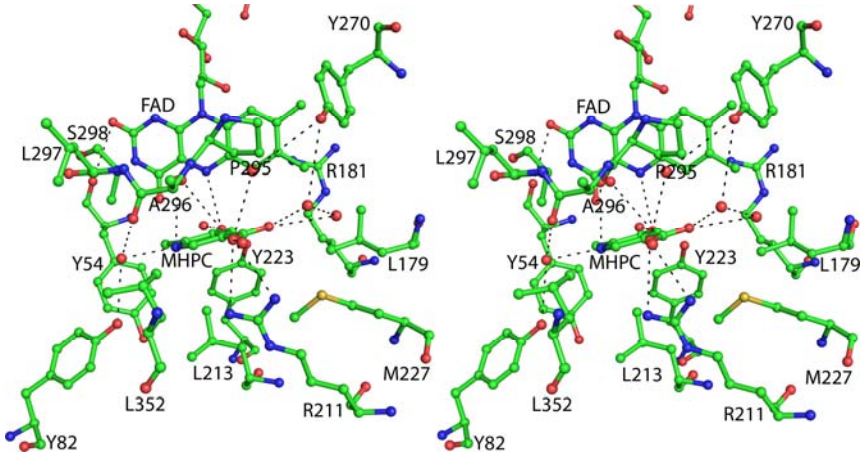


Figure 6.5. Stereoview diagram of the active site of MHPCO with FAD and MHPC bound. Water molecules are shown as nonbonded red spheres. Hydrogen bonds between MHPC and water molecules and water molecules and the protein are shown as dashed lines.

Section 6.4. Discussion

Oligomeric State of MHPCO. Previous work on MHPCO in both *Pseudomonas* MA-1 and *M. loti* has identified MHPCO as a homotetramer using gel

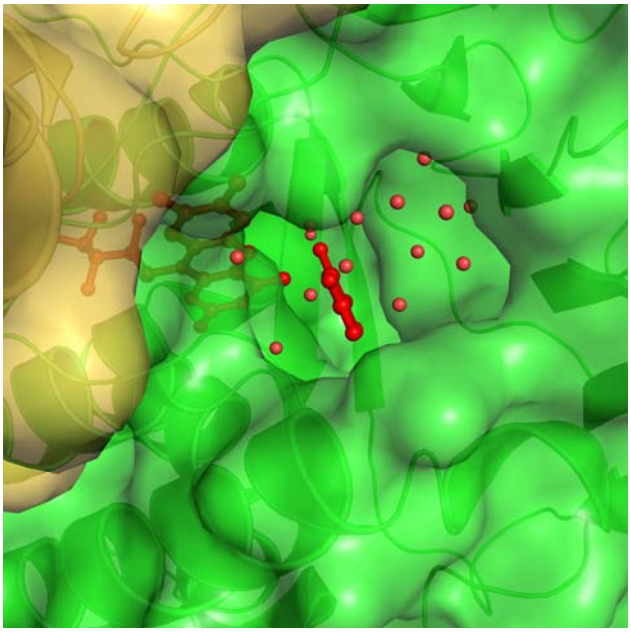


Figure 6.6. Space filling representation of the substrate tunnel leading to the binding pocket. FAD and MHPC are red and shown in ball and stick and the electrostatic surface of MHPCO is colored by protomer with green being the chain with FAD and MHPC bound and yellow being the opposing chain. Water molecules in the tunnel are shown as red spheres. The secondary structural elements are shown in cartoon representation.

filtration and analytical ultracentrifugation. The stoichiometry of FAD in each tetramer has been less conclusively determined with different experiments indicating either two or four FAD molecules per tetramer (*12-14, 33*). The crystal structure of *M. loti* MHPCO suggests that each active site contains an FAD molecule, and sample preparation did not require excess flavin to be added during purification or crystallization. However, the MHPCO is a dimer in the crystal structures reported here. A dimer has only been reported for *Pseudomonas* MA-1 MHPCO apoenzyme where the FAD had been chemically removed from the active site (*33*). Activity assays indicate that samples of *Ml*MHPCO with and without the His tag are equally active. It is still possible that the His tag disrupted the tetramer assembly as opposed to crystal packing. The His tag and the first ten residues of MHPCO are disordered in the crystal structures and Thr11 is solvent exposed at one end of the monomeric

structure. It is conceivable that the His tag could have adopted a conformation that prevents the coupling of two dimers to form the tetramer. Examination of crystal packing shows some weak interactions between opposing dimers in adjacent asymmetric units. Arg47 hydrogen bonds to Ser44*, the carboxylate side chain of Asp123 forms hydrogen bonds to the amino side chain of Lys263*, and several water-mediated hydrogen bonds occur between the three stranded antiparallel β-sheet (β10↑β9↓β8↑) of one protomer and the loop region between α8 and β17 of the second protomer. These interactions could be remnants of a tetramer disrupted by the His tag or by crystal packing. To investigate the role the N-terminal His tag may have played in disrupting the tetramer, analytical ultracentrifugation was performed. The presence of the His tag does not appear to be sufficient to prevent tetramer formation in solution as shown in Figure 6.3C; however, it is possible that the His tag at the N-terminus could weaken the interactions occurring at the interface. In combination with crystallization conditions and crystal lattice packing, the tetramer may have been disrupted, resulting in the dimer observed in the crystal structures solved.

Comparison of MHPCO to Other Enzymes. Previous work has identified MHPCO as a member of the flavin aromatic monooxygenases on the basis of cofactor requirement, substrate structure, and sequence similarity (*16, 34*). Both BLAST and DALI searches were performed to determine which enzymes in this family are the closest structural relatives of MHPCO (*35, 36*). The BLAST search revealed that MHPCO is most closely related to flavin monooxygenases. The most similar proteins were hypothetical proteins, salicylate 1-monooxygenases, and FAD monooxygenases. These proteins had sequence identity with MHPCO of 22-25% and sequence similarity below 50%. Because of the nucleotide binding motif, the DALI search identified many enzymes as being structurally similar to MHPCO (Table 6.3). As expected, these enzymes are all flavin-dependent aromatic monooxygenases that use a flavin

hydroperoxide intermediate to hydroxylate an aromatic ring (*37-40*). The overall fold of these enzymes is the same as that seen for MHPCO and first described for PHBH (*41*). The substrates for three of these enzymes, *p*-hydroxybenzoate hydroxylase (PHBH), *m*-hydroxybenzoate hydroxylase (MHBH), and 2,6-dihydroxypyridine hydroxylase (DHPH) are aromatic rings with a hydroxyl group *ortho* to the position being hydroxylated, as seen in MHPCO (*42*). The RMSD for each of these enzymes to MHPCO is less than 3 Å for at least 325 aligned residues with Z-scores of 26 or higher, and the structures are compared in Figure 6.7. Each of these enzymes bind FAD in a long narrow cleft stretching between two domains and bind very specifically to their respective substrates.

Table 6.3. Enzymes Identified as Structurally Similar to MHPCO through DALI

Protein	**PDB ID**	**Z Score**	**RMSD**	**% Identical**	**# aligned residues**
Pyocyanin biosynthetic enzyme, PhzS	2RGJ	35.1	2.3	25.7	325
p-Hydroxybenzoate hydroxylase (PHBH)	1D7L	33.7	2.7	17.2	338
2,6-Dihydroxypyridine hydroxylase (DHPH)	2VOU	33.3	2.8	21.9	336
Probable tryptophan hydroxylase, VioD	3C4A	32.6	2.8	20.1	329
m-Hydroxybenzoate hydroxylase (MHBH)	2DKI	31.9	2.6	13.1	332
Phenol hydroxylase	1FOH	31.9	2.8	14.0	341
Rebeccamycin biosynthetic enzyme, RebC	2R0G	25.7	2.9	15.8	334

The mechanism of substrate binding within these enzymes appears to be different. Some enzymes, such as PhzS and MHBH, have a tunnel that provides access to the binding pocket without requiring a conformational change in the FAD (*38, 43*). Other structurally similar enzymes do not have such a channel. For example, DHPH could not be crystallized with substrate because the ligand would not bind. PHBH is believed to require the flavin to adopt multiple, different

conformations over the cycle of reaction. The flavin is found in an "IN" conformation when *p*-hydroxybenzoate (*p*-OHB) is bound, an "OPEN" conformation before *p*-OHB binds, and an "OUT" conformation when undergoing reduction by NAD(P)H (*40, 44, 45*). RebC appears to have a helix that can undergo conformational change upon substrate binding (*39*). Examination of MHPCO suggests that MHPC most likely enters the active site through a tunnel (Figure 6.6) without inducing any large conformational changes of the protein. Tyr270, Trp173, Arg211, Met227, Asp177, and Arg350 line the channel and form the mouth of the active site. Despite the presence of a substrate tunnel in several of the flavin monooxygenases, in MHPCO the tunnel is in a different location compared to PhzS and MHBH. Moreover, the MHBH tunnel has two openings leading to the surface of the protein, but these are pointed in different directions. One opening of MHBH is suggested to be for molecular oxygen to enter the active site and the second, filled with water molecules in the structure, is postulated to be the tunnel for substrate entrance (*38*). The channel of MHPCO does not have two entrances and does not appear to be as wide as that of MHBH although it is also filled with water molecules.

The binding of FAD by MHPCO also shares many similarities with other flavin monooxygenases but has some significant differences. In all structures of flavin aromatic monooxygenases determined, the FAD binds lengthwise in a groove contained within one protomer. The adenine base in all structures is bound in a pocket between arginine and aspartate residues and forms hydrogen bonds to the amide nitrogen atom of the protein backbone. A conserved glutamate residue (Glu41) is pointed toward the hydroxyl groups of the ribose ring in all structures. The phosphate tail is coordinated by water molecules hydrogen bonded to the protein backbone and several residues along the FAD are conserved, including Gly18, Gly23, and Asp288. The isoalloxazine ring is held in the "IN" conformation when the substrate is bound to

the active site. In some structures where ligand is not bound, the isoalloxazine ring has been found in the "OUT" conformation. The crystal structures of PhzS, RebC, and MHBH (PDB ID 2rgj, 2roc, and 2dkh respectively) show the flavin in the "OUT" conformation. In contrast to these enzymes, the FAD of MHPCO remains in the "IN" conformation and shows very little movement, regardless of the presence or absence of substrate (Figure 6.4). Using PHBH as a model, it has been previously proposed that NADH could bind parallel to the FAD molecule for transfer of the hydride to yield the reduced flavin (*13*). However, the binding site for NADH in MHPCO has not been definitively identified. Additionally, for FAD in MHPCO to adopt the "OUT" conformation seen in PhzS, the loop connecting β2 and β3 would have to move 3 to 4 Å to accommodate the isoalloxazine ring. The "OUT" conformation of FAD in PhzS is stabilized largely through π stacking with a tryptophan residue found on a loop that extends to the active site. This loop structure in MHPCO, between β16 and α7, is much shorter with only 8 residues instead of 16 residues, when compared to the PhzS structure and does not extend to cover the active site of MHPCO. The π stacking between the flavin in MHBH occurs with a tyrosine residue. The equivalent aromatic residue in MHPCO, Tyr270, is one residue shifted and is the most likely residue to provide π stacking to FAD if the "OUT" conformation were adopted. However, studies done by Chaiyen *et al.* using FAD analogues substituted at the 8-position showed that the effects on rate of reduction were due solely to the changes in redox potential and that the size of the substituent had no effect (*46*). This suggests that flavin movement may not play a role in FAD reduction for MHPCO and raises the possibility that FAD could remain in the "IN" conformation during reduction by NADH. The FAD of MHPCO may be reduced by NADH in a manner different than for other flavin aromatic monooxygenases.

Active Site of MHPCO. The active site of MHPCO is a large pocket lined mostly by hydrophobic and aromatic residues (Figure 6.5). These residues include Tyr54, Tyr82, Tyr223, Phe358, Trp203, Asp 203, and Arg211. The hydroxyl groups of the tyrosine residues point into the active site and the side chain of Arg211 is also positioned within the active site. These residues are hydrogen bonded to the many water molecules that are found within the active site in both structures. Comparisons of the active sites of flavin aromatic monooxygenases whose structures are available show some similarities. All structures have an absolutely conserved proline residue above the FAD isoalloxazine ring containing C7 and C8 (Pro295 in MHPCO). The other commonality is the presence of several aromatic residues at the back of the binding pocket. MHPCO has three tyrosine residues, two tryptophan residues, and two phenylalanine residues in the rear of the active site and none form any direct contacts with the substrate MHPC. PHBH (PDB ID 1pbe) and DHPH have a similar wall in the active site. *p*-Hydroxybenzoate (*p*-OHB), the substrate of PHBH, forms a hydrogen bond to a tyrosine hydroxyl group, which can deprotonate *p*-OHB for activation. The carboxylate group of *p*-OHB forms a salt bridge with an arginine side chain, helping correctly orient the substrate within the active site. The binding of MHPC within the active site of MHPCO is quite different. Unlike PHBH, MHPCO makes no direct interactions with its substrate (Figure 6.5). The 3-hydroxyl group of MHPC forms a water-mediated hydrogen bond to Tyr223 and the N1 atom is hydrogen bonded to Tyr82 through a water molecule. The carboxyl oxygen atoms are each hydrogen bonded to two water molecules that lead into the tunnel. Arg211 is one of the few charged residues found in the active site and it is poorly positioned to help stabilize MHPC. The surprising lack of direct interactions between MHPC and MHPCO and the large open cavity of the active site is unusual and differentiates MHPCO from other members of this enzyme family.

