REVERSAL OF SUBSTRATE SPECIFICITY OF CMP *N*-GLYCOSIDASE TO dCMP AND CRYSTALLOGRAPHIC STUDIES ON BACIMETHRIN AND AZINOMYCIN B BIOSYNTHESIS ENZYMES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Megan Sikowitz

August 2013

REVERSAL OF SUBSTRATE SPECIFICITY OF CMP *N*-GLYCOSIDASE TO dCMP AND CRYSTALLOGRAPHIC STUDIES ON BACIMETHRIN AND AZINOMYCIN B BIOSYNTHESIS ENZYMES

Megan Sikowitz, Ph. D.

Cornell University 2013

The concept that structure defines function is fundamental to understanding biology and this theme is repeated again and again at the molecular, cellular, and organism levels. Remarkable advances in X-ray crystallography allow scientists to probe the structure of enzymes at atomic resolution. Protein crystal structures allow a greater understanding of the chemical and biological functions of these cellular machines. Enzymes from prokaryotes that are involved in the biosynthesis and degradation of various metabolites often present novel chemistry and curious functionalities. The enzymes classified in the nucleoside 2′-deoxyribosyltransferase superfamily have conserved catalytic residues and share a mechanism for the *N*-glycosidic bond cleavage of nucleosides, but demonstrate varied substrate specificity and biochemical functions. The structures of two members of this superfamily, *Streptomyces rimofaciens* MilB and *Clostridium botulinum* BcmB, provide novel information about how these enzymes confer specificity at their substrate 2′-position.

The biosynthetic pathways of bacterial natural products also frequently present enzymes with unique functions and specificities. To better understand the biology of the thiamin antimetabolite bacimethrin, the *C. botulinum* genes responsible for its

biosynthesis were identified and several enzymes implicated in its production were structurally characterized. Here, the structure of the bacimethrin kinase, BcmD is analyzed to better understand its specificity and relationship to other ribokinases. Another enzyme clustered with the bacimethrin genes, Cb-thiaminase I, was thought to serve as a bacimethrin-detoxifying enzyme. While this was not supported, structural solution and supporting kinetic characterization provides a new mechanistic understanding of the thiaminase I class of enzymes. Finally, the biosynthetic genes for the *Streptomyces sahachiroi* antitumor antibiotic, azinomycin B, were recently identified and reveal a complex network of enzymes involved in its assembly. Biochemical studies suggest that production of the azinomycin B naphthoate moiety requires a thioesterase enzyme in addition to a polyketide synthase megaenzyme. Structural characterization of the thioesterase SsAziG co-crystallized with a product analog supports this finding and provides a basis for its proposed mechanism.

BIOGRAPHICAL SKETCH

Megan was born and raised in the Maryland suburbs of Washington, D.C. She is the second of three children to her parents, Pam and Matt Sikowitz. She attended Eleanor Roosevelt High School in Greenbelt, MD, where she participated in the Science and Technology program and her favorite classes were Chemistry and Band. In 2004, she started undergraduate studies at St. Mary's College of Maryland; Maryland's public liberal arts honors college. Megan was initially a Biology major at St. Mary's until developing a fascination with Organic Chemistry. She switched departments and became a Biochemistry major in order to focus on more chemistry oriented courses. While at St. Mary's she also completed a Mathematics minor and competed on the rowing team.

Megan had the opportunity to gain research experience in a synthetic organic chemistry lab at St. Mary's. While interesting, enzymes intrigued her more. After spending a summer in the laboratory of Prof. Nicole Sampson at Stony Brook University studying enzymatic synthesis of sterol derivatives, Megan completed her senior thesis at the University of Maryland Chesapeake Biological Laboratory studying the marine enzyme DMSP lyase. Because of these research experiences and enjoying her work as a TA, she decided to pursue a PhD in order to be able to teach and conduct research with undergraduates. Megan joined the Chemistry and Chemical Biology department at Cornell University in 2008. After completing her coursework, she began studying enzyme structure in Prof. Steven Ealick's research group. Her work has focused primarily on enzymes involved in bacimethrin and azinomycin B metabolism. In her free time she has enjoyed exploring the parks and

wineries of upstate New York, running around Ithaca, and playing oboe with the Ithaca Concert Band. She is relocating to the Washington, DC area and will marry Tom Erb in May 2014. She is excited for continuing her adventures in a warmer climate.

In memory of my dear Aunt Sue, who nurtured my love of learning, and my cousin Jacob Vogelman, a true mensch.

ACKNOWLEDGMENTS

First I would like to thank Prof. Steven Ealick for serving as my research advisor. Thank you for allowing me to work on interesting projects, providing many opportunities to enhance my understanding of structural biology, and encouraging independent scientific thought. I would also like to thank Prof. Hening Lin and Prof. Richard Cerione for their advice on research and for serving on my committee. My thesis work would not have been possible without many wonderful collaborators. I would like to acknowledge Prof. Coran Watanabe, her former student, Dr. Jennifer Foulke, and former post-doctoral scientist, Dr. Dinesh Simkhada, for their biochemical work on the azinomycin B project and thank them for many helpful discussions. I also owe much gratitude to Prof. Tadgh Begley for his collaboration on the bacimethrin project. Many thanks to his former postdoctoral scientist, Dr. Lisa Cooper, for her characterization of the bacimethrin enzymes and MilB, as well as his graduate student, Brateen Shome, for his work on Cb-thiaminase I. Thanks also to Dr. Cynthia Kinsland for general advice on working with proteins and for providing countless clones and mutants. I would also like to acknowledge Ms. Leslie Kinsland for her kindness and help she has provided in so many ways. The NE-CAT staff has been incredibly helpful and offered much practical advice on data collection. I have enjoyed working with them. I thank current and former members of the Ealick group for their camaraderie and discussions that have both helped me with research and provided many enjoyable days in lab. Finally, thanks to my parents, siblings, family, and friends for their support. Special thanks to Tom, Erika, Angie, Katherine, and Erin for helping make Ithaca my home over the past five years.

TABLE OF CONTENTS

	Page
Biographical Sketch	v
Dedication	vii
Acknowledgements	viii
Table of Contents	ix
List of Figures	x
List of Tables	xii
List of Abbreviations	xiii
Chapter 1. Introduction	1
Chapter 2. Reversal of substrate specificity of CMP *N*-glycosidase to dCMP	17
Chapter 3. Structure of *Clostridium botulinum* thiaminase I-thiamin complex reveals biochemical details of binding and degradation	57
Chapter 4. Structure-function studies on *Clostridium botulinum* bacimethrin kinase, BcmD	88
Chapter 5. Structural characterization of SsAziG, a *Streptomyces sahachiroi* thioesterase involved in azinomycin B biosynthesis	106
Chapter 6. Summary	127

LIST OF FIGURES

	Page
Figure 1.1. General mechanism for nucleobase transfer or hydrolysis by NDT superfamily members	3
Figure 1.2. Substrate preferences of NDT superfamily members	4
Figure 1.3. General thiamin biosynthesis scheme	6
Figure 1.4. Thiamin degradation by thiaminase I and II	7
Figure 1.5. Bacimethrin biosynthesis in Clostridium botulinum	9
Figure 1.6. Chemical structure of the azinomycins	10
Scheme 2.1. MilB mechanism	46
Figure 2.1. Natural products with cytosine precursors	20
Figure 2.2. Structure of MilB	29
Figure 2.3. Substrate binding in MilB	31
Figure 2.4. BcmB/PO4 and MilB/CMP active site comparison	32
Figure 2.5. Substrate specificity of MilB WT and MilB F17Y	34
Figure 2.6. Comparison of enzyme dimerization	37
Figure 2.7. Sequence alignment of enzymes containing the nucleoside 2′-deoxyribosyltransferase motif	40
Figure 2.8. NDT family member active site comparison	44
Figure 2.9. Active site comparison of MilB and NDT	47
Figure S2.1. No hydrolysis of purines by MilB	49
Figure S2.2. BcmB structure	50
Figure S2.3. BcmB WT and F6Y reaction with CMP and dCMP	51
Figure 3.1. Thiamin cleavage reaction catalyzed by the thiaminases	57

Figure 3.2. Overall structure of C143S/Thiamin 67

Figure 3.3. C143S/Thiamin active site interactions 69

Figure 3.4 Thiamin binding in C143S 69

Figure 3.5. Comparison of C143S/Thiamin with Thiaminase I/ACDP structure from B. thiaminolyticus 72

Figure 3.6. Mechanism of thiamin degradation by Cb-thiaminase I 79

Figure 3.7. Sequence alignment of thiaminase Is 80

Figure 3.8. Comparison of thiamin binding in Thiaminase I and II 81

Figure 4.1. Routes to thiamin pyrophosphate (ThDP) and 2'-methoxythiamin pyrophosphate biosynthesis (MeOThDP). 89

Figure 4.2. Bacimethrin biosynthesis pathway 90

Figure 4.3. Structure of BcmD 96

Figure 4.4. Stereoview of BcmD active site aligned with StHMPP kinase/HMP and two sulfates 94

Figure 4.5. BcmD/StHMPP kinase sequence alignment. 98

Figure 4.6. HMPP kinase/BcmD active site binding flap movement upon substrate binding 102

Figure 5.1. Structure of the azinomycins and synthesis of the naphthoate 107

Figure 5.2. Monomeric structure of AziG 114

Figure 5.3. The biological tetramer formed by AziG 115

Figure 5.4. AziG active site 116

Figure 5.5. Model for 5-methyl-NPA binding 117

Figure 5.6. Proposed mechanism for 5-methyl-NPA-CoA hydrolysis by AziG 120

LIST OF TABLES

	Page
Table 2.1. Summary of MilB/BcmB data collection and refinement statistics.	26
Table 2.2. Summary of MilB/BcmB kinetic data	35
Table 2.3. Enzymes structurally similar to MilB	36
Table 2.4. Summary of NDT-superfamily enzymes	38
Table S2.1. Enzyme/substrate concentrations used in kinetic assays	48
Table 3.1. Summary of C143S/thiamin data collection and refinement statistics	65
Table 3.2. Kinetic properties of thiaminase activity for Cb-thiaminase I WT and active site mutants	70
Table 3.3. Enzymes structurally similar to Cb-thiaminase I	71
Table 4.1. Summary of BcmD data collection and refinement statistics	94
Table 4.2. Summary of proteins structurally similar to BcmD	100
Table 5.1. AziG data collection and refinement statistics	112
Table 5.2. Summary of proteins structurally similar to SsAziG, as analyzed by DALI	119

LIST OF ABBREVIATIONS

NDT, nucleoside 2′-deoxyribosyltransferase

hmCMP, 5-hydroxymethyl cytidine 5′-monophosphate

CMP, cytidine 5′-monophosphate; SeMet, selenomethionyl

AMP, adenosine 5′-monophosphate; IMP, inosine 5′-monophosphate

UMP, uridine 5′-monophosphate

dCMP, 2′-deoxycytidine 5′-monophosphate

GMP, guanosine 5′-monophosphate

SAD, single-wavelength anomalous diffraction

PTD, purine deoxyribosyltransferase

DMHA, *N*,*N*-dimethylhexylamine

5MD, 5-methyl-2'-deoxypseudouridine

ADCP, 4-amino-6-chloro-2, 5-dimethylpyridimine

PBP, periplasmic binding protein

ThDP, thiamin pyrophosphate

HMP-PP, 4-amino-5-hydroxymethyl-2-methylpyrimidine pyrophosphate

THZ-P, thiazole phosphate carboxylate

ThMP, thiamin monophosphate

BCM, bacimethrin

MeOThDP, 2′-methoxythiamin pyrophosphate

IPTG, isopropyl-β-D-1-thiogalactopyranoside

PKS, polyketide synthase

NPA, naphthoic acid

CHESS, Cornell High Energy Synchrotron Sournce

rmsd, root mean squared deviation

CHAPTER 1

INTRODUCTION

Section 1.1. Structural Biology

Structural biologists use a wide array of tools to probe the three-dimensional arrangement of macromolecules. While structural biology is a diverse field, the basic goal of all structural investigations is to reveal the basis of molecular interactions. Atomic structures offer a greater understanding of a protein's biochemistry, which provides a context for understanding its role in cellular function. Many tools have been developed for probing protein structure. Among the most popular are nuclear magnetic resonance, electron microscopy, small angle X-ray scattering, and X-ray crystallography.[1] While each of these methods is valuable, X-ray crystallography is the most versatile and powerful tool for solving atomic resolution protein structures currently available. The work of X-ray crystallographers is prominently recognized in the scientific community. In the past seven years alone, three Nobel Prizes in Chemistry have been awarded to protein crystallographers.[2-4] Their work has led to groundbreaking advances in understanding fundamental biology upon which much biomedical research has been built.

The X-ray crystallography experiment is straighforward in principle. First, protein crystals are grown and then irradiated with X-rays. The X-rays are scattered by the electrons in the crystal and the resultant diffracting X-rays are recorded by a detector. From the diffraction data, a three-dimensional map of the electron density is calculated using Fourier methods. Finally an atomic model of the molecule is built into the electron density. In practice, solving an X-ray crystal structure presents many

challenges. A major caveat of this method is known as the 'phase problem'. The phase of the diffracted X-ray is an essential component in calculating the Fourier transformations of the diffraction data.[5] Because X-rays cannot be focused, information about their phases cannot be measured during the experiment.[5] The development of anomalous dispersion[6], isomorphous replacement[7], and molecular replacement[8] methods provides tools for the determination of phases. Validation methods and data collection statistics exist to assess the quality of the model and reporting of this information is expected concurrent with publication of a protein structure. The field of X-ray crystallography is quite mature, with over 78,000 atomic coordinate files publicly shared by members of the X-ray crystallography community in the Protein Data Bank (PDB).

The enzyme active site is often the area of most interest in a biochemical investigation. Structural information is used to identify active site amino acids and cofactors. This data is coupled with complementary kinetic and mutant studies to predict the mechanism of action. These findings are then used to understand fundamental biochemistry and can be exploited for drug or inhibitor design if the enzyme target is involved in disease progression. In the following chapters, X-ray crystallography and supporting biochemical studies are utilized to study the function, substrate specificity, and/or mechanisms of enzymes in various bacterial pathways that play important roles in nucleotide, thiamin, bacimethrin, and azinomycin B metabolism, including MilB, BcmB, Cb-thiaminase I, BcmD, and AziG.

Section 1. 2. Nucleoside 2′-deoxyribosyltransferases

Enzymes in the nucleoside 2′-deoxyribosyltransferase (NDT) superfamily catalyze the cleavage of the *N*-glycosidic bond of either nucleosides or nucleotides. Members of this family have conserved active site nucleophilic amino acids and share a common mechanism in which the nucleobase is replaced by an acceptor nucleophile (Figure 1.1).[9, 10] However, they differ in their substrate preferences and final acceptor (Figure 1.2). They also fulfill diverse physiological functions and are found in a wide variety of organisms. Because of the varied functionalities of enzymes in this superfamily, their structures were analyzed to determine the basis of their substrate preferences.

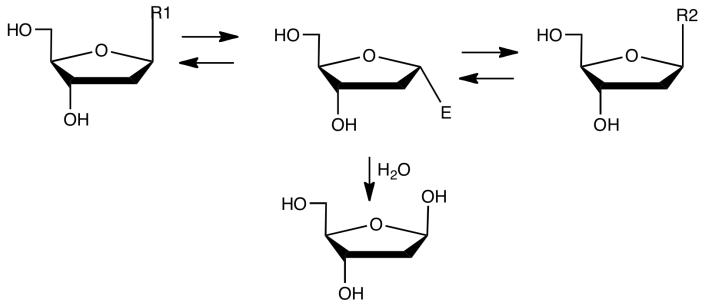
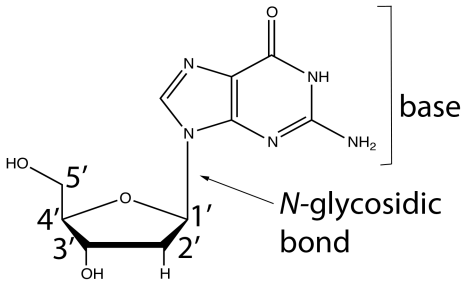


Figure 1.1. General mechanism for nucleobase transfer or hydrolysis by NDT superfamily members

The first member in this superfamily to be structurally characterized was *Lactobacillus leichmanii* NDT.[11] NDTs are implicated in nucleoside salvage in organisms that lack the common nucleoside biosynthesis pathways and in regulation of nucleoside availability.[12] They catalyze the reversible transfer of nucleobases from a deoxyribose, with highly flexible substrate specificity for both naturally occurring and synthetic nucleobases.[13-15] Similarly, purine 2′-deoxyribosyltransferases (PTD)

are found in some prokaryotes and catalyze a transferase reaction, but accept only substrates with purine nucleobases.[14] Rcl is a purine nucleotide hydrolase that provides much curiosity because it retains the conserved NDT mechanism, but utilizes water as the final acceptor.[16, 17] It also stands out in this class because it functions in the regulation of cell growth and is found exclusively in mammals. Unlike NDT and PTD, which only accept the 5′-hydroxyl containing nucleosides, Rcl shows specificity for the 5′-phosphate containing nucleotides.



Enzyme	Substrate preferences			Acceptor
	5'-position	2'-position	Base	
NDT	-OH	-H	Purine/pyrimidine	Purine/pyrimidine
PTD	-OH	-H	Purine	Purine
Rcl	phosphate	-H	Purine	Water
MilB/BcmB	phosphate	-OH	cytosine	Water

Figure 1.2. Substrate preferences of NDT superfamily members

In Chapter Two, the structures of two additional members of the NDT superfamily are analyzed. BcmB and MilB, with 47% sequence identity to each other, are responsible for the hydrolysis of the *N*-glycosidic bond of cytidine 5′-monophosphate (CMP) or 5-hydroxymethyl-CMP (hmCMP). This reaction releases the cytosine portion for incorporation into their respective natural product biosynthesis pathways.[18] CMP is a 2′-hydroxy containing substrate and is the first example of a

ribosylnucleotide substrate for an enzyme in this family. This specificity is especially interesting because previous studies show ribosylnucleotides act as inhibitors of both Rcl and NDT.[19] The structures of BcmB and MilB reveal that BcmB Phe6 and MilB Phe17 are near the substrate 2′-binding site and are the basis of their specificity towards the ribosyl containing substrates. In previously solved structures of other NDT family members, a tyrosine residue is positioned equivalently to this phenylalanine. Finally, MilB and BcmB mutants were engineered for the reversal of substrate specificity to the 2′-deoxy containing dCMP by mutation of the phenylalanine to tyrosine. The basis of the specificity of BcmB and MilB at the nucleotide 2′-position is further explored in Chapter Two.

Section 1.3. Thiamin Degradation

Thiamin (vitamin B_1) is important for many biological processes. It stabilizes carbanion intermediates in essential metabolic pathways and its deficiency can lead to illness or death.[20, 21] The biochemistry of thiamin has been of interest since the 1950's[22] and its metabolism still yields fascinating chemical and biological questions. The biosynthetic pathways that produce thiamin have been extensively characterized. While the individual steps differ in prokaryotes and eukaryotes, both follow a two-branched pathway in which the two heterocyclic halves of thiamin are synthesized separately, then ligated together to form thiamin.[20] Thiamin phosphate kinase then generates the active form of thiamin, thiamin diphosphate (ThDP)[23, 24] (Figure 1.3).

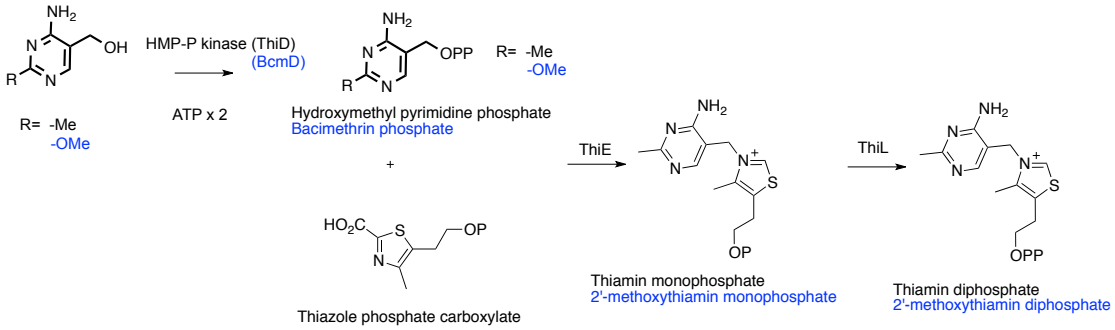


Figure 1.3. General thiamin biosynthesis scheme. Bacimethrin biosynthesis metabolites and enzymes are distinguished in blue.

The physiological necessity for thiamin biosynthesis is clear, but the physiological purpose for thiamin degradation is not as well understood. Two classes of thiamin degrading enzymes, thiaminase I and II, are known.[20] Members of both classes can catalyze the degradation of thiamin into its thiazole and pyrimidine moieties. Thiaminase Is utilize a variety of nucleophiles for CN bond cleavage,[25] whereas thiaminase IIs exclusively use water[26] (Figure 1.4). Thiaminase I tolerates substrate modifications on the thiamin thiazole moiety, including the phosphorylated thiamin forms ThDP and thiamin monophosphate (ThMP). Thiaminase IIs do not accept these phosphorylated thiamin forms that are abundant in the cell. The structural basis for this is established, as thiaminase II is an all α-helix bundle and its deep binding pocket limits accessibility.[26] While it is able to degrade thiamin, its true function is to serve as a thiamin salvage enzyme, with base degraded thiamin as its preferred substrate.[27] The structure and mechanism of thiaminase I differs from thiaminase II.