Mechanistic Implications of the MHPCO Structure. One possible mechanism for the oxidative ring opening of MHPC is shown in Figure 6.7. In this mechanism, the hydroxylation of MHPC first occurs by an electrophilic attack by the flavin hydroperoxide on C2 of the pyridine **7** to give **9**, which could then undergo an electrocyclic ring opening to produce the ketene **10**. Tautomerization to **11** followed by addition of water to the ketene would generate the product *E*-2AMS **8**. Several other variations on this proposal for the conversion of **9** to **8** are possible. (*34*). This mechanism is consistent with all previous $^{18}O_2$ and $H_2{}^{18}O$ labeling experiments, with substrate analogue studies, and with kinetics data (*16, 33, 47, 48*) and the current structure of MHPCO bound with MHPC seems to position the substrate optimally for this attack to occur.

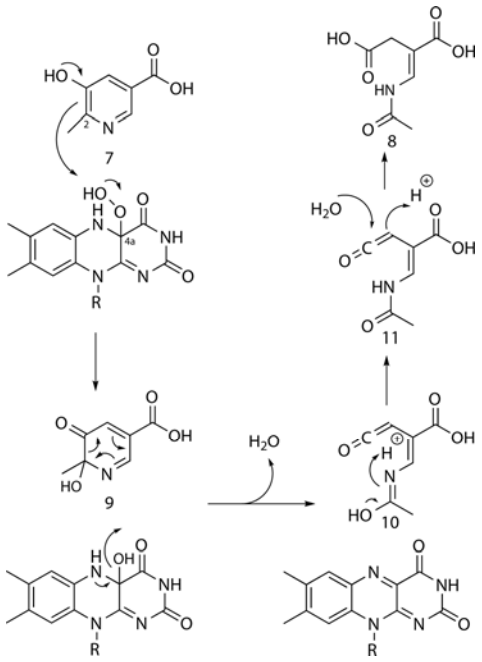


Figure 6.7. Possible mechanism for the reaction catalyzed by MHPCO. The aromatic ring is opened through an electrocyclic mechanism after pyridine hydroxylation by the flavin hydroperoxide.

The structure of MHPCO complexed with the substrate MHPC **7** clarifies many of the details of the active site and provides further insight into the catalytic

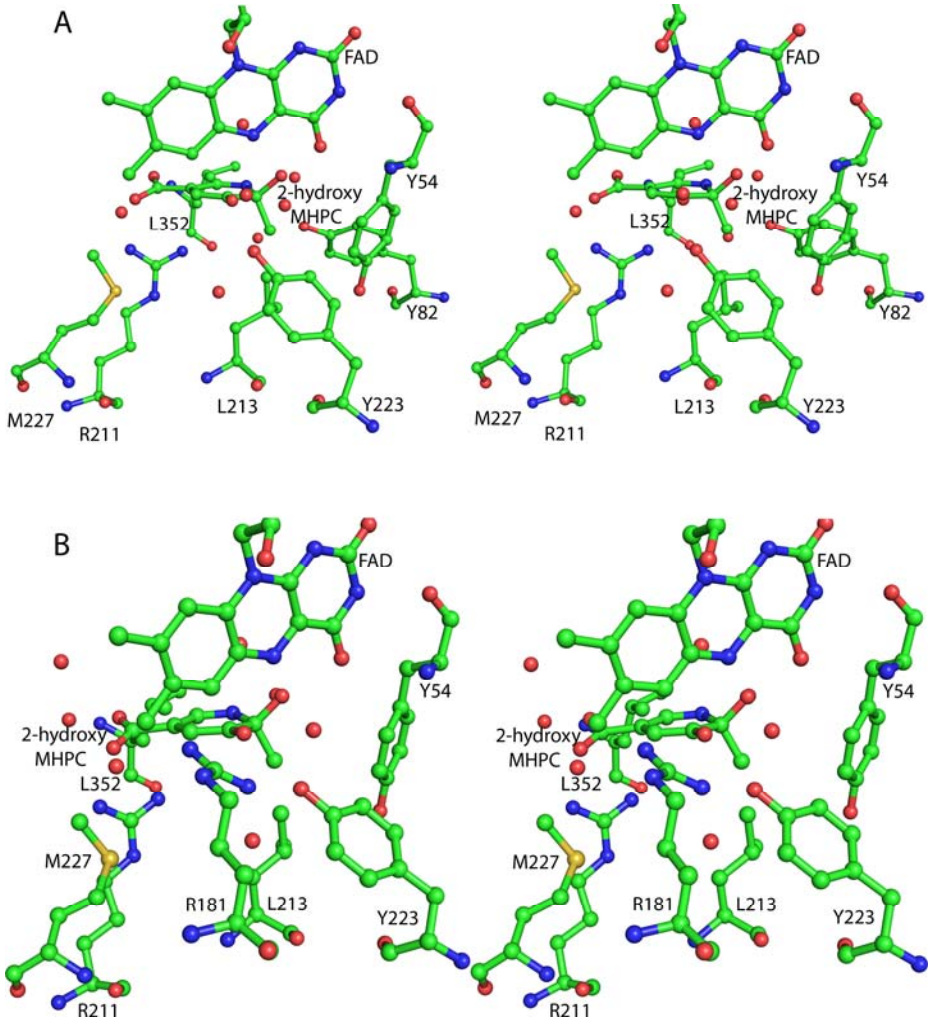


Figure 6.8. Stereoview diagrams of the active site with possible bases for the ring opening reaction. All residues and ligands are shown in ball and stick representation and colored according to atom type. Red nonbonded spheres are water molecules. The 2-hydroxy-MHPC intermediate **9** was manually modeled into the active site using the binding of MHPC as a guide. (A) Tyr223 as the active site acid. Tyr223 activates a water molecule for attack at the C3 carbonyl and could provide stabilization of the intermediate. (B) Arg181 as the active site base. Arg181 could protonate the carbonyl and stabilize the intermediate after attack of water at the carbonyl. Arg181 has adopted a different conformation than that seen in the crystal structure.

mechanism. The C2 of MHPC is located 5.3 Å from C4a of the isoalloxazine ring. This separation, in addition to the orientation of the planes of the pyridine and the isoalloxazine rings, gives a reasonable geometry for reaction of the distal oxygen atom of the flavin hydroperoxide with C2 of the substrate. The position of this oxygen is indicated by structurally conserved water molecules seen in both structures of MHPCO described here. Comparable substrate-C4a distances have been observed in other aromatic ring oxidizing flavin-dependent monooxygenases. Tyr223 and Tyr82 are likely catalytic residues (Figure 6.8A). Tyr223 could activate the water molecule located 2.7 Å from the substrate C3 oxygen and facilitate the conversion of **7** to **9**. Tyr82 is hydrogen bonded to the water molecule located 2.6 Å from the pyridine nitrogen and may facilitate the conversion of **10** to **11**. Arg181 is also a possible candidate for an active site base. In the crystal structure, this residue is oriented along the isoalloxazine ring but a conformational change of the side chain may position the guanidinium group within hydrogen bonding distance of the substrate C3 oxygen (modeled in Figure 6.8B). However, for Arg181 to adopt this conformation some conformational changes of the active site would have to occur to avoid close contacts between the isoalloxazine ring and the arginine. Removal of the hydroxyl group of Tyr223 (Y223F) or substitution of the guanidinium group of Arg181 with an amide group (R181Q) both render the enzyme inactive. Arg211 is located below the plane of the pyrimidine ring and its guanidinium group is facing the C5 carboxylate of MHPC. While poorly positioned to form a salt bridge in the enzyme substrate complex, conformational changes during the course of the reaction could increase the strength of this interaction.

REFERENCES

1. Snell, E. E., and Haskell, B. E. (1971) *The Metabolism of Vitamin B6, In: Comprehensive Biochemistry*, Vol. 21, Elsevier/North Holland, New York.
2. Burg, R. W., Rodwell, V. W., and Snell, E. E. (1960) Bacterial oxidation of vitamin B6. II. Metabolites of pyridoxamine, *J Biol Chem 235*, 1164-1169.
3. Rodwell, V. W., Volcani, B. E., Ikawa, M., and Snell, E. E. (1958) Bacterial oxidation of vitamin B6. I. Isopyridoxal and 5-pyridoxic acid, *J Biol Chem 233*, 1548-1554.
4. Ikawa, M., Rodwell, V. W., and Snell, E. E. (1958) Bacterial oxidation of vitamin B6. II. Structure of 260 compound, *J Biol Chem 233*, 1555-1559.
5. Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S., Watanabe, A., Idesawa, K., Ishikawa, A., Kawashima, K., Kimura, T., Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A., Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimpo, S., Sugimoto, M., Takeuchi, C., Yamada, M., and Tabata, S. (2000) Complete genome structure of the nitrogen-fixing symbiotic bacterium Mesorhizobium loti, *DNA Res 7*, 331-338.
6. Funami, J., Yoshikane, Y., Kobayashi, H., Yokochi, N., Yuan, B., Iwasaki, K., Ohnishi, K., and Yagi, T. (2005) 4-Pyridoxolactonase from a symbiotic nitrogen-fixing bacterium Mesorhizobium loti: cloning, expression, and characterization, *Biochim Biophys Acta 1753*, 234-239.
7. Ge, F., Yokochi, N., Yoshikane, Y., Ohnishi, K., and Yagi, T. (2008) Gene identification and characterization of the pyridoxine degradative enzyme 4-pyridoxic acid dehydrogenase from the nitrogen-fixing symbiotic bacterium Mesorhizobium loti MAFF303099, *J Biochem 143*, 603-609.

8. Mukherjee, T., Hilmey, D. G., and Begley, T. P. (2008) PLP catabolism: identification of the 2-(Acetamidomethylene)succinate hydrolase gene in Mesorhizobium loti MAFF303099, *Biochemistry 47*, 6233-6241.
9. Mukherjee, T., Kinsland, C., and Begley, T. P. (2007) PLP catabolism: identification of the 4-pyridoxic acid dehydrogenase gene in Mesorhizobium loti MAFF303099, *Bioorg Chem 35*, 458-464.
10. Mukherjee, T., McCulloch, K. M., Ealick, S. E., and Begley, T. P. (2007) Gene identification and structural characterization of the pyridoxal 5'-phosphate degradative protein 3-hydroxy-2-methylpyridine-4,5-dicarboxylate decarboxylase from mesorhizobium loti MAFF303099, *Biochemistry 46*, 13606-13615.
11. Yuan, B., Yokochi, N., Yoshikane, Y., Ohnishi, K., Ge, F., and Yagi, T. (2008) Gene identification and characterization of the pyridoxine degradative enzyme alpha-(N-acetylaminomethylene)succinic acid amidohydrolase from mesorhizobium loti MAFF303099, *J Nutr Sci Vitaminol (Tokyo) 54*, 185-190.
12. Yuan, B., Yokochi, N., Yoshikane, Y., Ohnishi, K., and Yagi, T. (2006) Molecular cloning, identification and characterization of 2-methyl-3-hydroxypyridine-5-carboxylic-acid-dioxygenase-coding gene from the nitrogen-fixing symbiotic bacterium Mesorhizobium loti, *J Biosci Bioeng 102*, 504-510.
13. Chaiyen, P., Ballou, D. P., and Massey, V. (1997) Gene cloning, sequence analysis, and expression of 2-methyl-3-hydroxypyridine-5-carboxylic acid oxygenase, *Proc Natl Acad Sci U S A 94*, 7233-7238.
14. Sparrow, L. G., Ho, P. P., Sundaram, T. K., Zach, D., Nyns, E. J., and Snell, E. E. (1969) The bacterial oxidation of vitamin B6. VII. Purification, properties,

and mechanism of action of an oxygenase which cleaves the 3-hydroxypyridine ring, *J Biol Chem 244*, 2590-2600.

15. van Berkel, W. J., Kamerbeek, N. M., and Fraaije, M. W. (2006) Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts, *J Biotechnol 124*, 670-689.
16. Chaiyen, P., Brissette, P., Ballou, D. P., and Massey, V. (1997) Unusual mechanism of oxygen atom transfer and product rearrangement in the catalytic reaction of 2-methyl-3-hydroxypyridine-5-carboxylic acid oxygenase, *Biochemistry 36*, 8060-8070.
17. Oonanant, W., Sucharitakul, J., Yuvaniyama, J., and Chaiyen, P. (2005) Crystallization and preliminary X-ray crystallographic analysis of 2-methyl-3-hydroxypyridine-5-carboxylic acid (MHPC) oxygenase from Pseudomonas sp. MA-1, *Acta Crystallogr Sect F Struct Biol Cryst Commun 61*, 312-314.
18. Ausubel, F. M., and Brent, F., (Eds.) (1987) *Current Protocols in Molecular Biology*, John Wiley and Sons, New York.
19. Sambrook, J., Fritsch, G. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Guide*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
20. Philo, J. S. (2006) Improved methods for fitting sedimentation coefficient distributions derived by time-derivative techniques, *Anal Biochem 354*, 238-246.
21. Matthews, B. W. (1968) Solvent content of protein crystals, *J. Mol. Biol. 33*, 491-497.
22. Otwinowski, Z., and Minor, W. (1997) Processing of x-ray diffraction data collected in oscillation mode, *Methods Enzymol. 276*, 307-326.