Thiaminase I is found in some species of plants, fish, and bacteria.[20] The first structure of a thiaminase I, from *Bacillus thiaminolyticus* with a mechanism based inhibitor, revealed its active site binding pocket.[28] Thiaminase Is feature an open V-

shaped binding cleft and display a conserved periplasmic binding protein (PBP) fold. PBPs have evolved to accommodate various substrates for transport and delivery of metabolites.[29] The PBP fold architecture enables flexibility in substrate binding of thiamin, ThMP, and ThDP, with additional accessibility for a variety of nucleophiles. Thiaminase I is infamous for its role in the deaths of Australian explorers in the 1860's who consumed the native nardoo fern.[20] Unbeknownst to them, this plant contains large amounts of thiaminase I. As a result of ingesting it, the explorers had insufficient cellular levels of thiamin, which led to their deaths by the thiamin deficiency disease, beriberi. Even after decades of study, the physiological purpose for producing thiaminase I remains mysterious.

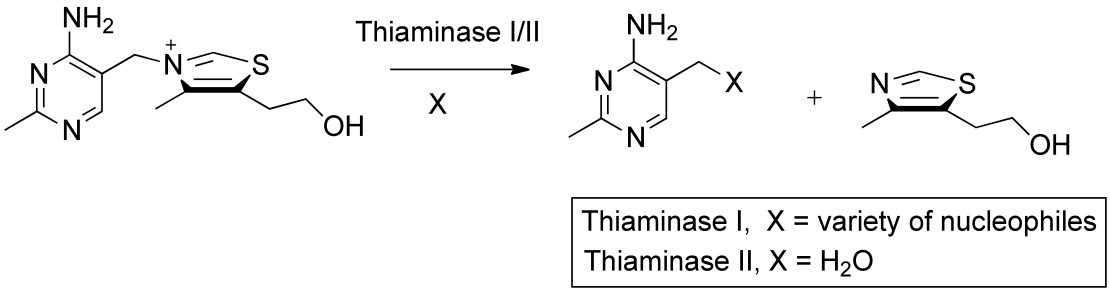


Figure 1.4. Thiamin degradation by thiaminase I and II

In Chapter Three, the structure of Cb-thiaminase I from *Clostridium botulinum* is explored. Cb-thiaminase I is clustered with the genes from the newly discovered bacimethrin (BCM) biosynthesis pathway. Introduced in detail in section 1.4, BCM is a thiamin antimetabolite. Briefly, BCM is an analog to the thiamin precursor hydroxymethyl pyrimidine (HMP) and it hijacks the thiamin biosynthetic machinery[30] to produce a biologically insufficient thiamin analog (Figure 1.3). The

original hypothesis that Cb-thiaminase I serves to detoxify BCM turned out to be incorrect (Cooper, et. al., Begley Lab- *manuscript in preparation*). However, the crystal structure of a catalytically inactive Cb-thiaminase I cocrystallized with thiamin aided in identification of key residues involved in catalysis of thiamin degradation. Based on these structural details along with supporting kinetic studies, a detailed mechanism for the thiaminase I reaction is proposed. The physiological purpose of thiamin degradation still remains to be discovered, but the details of thiaminase I chemistry are now clear.

Section 1.4. Bacimethrin

BCM is a thiamin antimetabolite toxic to yeast and bacteria. Initially thought to be an inhibitor of thiamin biosynthesis enzymes,[31] it was later shown that BCM is in fact a substrate for these enzymes.[30] BCM differs from the thiamin precursor HMP by the addition of a methoxy group at the O2 position. In *E. coli* treated with BCM or HMP, *in vitro* conversion of BCM to 2′-methoxythiamin is six times faster than HMP conversion to thiamin, suggesting BCM overwhelms the thiamin biosynthetic pathway.[30] The abundance of this biologically insufficient thiamin analog leads to toxic effects and toxicity is reversed by addition of thiamin to the growth media.[31]

The BCM biosynthetic pathway was recently identified in *C. botulinum* (Figure 1.5) (Cooper, et. al., Begley Lab- *manuscript in preparation*). In addition to providing a basis for understanding the origin of BCM, the functions of enzymes in this pathway raise interesting biochemical questions. BcmA is responsible for the addition of a 5-hydroxymethyl group to the CMP precursor. Next BcmB, a member of a small family of ribosylnucleotide hydrolases introduced in Section 1.2, hydrolyzes

the *N*-glycosidic of hmCMP. After further methylation at the cytosine O2 by BcmC, BCM is prepared for shuttling into the thiamin metabolic pathway by BcmD diphosphorylation.

BcmB and Cb-thiaminase I (found clustered with the BCM biosynthesis genes) are discussed in Chapters 2 and 3, respectively. BcmD is highly similar in sequence and structure to known HMPP kinases. It is structurally homologous to several ribokinases and is able to phosphorylate both BCM and HMP. Comparison of the BcmD structure to the thiamin biosynthesis HMPP kinase establishes a close relationship and conservation of active site architecture. The substrate flexibility and preferences of BcmD and its role in BCM toxicity are explored in Chapter Four.

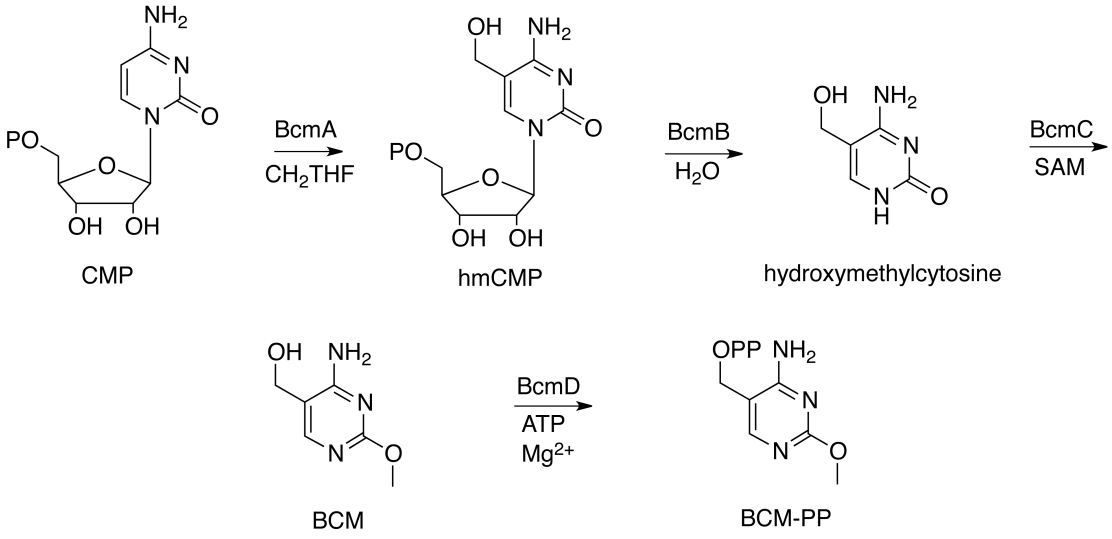


Figure 1.5 Bacimethrin biosynthesis in *Clostridium botulinum*

Section 1.5. Azinomycin B Biosynthesis

Complex natural products with antibiotic properties are frequently found in bacteria. Chemical biologists study their functionalities and mechanism of action in an effort to develop new drug compounds. The azinomycins are one class of natural products of interest[32] (Figure 1.6). Isolated from *Streptomyces sahachiroi* and *Steptomyces griseofuscus,*[33, 34] azinomycin B has antitumor antibiotic activity and shows activity at the nanomolar level against some cancers in mouse models.[35] Mechanistically, azinomycin B crosslinks DNA by electrophilic attack on to the N7 position of purines with apparent sequence selectivity. [36] Chemically, azinomycin B is of interest because of its unique tailoring and densely assembled functional groups. While fermentation methods have been exploited to obtain azinomycin B for some time,[33] the genes responsible for the biosynthesis of azinomycin B were only recently identified in *S. sahachiroi*.[34]

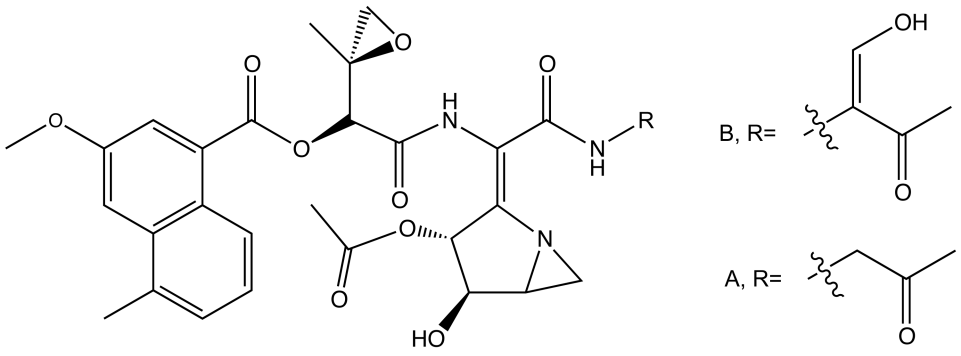


Figure 1.6. Chemical structure of the azinomycins

Initial reports suggested that the *S. sahachiroi* polyketide synthase encoded by the *aziB* gene, SsAziB, is the only enzyme in the pathway necessary for production of the azinomycin B naphthoate moiety.[34, 37] Recent findings suggest that SsAziB acts in concert with a thioesterase, SsAziG, to produce the naphthoate fragment (Watanabe

Lab- *manuscript in preparation*). SsAziB polymerizes and condenses acetyl-CoA and malonyl-CoA precursor molecules and then SsAziG hydrolyzes the 5-methyl-naphthate-CoA thioester bond to release the product. In Chapter 5, crystallographic studies of SsAziG with the bound product analog, 3-hydroxy-2-naphthoic acid, provide evidence that supports these new biochemical findings. Structural comparisons between SsAziG and related thioesterases support the function of SsAziG in naphthoate biosynthesis and a possible mechanism.

Section 1.6. Acknowledgements

This work reflects a collaborative effort and those collaborators deserve full recognition for their contributions. Dr. Lisa Cooper and Dr. Sean O'Leary in the Begley group at Texas A&M University identified and characterized the bacimethrin biosynthesis pathway. Lisa also conducted the substrate specificity and kinetic studies on MilB and BcmB. Brateen Shome, also in the Begley group, did the kinetic characterization and much of the mechanistic work on Cb-thiaminase I. My collaborators in the Watanabe group at Texas A&M University, Dr. Jennifer Foulke and Dr. Dinesh Simkhada, carried out the functional and biochemical studies on SsAziG and other azinomycin B associated proteins. I performed the structural work on these enzyme systems with these collaborators who performed the biochemical characterizations.

REFERENCES

1. Campbell, I. D. (2002) Timeline: the march of structural biology, *Nat Rev Mol Cell Biol 3*, 377-381.
2. Benovic, J. L. (2012) G-protein-coupled receptors signal victory, *Cell 151*, 1148-1150.
3. Landick, R. (2006) A long time in the making--the Nobel Prize for RNA polymerase, *Cell 127*, 1087-1090.
4. Hu, Y. L. (2009) Studies on Structure and Function of Ribosome: a Brief Introduction to 2009 Nobel Prize in Chemistry, *Prog Biochem Biophys 36*, 1239-1243.
5. Taylor, G. (2003) The phase problem, *Acta Crystallogr D 59*, 1881-1890.
6. Dodson, E. (2003) Is it jolly SAD?, *Acta Crystallogr D 59*, 1958-1965.
7. Perutz, M. F. (1956) Isomorphous Replacement and Phase Determination in Non-Centrosymmetric Space Groups, *Acta Crystallogr 9*, 867-873.
8. Rossmann, M. G., and Blow, D. M. (1962) Detection of Sub-Units within Crystallographic Asymmetric Unit, *Acta Crystallogr 15*, 24-&.
9. Porter, D. J., Merrill, B. M., and Short, S. A. (1995) Identification of the active site nucleophile in nucleoside 2-deoxyribosyltransferase as glutamic acid 98, *J Biol Chem 270*, 15551-15556.
10. Short, S. A., Armstrong, S. R., Ealick, S. E., and Porter, D. J. (1996) Active site amino acids that participate in the catalytic mechanism of nucleoside 2'-deoxyribosyltransferase, *J Biol Chem 271*, 4978-4987.

11. Armstrong, S. R., Cook, W. J., Short, S. A., and Ealick, S. E. (1996) Crystal structures of nucleoside 2-deoxyribosyltransferase in native and ligand-bound forms reveal architecture of the active site, *Structure 4*, 97-107.

12. Chawdhri, R. F., Hutchinson, D. W., and Richards, A. O. (1991) Nucleoside Deoxyribosyltransferase and Inosine Phosphorylase-Activity in Lactic-Acid Bacteria, *Archives of Microbiology 155*, 409-411.

13. Cardinaud, R., and Holguin, J. (1979) Nucleoside Deoxyribosyltransferase-Ii from Lactobacillus-Helveticus - Substrate-Specificity Studies - Pyrimidine-Bases as Acceptors, *Biochim Biophys Acta 568*, 339-347.

14. Becker, J., and Brendel, M. (1996) Rapid purification and characterization of two distinct *N*-deoxyribosyltransferases of Lactobacillus leichmannii, *Biol Chem H-S 377*, 357-362.

15. Fernandez-Lucas, J., Acebal, C., Sinisterra, J. V., Arroyo, M., and de la Mata, I. (2010) *Lactobacillus reuteri* 2 '-Deoxyribosyltransferase, a Novel Biocatalyst for Tailoring of Nucleosides, *Appl Environ Microb 76*, 1462-1470.

16. Ghiorghi, Y. K., Zeller, K. I., Dang, C. V., and Kaminski, P. A. (2007) The c-Myc target gene Rcl (C6orf108) encodes a novel enzyme, deoxynucleoside 5'-monophosphate *N*-glycosidase, *J Biol Chem 282*, 8150-8156.

17. Lewis, B. C., Shim, H., Li, Q., Wu, C. S., Lee, L. A., Maity, A., and Dang, C. V. (1997) Identification of putative c-Myc-responsive genes: characterization of rcl, a novel growth-related gene, *Mol Cell Biol 17*, 4967-4978.

18. Li, L., Xu, Z., Xu, X., Wu, J., Zhang, Y., He, X., Zabriskie, T. M., and Deng, Z. (2008) The mildiomycin biosynthesis: initial steps for sequential generation

of 5-hydroxymethylcytidine 5'-monophosphate and 5-hydroxymethylcytosine in Streptoverticillium rimofaciens ZJU5119, *Chembiochem 9*, 1286-1294.

19. Yang, Y., Padilla, A., Zhang, C., Labesse, G., and Kaminski, P. A. (2009) Structural characterization of the mammalian deoxynucleotide *N*-hydrolase Rcl and its stabilizing interactions with two inhibitors, *J Mol Biol 394*, 435-447.
20. 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.
21. Jordan, F. (2003) Current mechanistic understanding of thiamin diphosphatedependent enzymatic reactions, *Nat Prod Rep 20*, 184-201.
22. Breslow, R. (1958) On the Mechanism of Thiamine Action. 4. Evidence from Studies on Model Systems, *J Am Chem Soc 80*, 3719-3726.
23. Webb, E., and Downs, D. (1997) Characterization of thiL, encoding thiamin-monophosphate kinase, in Salmonella typhimurium, *J Biol Chem 272*, 15702-15707.
24. 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.
25. Mccleary, B. V., and Chick, B. F. (1977) Purification and Properties of a Thiaminase I Enzyme from Nardoo (Marsilea-Drummondii), *Phytochemistry 16*, 207-213.
26. 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.

27. 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.

28. Campobasso, N., Begun, J., Costello, C. A., Begley, T. P., and Ealick, S. E. (1998) Crystallization and preliminary X-ray analysis of thiaminase I from Bacillus thiaminolyticus: space group change upon freezing of crystals, *Acta Crystallogr D 54*, 448-450.

29. Dwyer, M. A., and Hellinga, H. W. (2004) Periplasmic binding proteins: a versatile superfamily for protein engineering, *Curr Opin Struct Biol 14*, 495-504.

30. Reddick, J. J., Saha, S., Lee, J. M., Melnick, J. S., Perkins, J., and Begley, T. P. (2001) The mechanism of action of bacimethrin, a naturally occurring thiamin antimetabolite, *Bioorg Med Chem Lett 11*, 2245-2248.

31. Drautz, H., Messerer, W., Zahner, H., Breidingmack, S., and Zeeck, A. (1987) Metabolic Products of Microorganisms .239. Bacimethrin Isolated from Streptomyces-Albus Identification, Derivatives, Synthesis and Biological Properties, *J Antibiot 40*, 1431-1439.

32. Casely-Hayford, M. A., Pors, K., James, C. H., Patterson, L. H., Hartley, J. A., and Searcey, M. (2005) Design and synthesis of a DNA-crosslinking azinomycin analogue, *Org Biomol Chem 3*, 3585-3589.

33. Nagaoka, K., Matsumoto, M., Oono, J., Yokoi, K., Ishizeki, S., and Nakashima, T. (1986) Azinomycin-a and Azinomycin-B, New Antitumor Antibiotics. 1. Producing Organism, Fermentation, Isolation, and Characterization, *J Antibiot 39*, 1527-1532.

34. Zhao, Q., He, Q., Ding, W., Tang, M., Kang, Q., Yu, Y., Deng, W., Zhang, Q., Fang, J., Tang, G., and Liu, W. (2008) Characterization of the azinomycin B biosynthetic gene cluster revealing a different iterative type I polyketide synthase for naphthoate biosynthesis, *Chem Biol 15*, 693-705.

35. Ishizeki, S., Ohtsuka, M., Irinoda, K., Kukita, K. I., Nagaoka, K., and Nakashima, T. (1987) Azinomycin-a and Azinomycin-B, New Antitumor Antibiotics .3. Antitumor-Activity, *J Antibiot 40*, 60-65.

36. Armstrong, R. W., Salvati, M. E., and Nguyen, M. (1992) Novel Interstrand Cross-Links Induced by the Antitumor Antibiotic Carzinophilin Azinomycin-B, *Journal of the American Chemical Society 114*, 3144-3145.

37. Ding, W., Deng, W., Tang, M., Zhang, Q., Tang, G., Bi, Y., and Liu, W. (2010) Biosynthesis of 3-methoxy-5-methyl naphthoic acid and its incorporation into the antitumor antibiotic azinomycin B, *Mol Biosyst 6*, 1071-1081.

CHAPTER 2

REVERSAL OF SUBSTRATE SPECIFICITY OF CMP *N*-GLYCOSIDASE TO dCMP

Section 2.1. Introduction

MilB is an enzyme involved in the biosynthesis of the natural product mildiomycin (Figure 2.1A), a commercially available antifungal agent isolated from *Streptomyces rimofaciens.*[1-3] MilB preferentially hydrolizes the *N*-glycosidic bond of 5-hydroxymethyl cytidine 5′-monophosphate (hmCMP) to release 5-hydroxymethyl cytosine, which is later incorporated into mildiomycin (Figure 2.1B).[1] Previous studies found that MilB also hydrolyzes cytidine 5′-monophosphate (CMP) to its cytosine and ribose 5′-phosphate (Figure 2.1B).[1] MilB is a member of the 2′-deoxynucleoside ribosyltransferase (NDT) superfamily. Typically, NDTs catalyze the reversible transfer of 2′-deoxyribose from a donor 2′-deoxynucleoside to an acceptor base. Their structural and functional properties have been investigated in detail.[4-6] NDTs are found in a limited number of organisms and are thought to be important for nucleoside recycling when alternative nucleoside salvage pathways do not exist.[5] In contrast to most NDTs, MilB participates in a hydrolysis reaction, in which the acceptor base is replace by a water molecule. While NDT exhibits hydrolysis activity in the absence of an acceptor base, the hydrolysis activity is about 100-fold lower than its transferase reaction.[6]

BLAST analyses starting with the MilB sequence suggest homology to BcmB from *Clostridium botulinum* (47% sequence identity). Like MilB, BcmB also shows

CMP hydrolase activity, but is utilized in a different biosynthetic pathway. BcmB is responsible for the production of the 5-hydroxymethylcytosine precursor for the biosynthesis of bacimethrin, a thiamin antimetabolite found in *C. botulinum* (Figure 2.1) (T.P. Begley, manuscript in preparation). While MilB and BcmB both belong to the NDT superfamily, they differ from other members of this family in their substrate preferences. Most enzymes containing the NDT-motif are specific for 2′-deoxynucleosides.[4, 7] MilB and BcmB hydrolize hmCMP, a ribonucleotide containing 5′-monophosphate and 2′-hydroxy groups.

In addition to MilB and BcmB, several other enzymes containing the NDT motif function as nucleoside transferases. BlsM from *Streptomyces griseochromogenes* (52% sequence identity to MilB) is a nucleotide hydrolase, with CMP as the preferred substrate. The cytosine produced from the BlsM reaction is incorporated into blasticidin S, a commercially available antibiotic compound (Figure 2.1A).[8] Rcl, another member of the NDT superfamily, catalyzes the hydrolysis of purine 2′-deoxyribosylnucleotides and is inhibited by ribonucleotides.[9] Its solution structure,[10, 11] and recent crystal structures with bound inhibitors,[12] provide highly detailed information about its substrate-enzyme interactions. Rcl is unusual because it is the only NDT family member found in mammals. Rcl is upregulated in some cancers and contributes to tumorigenesis; however, the details of its cellular function are unclear.[13, 14]

Because members of the NDT superfamily are not widely distributed and they show significant variation in substrate specificity, their evolution is not clearly

understood. Conservation of key active site residues suggests a common catalytic mechanism even though the substrates for different family members vary in their preference for ribosyl or 2′-deoxyribosyl groups, 5′-hydroxyl or 5′-phosphate groups, and the type of nucleobase. Here we report the X-ray structures of MilB and BcmB, and of MilB complexed with CMP. Comparisons with the structures of NDT[6] and Rcl[10-12] revealed a structural basis, which was not obvious from sequence comparisons alone, by which NDT family members differentiate between ribosyl and 2′-deoxyribosyl substrates. The selectivity is attributed to a phenylalanine residue found in MilB and BcmB that is replaced by tyrosine in NDT and Rcl. In support of this hypothesis, *in vitro* enzymatic activity assays showed that for both MilB and BcmB, mutation of the active site phenylalanine residue to tyrosine resulted in a reversal of substrate preference from CMP to dCMP.