23. Pape, T., and Schneider, T. R. (2004) HKL2MAP: a graphical user interface for phasing with SHELX programs, *J Appl Cryst 37*, 843-844.

24. Schneider, T. R., and Sheldrick, G. M. (2002) Substructure solution with SHELXD, *Acta Crystallog. D D58*, 1772-1779.

25. Sheldrick, G. M. (2002) Macromolecular phasing with SHELXE, *Z Kristallogr 217*, 644-650.

26. 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.

27. Terwilliger, T. C. (2003) Automated main-chain model building by template matching and iterative fragment extension, *Acta Crystallogr D Biol Crystallogr 59*, 38-44.

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

29. 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.

30. DeLano, W. L. (2002) The PyMOL Molecular Graphics Systems, DeLano Scientific, San Carlos, CA.

31. Hasel, W., Hendrickson, T. F., and Still, W. C. (1988) A rapid approximation to the solvent-accessible surface areas of atoms, *Tetrahedron Comp. Meth. 1*, 103-116.

32. Lebowitz, J., Lewis, M. S., and Schuck, P. (2002) Modern analytical ultracentrifugation in protein science: a tutorial review, *Protein Sci 11*, 2067-2079.

33. Kishore, G. M., and Snell, E. E. (1981) Interaction of 2-methyl-3-hydroxypyridine-5-carboxylic acid oxygenase with FAD, substrates, and analogues. Spectral and fluorescence investigations, *J Biol Chem 256*, 4234-4240.

34. Chaiyen, P., Brissette, P., Ballou, D. P., and Massey, V. (1997) Reaction of 2-methyl-3-hydroxypyridine-5-carboxylic acid (MHPC) oxygenase with N-methyl-5-hydroxynicotinic acid: studies on the mode of binding, and protonation status of the substrate, *Biochemistry 36*, 13856-13864.

35. Gibrat, J. F., Madej, T., and Bryant, S. H. (1996) Surprising similarities in structure comparison, *Curr Opin Struct Biol 6*, 377-385.

36. 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.

37. Entsch, B., and van Berkel, W. J. (1995) Structure and mechanism of para-hydroxybenzoate hydroxylase, *FASEB J 9*, 476-483.

38. Hiromoto, T., Fujiwara, S., Hosokawa, K., and Yamaguchi, H. (2006) Crystal structure of 3-hydroxybenzoate hydroxylase from Comamonas testosteroni has a large tunnel for substrate and oxygen access to the active site, *J Mol Biol 364*, 878-896.

39. Ryan, K. S., Howard-Jones, A. R., Hamill, M. J., Elliott, S. J., Walsh, C. T., and Drennan, C. L. (2007) Crystallographic trapping in the rebeccamycin biosynthetic enzyme RebC, *Proc Natl Acad Sci U S A 104*, 15311-15316.

40. Treiber, N., and Schulz, G. E. (2008) Structure of 2,6-dihydroxypyridine 3-hydroxylase from a nicotine-degrading pathway, *J Mol Biol 379*, 94-104.

41. Wierenga, R. K., de Jong, R. J., Kalk, K. H., Hol, W. G., and Drenth, J. (1979) Crystal structure of p-hydroxybenzoate hydroxylase, *J Mol Biol 131*, 55-73.

42. Joosten, V., and van Berkel, W. J. (2007) Flavoenzymes, *Curr Opin Chem Biol 11*, 195-202.

43. Greenhagen, B. T., Shi, K., Robinson, H., Gamage, S., Bera, A. K., Ladner, J. E., and Parsons, J. F. (2008) Crystal structure of the pyocyanin biosynthetic protein PhzS, *Biochemistry 47*, 5281-5289.

44. Wang, J., Ortiz-Maldonado, M., Entsch, B., Massey, V., Ballou, D., and Gatti, D. L. (2002) Protein and ligand dynamics in 4-hydroxybenzoate hydroxylase, *Proc Natl Acad Sci U S A 99*, 608-613.

45. Cole, L. J., Entsch, B., Ortiz-Maldonado, M., and Ballou, D. P. (2005) Properties of p-hydroxybenzoate hydroxylase when stabilized in its open conformation, *Biochemistry 44*, 14807-14817.

46. Chaiyen, P., Sucharitakul, J., Svasti, J., Entsch, B., Massey, V., and Ballou, D. P. (2004) Use of 8-substituted-FAD analogues to investigate the hydroxylation mechanism of the flavoprotein 2-methyl-3-hydroxypyridine-5-carboxylic acid oxygenase, *Biochemistry 43*, 3933-3943.

47. Chaiyen, P., Brissette, P., Ballou, D. P., and Massey, V. (1997) Thermodynamics and reduction kinetics properties of 2-methyl-3-hydroxypyridine-5-carboxylic acid oxygenase, *Biochemistry 36*, 2612-2621.

48. Kishore, G. M., and Snell, E. E. (1981) Kinetic investigations on a flavoprotein oxygenase, 2-methyl-3-hydroxypyridine-5-carboxylic acid oxygenase, *J Biol Chem 256*, 4228-4233.

CHAPTER 7

STRUCTURE DETERMINATION AND CHARACTERIZATION OF THE VITAMIN B6 DEGRADATIVE ENZYME (*E*)-2-(ACETAMIDOMETHYLENE)SUCCINATE HYDROLASE[1]

Section 7.1. Introduction

Vitamin B_6 is an essential vitamin whose active cofactor, pyridoxal 5′-phosphate (PLP), plays an important role within the cell to stabilize carbanion intermediates adjacent to an amino group (*1*). Vitamin B_6 degradation was first shown to occur in select microorganisms in the late 1950s by culturing bacteria on minimal media where the only carbon and nitrogen sources were different B_6 vitamers (*2*). Two pathways for the degradation of vitamin B_6 have been described; the first has been characterized in *Pseudomonas* sp. MA-1 and the second in *Pseudomonas* IA and *Arthrobacter* Cr-7 (*2*). None of the PLP catabolic genes in these microorganisms were identified. Gene identification was greatly facilitated when the genome of *Mesorhizobium loti* MAFF303099 was sequenced in 2000 (*3*). Shortly thereafter, the pyridoxine 4-oxidase gene was identified, suggesting that this bacterium was capable of degrading vitamin B_6 (*4*). Six other catabolic genes were rapidly identified completing the vitamin B_6 degradative pathway in *M. loti*, shown in Figure 7.1 (*5-11*). Crystal structures have been determined for three of these enzymes (*12-14*). The *M. loti* degradative pathway is slightly different from that observed in *Pseudomonas* sp. MA-1.

E-2-(acetamidomethylene)succinate (*E*-2AMS) hydrolase is the last enzyme to be biochemically characterized on the vitamin B_6 catabolic pathway. This enzyme catalyzes the hydrolysis of *E*-2AMS to succinic semialdehyde, ammonia, acetate, and

[1] Reproduced with permission from McCulloch, K.M., Mukherjee, T., Begley, T.P., and Ealick, S.E. (2010) *Biochemistry* 49: 1226-1235. Copyright 2010 American Chemical Society

carbon dioxide. The products of this reaction have been conclusively identified and the steady state kinetics of the enzyme have been characterized (*7, 10*). A similar reaction in nicotinate catabolism, catalyzing the amide hydrolysis of 2-(enamine)-glutarate, is performed by an enamidase, a member of the amidohydrolase superfamily (*15*). The primary sequence of *E*-2AMS hydrolase, however, indicated that this enzyme was most similar to members of the α/β hydrolase superfamily, although the sequence homology was low. This superfamily often utilizes a catalytic triad consisting of an active site nucleophile, an absolutely conserved histidine residue, and an acidic residue. Sequence alignments failed to identify this acidic residue and it was unclear how this enzyme catalyzed its metal-independent hydrolysis reaction (*10*). The structure reported here identifies this acidic residue and also suggests several other residues within the active site that could be involved in substrate binding and catalysis.

Section 7.2. Materials and Methods

Cloning, Overexpression, and Purification of E-2AMS Hydrolase. The molecular cloning of the mlr6787 gene from *M. loti* MAFF303099 has been previously described (*7*). The resulting pMl5331.XF1 plasmid contains the mlr6787 gene with an N-terminal polyhistidine tag and was transformed into *Escherichia coli* B834(DE3), a cell line auxotrophic for methionine. Overnight cultures were grown by transferring a single colony to 10 mL LB medium with 150 μg/mL ampicillin at 37 °C with shaking. Selenomethionyl-incorporated (SeMet) protein was grown using a minimal medium supplemented with 0.4% glucose, 20 mg/L of all amino acids except methionine, 50 mg/L L-selenomethionine, 1× MEM vitamin mix (Invitrogen), 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 25 mg/L $FeSO_4$, and 150 μg/mL ampicillin. The overnight culture, grown in LB, was harvested at 4 °C at 2,000 × *g* for 15 min then resuspended in the minimal medium for inoculation of 1 L cultures. The large culture was allowed

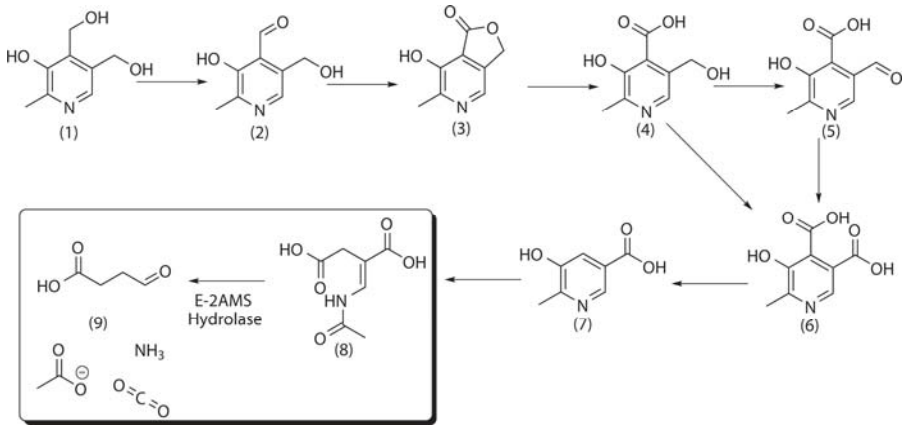


Figure7.1. Vitamin B_6 degradative pathway found in *M. loti*. 1, pyridoxine; 2, pyridoxal; 3, 4-pyridoxolactone; 4, 4-pyridoxic acid; 5, 5-formyl-2-methyl-3-hydroxypyridine-4-carboxylic acid; 6, 2-methyl-3-hydroxypyridine-4,5-dicarboxylic acid; 7, 2-methyl-3-hydroxypyridine-5-carboxylic acid; 8, 2-(acetamidomethylene)succinic acid; 9, succinic semialdehyde. Other final products are ammonia, acetate, and carbon dioxide

to grow at 37 °C with shaking to an OD_{595} of 0.6. The temperature was then lowered to 15 °C and overexpression of protein was induced overnight with 1 mM isopropyl β-D-thiogalactopyranoside (IPTG). After 16 h, cells were pelleted by centrifugation at 6,000 × *g* and stored frozen at -20 °C. For native protein overexpression, 1 L of LB medium with 150 μg/mL ampicillin was inoculated with 5 mL of overnight culture, then grown at 37 °C with shaking to an OD_{595} of 0.6 and induced with 1 mM IPTG. The cells were harvested as described for the SeMet protein preparation.

Purification of both native and SeMet protein followed the same protocol. Frozen cell pellet was resuspended in 30 mL of purification buffer (50 mM NaH_2PO_4 at pH 8.0, 300 mM NaCl, 3 mM β-mercaptoethanol) with 10 mM imidazole and lysed by sonication. After lysis, the cell extract was clarified by centrifugation at 40,000 × *g* for 1 h at 4 °C. The supernatant was then twice passed over a 2 mL Ni-NTA column (Qiagen) pre-equilibrated with purification buffer. The column was then washed with 50 mL of purification buffer and nonspecifically binding contaminants were removed by washing with 25 mL of purification buffer containing 20 mM imidazole. The

protein was eluted from the column using purification buffer containing 250 mM imidazole. The resulting sample was further purified using size exclusion chromatography (HiLoad 26/60 Superdex 75 pg, GE Healthcare) to greater than 95% homogeneity as judged by SDS-PAGE analysis (results not shown). The protein samples were concentrated to ~8 mg/mL as measured by Bradford assay and stored at -80 °C in storage buffer (20 mM Tris pH 8.0, 50 mM NaCl), with 1 mM dithiothreitol (DTT) added to the SeMet sample (*16*).