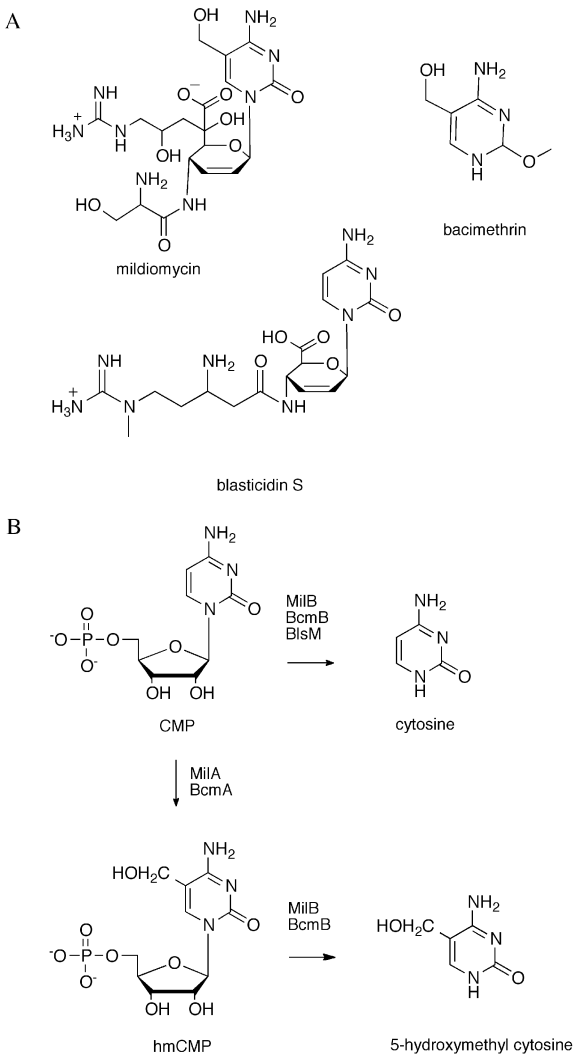


Figure 2.1. Natural products with cytosine precursors. (A) Cytosine, or a cytosine derivative, is incorporated in the natural products mildiomycin, bacimethrin and blasticidin S. (B) The first steps in their biosyntheses involve the enzymatic release of cytosine or hydroxymethylcytosine from a CMP precursor by a CMP hydrolase enzyme.

Section 2.2. Materials and Methods

Cloning, Expression, and Purification of MilB and BcmB. All cloning procedures followed standard DNA manipulation methods. The *S. rimofaciens* ZJU5119 *milB* gene was commercially synthesized by DNA2.0 and *bcmB* was cloned

using genomic DNA from *C. botulinum* A str. ATCC 19397. The *milB* and *bcmB* genes were inserted into pTHT, a modified pET-28 plasmid (Novagen) containing an N-terminal His_6-tag and TEV protease recognition site. All mutant plasmids were prepared at the Cornell Protein Production and Purification Facility by site-directed mutagenesis of the native gene following standard, PCR-based mutagenesis.[15]

Escherichia coli B834(DE3), a methionine auxotrophic cell line, was transformed with the pTHT-*milB* wild-type (WT) plasmid. *E. coli* BL21(DE3) cells were used to express all other proteins. Starter cultures were grown from a single colony in 10 mL of sterile LB media containing 30 μg/mL kanamycin (LB-kan) at 37 °C overnight with shaking. Selenomethionine (SeMet)-containing MilB was prepared in 1.5 L- cultures in minimal media with 0.4% glucose, 1x MEM vitamin mix, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 25 mg/L $FeSO_4$, 20 mg/L all amino acids except methionine, and 50 mg/L of L-selenomethionine. To remove any LB media during SeMet protein preparation the overnight culture was centrifuged at 1000*g* for 20 min before the cell pellet was collected and resuspended in minimal media. The 1.5 L volume was then inoculated with 5 mL of starter culture and allowed to grow at 37 °C with shaking until reaching an OD_{600} of 0.6, at which point the incubator temperature was reduced. Once reaching an induction temperature of 15 °C and OD_{600} ~0.8, protein expression was initiated by inoculating the cultures with 0.5 mM isopropyl β-D-thiogalactopyranoside. After 18 h, cells were centrifuged at 2000*g* for 20 min, collected, and frozen at -20 °C for storage. For native protein overexpression, 5 mL of

overnight culture was added directly to 1.5 L of LB-kan media. The cells were then grown, collected and stored as described for the preparation of SeMet-MilB protein.

After thawing and resuspending the cell pellet in 45 mL of lysis buffer (50 mM Tris, 300 mM NaCl, and 20 mM imidazole at a pH of 8.0), the cells were lysed on ice with two rounds of sonication. The cell extract was centrifuged for 30 min at 40,000*g* at 4 °C after sonication. The supernatant was then collected and loaded onto a 2 mL Ni-NTA column (Qiagen) preequilibrated with the lysis buffer. The column was washed with 50 mL of lysis buffer to remove any nonspecifically bound contaminants. To elute the protein, the column was washed with elution buffer (50 mM TRIS, 300 mM NaCl, and 250 mM imidazole at a final pH of 8.0). The elution sample was collected and further purified using size exclusion chromatography (HiLoad Superdex 200 PG; GE Healthcare). The purification procedures for WT and mutant BcmB, WT and mutant MilB, and SeMet MilB all followed this protocol. The native MilbWT polyhistidine-tag was removed for crystallization by incubation with TEV protease for 12 h at 4 °C. The sample was passed over a Ni-NTA column preequilibrated with lysis buffer and the flow-through volume containing MilB was collected. The purity of the sample was analyzed by SDS-PAGE before concentration to ~20 mg/mL. Protein was buffer exchanged by overnight dialysis into storage buffer (SeMet MilB: 50 mM Tris, pH 8.5, 500 mM NaCl; WT MilB: 10 mM Tris, pH 8.0, 100 mM NaCl; BcmB: 10 mM Tris pH 8, 50 mM NaCl) before flash freezing with liquid nitrogen for storage at -80 °C.

Crystallization of MilB and BcmB. The hanging-drop vapor diffusion method was used for crystallization of both MilB and BcmB. After mixing equal volumes of protein sample and reservoir solution, the samples were equilibrated overnight at 18 °C against a total reservoir volume of 500 μL. Protein concentrations for crystallization were 10 mg/mL for MilB and 15 mg/mL for BcmB. Initial crystallization conditions were determined using commercially available sparse matrix screens (Hampton Research, Emerald Biosystems).

SeMet-MilB crystals grew as hexagonal rods approximately 200 μm long and 40 μm wide under the optimized crystallization condition of 8% PEG 1000, 0.1 M phosphate-citrate pH 4.2, and 0.2 M Li_2SO_4 after two days. A second condition for MilB crystallization was found with the N-terminal polyhistidine tag removed. Square rods 400 μm long by 20 μm wide crystallized after two days in 0.1 M Tris pH 7.5, 0.2 M $MgCl_2$, and 18% PEG 8000. The ligand soaked into the crystals added using a solution containing the components of the reservoir solution supplemented with 10 mM cytosine and 10 mM ribose 5-phosphate for 5 min before cryoprotection.

BcmB crystals formed 200 μm by 150 μm plates with 10 μm thickness in 18% PEG MME 2000, 0.1 M phosphate-citrate pH 4.4, 0.01 M spermine, and 0.2 M $(NH_4)_2SO_4$ after five days. All crystals were cryoprotected with a solution consisting of their crystallization condition supplemented with 20% ethylene glycol before flash freezing with liquid nitrogen.

Data Collection and Processing. X-ray diffraction experiments were conducted at the Advanced Photon Source NE-CAT beamline 24-ID-C (Argonne

National Lab) using a Quantum315 detector (Area Detector Systems Corp.) using vitrified crystals. A fluorescence scan was conducted to determine the data collection wavelength corresponding to the maximum f'' for the SeMet-MilB crystals. All data collection runs used the oscillation method with 1.0° rotation per frame. HKL2000 was used for indexing, integrating and scaling the data.[16] Data collection and processing statistics are listed in Table 1.

Structure Determination, Model Building, and Refinement. The crystal structure of MilB was solved by single-wavelength anomalous diffraction (SAD) using SeMet-MilB. This crystals belong to space group $P6_222$ with a Matthews coefficient of 2.1 $Å^3$/Da and an estimated solvent content of 41%, assuming one protein chain per asymmetric unit.[17] Initial phase calculation and model building were carried out using PHENIX AutoSol.[18, 19] Subsequent model improvement was performed using COOT for manual manipulations and PHENIX.refine with default parameters for structure refinement.[19, 20] Water molecules were added during later rounds of refinement in PHENIX and the final structure was analyzed using COOT and MolProbity.[21] The structure of MilB complexed with CMP was determined in space group $C2$ with two protein chains per asymmetric unit. Under this new condition, MilB also crystallized with a Matthews coefficient of 2.1 $Å^3$/Da corresponding to 41% solvent content. SKETCHER was used for ligand generation before placement into $F_o - F_c$ electron density.[22] Model improvement was performed as described above.

BcmB crystallized in space group *C*2 with three chains per asymmetric unit and a Matthews number of 3.2 $Å^3$/Da, corresponding to a solvent content of approximately 61%. The structure of BcmB was determined using molecular replacement, with the newly solved MilB structure as the search model.[23] MolRep placed three chains in the asymmetric unit and further model building and refinement was carried out using COOT and Refmac.[20, 24] The three chains correspond to one dimer formed through noncrystallographic twofold symmetry and one dimer generated by crystallographic symmetry. Several rounds of refinement using Refmac and PHENIX.refine resulted in a BcmB model with R-factor and R_{free} values of 21.7% and 24.2%, respectively. Refinement statistics for SeMet-MilB, MilB-CMP and BcmB are summarized in Table 2.1.

Analytical HPLC. Reaction mixtures were analyzed by reverse-phase HPLC on an Agilent 1200 HPLC system equipped with a quaternary pump and thermostatted autosampler (4 °C). The stationary phase was a Supelcosil LC-18-T column (15 cm × 4.6 mm, 3 μm particles). The LC eluent (1 mL/min flow rate) consisted of a gradient of methanol and 6 mM *N*,*N*-dimethylhexylamine (DMHA) in water that was adjusted to pH 6.5 by addition of acetic acid. In the analytical method the percentages of DMHA (*D*) and methanol (*M*) balanced with water at time *t* varied according to the following scheme: (*t*,*M*,*D*): (0,0,100), (2,10,80), (12,20,60), (14,65,10), (16,0,100), (21,0,100). Chromatograms were detected using the absorbance at 270 nm. UV-Vis spectra of substrate and product peaks were also collected and assessed.

Table 2.1: Summary of MilB and BcmB data collection and refinement statistics.

	SeMet MilB-SO_4	MilB-CMP	BcmB-PO_4
beamline	APS 24-ID-C	APS 24-ID-C	APS 24-ID-C
resolution (Å)	1.95	1.55	2.99
wavelength (Å)	0.97918	0.97918	0.97918
space group	$P6_222$	$C2$	$C2$
a (Å)	97.8	45.2	177.9
b (Å)	97.8	100.4	40.2
c (Å)	61.9	71.5	98
β (°)	90.0	99.6	98.0
Matthews coefficient	2.1	2.1	3.2
% solvent	41	41	61
mol/asu	1	2	3
Measured reflections	137124	162974	54150
Unique reflections	23697	44996	14604
Average I/σ	37.1 (5.3)[a]	25.1(3.6)	17.8 (4.7)
Redundancy	5.7 (5.4)	3.6(3.6)	3.7(3.8)
Completeness (%)	98 (97)	99.8 (100)	99.9(100)
R_{sym} (%)[b]	5.4 (32.1)	5.7 (35.2)	10.0(34.2)
No. of protein atoms	1164	2369	2700
No. of ligand atoms	10	41	15
No. of water atoms	96	196	10
Reflections in working set	23686	42663	14582
Reflections in test set	1168	2269	737
R-factor/R_{free} (%)[c]	20.6/25.2	19.9/22.7	20.7/23.9
rms deviation from ideals			
bonds (Å)	0.007	0.006	0.015
angles (°)	1.072	1.061	1.975
average B factor for protein (Å^2)	28.1	18.0	49.1
average B factor for water (Å^2)	34.1	25.6	54.5
average B factor for ligand (Å^2)	31.3	17.0	84.2
Ramachandram plot			
most favored (%)	97.3	98.0	96.3
allowed (%)	2.7	2.0	3.7
disallowed (%)	0.0	0.0	0.0

[a]Values in parentheses are for the highest-resolution shell. [b]$R_{sym} = \Sigma\Sigma_i | I_i - \langle I \rangle | / \Sigma \langle I \rangle$, where $\langle I \rangle$ is the mean intensity of the N reflections with intensities I_i and common indices h,k,l. [c]$R_{work} = \Sigma_{hkl} | |F_{obs}| - k |F_{cal}| | / \Sigma_{hkl} |F_{obs}|$ where F_{obs} and F_{cal} are observed and calculated structure factors, respectively, calculated over all reflections used in the refinement. R_{free}, is similar to R_{work} but calculated over a subset of reflections (5%) excluded from all stages of refinement.

Substrate Specificity for MilB and BcmB. Substrate specificity studies for MilB and MilB F17Y were carried out with adenosine 5′-monophosphate (AMP), CMP, 2′-deoxycytidine 5′-monophosphate (dCMP), guanosine 5′-monophosphate (GMP), and inosine 5′-monophosphate (IMP) (Figure S2.1). Experiments were designed based on the previously described assay conditions used for MilA/MilB [25]. *In vitro* enzymatic assays were prepared with a final volume of 50 µL. AMP, CMP, dCMP, GMP, or IMP (1 mM) were added to MilB WT or MilB F17Y (10 µM) in 50 mM Tris buffer, pH 7.5 and then the samples were incubated at 37 °C for 3 h. Control samples omitting enzyme were also prepared. Reactions were quenched by boiling at 100 °C for 5 min. Samples were diluted with DMHA, filtered using a 3 kDa MWCO membrane, and analyzed by HPLC as described above. Reaction products were compared to commercially available standards that were analyzed under identical conditions. Substrate specificity studies for BcmB and BcmB F6Y were carried out with CMP and dCMP as described above.

Kinetics for MilB and BcmB. Kinetic parameters were monitored based on cytosine production from CMP or dCMP for MilB WT, MilB F17Y, BcmB WT, and BcmB F6Y. Each enzyme was incubated with varying substrate concentrations (see Table S2.1) in 50 mM Tris, pH 7.5 for 30 min at 37 °C and then the reactions were quenched by boiling at 100 °C for 5 min. Samples were diluted with DMHA, filtered using a 3 kDa MWCO membrane, and analyzed by HPLC as described above. Reaction products were compared to commercially available standards that were analyzed under identical conditions. The product absorbance signal was measured by

integration (ChemStation, Agilent Technologies) and compared to cytosine standards of known concentration. Kinetic parameters were calculated using the Michaelis-Menten function in KaleidaGraph 4.0 (Synergy Software).

Figure 2. Preparation. All figure 2.s were prepared using PyMOL [26], ChemDraw (Cambridge Biosoft), KaleidaGraph 4.0 (Synergy Software), or Excel (Microsoft), and compiled in Photoshop (Adobe).

Section 2.3. Results

Structure of MilB. MilB with bound SO_4 contains one protomer per asymmetric unit; however, a dimer is formed by crystallographic twofold symmetry. The final model at 1.95 Å resolution contains residues 11-162 of the possible 170 amino acid residues. The protomer of MilB contains an α/β fold with a five-stranded parallel β-sheet with a strand order of 21345 (flavodoxin-like). The β-sheet is flanked on both sides by α-helices in an asymmetric manner (Figure 2.2A and 2.2B). Dimer formation occurs through interactions between helices α4, α5, and α6 of the two protomers with an interface area of 1170 $Å^2$. Size exclusion chromatography confirmed the dimeric state of MilB (data not shown), which is consistent with the quaternary structures of both NDT (trimer of dimers) and Rcl.

The MilB/CMP complex also crystallizes with one dimer per asymmetric unit. The final model at 1.55 Å resolution contains residues 12-168 of chain A and residues 10-162 of chain B. Each protomer is occupied by one CMP molecule and the overall structure is essentially the same as that of the MilB/SO_4 complex. Each dimer features two equivalent active sites.

MilB Active Site. The MilB active site is located at the C-terminal edge of the β-sheet, between helices α4 and α5 (Figure 2.2C). All of the interactions necessary for substrate binding and catalysis are contained within a single MilB protomer, with no direct participation from the adjacent protomer. The phosphate/sulfate binding pocket contains hydrogen bonding interactions with the Ser97 sidechain, the amide groups of Gly99, Arg23, and Phe22, and two well ordered water molecules (Figure 2.2D). Ordered water molecules occupy the rest of the active site in the MilB/SO_4 structure.

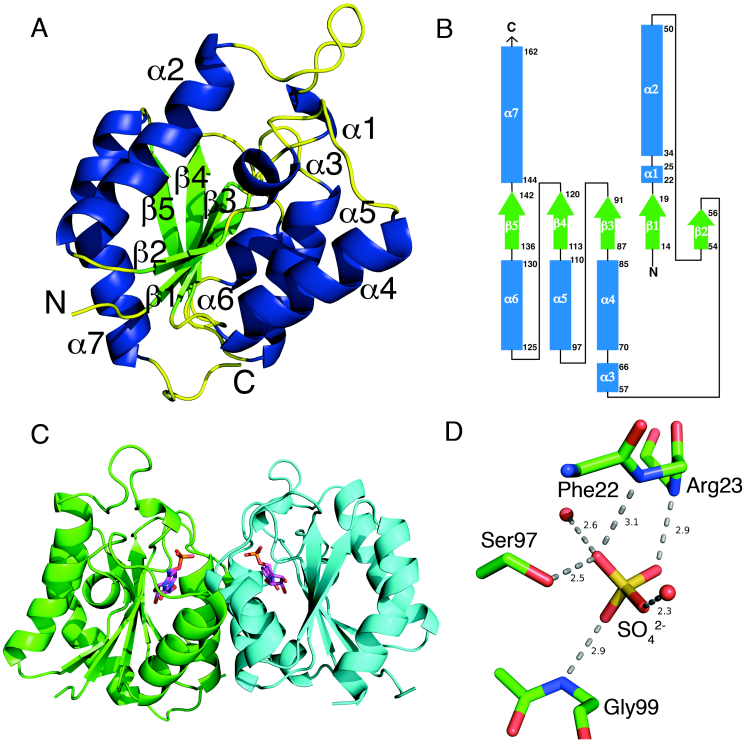


Figure 2.2. Structure of MilB. (A) The MilB protomer structure adopts an overall α/β-twist fold, depicted with β-strands in green and α-helices in blue. (B) The topology diagram of MilB with β-strands in light green and α-helices in light blue. (C) The dimeric structure of MilB with chain A in green and chain B in cyan and ligands bound in the active sites. (D) The active site phosphate-binding pocket is occupied by a sulfate ion in the MilB/SO_4 complex.

The structure of the MilB/CMP complex provides the molecular details of the nucleobase and ribosyl binding sites (Figure 2.3). The products of the CMP hydrolysis reaction, cytosine and ribose 5-phosphate, were soaked into the MilB crystals. Continuous electron density is observed in the active site representing CMP, the product of the reverse reaction, yet our *in vitro* studies did not show significant formation of CMP upon incubation of MilB, cytosine, and ribose 5-phosphate (data not shown). Thus, the presence of CMP in the MilB/CMP active site is believed to be an artifact of crystallization conditions. Glu92 accepts a hydrogen bond from the 4-amino group of the cytosine base and, if protonated, would donate a hydrogen bond to CMP N5. Glu103 accepts hydrogen bonds from the CMP 2′- and 3′-hydroxyl groups. Additionally, the side chain of Asp78 is within hydrogen bonding distance of the cytosine O2. Phe17 C4 is positioned 3.5 Å from the CMP 2′-oxygen in the MilB structure.

Structure of BcmB. BcmB crystallized in space group *C*2, with three protein chains in the asymmetric unit. The structure of BcmB was determined to 3.0 Å resolution using the MilB structure (47% sequence identity) described above as the molecular replacement search model. The protomeric structure of BcmB displays the same α/β-twist fold as MilB, with five parallel β-strands forming a parallel β-sheet flanked on either side by α-helices (Figures S2.2A and S2.2B). The final BcmB model contains 120 amino acids in chain A, 123 in chain B, and 110 in chain C. The missing regions include residues 11-20 and 110-113 in chain A, 11-21 and 112 in B, and 10-24, and 109-112 in C, which correspond to loop regions. Missing residues 47-

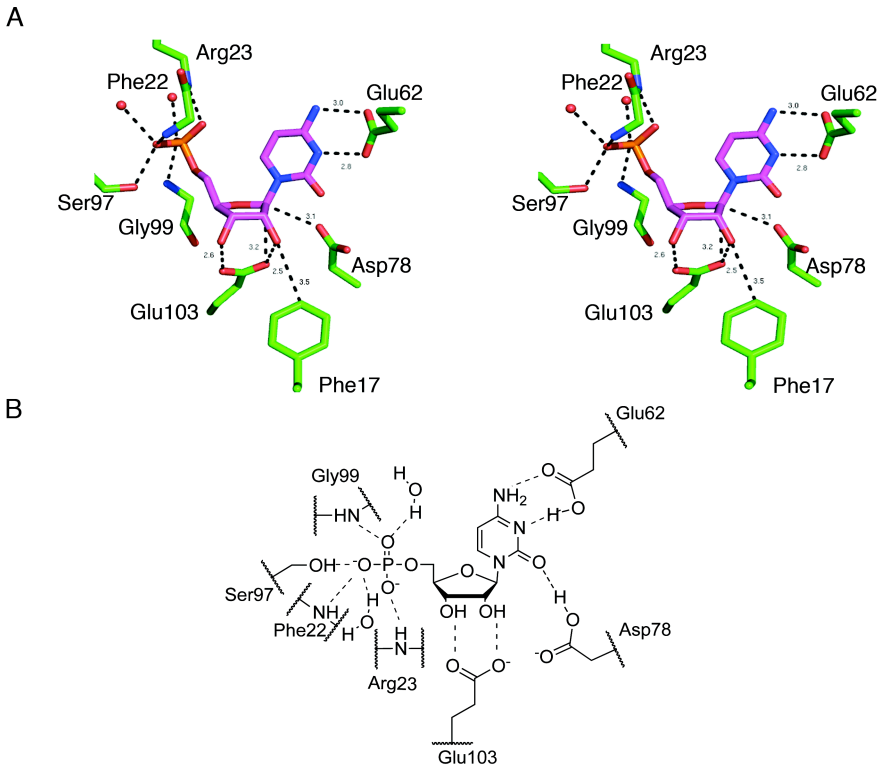


Figure 2.3. Substrate binding in MilB. (A) MilB-CMP binding interactions represented in stick view. (B) Schematic diagram of the MilB active site with CMP bound.

55, 48-56, and 46-58 in protomers A, B, and C, respectively, correspond to a short α-helix and loop when aligned with MilB.

Chains A and B form a dimer by twofold noncrystallographic symmetry. The buried surface area at the interface is 1320 Å^2. Chain C forms a dimer with an equivalent chain C using crystallographic twofold symmetry. The buried surface area of this interaction is 1210 Å^2. This small discrepancy between the A/B and C/C interface surfaces areas can be attributed to additional disordered regions in chain C. Residues 58 and 59 in A and B and residue 114 in chain A are present at the A/B dimer interface, but are disordered at the C/C dimer interface. The oligomeric state of

BcmB observed in the crystal structure is supported by size exclusion chromatography results (data not shown) and closely resembles the dimer formed by MilB (Figure S2.2C).

BcmB Active Site. The active site of BcmB is similar to that of MilB. Each active site contains a bound phosphate ion from the crystallization solution. The phosphate makes hydrogen bonds with the side chain of Ser86 and the amide backbone of Gly88 in the phosphate-binding pocket (Figure S2.2D). Based on structural comparison to MilB, the BcmB active site residues that participate in substrate binding include Asp67, Glu92, and Phe6 (Figure 2.4). Glu41 falls in a disordered region in the BcmB crystal structure, but is expected to be equivalent to Glu62 in MilB on the basis of sequence alignment.

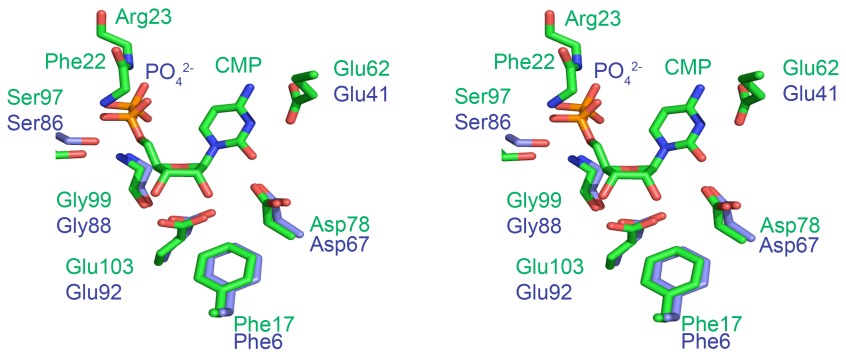


Figure 2.4. BcmB/PO_4 and MilB/CMP active site comparison. A stereoview of BcmB/PO_4 and MilB/CMP active sites superimposed. The active site residues of BcmB and MilB display a conserved architecture. Glu41 is disordered in the BcmB structure but conserved in sequence with MilB Glu62.

Kinetics and Substrate Specificity for MilB and BcmB. Initial studies determined the substrate specificity of MilB for cytosine containing nucleotides.[1]

MilB preferentially hydrolyzed CMP, but cytosine was also observed, to a lesser extent, from dCMP (Figure 2.5). MilB showed little to no hydrolysis of purine containing substrates (Figure S2.1). CMP or dCMP was incubated with MilB F17Y and the production of cytosine was monitored to measure *N*-glycosidase activity. MilB F17Y successfully cleaved dCMP but was no longer able to effectively hydrolyze CMP (Figure 2.5). Similar studies were conducted for both BcmB wild type (WT) and the corresponding active site Phe mutant (F6Y). As with MilB WT, BcmB WT can hydrolyze CMP and, to a lesser extent, dCMP (Table 2.2). Conversely, BcmB F6Y hydrolyzed dCMP with a 2′-deoxyribosyl group, while showing only minor activity with CMP (Figure S2.3).

Steady-state kinetic parameters were determined for MilB WT, MilB F17Y, BcmB WT, and BcmB F6Y using either CMP or dCMP as substrates (Table 2.2). The k_{cat}/K_m value for MilB WT with hmCMP was previously reported as 22.1 $M^{-1}s^{-1}$.[1] The k_{cat}/K_m values observed for MilB WT hydrolysis of CMP or dCMP were of 5.2 $M^{-1}s^{-1}$ and 1.1 $M^{-1}s^{-1}$, respectively.[1]

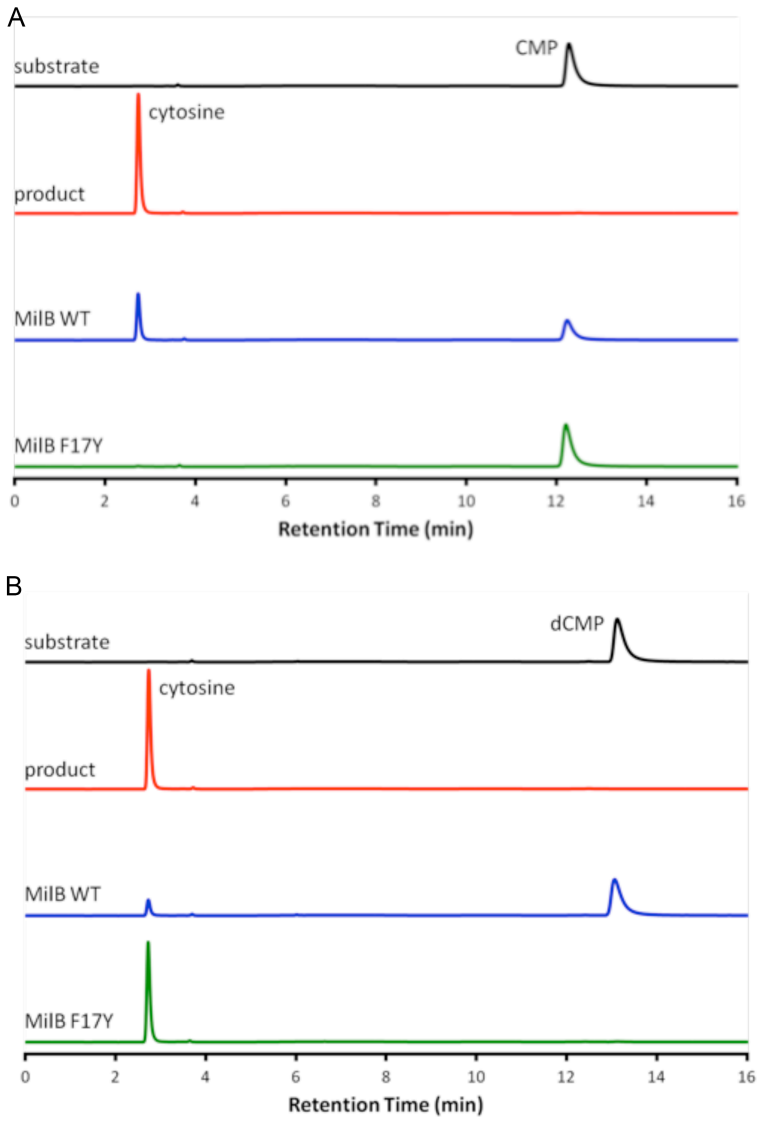


Figure 2.5. Substrate specificity of MilB WT and MilB F17Y. (A) When MilB WT is incubated with CMP a large peak representing free cytosine is observed. For MilB F17Y incubated with CMP, little hydrolysis is observed. (B) The opposite effect is observed with dCMP as substrate. MilB F17Y hydrolyzes dCMP producing free cytosine. There is comparably less dCMP hydrolysis by MilB WT.

Table 2.2: Summary of kinetic data

Protein (Substrate)	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m ($M^{-1}s^{-1}$)
MilB WT (CMP)	1.19 (0.06)	3.8 (0.5)	5.2 (0.7)
MilB WT (dCMP)	0.36 (0.02)	5.4 (0.8)	1.1 (0.2)
MilB F17Y (CMP)	0.0126 (0.0003)	4.5 (0.4)	0.046 (0.004)
MilB F17Y (dCMP)	7.6 (0.3)	1.2 (0.2)	100 (10)
BcmB WT (CMP)	0.090 (0.003)	0.19 (0.02)	7.8 (0.9)
BcmB WT (dCMP)	0.055 (0.001)	0.57 (0.05)	1.6 (0.1)
BcmB F6Y (CMP)	0.0048 (0.0004)	0.9 (0.2)	0.09 (0.02)
BcmB F6Y (dCMP)	0.081 (0.001)	0.042 (0.004)	32 (3)

Section 2.4. Discussion

Structure Comparison. MilB and BcmB both adopt an α/β-fold composed of five parallel β-strands forming a core β-sheet flanked on either side by several α-helices. This fold is highly conserved among the NDT superfamily enzymes. A pairwise DALI comparison between the MilB and BcmB structures results in a Z-score of 16.9 and rmsd of 1.7, confirming their highly conserved tertiary structure (Table 2.3).[27] A DALI structure search against the PDB starting with MilB reveals structural homology to the known 2′-deoxyribosyltransferases and Rcl. Though their sequence identities are relatively low (11-22%), the enzymes retain similar folds. Based on a Z-score of 15.5 with a rmsd value of 2.3Å, MilB has the highest structural similarity to *R. norvegicus* Rcl (PDB ID 4FYI).[12] Similar rmsd values of 2.5 Å and 2.9 Å are observed upon comparison of MilB to the Rcl solution structures with PDB

IDs 2KHZ[11] and 2KLH,[10, 11] respectively. Several 2′-deoxyribosyltransferases, such as the *Trypanosoma brucei*[30] and *L. leichmannii* [6, 28, 29] NDTs (PDB ID 2F2T and 1F8X) and *L. helveticus* purine nucleoside 2′-deoxyribosyltransferase (PTD) (PDB ID 1S2L)[29] are also among the most homologous structures.

While the enzymes in the NDT superfamily maintain a similar tertiary structure, their oligomeric assembly is not fully conserved. Both NDTs and PTD form a hexamer composed of a trimer of homodimers, while Rcl, MilB, and BcmB are observed to form only dimers.[6, 11, 28] Comparison of the dimers reveals a conserved mode of dimerization (Figure 2.6). Like many members of the NDT superfamily,[10] the interactions observed at the dimer interfaces in MilB and BcmB are mainly hydrophobic (>97%).

Table 2.3: Enzymes structurally similar to MilB

Protein	PDB ID	Z score	rmsd	%identity	No. aligned residues
C. botulinum BcmB	4JEL	16.9	1.7	46	123
R. norvegicus Rcl	4FYI	15.5	2.3	21	126
R. norvegicus Rcl (NMR)	2KHZ	14	2.5	22	134
R. norvegicus Rcl (NMR)	2KLH	13.7	2.9	21	131
T. brucei NDT	2F2T	12	2.8	13	126
L. helveticus PTD	1S2L	11.9	2.4	19	122
L. leichmannii NDT	1F8X	10.6	2.8	11	123

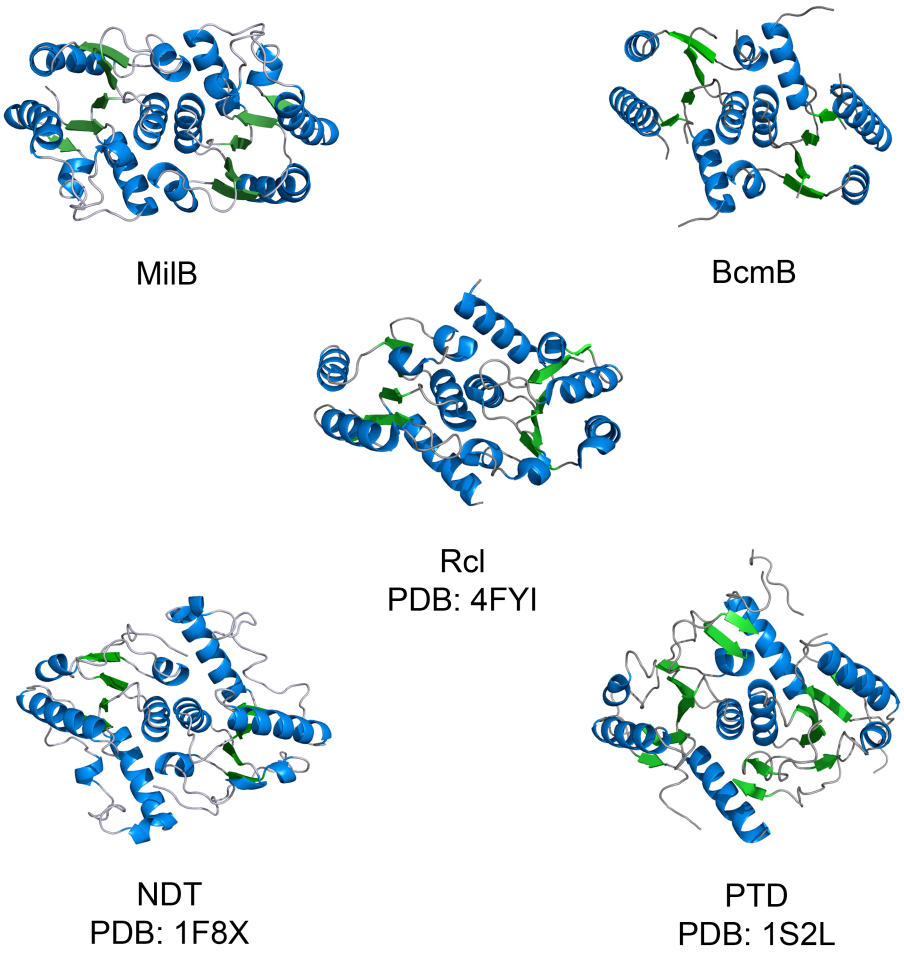


Figure 2.6. Comparison of enzyme dimerization. MilB and BcmB are observed to be dimers. Like MilB, BcmB and Rcl have a dimeric tertiary structure. In their functional forms, NDT and PTD are hexamers composed from a trimer of dimers, in which the dimer (shown) is similar to that of MilB.

Base Specificity. Structural analysis of MilB, BcmB, and related enzymes reveals the basis of their substrate specificities. MilB shows specificity for the pyrimidine containing hmCMP or CMP substrates. It does not hydrolyze the purine containing nucleotides, demonstrated by an inability to hydrolyze AMP, GMP, or IMP (Figure S2.1). While highly specific for hmCMP or CMP, its relatives display variable

nucleobase specificity (Table 2.4). NDT accepts all naturally occurring and some synthetic nucleosides, while PTD[30] and Rcl[7] are specific for purine containing nucleosides or nucleotides, respectively. In the structure of MilB/CMP, the Glu62 carboxylic group is positioned approximately 3 Å from the cytosine, forming hydrogen bonds with the cytosine N3 and 4-amine group (Figure 2.3). Helix α3 (on which Glu62 is located), helix α4, and the loop region formed by residues 67-70 shield the MilB active site. As a result, MilB has a relatively compact substrate-binding site that cannot accommodate the bulkier purine containing nucleotides. BcmB specificity is predicted to function similarly to MilB based on sequence and structural homology.

Table 2.4: Summary of NDT-superfamily enzymes

Enzyme	Preferred Base	5'-specificity	2'-specificity	Acceptor/ Nucleophile
MilB	hydroxymethylcytosine/ cytosine	phosphate	OH	water
BcmB	hydroxymethylcytosine/ cytosine	phosphate	OH	water
BlsM	cytosine	phosphate	OH	water
Rcl	purine	phosphate	H	water
NDT	purine/pyrimidine	OH	H	purine/ pyrimidine
PTD	purine	OH	H	purine

The carboxylic acid from the NDT C-terminal Tyr157′ and Gln46 amine interacts with its substrate nucleobase. NDT can accommodate a variety of nucleobases because of its flexibility in both accepting and donating hydrogen bonds.[6] Like NDT, PTD interacts with nucleobases via hydrogen bonds from the carboxylic

tail of the C-terminal tyrosine located in the active site. PTD diverges from NDT, as there are no equivalent interactions to those provided by Gln46 in NDT.[28] The complexed Rcl structures[10-12] reveal a large nucleobase binding area with no specific enzyme-nucleobase interactions, allowing Rcl to accommodate a variety of nucleobases in its active site. Because the Rcl active site is rather open, it is solvent accessible and favors the larger purine containing substrates that can aid in shielding the active site from the solvent. As evidenced by the structures of MilB and other NDT superfamily members, their nucleobase specificities are derived from a combination of favorable enzyme-substrate hydrogen bonding interactions and active site accessibility.

Specificity at the 5′-position. The NDT family members differ in their substrate preference for either a hydroxl or phosphate group at the 5′-position (Table 2.4). The high-resolution MilB/CMP structure reveals key information about the enzyme-substrate binding interactions at the CMP 5′-position, including seven hydrogen bonds. Two ordered water molecules each contribute one hydrogen bond, while Ser97 is hydrogen bonding distance from either of the two 5′-phosphate oxygen atoms. While the active site serine residue observed in MilB is conserved in both BcmB and BlsM (Figure 2.7), the remaining interactions in the MilB phosphate-binding pocket come from amide backbone groups and ordered water molecules. The amide backbones of Phe22, Arg23, and Gly99 act as hydrogen bond donors, completing the phosphate-binding pocket (Figure 2.8A). Although the identities of the backbone residues are not conserved in sequence, the phosphate ion in BcmB/PO_4 is

observed in an analogous binding pocket to MilB (Figure 2.8B). This pocket is also rich with hydrogen bond donors resulting in preference for phosphate groups at the 5′-position.

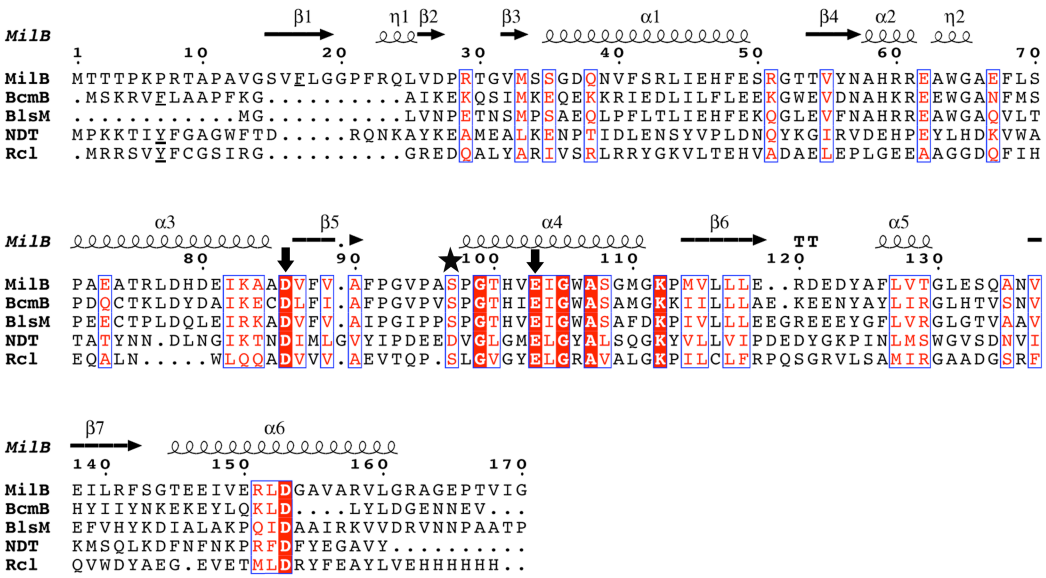


Figure 2.7. Sequence alignment of enzymes containing the nucleoside 2′-deoxyribosyltransferase motif. Sequence comparison between MilB/BcmB and other family members reveals conservation of the key catalytic residues (arrows) but not of those conferring substrate-binding specificity. The serine residue involved in phosphate binding is indicated with a star. Phenylalanine and tyrosine residues known to be positioned in the active site are underlined. Representative sequences for NDT from *L. leichmannii* and Rcl from *R. norvegicus* were used.

A 5′-phosphate binding pocket utilized for nucleotide binding was also observed in Rcl.[10-12] The structure of Rcl shows hydrogen bonding contacts between the GMP phosphate group and Ser17, Arg19, Ser87, and Ser117′ (Figure 2.8B). This motif is highly conserved in the sequences of 13 different species in Rcl.[10] While MilB does not share this motif, the positively charged pocket is consistent with

preference for nucleotide monophosphate binding. In NDTs and PTD, an aspartate residue plus either an asparagine or glutamine residue stabilizes nucleoside 5′-hydroxyl groups, favoring nucleoside binding (Figure 2.8C)[6, 28] and preventing the binding of 5′-phosphate containing nucleotides preferred by MilB and BcmB.[4]

Specificity at the 2′-position. While MilB and BcmB contain the conserved NDT motif, their ability to hydrolyze the *N*-glycosidic bonds of ribosylnucleotides makes them unique from other enzymes in this class. NDTs, PTD, and Rcl show no glycosidase activity towards ribosyl-containing substrates, which in some cases are inhibitory.[9] Conversely, MilB and BcmB show higher activity towards substrates containing 2′-hydroxy groups (Table 2.2). MilB and BcmB both have very low levels of activity for the hydrolysis of dCMP. In the related Rcl and NDT structures, a tyrosine residue is observed near the substrate 2′-binding position (Figure 2.8). Not obvious from sequence analysis alone, the structures of MilB and BcmB reveal a phenylalanine in the position equivalent to this tyrosine near the conserved catalytic glutamate (Figure 2.7). Compared to Rcl or NDTs, this phenylalanine residue is the only residue near the 2′-binding position that differs in either MilB or BcmB.