Activity Assay for E-2AMS Hydrolase. The assay used for determining the activity of *E*-2AMS hydrolase has previously been described (*7*). Briefly, freshly purified enzyme was buffer exchanged into 50 mM phosphate buffer containing a quantified amount of *E*-2AMS and the decrease in absorbance at 261 nm is monitored. This assay was used to determine the steady state kinetic parameters previously reported (*7, 10*).

Crystallization of E-2AMS Hydrolase. The hanging-drop vapor diffusion method was used to crystallize *E*-2AMS hydrolase by mixing equal volumes of protein sample with reservoir solution and allowing the drop to equilibrate against 500 μL of reservoir solution at 22 °C. Initial hits were determined for the native protein using commercially available sparse matrix screens (Hampton Research, Emerald BioSystems). Optimized reservoir conditions consisted of 8-12% PEG 8000, 200 mM $MgCl_2$, and 100 mM Tris buffer, pH 7.4-8.0 and 1.5 μL of reservoir solution was mixed with an equal volume of protein solution. Tetragonal bipyramidal crystals grew to 200 μm × 150 μm within two days. SeMet *E*-2AMS hydrolase crystals grew under the same conditions, but required an additional microseeding step. Briefly, native crystals were transferred to a seed stabilization solution of 9% PEG 8000, 200 mM $MgCl_2$, 100 mM Tris at pH 7.8, then crushed using Seed Bead (Hampton Research). The freshly prepared seeds were then serially diluted and 0.5 μL of seed solution was

added to SeMet hanging drops pre-equilibrated overnight. Crystals began growing within 20 min of seeding and reached full size overnight. Native and SeMet *E*-2AMS hydrolase crystals both belong to the space group $P2_12_12_1$ with average unit cell parameters of a = 115 Å, b = 179 Å, and c = 189 Å. The asymmetric unit consists of twelve chains, corresponding to a Matthew's coefficient of 2.40 $Å^3$/Da and a solvent content of 49% (*17*).

X-Ray Data Collection and Processing. Protein crystals were cryoprotected in the crystallization solution supplemented with 17% glycerol and then flash frozen by plunging into liquid nitrogen. A single wavelength anomalous dataset was collected on a single SeMet *E*-2AMS hydrolase crystal on the NE-CAT 24-ID-C beamline at the Advanced Photon Source. A fluorescence scan was performed on a SeMet crystal to determine the maximum f'' for selenium, and the data were collected at this wavelength to a maximum resolution of 2.7 Å over a total rotation range of 180°. A native crystal diffracted to 2.3 Å resolution and data were collected over 110° on the NE-CAT 24-ID-E beamline. Both datasets were collected using Quantum315 detectors (Area Detector Systems Corp.) and used a 0.5° oscillation range to minimize reflection overlaps. The data were then indexed, integrated, and scaled using the HKL2000 suite of programs (*18*). Data collection statistics are given in Table 1.

Structure Determination, Model Building, and Refinement. Positions for 36 of a possible 60 selenium atoms were determined using the SAD dataset to 2.7 Å and the program hkl2map (*19-21*). These heavy atom coordinates were then used for SAD phasing, followed by density modification and calculation of initial electron density maps in CNS (*22*). RESOLVE was then used to perform phase extension to 2.3 Å with the native data and automated model building (*23-24*). The initial model, consisting of 1511 out of 3336 residues in the asymmetric unit, required extensive manual model building in COOT (*25*). Once twelve chains had been positioned in the

asymmetric unit relying solely on the experimental phase information, iterative rounds of refinement were performed in CNS followed by additional manual model building. Strict non-crystallographic symmetry (NCS) restraints were utilized and gradually lessened throughout refinement (*26*). Water molecules were added during later rounds of refinement using CNS. The final model, consisting of residues 11-278 for twelve chains, was analyzed using PROCHECK; Ser106 and Asp257 were each identified as having geometry that falls within the disallowed region, although each residue has clear electron density (*27*). Final refinement statistics are summarized in Table 2. Figures were generated using PYMOL and ChemDraw (*28*).

Table 7.1. Summary of Data Collection Statistics

	SeMet	Native
Source	APS 24-ID-C	APS 24-ID-E
Resolution (Å)	2.70	2.26
Wavelength (eV)	12662	12667
Space Group	$P2_12_12_1$	$P2_12_12_1$
a (Å)	115.2	115.2
b (Å)	179.0	178.5
c (Å)	189.2	189.3
Matthew's Coef.	2.4	2.4
% solvent	49	49
Mol/a.s.u.	12	12
Measured Reflections	531900	727191
Unique reflections[a]	207285 (20629) [b]	168768 (12441)
Average I/σ	11.7 (2.4)	13.1 (1.7)
Redundancy	2.6 (2.5)	4.3 (2.1)
Completeness (%)	99.9 (99.4)	91.6 (68.2)
Rsym[c] (%)	7.9 (35.7)	10.3 (45.8)

[a]Unique reflections include Bijvoet pairs for anomalous data. [b]Values in parentheses are for the highest resolution shell. [c]$R_{sym} = \Sigma\Sigma_i|I_i - \langle\mathbf{I}\rangle|/\Sigma\langle\mathbf{I}\rangle$, where $\langle\mathbf{I}\rangle$ is the mean intensity of the N reflections with intensities I_i and common indicies h,k,l.

Mutagenesis of E-2AMS Hydrolase. Mutants of *E*-2AMS hydrolase were prepared using standard procedures for site-directed mutagenesis (*29*). The plasmid p5331.XFI in *E. coli* Mach1 cell line was used as template DNA. Primers used for generating S106A are as follows: 5′-(for)CCA TCC TCG TCG GAC ACG CGC TTG

GTG CTC GAA ATT CGG-3′; 5′-(rev)CCG AAT TTC GAG CAC CAA GCG CGT GTC CGA CGA GGA TGG-3′. The D130N mutant was prepared using the following primers: 5′-(for)GGT GCG GTC GGT CGT CGC GAT TAA CTT TAC GCC GTA CAT CGA G-3′; 5′-(rev)C TCG ATG TAC GGC GTA AAG TTA ATC GCG ACG ACC GAC CGC ACC-3′. To generate the S230A mutant, the following primers were used: 5′-(for)CGT TCG GGG CGA GTC CGC CAA GTT GGT TTC TGC G-3′; 5′-(rev)C GCA GAA ACC AAC TTG GCG GAC TCG CCC CGA ACG-3′. Additionally, the S230C and S230N mutants were generated by replacing the underlined bases in the forward primer with CTG and TTG and CAG and CAA, respectively. These mutants were each overexpressed and purified as described for the wild-type *E*-2AMS hydrolase. The S106A mutant was crystallized using the seeding procedure required for the SeMet sample.

Section 7.3. Results

Packing of E-2AMS Hydrolase. A typical diffraction pattern from an *E*-2AMS hydrolase dataset is shown in Figure 7.2A. At low resolution, it is clear that every third row of diffraction maxima exhibited a much greater intensity than the intervening two rows of maxima. This pattern indicated the possibility of pseudo-translational symmetry (*30*), which was confirmed by the structure. Twelve chains were built into the asymmetric unit and resulted in 49% solvent content. Figure 7.2B shows one layer of monomers in the unit cell, which consists of three unique protomers from the asymmetric unit and their crystallographic equivalents. The pseudo-translational symmetry occurs along the B-axis, where the spacing between protomers is nearly equal for two of the three pairs and nearly 10 Å further for the final spacing.

Overall Structure of E-2AMS Hydrolase. The structure of *E*-2AMS hydrolase was solved using SeMet SAD phasing at 2.7 Å resolution followed by phase extension

with native data to 2.3 Å. Each of the twelve protomers within the asymmetric unit consists of residues 11-278; the first ten residues and the polyhistidine tag are not visible in the electron density. Figure 7.3 shows the final model of *E*-2AMS hydrolase. The overall fold of this enzyme indicates that *E*-2AMS hydrolase belongs to the α/β hydrolase superfamily. The core of the enzyme is an eight-stranded mostly parallel β-sheet where β2 is the only strand running antiparallel (β1↑β2↓β4↑β3↑β5↑β6↑β9↑β10↑). The β-sheet has a very tight twist so that β10 is oriented nearly perpendicular to β1 when looking down the β-sheet. One side of the

Table 7.2. Summary of Data Refinement Statistics

	E-2AMS Hydrolase
Resolution (Å)	50.00 – 2.30
No. of protein atoms	24444
No. of ligand atoms	12
No. of water atoms	1387
Reflections in working set	279026
Reflections in test set	18821
R factor[a] (%)	20.4
R_{free}[b] (%)	24.5
Rmsd from ideal	
Bonds (Å)	0.0058
Angles (º)	1.2
Average B-factor (Å^3)	28.4
Ramachandran Plot	
Most favored (%)	87.6
Additionally allowed (%)	11.9
Generously allowed (%)	0.1
Disallowed (%)	0.4

[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

β-sheet is flanked by three α-helices, α2, α10, and α11. A bundle of α-helices, consisting of α1, α3, α8, and α9, packs on the other side of the β-sheet. A cap domain is inserted between β6 and β9 and sits at the N-terminal side of the β-sheet. This lid is made up of four α-helices (α4, α5, α6, and α7) and a small two-stranded antiparallel β-sheet (β7↑β8↓).

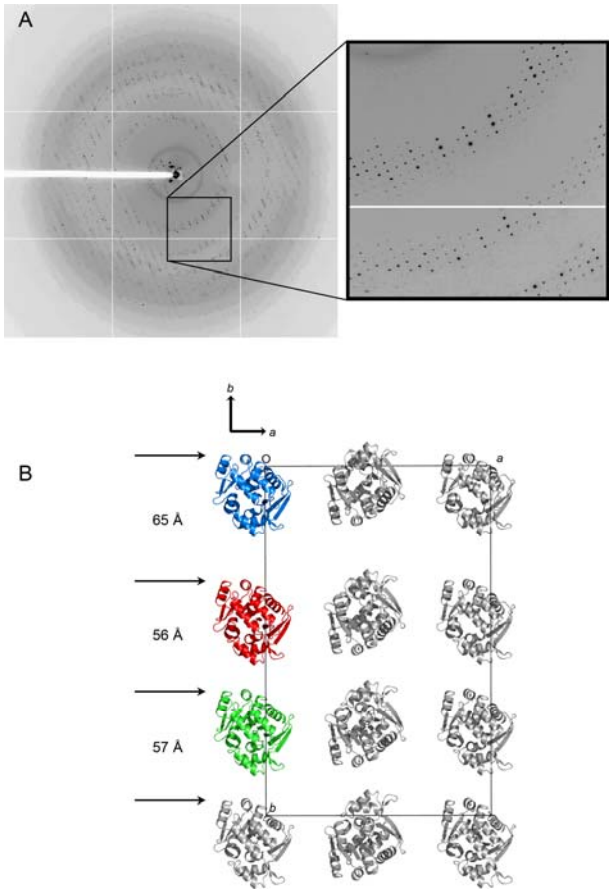


Figure 7.2. Pseudo-translational symmetry in *E*-2AMS hydrolase crystals. (A) A typical diffraction pattern for *E*-2AMS hydrolase. At low resolution, every third row of reflections is much more intense than other rows. (B) Crystal packing of *E*-2AMS hydrolase. For clarity only one third of the unit cell is shown, with the three unique chains colored blue, red and green. The crystallographically related protomers are shown in grey.