The identity of the active site phenylalanine or tyrosine residue proves to be important in differentiating between ribosyl or deoxyribosyl containing substrates by MilB and BcmB. Structural studies of PTD,[28] which has a tyrosine residue near the 2′-binding position, suggested that ribosylated substrates form hydrogen bonds with the catalytic glutamate, rendering it unreactive. This rational contradicts our findings in which MilB and BcmB react with ribosylated substrates, despite also showing

interactions between their catalytic glutamate and the substrate 2′-hydroxy group. When both a hydroxy containing substrate and active site tyrosine are present, the glutamate may be too constricted by hydrogen bonds to position itself for catalysis. In the absence of an active site tyrosine, the catalytic glutamate is not restricted by the ribosylated substrate. Rather, the ribosylated substrate may help orient the glutamate for catalysis.

Acceptor/Nucleophile Binding Site. Previous studies of NDT-superfamily members have explored the structural basis of nucleophile binding at the active site. In NDT, the interaction between Gln46 and the nucleobase positions the loop region, on which Gln46 is located, shielding the active site from solvent.[28] Along with a hydrophobic binding pocket, this excludes water as a nucleophile. After initial *N*-glycosidic bond cleavage and release of the base, NDT binds the incoming acceptor base, positioning it for the second half of the NDT transferase reaction. Due to more limited substrate-enzyme interactions in PTD, the corresponding loop region does not shield its active site[28] and bulkier purine bases that shield the active site are favored over smaller pyrimidines, both as substrate and acceptor nucleobases. Compared to NDT, Rcl has a longer loop region that does not shield the active site from solvent. It also does not have any residues that specifically interact with and bind acceptor nucleobases.[10-12] These structural features allow a water molecule to act as the final acceptor. Because the MilB active site is shielded by α-helices, it is less solvent accessible as compared to PTD or Rcl. Compared to NDT, the MilB active site is similarly shielded, but is not as hydrophobic. As a result, an active site water

molecule is available to act as the final acceptor. Accordingly, the MilB/CMP structure reveals a water molecule in the active site positioned 4.9 Å from the 1′-ribosyl position.

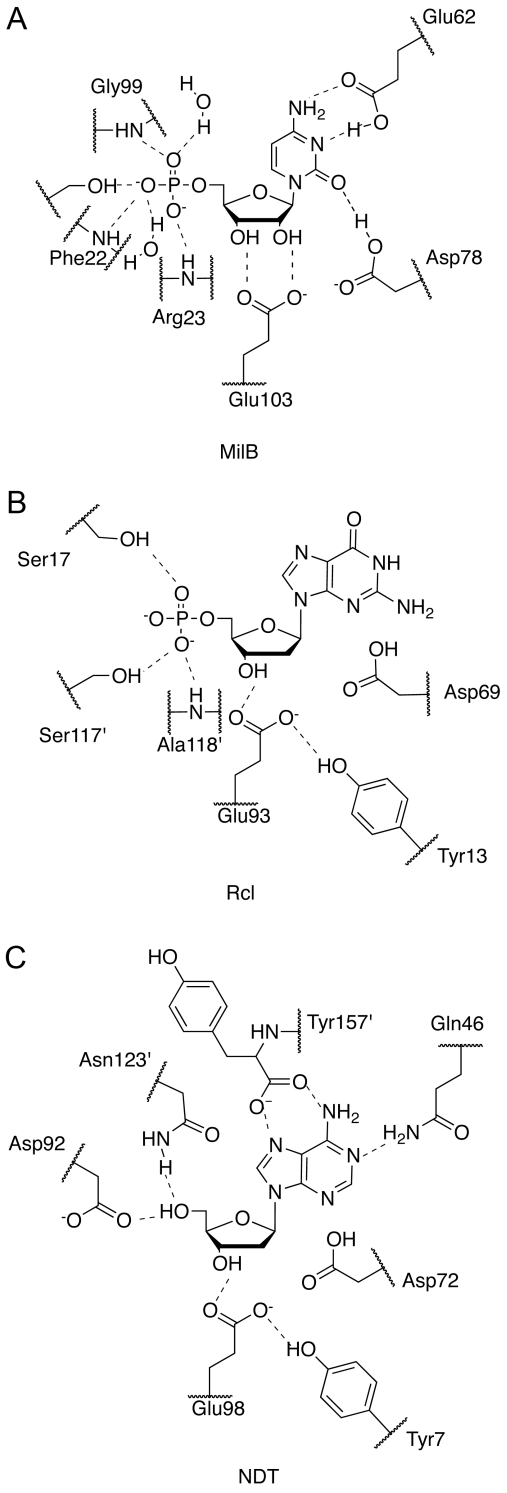
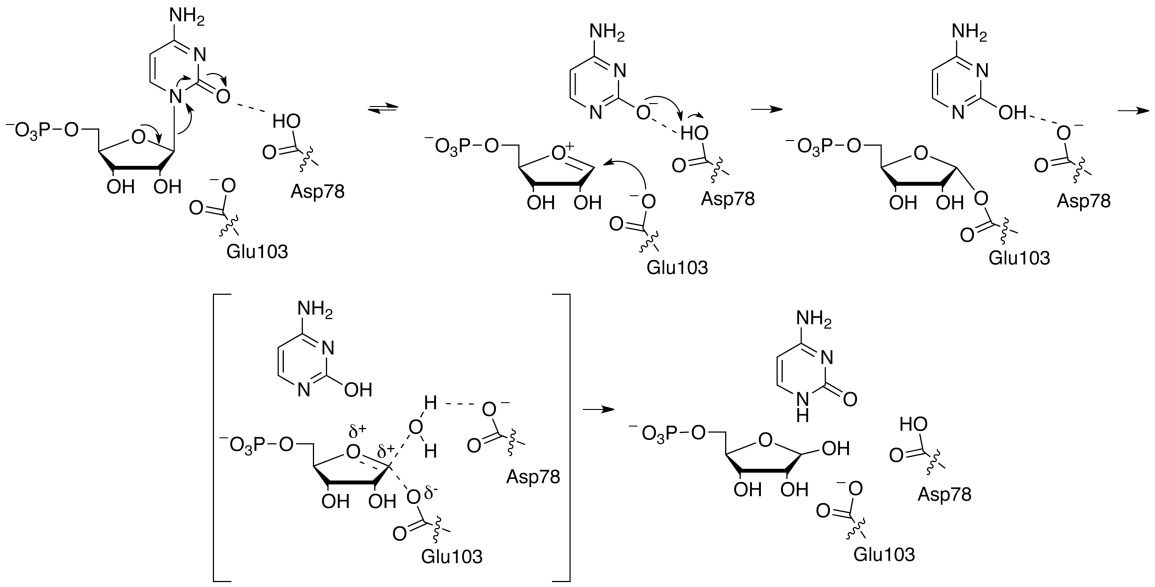


Figure 2.8. Active site comparison. (A) Observed MilB interactions with CMP. (B) Rcl active site with dGMP. (C) NDT active site with deoxyadenosine.

Mechanism of Hydrolysis. Previous studies showed that NDT family members catalyze transferase reactions utilizing a covalent enzyme-(deoxy)ribosyl intermediate.[6, 9, 31] In the first half reaction the covalent intermediate forms between a glutamate side chain and C1′ of the (deoxy)ribosyl group. The second half reaction takes place after the purine/pyrimidine base is released and a nucleophile (either a different base or a water molecule) displaces the glutamate side chain. The catalytic glutamate residue is absolutely conserved among the NDTs, Rcl, MilB, BcmB, and BlsM (Figure 2.8). Based on this mechanism, Glu103 of MilB forms a covalent attachment to CMP at the ribosyl 1′-position in the first half of the hydrolysis reaction (Scheme 2.1). Previous studies on xylanases, which utilize a glutmate/glycoside covalent intermediate in a hydrolysis reaction similar to that of MilB, showed that each half reaction proceeds through an oxocarbenium-like transition state.[32] In the crystal structure of the MilB/CMP complex the glutamate oxygen atom is not correctly oriented for a strictly S_N2 mechanism suggesting the formation of an oxocarbenium ion intermediate followed by trapping of the intermediate by Glu103. The second half reaction, in which water displaces the glutamate side chain, likely proceeds through an oxocarbenium-like transition state. In MilB, Asp78 is positioned to protonate the cytosine base leaving group and to activate a water molecule for the second half of the reaction. This aspartate residue is conserved in MilB, BcmB, BlsM, and Rcl, also suggesting a conserved mechanism of hydrolysis.



Scheme 2.1

Reversal of Substrate Specificity. Based on the structural data, we hypothesized that the active site phenylalanine is the key residue that confers specificity for either 2′-hydroxyl or 2′-deoxy groups. In MilB, Phe17 C4 is 3.6 Å from the catalytic glutamate residue and 3.5 Å from the CMP 2′-hydroxyl group. A tyrosine is at a position equivalent to MilB Phe17 in all previously solved structures of enzymes in the NDT superfamily, which prefer 2′-deoxy substrates (Figure 2.9). MilB F17Y was constructed using site-directed mutagenesis and used to test if the phenylalanine to tyrosine change is sufficient to increase the efficiency of dCMP hydrolysis. While MilB WT preferentially hydrolyzes CMP, with a five-fold lower efficiency towards dCMP, mutation of F17Y MilB resulted a 100-fold decrease in the hydrolysis of CMP and a 100-fold increase in the hydrolysis of dCMP, corresponding

to a >10,000-fold inversion of substrate specificity (Table 2.2). A similar, but slightly smaller affect in substrate specificity was observed for BcmB.

The largest change is seen in the preference of F17Y MilB for dCMP compared to CMP. The value of k_{cat} for WT MilB with CMP is about 3-fold larger than for that of dCMP. For F17Y MilB, the value of k_{cat} with dCMP is about 600-fold larger than for CMP. For both WT and F17Y MilB, the K_M values for both substrates are in a similar range with the value for F17Y MilB with dCMP being the lowest. The value of k_{cat} for WT BcmB with CMP is about 20-fold larger than for that of dCMP. For F6Y BcmB, the k_{cat} with dCMP is about 20-fold larger than for CMP. The K_M values for WT and mutant BcmB show a larger range compared to those of MilB. The lowest value is 0.042 mM for F6Y BcmB with dCMP and the highest value is 0.9 mM for WT BcmB with dCMP.

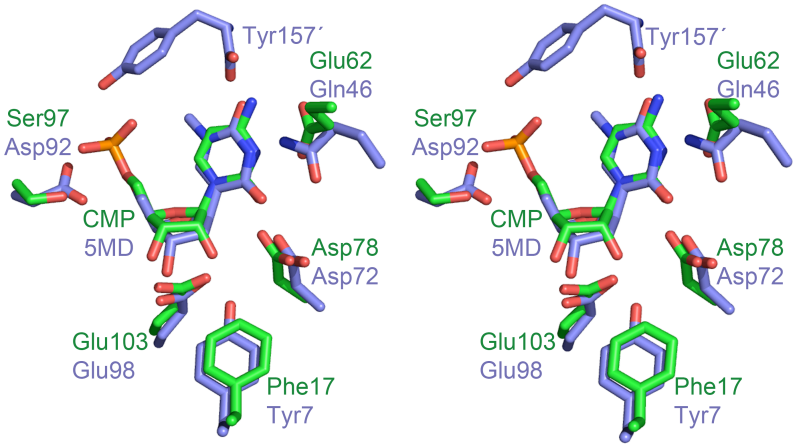


Figure 2.9. Active site comparison of MilB and NDT. A stereoview active site comparison between MilB with CMP bound (green) and NDT with 5-methyl-2'-deoxypseudouridine (5MD) (blue) reveals conserved positioning of the catalytic glutamate and aspartate residues. Other key residues for substrate binding deviate between the two structures.

In the structure of MilB/CMP, the CMP 2′-hydroxyl group interacts with the catalytic Glu103, orienting the molecule in the active site. Based on the location of the Phe17 in MilB/CMP, the MilB F17Y Tyr17 hydroxyl group would be 3.1 Å from the 2′-hydroxyl group of bound CMP. This may cause unfavorable interactions for CMP binding or cause electronic repulsions effecting hydrolysis by the catalytic glutamate residue. The latter is consistent with the observed values of k_{cat}, which is significantly enhanced while the K_M values are in a similar range. Additionally, WT MilB less efficiently hydrolyzes dCMP possibly because the 2′-group of dCMP cannot form any such hydrogen bonding interaction to orient it for catalysis. Thus, WT MilB catalytic residues are best positioned for hydrolysis of CMP with the aid of the 2′-hydroxyl group orienting CMP in the binding pocket. These observations would also apply to WT and F6Y BcmB, which show patterns similar to WT and F17Y MilB in efficiency for hydrolysis of CMP and dCMP.

Section 2.5. Supplementary Information

Table S2.1. Enzyme and substrate concentrations used in kinetics assays

Protein (Substrate)	[Enzyme] (μM)	Substrate Range (mM)
BcmB WT (CMP)	20	0.02 to 1.5
BcmB WT (dCMP)	20	0.05 to 6
BcmB F6Y (CMP)	80	0.1 to 6
BcmB F6Y (dCMP)	10 or 20	0.03 to 1.2
MilB WT (CMP)	10 or 20	0.05 to 18
MilB WT (dCMP)	60	0.3 to 30
MilB F17Y (CMP)	175	0.3 to 30
MilB F17Y (dCMP)	1	0.1 to 8

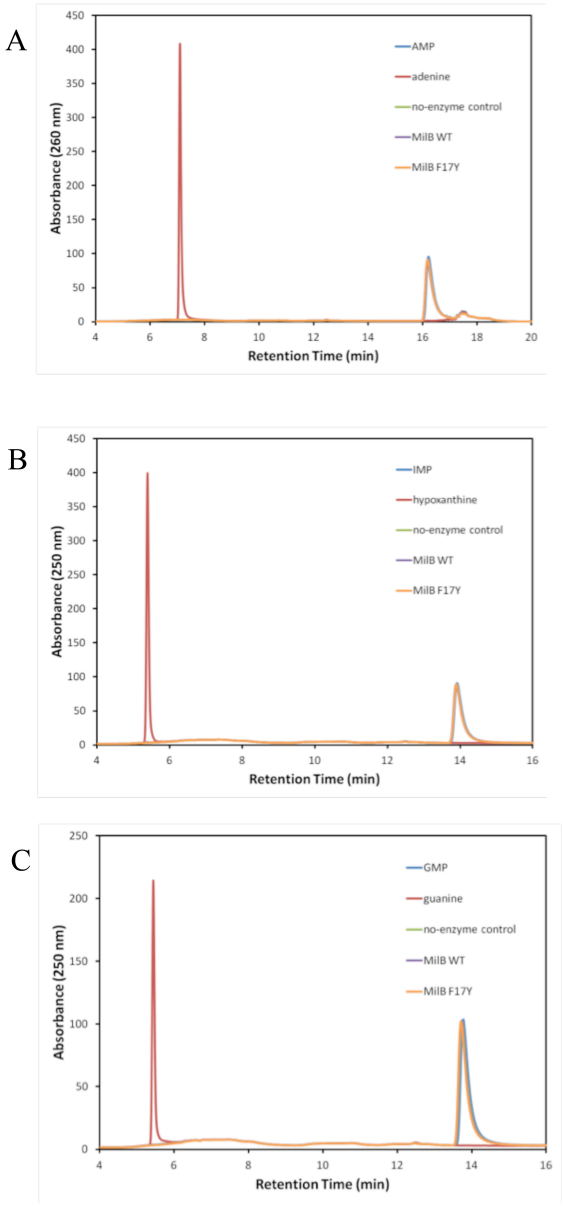


Figure S2.1. No hydrolysis of AMP, IMP, or GMP is observed when incubated with MilB or MilB F17Y.

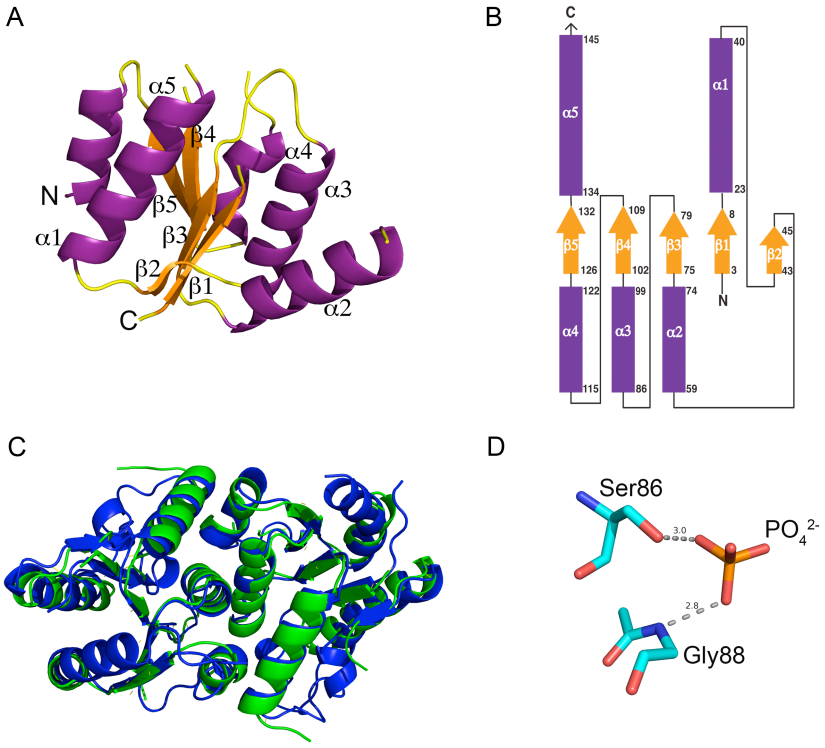
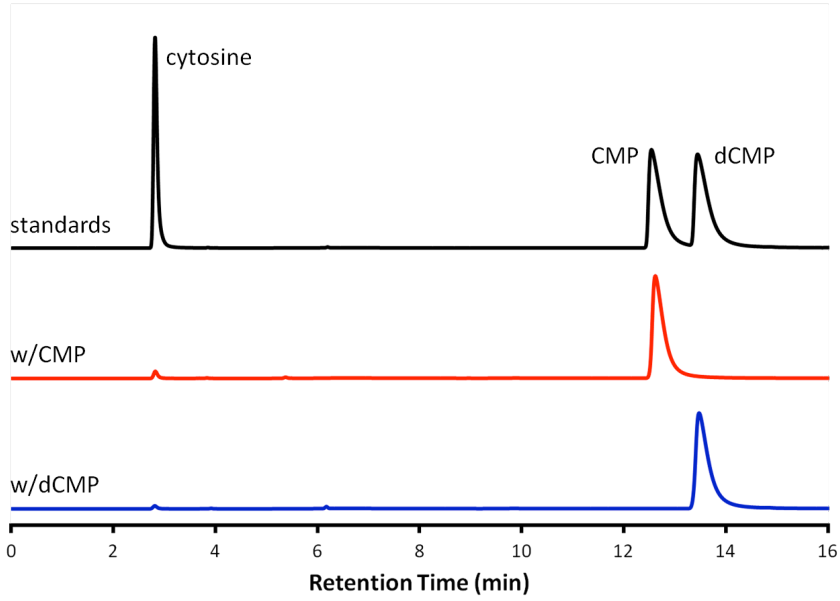


Figure S2.2. Structure of BcmB. (A) The BcmB protomer adopts the same α/β-twist fold as MilB. (B) The topology diagram of BcmB shows α-helices in purple and β-strands in orange. (C) The MilB dimer (blue) is shown superimposed with the BcmB dimer (green). The two molecules have a rmsd of 1.7 and display a highly similar overall fold. (D) A phosphate ion is bound in the active site and makes hydrogen bond interactions with a serine residue and an amide group.