The twelve protomers of *E*-2AMS hydrolase in the asymmetric unit form six pairs of dimers with identical interfaces. The two protomers are related to each other by local twofold symmetry. The β-sheets of each chain face in the same direction, but twist in opposite directions, as seen in Figure 7.4. The dimer interface involves the loop region between α9 and β10, residues from β10, and the C-terminal end of α10 from both protomers. Each dimer is about 95 Å long with a total surface area of nearly 20,000 $Å^2$. Roughly 2000 $Å^2$ is buried at the dimer interface (*31*). Two hydrogen bonds are formed between the carbonyl oxygen and amide nitrogen atoms of

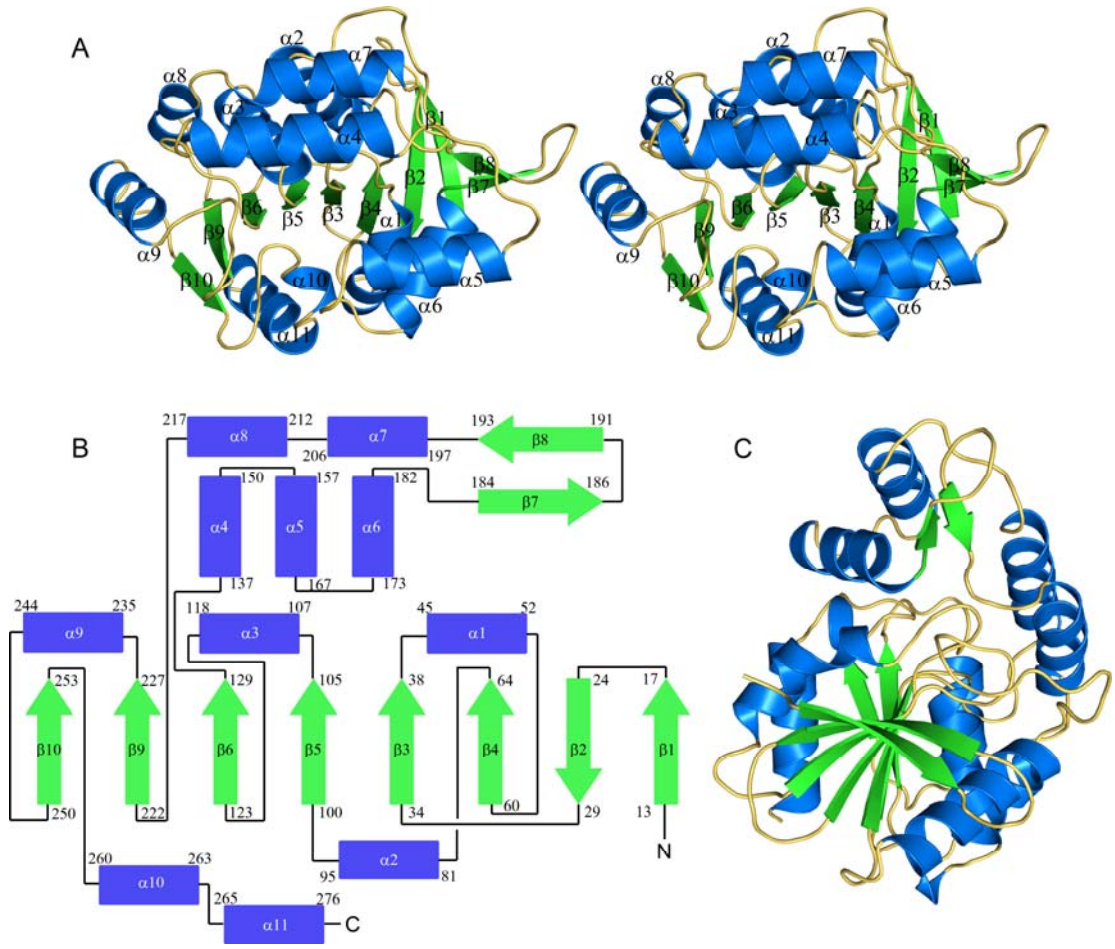


Figure 7.3. Monomeric structure of *E*-2AMS hydrolase. (A) Stereoview ribbon diagram of *E*-2AMS hydrolase with the secondary structures labelled. β-strands are shown in green and α-helices are shown in blue. (B) Topology diagram of *E*-2AMS hydrolase, using the same color scheme as A. (C) Ribbon diagram of *E*-2AMS hydrolase illustrating the sharp twist in the β-sheet running through the core of the enzyme and the cap domain.

the protein backbone at Val250 and Val250* (where * indicates the second protomer of the dimer). A salt bridge is formed between Arg243 and Glu228*, a hydrogen bond is found between Asn274 and Aps247*, and Arg124 forms interactions with the oxygen atoms of the carboxy terminus at Ala278*. Additionally, Phe275, Pro221, Leu223, Pro 249, Val250, Val251, and Ala271 from each protomer form an extended hydrophobic patch at the dimer interface.

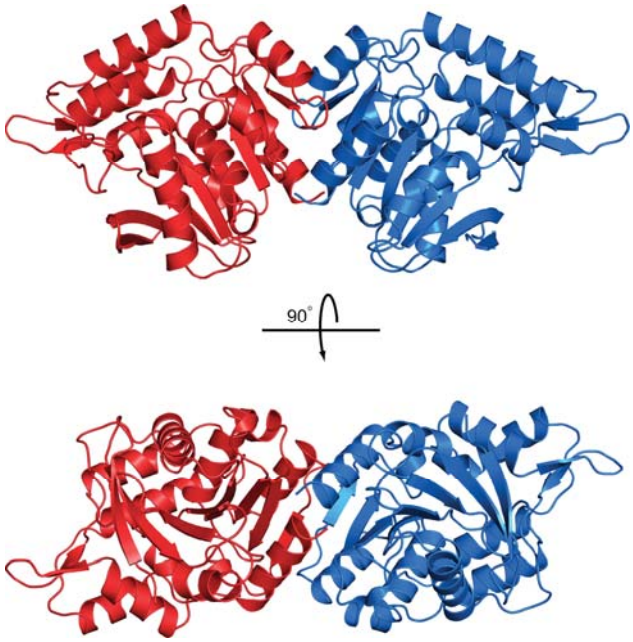


Figure 7.4. Dimeric structure of *E*-2AMS hydrolase. Each chain is colored differently, one red and the second blue. The second orientation obtained by rotating the first orientation by 90° along the horizontal axis.

Active Site of E-2AMS Hydrolase. The putative active site was identified through sequence and structural homology with other α/β hydrolases and is located at the C-terminal end of the β-sheet and is shielded from solvent by the cap domain. Three positively charged residues, Arg146, Arg167, and Lys231, are present at the entrance to the active site. The active site pocket has two distinct environments, as seen in Figure 7.5. The narrow cleft adjacent to Ser106 is lined with primarily hydrophobic residues, including Ile41, Leu207, Leu143, Leu232, and Phe131. The second half of the active site, however, is more hydrophilic. This pocket, roughly 10 Å deep and 6 Å wide, is filled with water molecules coordinated to hydrophilic residues. Arg179 forms hydrogen bonds to two water molecules and is oriented in the active site by Glu262. Three tyrosine residues, Tyr184, Tyr168, and Tyr259, are positioned so the hydroxyl group faces into the active site and forms hydrogen bonds with water molecules. Other hydrophilic residues in the active site include Thr42, Ser43, and His105.

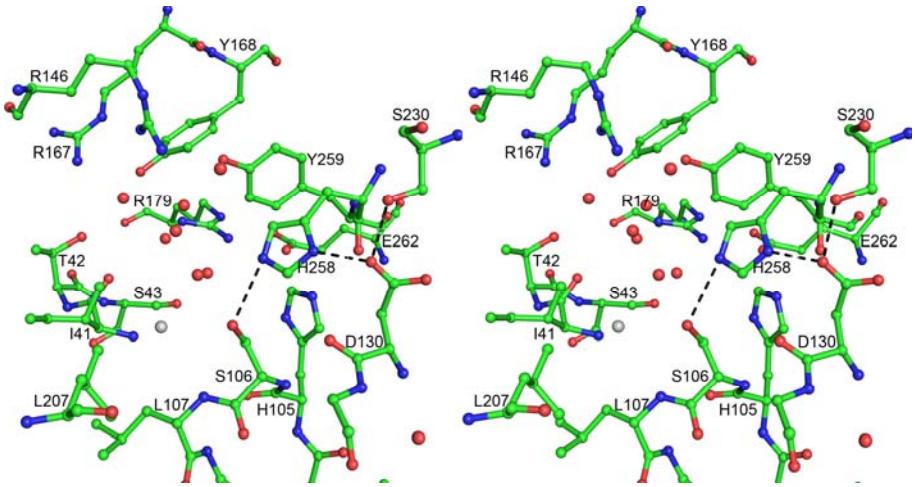


Figure 7.5. Stereoview ball and stick diagram of the active site of *E*-2AMS hydrolase. Water molecules are shown as red nonbonded spheres and the chloride ion is shown as a nonbonded grey sphere. Key protein interactions are shown as dashed lines.

The putative nucleophilic residue, Ser106, adopts a strained conformation at the C-terminal end of β5 and is hydrogen bonded to a structurally conserved water molecule, which is also coordinated to the backbone amide nitrogen atoms of Leu107 and Leu41. A chloride ion is bound in a pocket next to Ser106. The chloride ion is bound in an oxyanion hole, coordinated to Ser106 2.9 Å away and is also forming interactions with two backbone amide nitrogen atoms from Ile41 and Leu107 (3.2 and 3.3 Å, respectively). His258, located after β10, is 3.0 Å from Ser106 and is itself positioned by Asp130, 2.5 Å away. Asp130 is found at the end of β6 and Ser230 occurs after β9. Ser230 forms a strong hydrogen bond with Asp130 (2.7 Å) and a second hydrogen bond with the amide nitrogen atom from the protein backbone of Leu232. These interactions are shown in Figure 7.5.

Activity of E-2AMS Hydrolase Mutants. The extinction coefficient at 261 nm for *E*-2AMS is 18600 M^{-1} cm^{-1} as previously determined, and the K_M and k_{cat} values for *E*-2AMS hydrolase have also been reported (*7, 10, 32*). The activity of the active site mutants S106A, D130N, and S230C were assayed by measuring the decrease in

the absorbance at 261 nm, the λ_{max} for *E*-2AMS as previously described (*7, 10*). Each of these mutants was found to be completely inactive relative to the native enzyme, which retained an activity comparable to published results. The active site mutants S230A and S230N could not be assayed because these proteins precipitated in the assay buffer. The inactive S106A mutant was then used for cocrystallization and soaking experiments with *E*-2AMS, which were unsuccessful in producing an enzyme-substrate complex.

Section 7.4. Discussion

Comparison of E-2AMS Hydrolase to Other Members of the α/β Hydrolase Superfamily. The α/β hydrolase superfamily comprises a very large class of enzymes, catalyzing myriad reactions and related through similar tertiary structures. This family has been structurally well characterized and includes esterases, lipases, epoxidases, alkane dehalogenases, carbon-carbon bond hydrolases, amidohydrolases, and thioesterases (*33-34*). Previous analysis of the primary sequence for *E*-2AMS hydrolase using BLAST suggested that this enzyme was a member of the α/β hydrolase superfamily, despite a low sequence identity to this family of enzymes (*10*). A DALI search was performed using the structure of *E*-2AMS hydrolase and these results confirmed that *E*-2AMS hydrolase is a member of the α/β hydrolase family. Representative members of this superfamily are shown in Table 3.3 and their similarity to *E*-2AMS hydrolase is given by their Z-score, rmsd, sequence identity, and the number of residues aligned (*34-40*). The core of *E*-2AMS hydrolase is formed by an eight stranded β-sheet flanked on both sides by α-helices with a strand order that agrees with the other members of this superfamily (*41-42*). The cap domain is an insertion between β6 and β9, with a short two stranded antiparallel β-sheet and four α-helices. The cap domain acts to shield the active site from solvent. While this second domain is not part of the α/β hydrolase fold, insertions are common after the fifth β-

strand of the core β-sheet (*42*). Additionally, the cap domain is not conserved. For example, the cap domains of the serine hydrolase CarC and the carbon-carbon hydrolase MhpC (*34, 37*) are composed of three long α-helices and the cap domain from an α/β hydrolase of *Novosphingobium aromaticivorans* DSM 12444 has a three stranded antiparallel β-sheet flanked by four α-helices (PBD ID: 3BWX).

Table 7.3. Enzymes Identified as Structurally Similar to *E*-2AMS Hydrolase by DALI

Protein	**PDB ID**	**Z Score**	**RMSD**	**% ID**	**# aligned residues**
α/β Hydrolase, YP_496220.1	3BWX	30.4	2.4	23	263
Serine hydrolase (CarC)	1J1I	29.1	2.2	21	238
Chloroperoxidase L	1A88	28.7	2.4	21	246
Arylesterase (PFE)	1VA4	28.7	2.5	23	247
Esterase (EST)	1ZOI	28.5	2.5	21	247
Esterase (ybfF)	3BF7	27.5	2.5	21	242
Fluoroacetate dehalogenase	1Y37	27.3	2.7	24	246
C-C bond hydrolase (MhpC)	1U2E	26.3	2.6	18	251

Figure 7.6 shows a sequence alignment of similar enzymes identified through DALI (*43-44*). The "nucleophilic elbow" begins at the C-terminal end of β5 in *E*-2AMS hydrolase and is identified using the fingerprint motif Gly-X-Nu-X-Gly, where X is any amino acid and Nu is the nucleophilic amino acid necessary for catalysis. In *E*-2AMS hydrolase, the nucleophile is Ser106. The sequence alignment identified few conserved residues; the majority being glycine residues, presumably important for proper folding. An absolutely conserved histidine residue required for catalysis in the α/β hydrolase family is found after the last β-strand in the core domain. Additionally, the seventh β-strand of the eight stranded β-sheet is mostly conserved, with hydrophobic residues forming interactions with neighboring β-strands.

Active Site Comparison. The α/β hydrolase superfamily utilizes a catalytic triad consisting of a nucleophile, a strictly conserved histidine residue, and an acidic

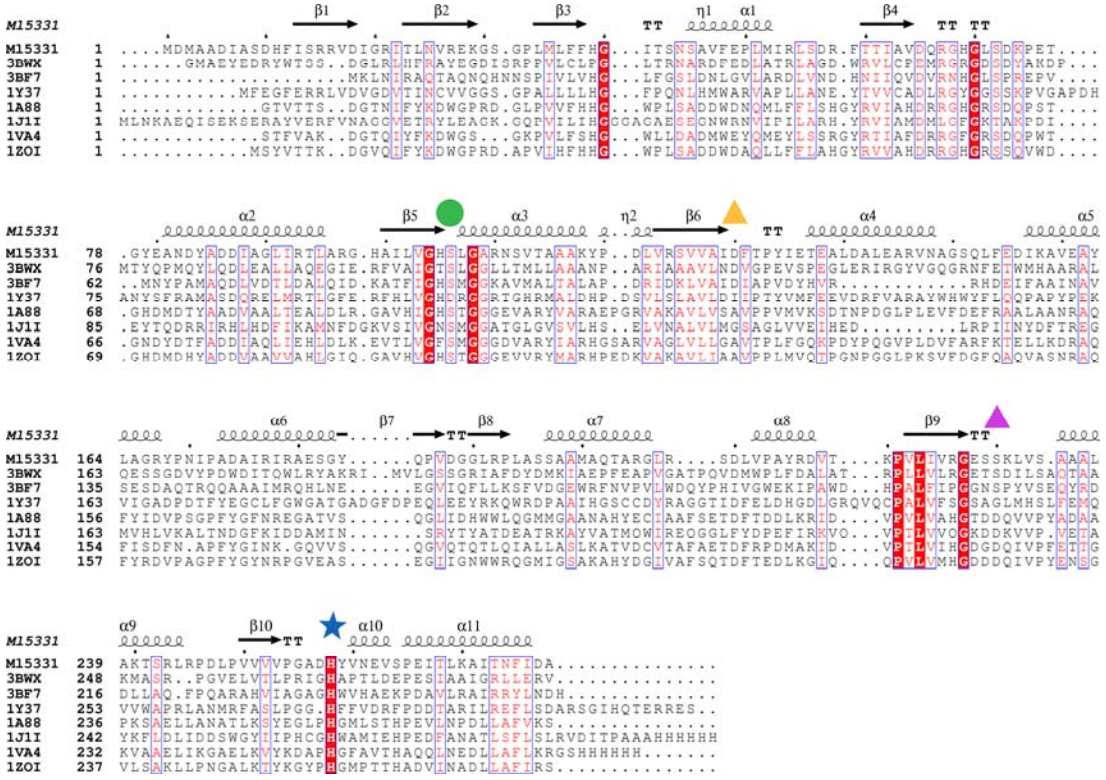


Figure 7.6. Sequence alignment of *E*-2AMS hydrolase with other enzymes identified as structurally homologous by DALI. Included enzymes are the α/β hydrolase YP_496220.1 (3BWX), CarC (1J1I), chloroperoxidase L (1A88), the PFE arylesterase (1VA4), the EST esterase (1ZOI), ybfF (3BF7), fluoroacetate dehalogenase (1Y37), and MhpC (1U2E). The nucleophile is marked with a green circle, the catalytic histidine is labelled with a blue star, and the different active site acids are labelled using triangles. The expected acidic residues have a purple triangle, while the location of the acidic residue in *E*-2AMS hydrolase has a yellow triangle.

residue, usually an aspartate residue. The nucleophile, identified as Ser106, and the catalytic histidine residue, His258, are located in structurally conserved positions at the interface of the two domains of *E*-2AMS hydrolase. However, the third residue of the catalytic triad is not located at the C-terminal end of the seventh β-strand as expected. Instead, His258 forms a hydrogen bond with Asp130 found after β6 with Ser230 occupying the turn between β9 and α9. Figure 7.7 shows the superposition of the active site of *E*-2AMS hydrolase (green) with the esterase ybfF (white) from *E. coli* (PDB ID: 3BF7) and the carbon-carbon bond hydrolase MhpC (purple) from *E.*

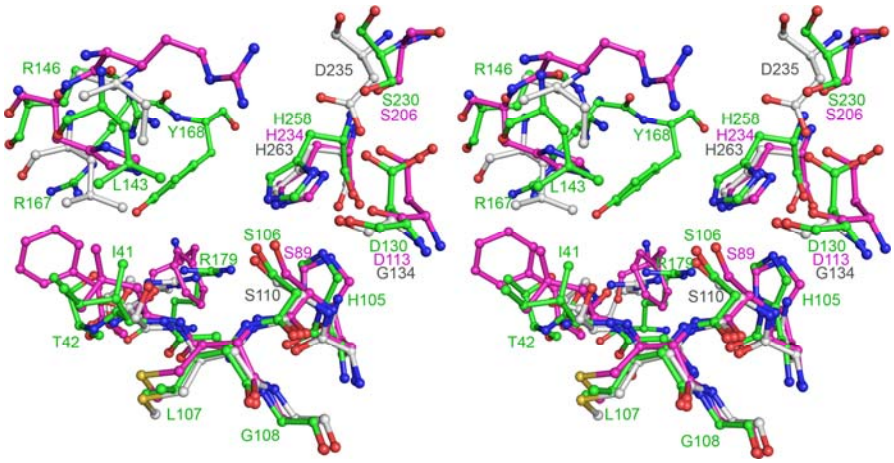


Figure 7.7. Stereoview ball and stick diagram comparing the active site of *E*-2AMS hydrolase to other members of the α/β hydrolase superfamily. The *E*-2AMS hydrolase structure is shown with green carbon atoms, ybfF is shown using magenta carbon atoms, and the carbon-carbon bond hydrolase MhpC has white carbon atoms. All residues of *E*-2AMS hydrolase are labelled and the catalytic residues of ybfF and MhpC are labelled.

coli (PDB ID: 1U2E) (*34, 40*). All three structures have an oxyanion hole formed by the amide nitrogen atoms of the protein backbone using the residue adjacent to the nucleophile (Leu107 in *E*-2AMS hydrolase) and a residue at the C-terminal end of the fourth β-strand (Ile41 at the end of β3 in *E*-2AMS hydrolase). The catalytic triads observed in *E*-2AMS hydrolase and ybfF are similar. The histidine responsible for activation of the nucleophile is 3.0 Å from Ser106 and 2.9 Å in the ybfF structure. The histidine residues in these two structures are each then coordinated to one oxygen atom of the carboxylate group of the acidic residue at the end of β6. The catalytic triad seen in MhpC, shown in purple in Figure 7.7, adopts the more common arrangement. The nucleophile, Ser110, is positioned 3.3 Å from the Nε2 atom of the side chain of His263. The aspartate residue responsible for activation of His263, Asp235, is coordinated to the Nδ1 atom of the histidine side chain using both oxygen atoms of the carboxylate side chain. Asp235 is located at the end of the seventh β-strand of the β-sheet. The remaining residues lining the active site are not conserved

and show no sequence similarity. However, careful sequence analysis has identified several residues apart from the catalytic triad that can help identify the reaction catalyzed and suggest *E*-2AMS hydrolase is quite similar to the esterase subclass of the α/β hydrolase superfamily (*45-46*). The residue preceding the active site nucleophile is a histidine and two residues before the catalytic histidine is a hydrophobic residue in esterases (His105 and Ala261, respectively, in *E*-2AMS hydrolase). While the reaction catalyzed is a hydrolysis of an amide bond instead of an ester linkage, the requirements and chemistry are very similar. Additionally, the arrangement of active site residues in which the acidic residue is found at the end of β6 and a fourth catalytic residue, a serine at the C-terminal end of the seventh β-strand, is not uncommon in esterases (*40*).

Mechanistic Implications for E-2AMS Hydrolase. Previous characterization of *E*-2AMS hydrolase has suggested a possible mechanism for the hydrolysis of *E*-2AMS to form succinic semialdehyde, acetate, ammonia, and carbon dioxide, shown in Figure 7.8 (*7*). This mechanism utilizes an active site base to activate a water molecule for nucleophilic attack on the carbonyl carbon atom of the amide group. Collapse of the tetrahedral intermediate then leads to the loss of acetate. Tautomerization of intermediate 11 is followed by loss of carbon dioxide and release of ammonia through the addition of water. The last two steps could also occur in the reverse order. Direct addition of an active site nucleophile to the amide carbonyl carbon atom is also a possibility (Figure 7.8, Mechanism 2). Direct attack by the active site nucleophile is most commonly observed in serine proteases, which are not members of the α/β superfamily but have a similar active site triad, composed of serine, histidine, and aspartate (*47*). However, 2,4-dioxygenases, which have the α/β hydrolase fold, are believed to function through direct attack of the nucleophilic serine residue at a carbonyl carbon of the substrate to initiate catalysis (*48*). Another

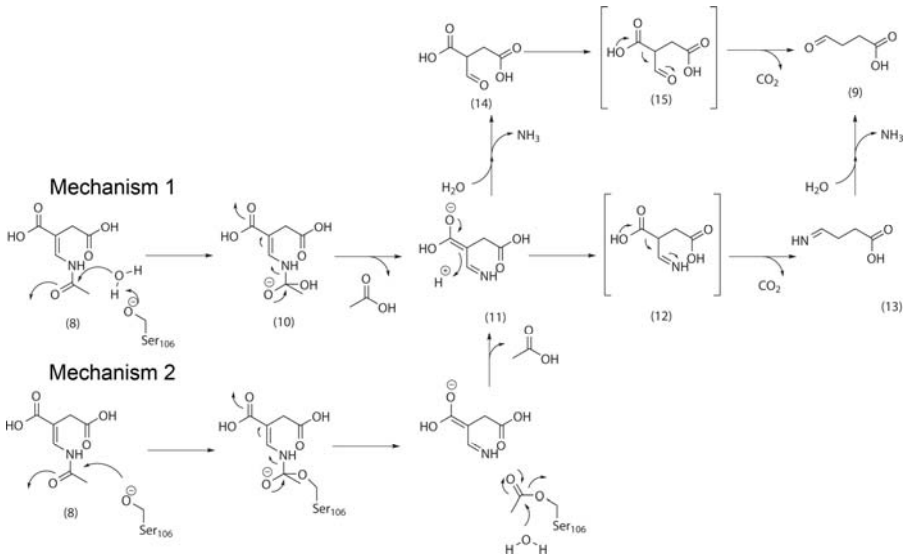


Figure 7.8. Proposed mechanisms for the hydrolysis of *E*-2AMS to produce succinic semialdehyde, acetate, ammonia, and carbon dioxide. Mechanism 1 utilizes Ser106 to activate a water molecule for attack at the carbonyl carbon of the amide bond, while Mechanism 2 utilizes a direct attack on *E*-2AMS by Ser106.

example of direct attack by the nucleophilic serine is found in an aryl esterase from *Pseudomonas fluorescens* capable of catalyzing both ester and bromoperoxide bonds (*35*). A third possibility for the hydrolysis of *E*-2AMS to form intermediate 10 utilizes a general base mechanism, where His258 deprotonates a water molecule for attack on the carbonyl carbon, leading to loss of acetate and intermediate 11 through tautomerization. In this mechanism, Ser106 could serve to help stabilize the negatively charged tetrahedral intermediate found in the oxyanion hole. The general base mechanism has precedence; the carbon-carbon bond hydrolase MhpC from *E. coli* is believed to use His263 for activation of water and Ser110 for stabilization of the intermediate after tautomerization of its substrate, 2-hydroxy-6-keto-nona-2,4-diene-1,9-dioic acid (*34, 49-50*).

Although all attempts at cocrystallization or soaking *E*-2AMS into crystals of *E*-2AMS hydrolase or into crystals of the *E*-2AMS hydrolase S106A mutant were unsuccessful, examination of the active site suggests several residues that could play

important roles in both the binding of *E*-2AMS and the catalysis of *E*-2AMS hydrolase (Figure 7.9). The catalytic triad, shown in Figure 7.4, consists of Ser106, His258, and Asp130 and most likely functions as seen in other α/β hydrolase enzymes (*42*). Briefly, Asp130 activates His258 for deprotonation of Ser106, which could act as a nucleophile for direct attack on the *E*-2AMS, or activate a water molecule for attack on the substrate. Mutation of Ser106 to alanine abolished all catalytic activity, confirming this residue's critical importance for activity. The D130N mutant also showed no activity, indicating that the activation and stabilization of the catalytic histidine residue is necessary for catalysis. Ser230, found after β9 in the position most commonly occupied by the acidic residue of the catalytic triad, is more than 4 Å from His258 but forms a strong hydrogen bond (2.7 Å) with Asp130. Ser230 may play an important role in the structural integrity of the enzyme because the S230A, S230C and S230N mutants were all unstable. It is possible the slightly larger sulphur atom of cysteine in the S230C mutant could disrupt this sharp turn between β9 and α9. Other interactions involving Ser230 include two hydrogen bonds to amide nitrogen atoms of the protein backbone of Leu232 and Val233 in a β-turn.