A)



B)

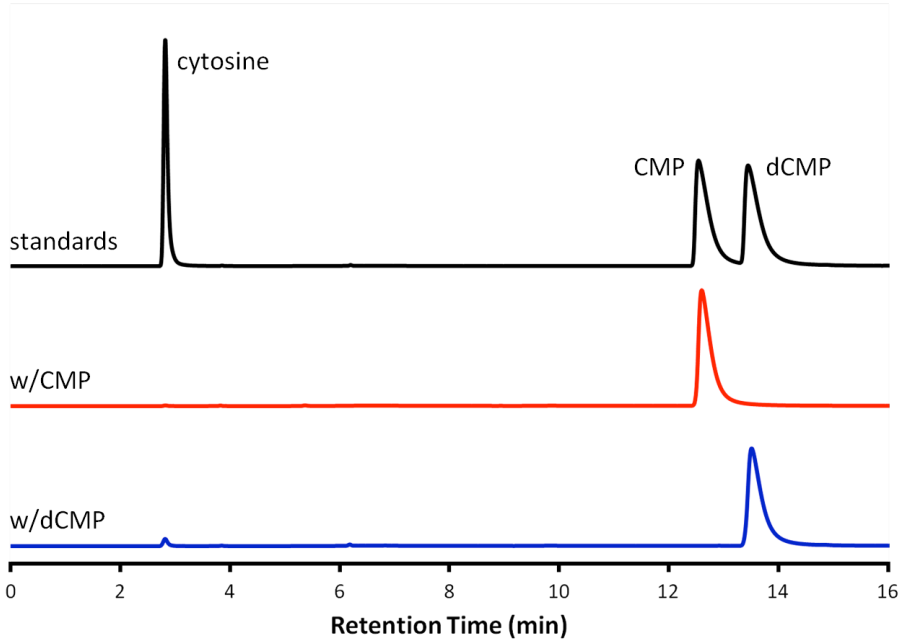


Figure S2.3. Substrate specificity studies with BcmB. A) BcmB WT with CMP and dCMP. B) BcmB F6Y with CMP and dCMP

REFERENCES

1. Li, L., Xu, Z., Xu, X., Wu, J., Zhang, Y., He, X., Zabriskie, T. M., and Deng, Z. (2008) The mildiomycin biosynthesis: initial steps for sequential generation of 5-hydroxymethylcytidine 5'-monophosphate and 5-hydroxymethylcytosine in Streptoverticillium rimofaciens ZJU5119, *Chembiochem 9*, 1286-1294.
2. Feduchi, E., Cosin, M., and Carrasco, L. (1985) Mildiomycin: a nucleoside antibiotic that inhibits protein synthesis, *J Antibiot (Tokyo) 38*, 415-419.
3. Harada, S., and Kishi, T. (1978) Isolation and characterization of mildiomycin, a new nucleoside antibiotic, *J Antibiot (Tokyo) 31*, 519-524.
4. Macnutt, W. S. (1952) The enzymically catalysed transfer of the deoxyribosyl group from one purine or pyrimidine to another, *Biochem J 50*, 384-397.
5. Chawdhri, R. F., Hutchinson, D. W., and Richards, A. O. (1991) Nucleoside Deoxyribosyltransferase and Inosine Phosphorylase-Activity in Lactic-Acid Bacteria, *Archives of Microbiology 155*, 409-411.
6. Armstrong, S. R., Cook, W. J., Short, S. A., and Ealick, S. E. (1996) Crystal structures of nucleoside 2-deoxyribosyltransferase in native and ligand-bound forms reveal architecture of the active site, *Structure 4*, 97-107.
7. Ghiorghi, Y. K., Zeller, K. I., Dang, C. V., and mKaminski, P. A. (2007) The c-Myc target gene Rcl (C6orf108) encodes a novel enzyme, deoxynucleoside 5'-monophosphate N-glycosidase, *J Biol Chem 282*, 8150-8156.

8. Grochowski, L. L., and Zabriskie, T. M. (2006) Characterization of BlsM, a nucleotide hydrolase involved in cytosine production for the biosynthesis of blasticidin S, *Chembiochem 7*, 957-964.

9. Dupouy, C., Zhang, C., Padilla, A., Pochet, S., and Kaminski, P. A. (2010) Probing the active site of the deoxynucleotide N-hydrolase Rcl encoded by the rat gene c6orf108, *J Biol Chem 285*, 41806-41814.

10. Doddapaneni, K., Mahler, B., Pavlovicz, R., Haushalter, A., Yuan, C., and Wu, Z. (2009) Solution structure of RCL, a novel 2'-deoxyribonucleoside 5'-monophosphate N-glycosidase, *J Mol Biol 394*, 423-434.

11. Yang, Y., Padilla, A., Zhang, C., Labesse, G., and Kaminski, P. A. (2009) Structural characterization of the mammalian deoxynucleotide N-hydrolase Rcl and its stabilizing interactions with two inhibitors, *J Mol Biol 394*, 435-447.

12. Padilla, A., Amiable, C., Pochet, S., Kaminski, P.A., Labesse, G. (2013) X-ray structure of the oncoprotein Rcl bound to three nucleotide analogs, *Acta Crystallographia Section D D69*, 247-255.

13. Lewis, B. C., Shim, H., Li, Q., Wu, C. S., Lee, L. A., Maity, A., and Dang, C. V. (1997) Identification of putative c-Myc-responsive genes: characterization of rcl, a novel growth-related gene, *Mol Cell Biol 17*, 4967-4978.

14. Shin, S., Bosc, D. G., Ingle, J. N., Spelsberg, T. C., and Janknecht, R. (2008) Rcl is a novel ETV1/ER81 target gene upregulated in breast tumors, *J Cell Biochem 105*, 866-874.

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

16. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode, In *Methods Enzymol.* (Carter, C. W. J., and Sweet, R. M., Eds.), pp 307-326, Academic Press, New York.

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

18. Zwart, P. H., Afonine, P. V., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., McKee, E., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., Storoni, L. C., Terwilliger, T. C., and Adams, P. D. (2008) Automated structure solution with the PHENIX suite, *Methods Mol Biol 426*, 419-435.

19. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr D Biol Crystallogr 66*, 213-221.

20. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot, *Acta Crystallogr D Biol Crystallogr 66*, 486-501.

21. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Cryst. D 66*, 12-21.

22. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments, *Acta Crystallogr. D 67*, 235-242.

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

24. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures, *Acta Crystallogr. D Biol. Crystallogr. 67*, 355-367.

25. Li, L., Xu, Z., Xu, X., Wu, J., Zhang, Y., He, X., Zabriskie, T. M., and Deng, Z. (2008) The mildiomycin biosynthesis: initial steps for sequential generation of 5-hydroxymethylcytidine 5'-monophosphate and 5-hydroxymethylcytosine in *Streptoverticillium rimofaciens* ZJU5119, *ChemBioChem 9*, 1286-1294.

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

27. Holm, L., and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D, *Nucleic Acids Res. 38 Suppl*, W545-549.

28. Anand, R., Kaminski, P. A., and Ealick, S. E. (2004) Structures of purine 2'-deoxyribosyltransferase, substrate complexes, and the ribosylated enzyme intermediate at 2.0 A resolution, *Biochemistry 43*, 2384-2393.

29. Bosch, J., Robien, M. A., Mehlin, C., Boni, E., Riechers, A., Buckner, F. S., Van Voorhis, W. C., Myler, P. J., Worthey, E. A., DeTitta, G., Luft, J. R., Lauricella, A., Gulde, S., Anderson, L. A., Kalyuzhniy, O., Neely, H. M., Ross, J., Earnest, T. N., Soltis, M., Schoenfeld, L., Zucker, F., Merritt, E. A., Fan, E., Verlinde, C. L., and Hol, W. G. (2006) Using fragment cocktail crystallography to assist inhibitor design of Trypanosoma brucei nucleoside 2-deoxyribosyltransferase, *J Med Chem 49*, 5939-5946.

30. Kaminski, P. A. (2002) Functional cloning, heterologous expression, and purification of two different N-deoxyribosyltransferases from Lactobacillus helveticus, *J Biol Chem 277*, 14400-14407.

31. Porter, D. J., Merrill, B. M., and Short, S. A. (1995) Identification of the active site nucleophile in nucleoside 2-deoxyribosyltransferase as glutamic acid 98, *J Biol Chem 270*, 15551-15556.

32. Tull, D., and Withers, S. G. (1994) Mechanisms of cellulases and xylanases: a detailed kinetic study of the exo-beta-1,4-glycanase from Cellulomonas fimi, *Biochemistry 33*, 6363-6370.

CHAPTER 3

STRUCTURE OF *CLOSTRIDIUM BOTULINUM* THIAMINASE I-THIAMIN COMPLEX REVEALS BIOCHEMICAL DETAILS OF BINDING AND DEGRADATION

Section 3.1. Introduction

Thiamin (vitamin B_1) is an essential vitamin in all living organisms. Its active form, thiamin pyrophosphate, is a cofactor in many biological processes including carbohydrate metabolism and amino acid biosynthesis. Most plants, fungi, and bacteria can synthesize thiamin, but animals must obtain it from their diets. Numerous investigations over the past sixty years provide a nearly complete understanding of thiamin biosynthesis,[1] yet the physiological basis for its degradation and the fate of the breakdown products is still unknown. Thiaminases catalyze the degradation of thiamin into its thiazole and pyrimidine components (Figure 3.1). These enzymes are found in a wide variety of organisms, including plants, fish, and bacteria.[2-10] Consequently, ingestion of thiaminase containing foods by humans or other animals can cause symptoms of thiamin deficiency.[11, 12] In humans, this deficiency can affect the cardiovascular system (wet beriberi) or nervous system (dry beriberi) with potentially fatal consequences.[13, 14]

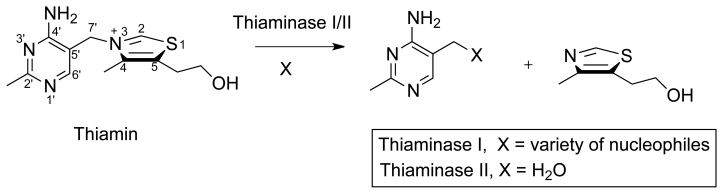


Figure 3.1. Thiamin cleavage reaction catalyzed by the thiaminases. Thiamin is degraded by thiaminase I or thiaminase II. Thiaminase I utilizes a variety of molecules, whereas thiaminase II exclusively uses water, as the nucleophile.

Biochemical analyses have established two distinct classes of thiaminases. Thiaminase I utilizes a variety of nucleophiles,[15, 16] whereas thiaminase II exclusively utilizes water as the nucleophile.[17] A crystal structure of the thiaminase I from *Bacillus thiaminolyticus* (Bt-thiaminase I) has been reported.[18] Bt-thiaminase I is a monomer composed of two domains joined by three crossover segments and is structurally homologous to the group II periplasmic binding proteins (PBPs).[18, 19] PBPs bind small molecules and deliver them to ABC transporters for eventual uptake into the cytoplasm. Structural studies reveal group II PBPs in an open conformation when no ligand is bound and in a closed conformation when a ligand is bound.[19] Bt-thiaminase I is a rare example of an enzyme in the group II PBP superfamily. Interestingly, THI5, the thiamin pyrimidine synthase in eukaryotes, is also a member of the group II PBP superfamily.[20] Bt-thiaminase I is structurally similar to TbpA,[21] which is the PBP in many prokaryotes for thiamin, thiamin phosphate, and thiamin pyrophosphate, and THI5 is structurally similar to ThiY,[20] which is the PBP in some prokaryotes for the thiamin degradation product formylaminopyrimidine. These observations raise curiosities about the evolution of thiaminase I and its relationship to the PBPs. Thiaminase II diverges in both structure and sequence from thiaminase I.

Previous mechanistic and structural studies established that an active site cysteine residue is involved in thiaminase I catalysis.[15, 16, 18] A crystal structure of Bt-

thiaminase I with the mechanism-based, irreversible inhibitor, 4-amino-6-chloro-2, 5-dimethylpyrimidine (ACDP), showed that the active site is located in a V-shaped cleft created between the two Bt-thiaminase I domains. The active site location is structurally similar to the ligand binding site in the PBPs. The active site cysteine residue is activated by a nearby glutamate residue to initiate the thiaminase I reaction. The activated cysteine first forms a covalent bond to the thiamin pyrimidine C6′. Subsequently, a nucleophile adds to the 5-methylene carbon of the pyrimidine moiety and the thiazole moiety is displaced. Despite their highly divergent sequences and structures, thiaminase I and most thiaminase IIs utilize an activated cysteine residue to form a covalently bound intermediate.[17, 18]

Cb-thiaminase I is a 46.3 kDa enzyme from *Clostridium botulinum*. The catalytic cysteine residue was predicted by sequence alignment with other thiaminase Is and mutated to generate a catalytically inactive Cb-thiaminase I (C143S). C143S was co-crystallized with thiamin to obtain the C143S/thiamin complex reported here. This is the first structure of a thiaminase I/substrate complex and reveals atomic level details of thiamin binding prior to degradation. This structure was used to identify the key amino acids involved in the Cb-thiaminase I reaction. Kinetic studies of thiamin degradation by Cb –thiaminase I and six active site mutants determined the roles of these amino acids in thiamin degradation. Together the C143S/thiamin structure and subsequent kinetic studies lead to a detailed Cb-thiaminase I mechanistic proposal. The physiological role of the thiaminases is explored once again with new insights from the atomic level interactions revealed by the C143S/thiamin complex.

Section 3.2. Materials and Methods

Cloning, Overexpression and Purification of Cb-thiaminase I. The thiaminase I gene, *bcmE*, was cloned from *C. botulinum* A str. ATCC 19397 genomic DNA. Standard DNA manipulation methods were used for all of the cloning procedures. The gene was inserted into pTHT, a modified pET-28 plasmid with an N-terminal His_6 tag, followed by a TEV protease cleavage site. All active site mutants were obtained by site-directed mutagenesis of the native gene using standard, PCR-based mutagenesis.[22]

After DNA sequencing to verify plasmid accuracy, the plasmids were transformed into *Escherichia coli* BL21(DE3) cells and grown overnight at 37 °C on selective kanamycin (30 μg/mL) containing agar. Starter cultures were then grown from a single colony. A selected colony was placed in 10 mL sterile LB Lennox media containing 30 μg/mL kanamycin at 37 °C with shaking overnight. The following day 5 mL of the overnight starter culture were added directly to 1.5 L volumes of sterile LB media. Cells grew with shaking at 37 °C until reaching an OD600 of 0.6, at which point the incubator temperature was reduced to 15 °C. Upon reaching an induction temperature of 15 °C and an OD600 of 0.8, protein overexpression was initiated by adding IPTG to the cultures with a final concentration of 0.5 mM IPTG. Cells were harvested 18 h after induction of protein overexpression by centrifugation at 2000*g* for 20 min. The cell pellet was collected and frozen at -20 °C for storage.

The cell pellet was later thawed and resuspended in 45 mL of lysis buffer (50 mM Tris-HCl, 300 mM NaCl, and 20 mM imidazole at pH 8.0) before lysing by sonication. The lysed cell extract was centrifuged at 40,000*g* for 30 min at 4 °C. The

supernatant was collected and loaded on to a 2 mL NiNTA column (Qiagen), preequilibrated with lysis buffer. The column was then washed with 45 mL of lysis buffer to remove any nonspecifically bound contaminants from the column. The protein was eluted from the column by passing 20 mL of elution buffer (50 mM Tris-HCl, 300 mM NaCl, and 250 mM imidazole at a final pH of 8.0) through the column. The elution volume containing Cb-thiaminase I was collected. This procedure was followed for purification of Cb-thiaminase I and all active site mutants.

To remove the His_6 tag for crystallization, the purified C143S protein was incubated with TEV protease during dialysis into 3.5 L of 10 mM Tris HCl at pH 8.0, 150 mM KCl and 1 mM dithiothreitol for 18 h at 4 °C, utilizing a ratio of 0.5 mg of TEV for every 10 mg of C143S. The sample was then passed over a NiNTA column preequilibrated with lysis buffer. The elution volume containing C143S with the His_6 tag removed was collected in the flow through. Complete cleavage and purity were assessed by SDS-PAGE analysis. The sample was buffer exchanged by overnight dialysis into 10 mM Tris-HCl at pH 8.0 and 150 mM KCl, then concentrated to 15 mg/mL using a centrifugal filter device with a molecular weight cutoff of 10,000 Da. The purification process yielded approximately 3.5 mg protein per L of cell culture. The purified protein was aliquoted and flash frozen in liquid nitrogen for storage at -80 °C.

Co-crystallization of C143S/Thiamin. For crystallization using the hanging drop vapor diffusion method, the protein was diluted with a solution of 10 mM Tris pH 8.0, 150 mM KCl and 10 mM thiamin with final concentrations of 7.5 mg/mL protein and 5 mM thiamin. Equal volumes of protein and reservoir solutions were

mixed and equilibrated at 18 °C against a total volume of 500 μL well solution. The initial crystallization condition was determined using the commercially available Wizard III sparse matrix screen (Emerald Biosystems). The optimized crystallization conditions were 22% (w/v) polyethylene glycol 10,000, 0.1 M sodium citrate pH 4.4, and 2% (v/v) dioxane. Needle-like C143S/thiamin crystals grew approximately 200 μm long and 10-20 μm thick in three to five days. Crystals were cryoprotected for data collection in a solution composed of the mother liquor supplemented with 20% (v/v) ethylene glycol.

X-ray Data Collection and Processing. X-ray diffraction experiments were conducted at APS NE-CAT beamline 24-ID-C (Argonne National Laboratory) using a Quantum 315 detector (Area Detector Systems Corp.) at a distance of 200 mm. The oscillation method was used with 1.0º rotation per frame for 180º with cryocooling. Diffraction data was collected at wavelength of 0.97918 Å. The NE-CAT in-house automated RAPD data collection and processing system, which utilizes XDS,[23, 24] was used for indexing, integration, and scaling of the data. Data collection and processing statistics are listed in Table 3.1.

Structure Determination, Model Building, and Refinement. The structure of C143S/thiamin was determined using the molecular replacement method. The crystal belongs to space group $P2_1$ and contains two molecules per asymmetric unit corresponding to a Matthews coefficient[25] of 1.96 $Å^3$/Da and estimated solvent content of 37.4%. The thiaminase I from *B. thiaminolyticus* (PDB ID: 2THI), with a sequence identity of 51% to C143S, was the molecular replacement model. The automated search program MolRep[26] positioned two chains in the asymmetric unit. Iterative

rounds of model refinement were conducted using COOT[27] for manual model building followed by refinement with PHENIX.refine[28] using default parameters. The thiamin substrate was added using PHENIX.ligandfit for placement into $F_o - F_c$ electron density. Water molecules were added during later rounds of refinement. The quality of the structure was analyzed using MolProbity[29] and COOT.[27] Structure refinement converged with a final R-factor value of 18.7% and R_{free} of 23.5%. Complete refinement statistics are listed in Table 3.1.

Determination of Kinetic Parameters. Steady state kinetic analysis was performed for six Cb-thiaminase I mutant proteins selected based on the crystal structure of C143S/thiamin. Cb-thiaminase I reactions were carried out in 100 mM potassium chloride, 713 mM β-mercaptoethanol, and 50 mM phosphate buffer at pH 8.0. All reactions were incubated at 25 °C except the E271Q mutant, which was measured at 37 °C due to low activity levels. The reactions were initialized by the addition of enzyme to the reaction volume containing thiamin, with thiamin concentrations ranging from 100 μM to 10 mM. Aliquots were taken from the reaction mixture after definite time intervals and quenched with 1 M HCl. The aliquot was filtered through a 10,000 Da cutoff membrane to remove the enzyme and the filtrate was analyzed by HPLC.

In the HPLC method, the ratio of 100 mM phosphate buffer at pH 6.6 (P) to methanol (M)/water mixture was varied over time in seconds (t). The gradient was carried out following the scheme (t, M%, P%): (0, 0, 100), (5, 0, 90), (9, 15, 60), (14, 65, 15), (19, 0, 100), (25, 0, 100). HPLC peak areas were measured for varying concentrations of pure 4-methyl-5-thiazoleethanol. A calibration curve relating peak area to 4-methyl-5-

thiazoleethanol concentration was used to create a standard plot. Peak areas of the product in the enzyme assays were measured and the concentrations calculated based on the standard plot. The initial rates were calculated from plots of concentration versus time. Initial rates were then plotted and fit to the Michaelis–Menten equation for calculation of kinetic parameters.

Figure preparation. All figures were prepared using PyMOL[30] or ChemDraw (Cambridge Biosoft) and compiled in Photoshop (Adobe).

Table 3.1. Summary of data collection and refinement statistics

	C143S/Thiamin
beamline	APS 24-ID-C
resolution (Å)	2.18
wavelength (Å)	0.97918
space group	$P2_1$
a (Å)	63.5
b (Å)	36.3
c (Å)	156.8
β (°)	91.4
Matthews coefficient	1.96
% solvent	37.4
mol/asu	2
Measured reflections	132,181
Unique reflections	37,678
Average I/σ	16.3 (5.0)[a]
Redundancy	3.4 (3.0)
Completeness (%)	99 (92)
R_{sym} (%)[b]	6.5 (29.4)
No. of protein atoms	5643
No. of ligand atoms	36
No. of water atoms	496
Reflections in working set	37.537
Reflections in test set	1877
R-factor/R_{free} (%)[c]	18.7/23.5
rms deviation from ideals	
bonds (Å)	0.010
angles (°)	1.29
average *B* factor for protein (Å^2)	19.3
average *B* factor for water (Å^2)	26.8
average *B* factor for ligand (Å^2)	19.2
Ramachandram plot	
most favored (%)	98.2
allowed (%)	1.8
disallowed (%)	0.0

[a]Values in parentheses are for the highest-resolution shell.
[b]$R_{merge} = \Sigma\Sigma_i | I_i - \langle I \rangle | / \Sigma \langle I \rangle$, where $\langle I \rangle$ is the mean intensity of the N reflections with intensities I_i and common indices h,k,l.
[c]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, calculated over all reflections used in the refinement. R_{free}, is similar to R_{work} but calculated over a subset of reflections (5%) excluded from all stages of refinement.

Section 3.3. Results

Overall structure of Cb-thiaminase I. The structure of the C143S/thiamin complex, with two molecules per asymmetric unit, was determined at 2.2 Å resolution. The two molecules (A and B) make a total of 16 non-bonded interactions along an interface area of approximately 210 $Å^2$, suggesting only crystal packing contacts. Size exclusion chromatography (data not shown) predicted that Cb-thiaminase I is a monomer in solution. This is also consistent with the oligomeric state of Bt-thiaminase I.[18] Of the 404 possible amino acids in the Cb-thiaminase I sequence, the structure contains residues 39-403, except for 313-314, of molecule A, and residues 40-403 of molecule B, except for a missing loop area of residues 310-314, of chain B.

Cb-thiaminase I is divided into two distinct domains (Figure 3.2). The N-terminal domain is composed of residues 39-120 and 297-384. Residues 145-296 and 385-403 form the C-terminal domain. Both domains have a central β-sheet flanked on either side by α-helices with three crossover segments connecting the N- and C-domains. Residues 144-145, 287-296, and 375-385 comprise the three domain linking regions. The crossover segment formed by residues 287-296 connects β_{10} to β_{11} and the crossover segment created by residues 375-384 connects α_{14} and α_{15}. The third crossover connects β_6 to β_7 at residue 144. The active site is located at a cleft between the two domains.