The active site has several positively charged residues which could be involved in coordination to the carboxylate groups of *E*-2AMS (Figure 7.4). Three arginine residues, Arg146, Arg167, and Arg179, face into the active site. Additionally, Arg179 forms two salt bridges with Glu262 to help orient Arg179 within the active site. Manual placement of the substrate, such that the carbonyl carbon of the amide group is located near Ser106 in the active site and the carbonyl oxygen points towards the oxyanion hole, suggests Arg146 and Arg179 are both likely candidates for coordination to the carboxylate groups and could play a role in the proper orientation of *E*-2AMS within the active site (Figure 7.9). Either arginine residue could also stabilize the enol intermediate and serve as a proton donor in the first mechanism.

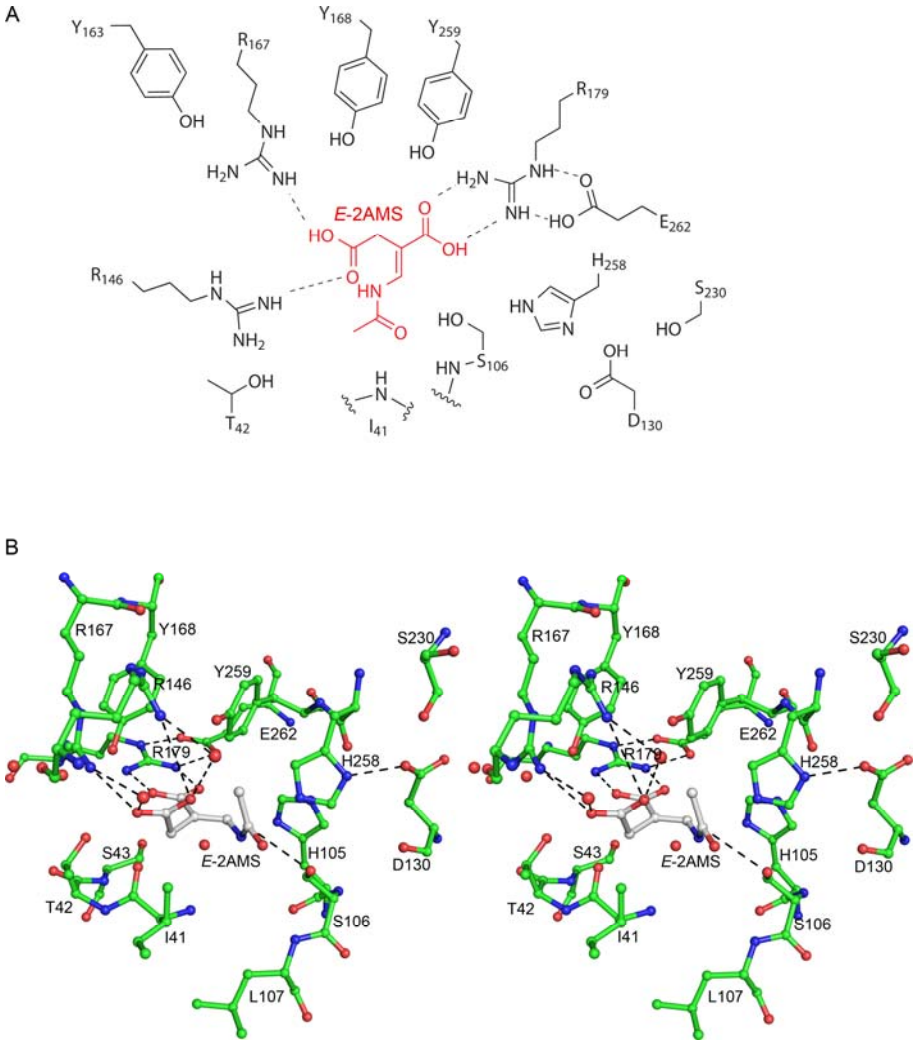


Figure 7.9. Possible coordination of *E*-2AMS within the active site. (A) Schematic diagram where *E*-2AMS (shown in red) was manually positioned such that either Ser106 or an activated water molecule could attack the carbonyl carbon of the amide group. Potential hydrogen bonds are shown as dashed lines. (B) Stereoview ball and stick diagram of the *E*-2AMS hydrolase active site with *E*-2AMS manually positioned in the active site. Carbon atoms of *E*-2AMS hydrolase are colored green and carbon atoms of *E*-2AMS are colored gray. Possible interactions and hydrogen bonds important for active site orientation are shown in dashed lines.

Protonation of the enol intermediate 11 could utilize Tyr166, Tyr259, or His105, either directly or using an activated water molecule. These same residues could also be involved in catalyzing the decarboxylation reaction leading to succinic semialdehyde.

REFERENCES

1. Eliot, A. C., and Kirsch, J. F. (2004) Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations, *Annu. Rev. Biochem. 73*, 383-415.
2. Snell, E. E., and Haskell, B. E. (1971) *The Metabolism of Vitamin B6, In: Comprehensive Biochemistry*, Vol. 21, Elsevier/North Holland, New York.
3. Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S., Watanabe, A., Idesawa, K., Ishikawa, A., Kawashima, K., Kimura, T., Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A., Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimpo, S., Sugimoto, M., Takeuchi, C., Yamada, M., and Tabata, S. (2000) Complete genome structure of the nitrogen-fixing symbiotic bacterium Mesorhizobium loti, *DNA Res. 7*, 331-338.
4. Yuan, B., Yoshikane, Y., Yokochi, N., Ohnishi, K., and Yagi, T. (2004) The nitrogen-fixing symbiotic bacterium Mesorhizobium loti has and expresses the gene encoding pyridoxine 4-oxidase involved in the degradation of vitamin B6, *FEMS Microbiol. Lett. 234*, 225-230.
5. Funami, J., Yoshikane, Y., Kobayashi, H., Yokochi, N., Yuan, B., Iwasaki, K., Ohnishi, K., and Yagi, T. (2005) 4-Pyridoxolactonase from a symbiotic nitrogen-fixing bacterium Mesorhizobium loti: cloning, expression, and characterization, *Biochim. Biophys. Acta 1753*, 234-239.
6. Ge, F., Yokochi, N., Yoshikane, Y., Ohnishi, K., and Yagi, T. (2008) Gene identification and characterization of the pyridoxine degradative enzyme 4-pyridoxic acid dehydrogenase from the nitrogen-fixing symbiotic bacterium Mesorhizobium loti MAFF303099, *J. Biochem. 143*, 603-609.

7. Mukherjee, T., Hilmey, D. G., and Begley, T. P. (2008) PLP catabolism: identification of the 2-(Acetamidomethylene)succinate hydrolase gene in Mesorhizobium loti MAFF303099, *Biochemistry 47*, 6233-6241.
8. Mukherjee, T., Kinsland, C., and Begley, T. P. (2007) PLP catabolism: identification of the 4-pyridoxic acid dehydrogenase gene in Mesorhizobium loti MAFF303099, *Bioorg. Chem. 35*, 458-464.
9. Yokochi, N., Nishimura, S., Yoshikane, Y., Ohnishi, K., and Yagi, T. (2006) Identification of a new tetrameric pyridoxal 4-dehydrogenase as the second enzyme in the degradation pathway for pyridoxine in a nitrogen-fixing symbiotic bacterium, Mesorhizobium loti, *Arch. Biochem. Biophys. 452*, 1-8.
10. Yuan, B., Yokochi, N., Yoshikane, Y., Ohnishi, K., Ge, F., and Yagi, T. (2008) Gene identification and characterization of the pyridoxine degradative enzyme alpha-(N-acetylaminomethylene)succinic acid amidohydrolase from Mesorhizobium loti MAFF303099, *J. Nutr. Sci. Vitaminol. (Tokyo) 54*, 185-190.
11. Yuan, B., Yokochi, N., Yoshikane, Y., Ohnishi, K., and Yagi, T. (2006) Molecular cloning, identification and characterization of 2-methyl-3-hydroxypyridine-5-carboxylic-acid-dioxygenase-coding gene from the nitrogen-fixing symbiotic bacterium Mesorhizobium loti, *J. Biosci. Bioeng. 102*, 504-510.
12. Mukherjee, T., McCulloch, K. M., Ealick, S. E., and Begley, T. P. (2007) Gene identification and structural characterization of the pyridoxal 5'-phosphate degradative protein 3-hydroxy-2-methylpyridine-4,5-dicarboxylate decarboxylase from mesorhizobium loti MAFF303099, *Biochemistry 46*, 13606-13615.

13. Yoshikane, Y., Yokochi, N., Yamasaki, M., Mizutani, K., Ohnishi, K., Mikami, B., Hayashi, H., and Yagi, T. (2008) Crystal structure of pyridoxamine-pyruvate aminotransferase from Mesorhizobium loti MAFF303099, *J. Biol. Chem. 283*, 1120-1127.
14. McCulloch, K. M., Mukherjee, T., Begley, T. P., and Ealick, S. E. (2009) Structure of the PLP degradative enzyme 2-methyl-3-hydroxypyridine-5-carboxylic acid oxygenase from Mesorhizobium loti MAFF303099 and its mechanistic implications, *Biochemistry 48*, 4139-4149.
15. Kress, D., Alhapel, A., Pierik, A. J., and Essen, L. O. (2008) The crystal structure of enamidase: a bifunctional enzyme of the nicotinate catabolism, *J. Mol. Biol. 384*, 837-847.
16. 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.
17. Matthews, B. W. (1968) Solvent content of protein crystals, *J. Mol. Biol. 33*, 491-497.
18. Otwinowski, Z., and Minor, W. (1997) Processing of x-ray diffraction data collected in oscillation mode, *Methods Enzymol. 276*, 307-326.
19. Pape, T., and Schneider, T. R. (2004) HKL2MAP: a graphical user interface for phasing with SHELX programs, *J. Appl. Cryst. 37*, 843-844.
20. Schneider, T. R., and Sheldrick, G. M. (2002) Substructure solution with SHELXD, *Acta Cryst. D58*, 1772-1779.
21. Sheldrick, G. M. (2002) Macromolecular phasing with SHELXE, *Z Kristallogr 217*, 644-650.
22. 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 Cryst. D54*, 905-921.

23. Terwilliger, T. C. (2000) Maximum-likelihood density modification, *Acta Cryst. D56*, 965-972.
24. Terwilliger, T. C. (2003) Automated main-chain model building by template matching and iterative fragment extension, *Acta Cryst. D59*, 38-44.
25. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Cryst. D60*, 2126-2132.
26. Vellieux, F. M., and Read, R. J. (1997) Noncrystallographic symmetry averaging in phase refinement and extension, *Methods Enzymol. 277*, 18-53.
27. 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. Cryst. 26*, 283-291.
28. DeLano, W. L. (2002) The PyMOL Molecular Graphics Systems, DeLano Scientific, San Carlos, CA.
29. Sambrook, J., Fritsch, G. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Guide*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
30. Chook, Y. M., Lipscomb, W. N., and Ke, H. (1998) Detection and use of pseudo-translation in determination of protein structures, *Acta Cryst. D54*, 822-827.
31. Hasel, W., Hendrickson, T. F., and Still, W. C. (1988) A rapid approximation to the solvent-accessible surface areas of atoms, *Tetrahedron Comp. Meth. 1*, 103-116.

32. Huynh, M. S., and Snell, E. E. (1985) Enzymes of vitamin B6 degradation. Purification and properties of two N-acetylamidohydrolases, *J. Biol. Chem. 260*, 2379-2383.

33. Du, X., Wang, W., Kim, R., Yakota, H., Nguyen, H., and Kim, S. H. (2001) Crystal structure and mechanism of catalysis of a pyrazinamidase from Pyrococcus horikoshii, *Biochemistry 40*, 14166-14172.

34. Dunn, G., Montgomery, M. G., Mohammed, F., Coker, A., Cooper, J. B., Robertson, T., Garcia, J. L., Bugg, T. D., and Wood, S. P. (2005) The structure of the C-C bond hydrolase MhpC provides insights into its catalytic mechanism, *J. Mol. Biol. 346*, 253-265.

35. Cheeseman, J. D., Tocilj, A., Park, S., Schrag, J. D., and Kazlauskas, R. J. (2004) Structure of an aryl esterase from Pseudomonas fluorescens, *Acta Cryst. D60*, 1237-1243.

36. Elmi, F., Lee, H. T., Huang, J. Y., Hsieh, Y. C., Wang, Y. L., Chen, Y. J., Shaw, S. Y., and Chen, C. J. (2005) Stereoselective esterase from Pseudomonas putida IFO12996 reveals alpha/beta hydrolase folds for D-beta-acetylthioisobutyric acid synthesis, *J. Bacteriol. 187*, 8470-8476.

37. Habe, H., Morii, K., Fushinobu, S., Nam, J. W., Ayabe, Y., Yoshida, T., Wakagi, T., Yamane, H., Nojiri, H., and Omori, T. (2003) Crystal structure of a histidine-tagged serine hydrolase involved in the carbazole degradation (CarC enzyme), *Biochem. Biophys. Res. Commun. 303*, 631-639.

38. Hofmann, B., Tolzer, S., Pelletier, I., Altenbuchner, J., van Pee, K. H., and Hecht, H. J. (1998) Structural investigation of the cofactor-free chloroperoxidases, *J. Mol. Biol. 279*, 889-900.