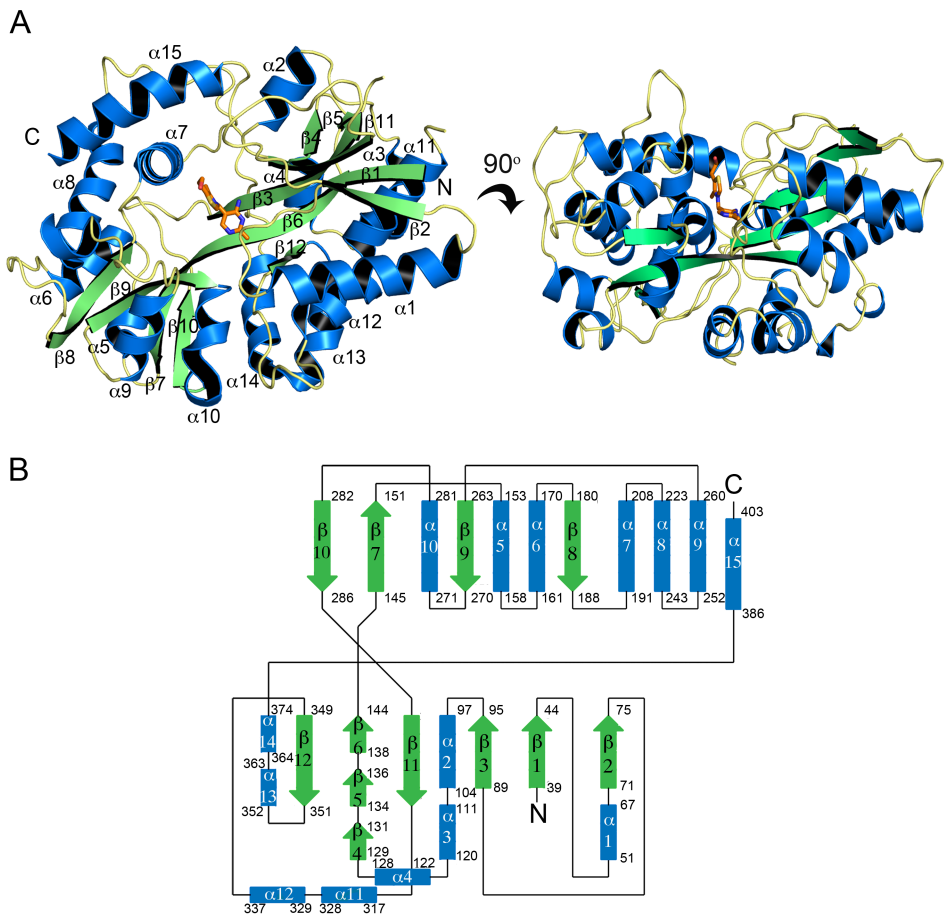


Figure 3.2. Overall structure of C143S/Thiamin.
(A) The ribbon diagram shows the α/β fold of C143S with α-helices depicted in blue and β-strands in green. The two domains are composed of β-sheets flanked by α-helices on both sides, connected by a central β-strand. The thiamin-binding site is located in the central groove between the N- and C- terminal domains. The 90° rotation shows the orientation of the thiamin molecule in the binding cleft. (B) The topology diagram of C143S following the same coloring scheme as in (A).

Cb-thiaminase I Active Site and Substrate Binding. The structure of C143S/thiamin reveals the substrate binding site and atomic level details of binding interactions (Figure 3.3). There is one thiamin molecule observed per protein molecule. This cleft is approximately 17 Å deep, 15 Å long, and 12 Å wide. Six tyrosine residues, Tyr46, Tyr48, Tyr80, Tyr252, Tyr269, and Tyr300, form the outer collar of the V-shaped cleft with the catalytic residues located at the bottom of the

cleft. The catalytic cysteine residue Cys143, which is represented by Ser143, is located on β_6.

Thiamin binds to the active site with a 108° angle between its pyrimidine C5′ and thiazole N3. The thiamin pyrimidine and thiazole moieties are approximately perpendicular with a dihedral angle of 85.9° between their two planes. With torsion angles of ϕ_T= -10° (C5′-C7′-N3-C2) and ϕ_P= -93.6° (N3-C7′-C5′-C4′) the bound thiamin most closely matches the standard thiamin F-conformation.[31] The pyrimidine portion of the thiamin is positioned towards the bottom of the V-shaped binding pocket. Hydrogen bonds form between the thiamin N1′ atom and the side chains of Ser143 and Glu271 (Figure 3.4). The thiamin N3′ atom is positioned 2.7 Å from the Asp302 carboxylate group and within hydrogen bond distance. The 4′-amino group is within hydrogen bond distance of Asp302 (3.0 Å), Asp94 (2.9 Å) and Tyr46 (3.1 Å). The Tyr80 hydroxyl group is 3.5 Å from the positivity charged N3 atom and 2.7 Å from the Tyr46 hydroxyl group. The thiazole group does not interact with the enzyme and is extended towards the top of the binding groove with the 5-hydroxyethyl group extended out toward the solvent region.

Steady-state kinetics of Cb-thiaminase I WT and mutants. Thiamin degradation by Cb-thiaminase I can be quantified using HPLC to analyze the reaction products. Bt-thiaminase I WT catalyzes the thiaminase reaction with a turnover rate of 2.31 x 10^2 s^{-1} and a k_{cat}/K_m value of 5.10 x 10^5 M^1s^{-1}. All activity was abolished by mutation of the catalytic cysteine in C143S. The kinetic parameters for five other Cb-thiaminase I active site mutants are reported in Table 3.2.

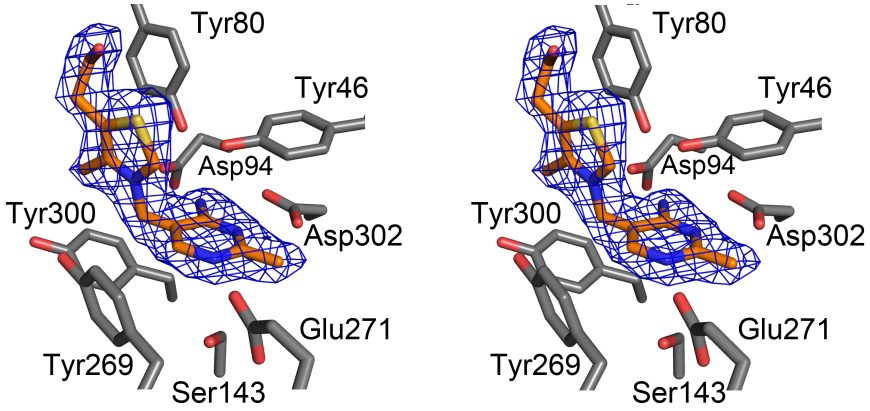


Figure 3.3. C143S/Thiamin active site interactions. The stereodiagram displays active site residues near the thiamin-binding site with the thiamin molecule modeled into the final $2F_o$-F_c electron density map contoured at 1σ.

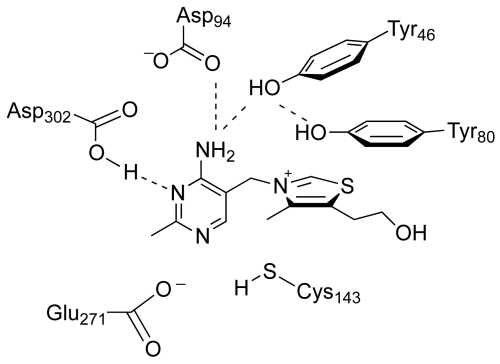


Figure 3.4. Thiamin binding in C143S. The schematic representation of thiamin in the C143S active site displays the enzyme-substrate active site interactions. All distance values given are in angstroms (Å).

Table 3.2. Kinetic properties of thiaminase activity for Cb-thiaminase I WT and active site mutants

Mutant	$k_{cat}(s^{-1})$	K_M (M)	$k_{cat}/K_M(M^{-1}s^{-1})$
WT	$(2.31 \pm 1.10) \times 10^{2}$	$(4.52 \pm 0.51) \times 10^{-4}$	5.10×10^{5}
C143S	0	-	-
Y80F	1.07 ± 0.12	$(2.11 \pm 0.81) \times 10^{-3}$	4.96×10^{2}
D302N	8.14 ± 0.74	$(3.07 \pm 1.09) \times 10^{-5}$	2.64×10^{4}
E271Q	$(1.1 \pm 0.03) \times 10^{-2}$	$(7.47 \pm 1.25) \times 10^{-5}$	1.50×10^{2}
Y46F	$6.09 \pm 0.384 \times 10^{1}$	$(2.45 \pm 0.51) \times 10^{-4}$	2.52×10^{5}
D94N	$(2.3 \pm 0.2) \times 10^{2}$	$(3.49 \pm 0.76) \times 10^{-3}$	6.61×10^{4}

Section 3.4. Discussion

Structure Comparison. Cb-thiaminase I is a group II PBP, as characterized by the arrangement of the three crossover segments characteristic of group II PBP members. A DALI[32] search against the PDB found proteins with high structural similarity to Cb-thiaminase I (Table 3.3). Cb-thiaminase I has high structural homology to various PBPs despite shraring relatively low sequence identity with them. The highest homology result found by the DALI search has a rmsd of 0.9 between Cb-thiaminase I and Bt-thiaminase I (PDB ID 2THI) [32]. This extremely low rmsd value is expected, as they are orthologs and have a high 51% sequence identity. Accordingly, the structure of Bt-thiaminase I was the first to reveal that thiaminase Is have a PBP fold. [18, 21]

The structures of Cb-thiaminase I and Bt-thiaminase I reveal rare examples of proteins with a PBP fold that have an enzymatic function. Most members of the PBP superfamily have evolved for various binding functions. PBPs bind a variety of molecules including metals, vitamins, sugars, or amino acids [33]. This adaptability is attributed to the binding site location near the protein surface, which increases accessibility, and the conformational change that then closes off the active site upon

ligand binding. [33] This arrangement allows easy access to the protein active site for environmental ligands and new binding capabilities as mutation of the binding residues occurs. The structural similarity of Cb-thiaminase I to PBPs suggests that thiaminase Is evolved from a PBP ancestor to accommodate thiamin and later developed a thiamin degrading function.

Table 3.3. Enzymes structurally similar to Cb-thiaminase I

Protein	PDB ID	Z score	rmsd	% identical	no. of aligned residues
Thiaminase I	2THI	57.2	0.9	51	361
Maltose binding protein	1EU8	28.6	3.3	17	328
ABC transporter binding, acbh	3OO6	26.2	3.4	11	390
Periplasmic binding protein	2DFZ	25.2	3.4	16	323
Iron binding protein	1O7T	21.2	3.0	10	287
Thiamin binding protein, TbpA	2QRY	18.3	3.9	14	308

Conformational Changes. As a result of their structural similarity, comparisons between the liganded Cb-thiaminase I and unliganded Bt-thiaminase I can be drawn. Structural alignment of C143S/Thiamin, unliganded Bt-thiaminase I (2THI), and liganded Bt-thiaminase I/ACDP (4THI) reveals no hinged movement of the two domains upon thiamin binding (Figure 3.5A). Both the liganded and unliganded forms appear to most closely resemble the closed conformation of group II PBPs. Thiaminase Is may have lost the characteristic PBP conformational change as it evolved. Because the PBPs display an open and closed conformation in order to bind, deliver, and release their substrates, for Cb-thiaminase I and Bt-thiaminase I this feature became unnecessary as they acquired a specialized enzymatic function.

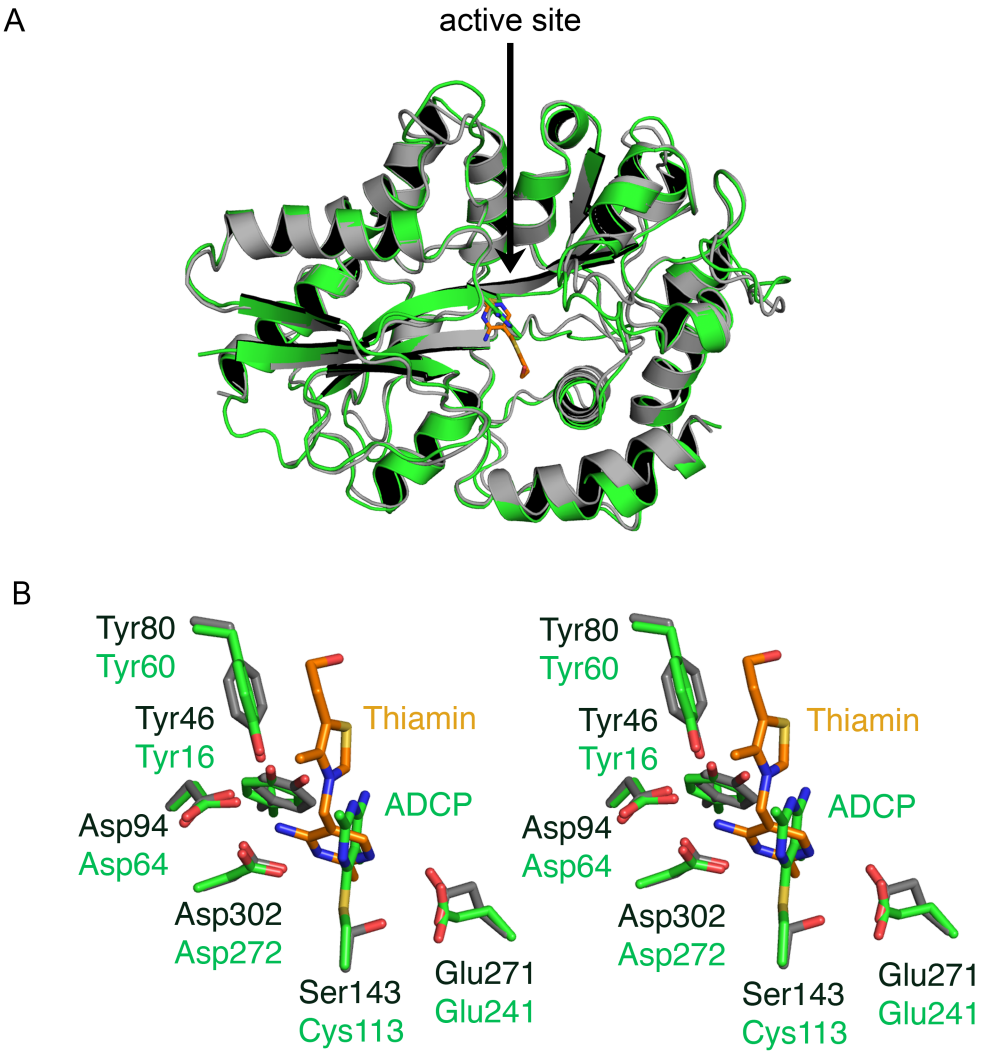


Figure 3.5. Comparison of C143S/Thiamin with Thiaminase I/ACDP structure from *B. thiaminolyticus*. (A) C143S structure (gray) with thiamin (orange) superimposed with Bt-thiaminase I with covalently bound inhibitor (green, PDB ID: 4THI). (B) Stereoview active site comparison between C143S and Bt-thiaminase I following the same coloring scheme as in (A). The pyrimidine portion of thiamin from the C143S/thiamin complex is oriented perpendicular to the 2,5-dimethylpyrimidin-4-ylamine (ADCP) inhibitor covalently bound to the active site cysteine in Bt-thiaminase I.

Comparison to Thiamin Binding Sites in Other Proteins. In addition to thiaminases, proteins that bind thiamin, or its mono- or diphosphorylated forms, include enzymes that use ThDP as a cofactor, thiamin biosynthetic enzymes, and thiamin transport proteins. A common thiamin-binding motif is found in enzymes that utilize ThDP as a cofactor. This 30-residue motif begins with –GDG- and concludes with –NN. ThDP binding is facilitated by phosphate binding interactions with a divalent metal, an aspartate, and an asparagine.[1] When used a cofactor, ThDP binds in the V-conformation with $\phi_T \approx \pm 90$ and $\phi_P \approx 90°$,[2] with torsion angles defined as ϕ_T= C5′-C7′-N3-C2 and ϕ_P= N3-C7′-C5′-C4′. The alternative F- ($\phi_T \approx 0°$, $\phi_P \approx \pm 90°$) and S- ($\phi_T \approx \pm 100°$, $\phi_P \approx \pm 150°$) thiamin conformations are less frequently observed in enzyme-substrate complexes although the F-conformation is the favored structure in solution. The C143S/thiamin complex exhibits a rare enzyme-bound thiamin F-conformation. Analysis of enzymes that bind thiamin or ThDP, but do not use it as a cofactor, reveals several other examples of structures with thiamin that is not in the common V-conformation.

The structure of thiamin pyrophosphokinase (TPK) contains a bound thiamin in the F-conformation (ϕ_T= -6°, ϕ_P= -78°) (PDB ID: 1IG3). This enzyme, which phosphorylates ThMP to form the active form ThDP, binds ThMP through multiple interactions.[3] In addition to stacking interactions with a nearby tryptophan, the thiamin pyrimidine moiety makes hydrogen-bonding contacts between its 4′- amine with an aspartate and main chain carbonyl group, with an addition contact resulting between interaction between its N1′ and a nearby serine. The TPK substrate is further

anchored through a phosphate-binding pocket. Similarly, ThMP kinase in complex with ThDP (PDB ID: 3C9T) displays the less common thiamin F-conformation (ϕ_T= -20°, ϕ_P= -64°).[4] ThDP is bound in a cleft with stacking interactions between the thiazole and a nearby tryptophan. In addition to phosphate binding pocket interactions, ThDP binding is further facilitated by 4′-amine hydrogen-bonding contacts with a glutamate residue. While TPK, ThMP, and Cb-thiaminase I structures all contain F-conformation thiamins, their thiamin-binding motifs show little conservation. Aside from all containing a common 4′-amine group anchored by either Glu or Asp, they share few other thiamin binding features. TPK and ThMP contain phosphate-binding pockets that can help facilitate binding of ThMP or ThDP. In Cb-thiaminase I, the phosphate end of ThMP or ThDP extends outward from the binding cleft toward the solvent. ThMP in thiamin phosphate synthase (TPS) displays a non-standard "cis" V-conformation (ϕ_T= -95.6°, ϕ_P= -113°) (PDB ID: 2TPS).[5] Like Cb-thiaminase I, its pyrimidine is located at the bottom of its binding cleft. There thiamin binding is aided by hydrogen bonding between its 4′-amine and N3′ group with a Gln within an otherwise hydrophobic pyrimidine-binding pocket.

Despite sharing only 14% sequence identity, Cb-thiaminase I and *E. coli* thiamin binding protein, TbpA, have the group 2 PBP fold and have similar binding site locations; however, their thiamin binding motifs differ. The TbpA ligand binding cleft is located in an analogous location between its two structural domains as in Cb-thiaminase I. In the reported structure, (PDB ID 2QRY)[6], ThMP spans the length of the V-shaped cleft in an F- conformation (ϕ_T= 0°, ϕ_P= -83°). TbpA binds ThMP by anchoring the phosphate group within a hydrogen bond rich pocket and sandwiching

the thiazole group between two tyrosine residues. The ThMP pyrimidine interacts mainly through water-mediated hydrogen bonding with one hydrogen bond to a serine residue side chain. In Cb-thiamianse I thiamin pyrimidine is buried in the bottom of cleft, with multiple hydrogen bonds and with C6′ near Cys143, and the thiazole portion extending upward towards the solvent making only van der Waals interactions with the protein. The similarities between thiamin binding modes between Cb-thiaminase I and other non-cofactor thiamin binding proteins are limited, suggesting divergent evolution of their binding sites. When used as a cofactor, the V-conformation is necessary for its chemistry but sterically unfavorable. It is unsurprising that in proteins that bind thiamin for non-cofactor functions the sterically favorable conformations are observed. The ThMP riboswitch is another example of a thiamin binding mode in which ThMP binds in an alternative conformation (PDB ID: 2HOM).[7] This highly specific mRNA element binds ThMP in the S-conformation (ϕ_T= -85°, ϕ_P= -173°), further supporting the observation that thiamin binding sites in these enzymes evolved to bind thiamin in a sterically favorable conformation, although not necessarily convergently.

Cb-thiaminase I Thiamin Binding and Reaction Mechanism. The C143S/thiamin structure served as a guide for investigation of the Cb-thiaminase I mechanism. To bind thiamin, Asp302 interacts with the thiamin pyrimidine through multiple hydrogen bonds. The D302N k_{cat} value is approximately two orders of magnitude less than that of Cb-thiaminase I WT, implicating the importance of the anchoring interaction between the thiamin 4′-amino group and Asp302. Other amino acids involved in thiamin binding include Tyr46, which hydrogen bonds with both

Tyr80 an the 4′-amino group, and mutational analysis of Y46F shows modest effects on both k_{cat} and K_M. Tyr80 is within hydrogen bond distance to Tyr46 and mutation of Y80F mutation decreases the turnover rate more than 100 times. The K_M is also affected by this substitution, which is double for Y80F compared to WT, suggesting that Tyr80 may interact electrostatically with the thiazolium ion for binding.

A detailed mechanistic proposal illustrates the complete series of steps involved in the reaction catalyzed by Cb-thiaminase I (Figure 3.6). In the proposed mechanism, Glu271 deprotonates Cys143 to activate it for catalysis. The deprotonated cysteine initiates the reaction by addition at the thiamin C6′ position. The covalent addition by Cys143 is critical and the C143S mutant shows complete loss of thiaminase activity. Demonstrating the importance of Glu271 for this step, the E271Q mutant k_{cat} of 1.1 x 10^{-2} s^{-1} is significantly reduced compared to Cb-thiaminase WT with a k_{cat} of 2.31 x 10^{2} s^{-1}. After deprotonation by Glu271, Cys143 forms a covalent enzyme-substrate intermediate and the thiazole half acts as a leaving group with assistance from Asp94. Structural and kinetic evidence supports the importance of Asp94 for catalysis, as well as thiamin binding. Cb-thiaminase I efficiency is reduced by D94N mutation, with a k_{cat}/K_M value 10-fold lower than that of Cb-thiaminase I WT. Upon addition of the reaction nucleophile, Asp94 acts as a general acid to reprotonate the 4′-amino group, and rearrangement allows the covalent enzyme-substrate bond to break. Cys143 is then reprotonated by Glu271 to regenerate the enzyme.