39. Holm, L., and Sander, C. (1998) Touring protein fold space with Dali/FSSP, *Nucleic Acids Res. 26*, 316-319.

40. Park, S. Y., Lee, S. H., Lee, J., Nishi, K., Kim, Y. S., Jung, C. H., and Kim, J. S. (2008) High-resolution structure of ybfF from Escherichia coli K12: a unique substrate-binding crevice generated by domain arrangement, *J. Mol. Biol. 376*, 1426-1437.

41. Heikinheimo, P., Goldman, A., Jeffries, C., and Ollis, D. L. (1999) Of barn owls and bankers: a lush variety of alpha/beta hydrolases, *Structure 7*, R141-146.

42. Nardini, M., and Dijkstra, B. W. (1999) Alpha/beta hydrolase fold enzymes: the family keeps growing, *Curr. Opin. Struct. Biol. 9*, 732-737.

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

44. 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.

45. Li, C., Hassler, M., and Bugg, T. D. (2008) Catalytic promiscuity in the alpha/beta-hydrolase superfamily: hydroxamic acid formation, C--C bond formation, ester and thioester hydrolysis in the C--C hydrolase family, *ChemBioChem 9*, 71-76.

46. Li, C., Li, J. J., Montgomery, M. G., Wood, S. P., and Bugg, T. D. (2006) Catalytic role for arginine 188 in the C-C hydrolase catalytic mechanism for Escherichia coli MhpC and Burkholderia xenovorans LB400 BphD, *Biochemistry 45*, 12470-12479.

47. Hedstrom, L. (2002) Serine protease mechanism and specificity, *Chem. Rev. 102*, 4501-4524.

48. Fischer, F., Kunne, S., and Fetzner, S. (1999) Bacterial 2,4-dioxygenases: new members of the alpha/beta hydrolase-fold superfamily of enzymes functionally related to serine hydrolases, *J. Bacteriol. 181*, 5725-5733.

49. Li, C., Montgomery, M. G., Mohammed, F., Li, J. J., Wood, S. P., and Bugg, T. D. (2005) Catalytic mechanism of C-C hydrolase MhpC from Escherichia coli: kinetic analysis of His263 and Ser110 site-directed mutants, *J. Mol. Biol. 346*, 241-251.

50. Li, J. J., and Bugg, T. D. (2007) Investigation of a general base mechanism for ester hydrolysis in C-C hydrolase enzymes of the alpha/beta-hydrolase superfamily: a novel mechanism for the serine catalytic triad, *Org. Biomol. Chem. 5*, 507-513.

CHAPTER 8

SUMMARY

The examination of structure through the use of X-rays and macromolecular crystals has been providing insight into mechanistic enzymology, rational drug design, and enzyme activity identification for decades. The elegant experiment of taking a pure protein solution, growing crystals, collecting the diffraction patterns, and determining the structure of a protein molecule at atomic resolution has revolutionized our understanding of the molecular machinery of the cell and the various pathways which exist within the cell. Knowing the precise positioning of each atom within a protein structure or complex often allows the determination of key catalytic residues, which can in turn suggest mechanistic possibilities. An alternative scenario in which structural biology can be very helpful is when two distinct mechanisms appear plausible. Often, the structure of the enzyme in question can provide evidence for one mechanism over another. An example of this is the determination that thiamin phosphate synthase utilizes a dissociative mechanism (*1*). Before the crystal structures were available both S_N1 and S_N2 type mechanisms were being considered; the complexes were able to illuminate the reaction coordinate by trapping intermediate species which represented the carbocation formed during the dissociative mechanism.

The thiamin biosynthetic pathway has been extensively probed over the last twenty years, yet interesting and novel chemistry, and surprising variations of this pathway, is still being discovered (*2*). The thiazole and pyrimidine moieties of thiamin are biosynthesized separately and then coupled together (Figure 1.1). The last portion of the thiamin biosynthetic pathway consists of this coupling reaction, followed by a phosphorylation event to produce the biologically active form of vitamin B_1, thiamin pyrophosphate. Recent work has shown that these final steps in thiamin biosynthesis contain interesting chemistry. The true product of the thiazole

biosynthetic branch is not the expected 4-methyl-5-(hydroxyethyl)thiazole phosphate, but a non-aromatic carboxylated thiazole phosphate tautomer (*3*). TenI, a protein closely related to thiamin phosphate synthase, has been implicated as an aromatase to generate 2-carboxyl-4-methyl-5-(hydroxyethyl)thiazole phosphate (Figure 1.1). The work presented in Chapter Two presents the structure determination of *B. subtilis* thiamin phosphate synthase obtained through a brief soak with a trifluorinated pyrimidine analogue and the carboxylated thiazole phosphate. The resulting structure clearly showed three species with disconnected electron density, representing an intermediate complex along the reaction coordinate toward thiamin phosphate. This data, in conjunction with NMR and steady state kinetics, demonstrated that the carboxylated thiazole phosphate is a substrate for thiamin phosphate synthase and is suggestive that the coupling of the pyrimidine and thiazole moieties occurs before the decarboxylation event takes place. An absolutely conserved active site lysine residue may be involved in the decarboxylation reaction; structural studies are underway to attempt to crystallize the K159A mutant with the carboxylated thiazole phosphate to determine if the carboxylated thiamin phosphate can be structurally characterized. Chapter Three presents a second structure of thiamin phosphate synthase, from *M. tuberculosis*. Because *M. tuberculosis* lacks a thiamin uptake or salvage pathway, the bacterium is reliant upon the *de novo* biosynthetic pathway to generate thiamin (*4*). The thiamin biosynthetic pathway is also a potential pathway to target as humans do not contain these enzymes and thus would not need to be concerned about inhibition of thiamin biosynthesis is man. This crystal structure of thiamin phosphate synthase was determined to 2.3 Å resolution, solved by molecular replacement, and is similar to the structures of other thiamin phosphate synthases. The active site contains two cysteine residues, however, that are not observed in other thiamin phosphate synthases. This structure was then used to screen a virtual library of small molecules

and the top hits were then scored based on their virtual docking and the free binding energy calculated for each molecule. The best 25 compounds were then subjected to a novel *in vivo* thiamin detection assay in which β-galactosidase becomes upregulated by the thiamin riboswitch in response to low thiamin levels (Table 3.3). Three compounds first identified by the virtual screening showed some inhibition of thiamin phosphate synthase. These compounds will be pursued structurally.

The final enzyme involved in thiamin biosynthesis is thiamin monophosphate kinase (ThiL), which catalyzes the transfer of the γ-phosphate group from ATP to thiamin monophosphate to generate thiamin pyrophosphate. Previous work on the purine biosynthetic pathway identified a novel ATP binding superfamily which consists of only five members: aminoimidazole ribonucleotide synthetase (PurM), *N*-formylglycinamidine ribonucleotide synthetase (PurL), HypE, a nickel-iron hydrogenase maturation protein, selenophosphate synthetase (SelD), and ThiL. The enzymes were hypothesized to utilize similar mechanisms and form a phosphoimidate intermediate, as seen in Figure 4.10 (*5*). However, it was unclear why a simple phosphate transfer would require this type of intermediate or if ThiL instead formed a phosphoenzyme intermediate. The crystal structures presented in Chapter Four describe four complexes: AaThiL with ATP, AaThiL with AMP-PCP, AaThiL with AMP-PCP and thiamin phosphate, and AaThiL with ADP and thiamin pyrophosphate at resolutions ranging from 1.5 Å to 2.6 Å. An active site serine residue was the only available residue that could be phosphorylated, however, the geometry was more suggestive of a direct, in-line transfer of the γ-phosphate group. Additionally, the presence of five magnesium ions in the complexes containing a thiamin moiety provides significant positive charge to stabilize the phosphate groups and the direct inline transfer of the phosphate group is more likely than an indirect transfer utilizing a phosphoenzyme intermediate.

Vitamin B_6 metabolism is another pathway containing interesting chemistry and novel enzymatic activities. The first pyridoxine catabolic pathway was identified in the 1950's and contains seven enzymatic steps to reduce PLP to succinic semialdehyde, ammonia, acetate and carbon dioxide, shown in Figure 1.2 (*6*). In the early 2000's a similar pathway was discovered in *M. loti* MAFF303099, a symbiotic bacterium that grows on the roots of legumes. This pathway contains several mechanistically interesting reactions, such as the decarboxylation of a substituted benzene ring without requiring any organic cofactor, or the oxidative ring opening reaction that is dependent only on flavin and does not need a non-heme iron for catalysis. Chapter Five presents the gene identification, kinetic characterization, and structural characterization of 3-hydroxy-2-methylpyridine-4,5-dicarboxylate decarboxylase (HMPDdc). The structure was determined by SAD phasing to 1.9 Å resolution and the structural superfamily HMPDdc belongs to offered a potential mechanism for catalysis (Figure 5.8). HMPDdc forms a tetramer using a fourfold crystallographic axis with the active site located at a dimer interface and a manganese ion coordinated by three conserved histidine residues. The hydroxyl group ortho to C4, the position being decarboxylated, coordinates to the active site manganese ion and forms a tetrahedral intermediate at C4 by removing a proton from the active site glutamate. Collapse of this intermediate leads to the loss of carbon dioxide, then product release and recoordination of the glutamate residue to the manganese ion. This type of chemistry is commonly found in class II aldolases, with which HMPDdc has high structural homology.

The ensuing enzyme, 2-methyl-3-hydroxypyridine-5-carboxylic acid oxygenase (MHPCO) has been extensively studied due to the highly unusual oxidative ring-opening of a substituted aromatic ring without requiring a nonheme iron for catalysis. MHPCO catalyzes two oxidation reactions where the first incorporates one

atom of oxygen from diatomic oxygen at the C2 of MHPC in a flavin dependent reaction, leading to a nonaromatic intermediate. An oxygen atom from water is then added, triggering the ring to break open (Figure 6.7). Despite many studies on MHPCO, the mechanism for this second step remained unclear. The solution of the MHPCO structure with FAD, both with and without MHPC bound, allowed for the identification of additional residues that could be involved in the second step (Figure 6.8). Mutagenesis studies of Tyr223 and Arg181 have demonstrated that loss of either residue renders this enzyme inactive.

The final enzyme in the PLP degradative pathway, (*E*)-2-(acetamidomethylene)succinate hydrolase (*E*-2AMS hydrolase) is responsible for degrading *E*-2AMS from MHPCO and producing ammonia, succinic semialdehyde, acetate, and carbon dioxide. This gene had been successfully cloned and kinetically characterized and seemed to belong to the α/β hydrolase superfamily based on catalytic activity and sequence homology with other members of this structural superfamily. The Ser-His-Asp catalytic triad, however, appeared to be missing the aspartate when sequence alignments were examined. To address catalysis, the structure of *E*-2AMS hydrolase is reported to 2.3 Å in Chapter Seven. These crystals suffered from pseudo-translational symmetry, producing a unit cell that was three times the expected volume and containing six dimers in the asymmetric unit. *E*-2AMS hydrolase is a true member of the α/β hydrolase superfamily, with the third catalytic residue, Asp130, within hydrogen bonding distance of His258 but not in the structurally conserved position (Figure 7.7). This structure has allowed a mechanism to be proposed for catalysis (Figure 7.8).

The thiamin biosynthetic pathway and the PLP catabolic pathway each offered unique opportunities to probe the structure, function, and mechanistic action of interesting enzymes. As collaboration between structural biologists, mechanistic

enzymologists, biochemists, and organic chemists continue to work at the interface between chemistry and biology, the interesting and difficult mechanistic questions will continue to be addressed and elucidated.

REFERENCES

1. Peapus, D. H., Chiu, H. J., Campobasso, N., Reddick, J. J., Begley, T. P., and Ealick, S. E. (2001) Structural characterization of the enzyme-substrate, enzyme-intermediate, and enzyme-product complexes of thiamin phosphate synthase, *Biochemistry 40*, 10103-10114.
2. 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.
3. Hazra, A., Chatterjee, A., and Begley, T. P. (2009) Biosynthesis of the Thiamin Thiazole in Bacillus subtilis: Identification of the Product of the Thiazole Synthase-Catalyzed Reaction, *J. Am. Chem. Soc. 131*, 3225-3229.
4. Rodionov, D. A., Vitreschak, A. G., Mironov, A. A., and Gelfand, M. S. (2002) Comparative genomics of thiamin biosynthesis in procaryotes. New genes and regulatory mechanisms, *J. Biol. Chem. 277*, 48949-48959.
5. Li, C., Kappock, T. J., Stubbe, J., Weaver, T. M., and Ealick, S. E. (1999) X-ray crystal structure of aminoimidazole ribonucleotide synthetase (PurM), from the *Escherichia coli* purine biosynthetic pathway at 2.5 Å resolution, *Structure 7*, 1155-1166.
6. Snell, E. E., and Haskell, B. E. (1971) *The Metabolism of Vitamin B6, In: Comprehensive Biochemistry*, Vol. 21, Elsevier/North Holland, New York.