Alignment of thiaminase I sequences from several organisms shows conservation of residues equivalent to Cb-thiaminase I Tyr80, Asp94, Cys143, Glu271

and Asp302 (Figure 3.7). In the proposed mechanism, these residues are critical for binding and catalysis as described. The conservation of the catalytic cysteine and glutamate residues is consistent with the previously proposed mechanism.[15, 18, 35] In addition, the importance of Asp94, Tyr80, and Asp302 for binding and catalysis is now established through structural evidence and kinetic data. While Tyr46 mutation shows only modest effects on Cb-thiaminase I activity, Tyr46 is not fully conserved in sequence across these organisms. This suggests that Tyr46 is involved in the thiamin-binding hydrogen bond network, but is not essential for activity.

Relationship to Thiaminase II. TenA is the thiaminase II from *Bacillus subtilus* (PDB ID IYAK). Like Cb-thiaminase I, TenA degrades thiamin into its pyrimidine and thiazole moieties.[17] The monophosphorylated and diphosphorylated forms of thiamin (ThMP and ThDP) are not substrates for thiaminase IIs, whereas thiaminase Is can tolerate modification of the thiazole moiety, degrading thiamin and many thiamin derivatives.[17, 34] Interestingly, thiaminase II is often clustered with the thiamin biosynthetic machinery. This grouping of the biosynthetic and degradation enzymes was initially thought to be counterproductive. It was later discovered that thiamin does not appear to be the preferred physiological substrate for thiaminase II. Rather, thiaminase II salvages base-degraded pyrimidine from the environment for incorporation into thiamin.[36-38]

TenA adopts an overall α-fold that shows no homology to the Cb-thiaminase I structure. The active site of TenA is buried and fairly inaccessible to potential nucleophiles, leading to exclusive use of water as the nucleophile. This contrasts the Cb-thiaminase I two-domain α/β-fold and accessible active site location near the

surface of the enzyme, allowing a larger variety of nucleophiles to aid in catalysis. Despite their major structural differences, TenA and Cb-thiaminase share some active site features. Both feature a catalytic cysteine and activating glutamate with an aspartate residue anchoring the pyrimidine.[37] In the absence of a TenA structure with bound thiamin, the position of thiamin was previously predicted based on the structure of TenA with 4-amino-2-methyl-5-hydroxymethylpyrimidine (HMH) (PDB ID: 1YAK). The thiamin F-conformation observed in C143S/thiamin and the position of HMH were used to model thiamin into the TenA active site (Figure 3.8). In this model, thiamin is anchored by interaction between Asp44 with N3′and the 4′-amino group, Glu205 and N1′, and Tyr163 and the 4′-amino group. The pyrimidine half of thiamin stacks above Tyr139 and the thiazole portion extends across the length of the binding pocket. It is further anchored by interactions between the thiazole 5-ethanol and Gln55 and Thr78. The nearby catalytic cysteine, Cys135, is positioned nearby the site of covalent addition to thiamin C6′.

In TenA the glutamate residue is not close enough to directly deprotonate the cysteine and must act through two intermediate tyrosine residues, in contrast to Cb-thiaminase, where Glu271 directly deprotonates cysteine. Unlike TenA, there are no nearby tyrosine residues in Cb-thiaminase I that could possibly participate in a similar interaction. While both thiaminase Is and IIs share a similar catalytic mechanism for thiamin degradation, their major structural differences and substrate preferences suggest that they evolved independently and may serve different physiological purposes.

The atomic level details of the Cb-thiaminase I enzyme-substrate complex provide a complete substrate binding model for comparison to functionally and structurally similar enzymes. Structural comparisons suggest that thiaminase Is evolved from a PBP ancestor and later acquired a thiamin degradation function. Given that they can accept a variety of thiamin analogs, thiaminase Is may function in pyrimidine salvage, like thiaminase II. An alternate hypothesis suggests that thiaminase I is used as a defense mechanism against other organisms, as observed in the toxic effect of consuming thiaminase I containing food. It remains to be determined if thiaminase I containing organisms can utilize thiamin analogs for thiamin dependent reactions. If so, depleting thiamin levels through thiaminase I activity would not be detrimental to the vitality of thiaminase I containing organisms. As identification of additional substrates and organisms containing thiaminase I genes continues, the structure of the Cb-thiaminase I co-crystallized with thiamin will assist in understanding their chemistry.

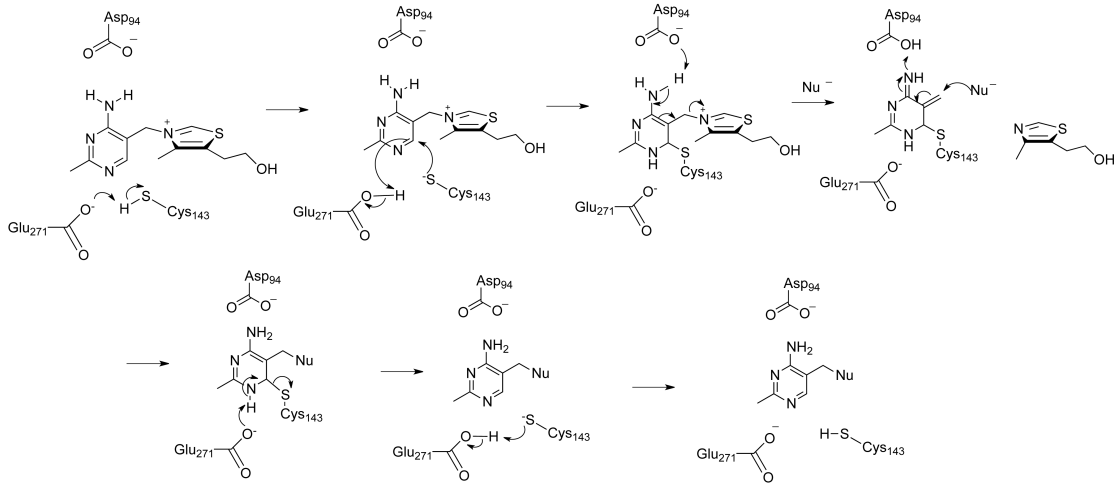


Figure 3.6. Mechanism of thiamin degradation by Cb-thiaminase I.
The mechanism proposed for thiamin degradation by Cb-thiaminase I based on kinetic studies of Cb-thiaminase I mutants and structural data.

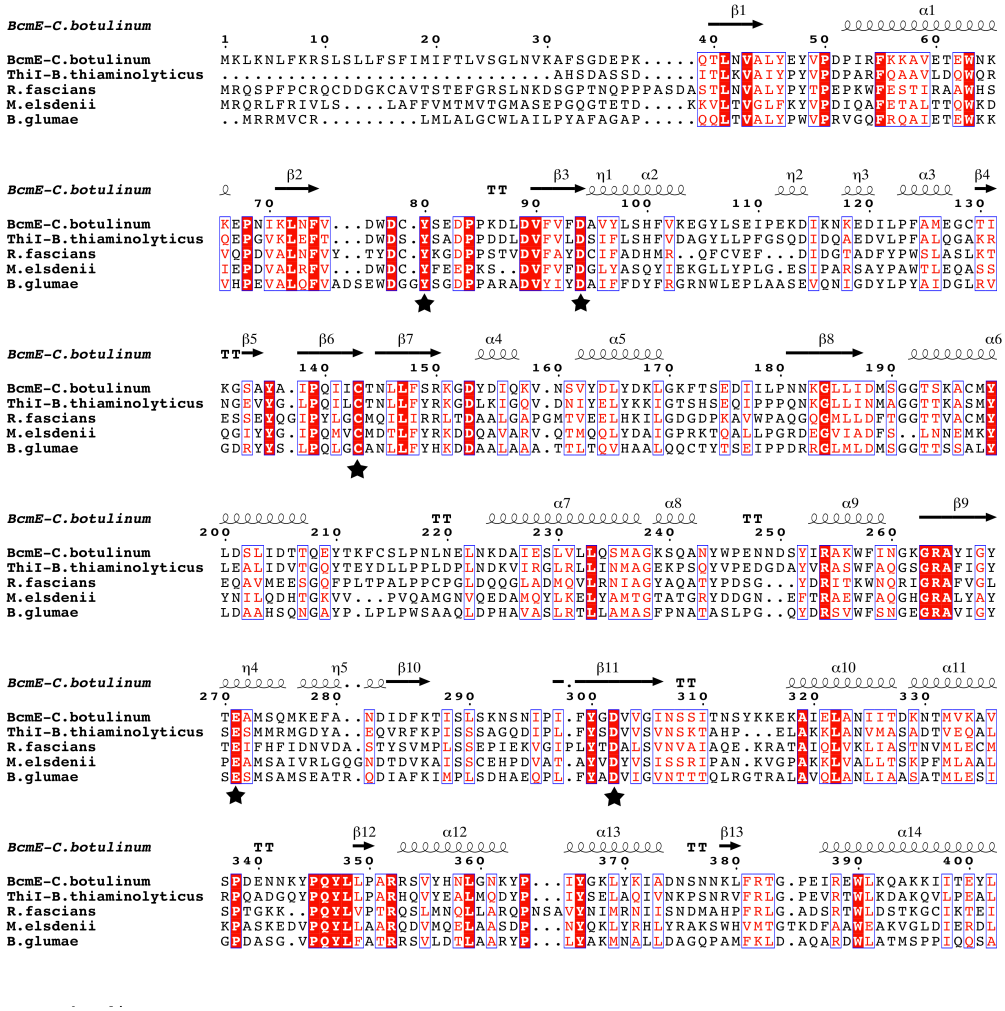


Figure 3.7. Sequence alignment of thiaminase Is. A multiple sequence alignment of Cb-thiaminase I, Bt-thiaminase I, and thiaminase I from additional organisms showing conservation of residues implicated for catalysis. Residues corresponding to Cb-thiaminase I Tyr80, Asp94, Cys143, Glu271, and Asp302 are conserved and highlighted with a star.

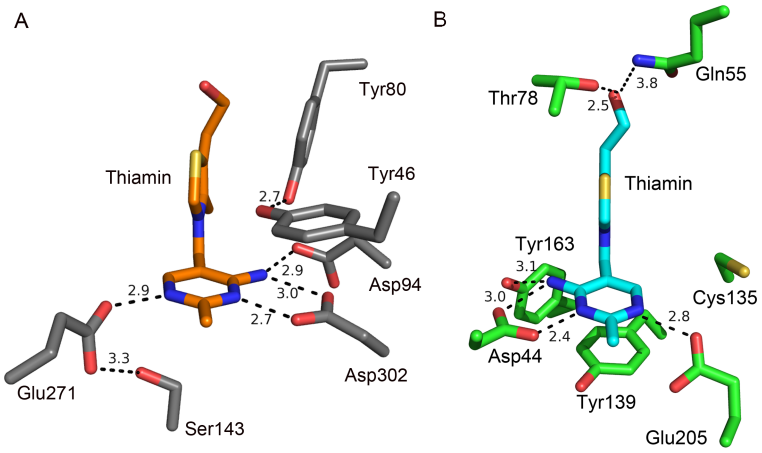


Figure 3.8. Comparison of thiamin binding in Thiaminase I and II. (A) Thiamin binding interactions observed in C143S/Thiamin. (B) Model for thiamin binding in TenA based on C143S/Thiamin structure and TenA PDB ID: 2QCX.

REFERENCES

1. 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.
2. Carvalho, P. S. M., Tillitt, D. E., Zajicek, J. L., Claunch, R. A., Honeyfield, D. C., Fitzsimons, J. D., and Brown, S. B. (2009) Thiamine Deficiency Effects on the Vision and Foraging Ability of Lake Trout Fry, *J Aquat Anim Health 21*, 315-325.
3. Honeyfield, D. C., Tillitt, D. E., Fitzsimons, J. D., and Brown, S. B. (2010) Variation in Lake Michigan Alewife (Alosa pseudoharengus) Thiaminase and Fatty Acids Composition, *J Freshwater Ecol 25*, 65-71.
4. Jaroszewska, M., Lee, B. J., Dabrowski, K., Czesny, S., Rinchard, J., Trzeciak, P., and Wilczynska, B. (2009) Effects of vitamin B-1 (thiamine) deficiency in lake trout (Salvelinus namaycush) alevins at hatching stage, *Comp Biochem Phys A 154*, 255-262.
5. Muller, I. B., Bergmann, B., Groves, M. R., Couto, I., Amaral, L., Begley, T. P., Walter, R. D., and Wrenger, C. (2009) The Vitamin B1 Metabolism of Staphylococcus aureus Is Controlled at Enzymatic and Transcriptional Levels, *Plos One 4*.
6. Ketola, H. G., Isaacs, G. R., Robins, J. S., and Lloyd, R. C. (2008) Effectiveness and retention of thiamine and its analogs administered to Steelhead and landlocked Atlantic salmon, *J Aquat Anim Health 20*, 29-38.

7. Kimura, Y., and Iwashima, A. (1987) Occurrence of Thiaminase-Ii in Saccharomyces-Cerevisiae, *Experientia 43*, 888-890.

8. Puzach, S. S., and Gorbach, Z. V. (1989) Characteristics of Products Developed after Degradation of Thiamin by Means of Mollusk Thiaminase-I, *Vop Med Khim 35*, 82-84.

9. Sato, M., Hayashi, S., and Nishino, K. (1994) Subcellular-Localization of Thiaminase-I in the Kidney and Spleen of Carp, Cyprinus-Carpio, *Comparative Biochemistry and Physiology a-Physiology 108*, 31-38.

10. Mccleary, B. V., and Chick, B. F. (1977) Purification and Properties of a Thiaminase I Enzyme from Nardoo (Marsilea-Drummondii), *Phytochemistry 16*, 207-213.

11. Earl, J. W., and Mccleary, B. V. (1994) Mystery of the Poisoned Expedition, *Nature 368*, 683-684.

12. Honeyfield, D. C., Brown, S. B., Fitzsimons, J. D., and Tillitt, D. E. (2005) Early mortality syndrome in great lakes salmonines, *J Aquat Anim Health 17*, 1-3.

13. Hazell, A. S., Todd, K. G., and Butterworth, R. F. (1998) Mechanisms of neuronal cell death in Wernicke's encephalopathy, *Metab Brain Dis 13*, 97-122.

14. Singleton, C. K., and Martin, P. R. (2001) Molecular mechanisms of thiamine utilization, *Curr Mol Med 1*, 197-207.

15. Costello, C. A., Kelleher, N. L., Abe, M., McLafferty, F. W., and Begley, T. P. (1996) Mechanistic studies on thiaminase I - Overexpression and identification of the active site nucleophile, *J Biol Chem 271*, 3445-3452.

16. Lienhard, G. E. (1970) Kinetic Evidence for a (4-Amino-2-Methyl-5-Pyrimidinyl)Methyl-Enzyme Intermediate in Thiaminase-I Reaction, *Biochemistry-Us 9*, 3011-&.

17. 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-Us 44*, 2319-2329.

18. Campobasso, N., Costello, C. A., Kinsland, C., Begley, T. P., and Ealick, S. E. (1998) Crystal structure of thiaminase-I from Bacillus thiaminolyticus at 2.0 angstrom resolution, *Biochemistry-Us 37*, 15981-15989.

19. Spurlino, J. C., Lu, G. Y., and Quiocho, F. A. (1991) The 2.3-A resolution structure of the maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis, *J Biol Chem 266*, 5202-5219.

20. Bale, S., Rajashankar, K. R., Perry, K., Begley, T. P., and Ealick, S. E. (2010) HMP Binding Protein ThiY and HMP-P Synthase THI5 Are Structural Homologues, *Biochemistry-Us 49*, 8929-8936.

21. 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-Us 47*, 1346-1357.

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

23. Kabsch, W. (2010) Xds, *Acta Crystallogr D 66*, 125-132.

24. Kabsch, W. (2010) Integration, scaling, space-group assignment and post-refinement, *Acta Crystallogr D 66*, 133-144.

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

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

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

28. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr D Biol Crystallogr 66*, 213-221.

29. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Cryst. D 66*, 12-21.

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

31. Shin, W., Oh, D. G., Chae, C. H., and Yoon, T. S. (1993) Conformational analyses of thiamin-related compounds. A stereochemical model for thiamin catalysis, *J Am Chem Soc 115*, 12238-12250.

32. Holm, L., and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D, *Nucleic Acids Res. 38 Suppl*, W545-549.

33. Dwyer, M. A., and Hellinga, H. W. (2004) Periplasmic binding proteins: a versatile superfamily for protein engineering, *Curr Opin Struct Biol 14*, 495-504.

34. Bos, M., and Kozik, A. (2000) Some molecular and enzymatic properties of a homogeneous preparation of thiaminase I purified from carp liver, *J Protein Chem 19*, 75-84.

35. Costello, C. A., and Begley, T. (1994) Thiaminase-I - Mechanistic Studies on a Thiamine Degrading Enzyme, *Abstr Pap Am Chem S 208*, 33-BIOL.

36. Jenkins, A. H., Schyns, G., Potot, S., Sun, G. X., and Begley, T. P. (2007) A new thiamin salvage pathway, *Nat Chem Biol 3*, 492-497.

37. 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.

38. Barison, N., Cendron, L., Trento, A., Angelini, A., and Zanotti, G. (2009) Structural and mutational analysis of TenA protein (HP1287) from the

Helicobacter pylori thiamin salvage pathway - evidence of a different substrate specificity, *Febs J 276*, 6227-6235.

CHAPTER 4

STRUCTURE-FUNCTION STUDIES ON *CLOSTRIDIUM BOTULINUM* BACIMETHRIN KINASE, BcmD

Section 4.1. Introduction

Thiamin (vitamin B_1) is a cofactor in many essential biological processes. Numerous enzymes involved in primary metabolism rely on the active form, thiamin pyrophosphate (ThDP), to function. As a result, severe thiamin deficiency causes illnesses such as beriberi or Wernicke-Korsakoff syndrome. Humans are among those that obtain thiamin from dietary sources, while some organisms are capable of thiamin biosynthesis. These biosynthetic routes differ between eukaryotes and bacteria, but both follow a two-branched path where 4-amino-5-hydroxymethyl-2-methylpyrimidine pyrophosphate (HMP-PP) and thiazole phosphate carboxylate (THZ-P) are synthesized separately, then coupled by thiamin phosphate synthase to form thiamin monophosphate (ThMP).[1] ThMP is then phosphorylated again to create ThDP (Figure 4.1A).

Bacimethrin (BCM), 4-amino-5-hydroxymethyl-2-methoxypyrimidine, is a structural homolog to HMP that acts as a thiamin antimetabolite. It has previously been isolated from *Bacillus magaterium* and *Streptomyces albus* and is shown to be toxic to yeast and bacteria.[2, 3] BCM toxicity is alleviated when the growth media is supplemented with HMP, thiamin, or pyridoxine.[2-4] Initially thought to act as an inhibitor of thiamin biosynthesis, studies show that BCM acts as a substrate for thiamin biosynthetic enzymes,[5] as BCM differs from HMP only by addition of a methoxy group at the 2-position. In *E. coli* treated with BCM, the *in vitro* conversion

to MeOThDP is six times faster than ThDP synthesis when fed HMP.[5] BCM hijacks the thiamin biosynthetic pathway, creating MeOThDP (Figure 1B) and diminished levels of ThDP. Without this necessary cofactor, ThDP dependent reactions are halted, leading to cell death.[5]

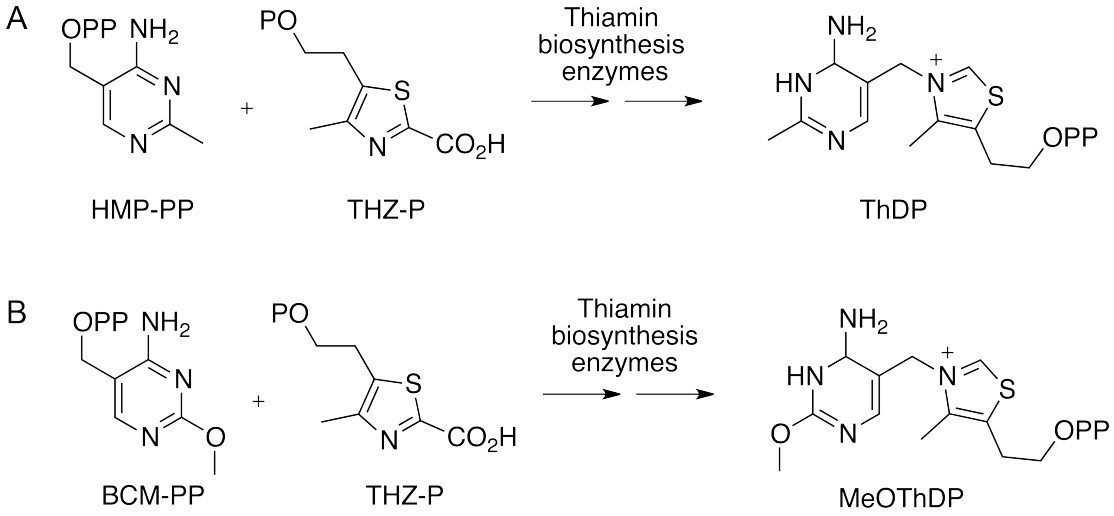


Figure 4.1. Routes to thiamin pyrophosphate (ThDP) and 2'-methoxythiamin pyrophosphate biosynthesis (MeOThDP). A) Hydroxymethylpyrimidine pyrophosphate (HMP-PP) and thiazole phosphate carboxylate (THZ-P) are synthesized separately then ligated together and phosphorylated by thiamin biosynthetic machinery to produce ThDP. B) Bacimethrin pyrophosphate (BCM-PP) is ligated to THZ-P by the thiamin biosynthesis enzymes to produce the thiamin analog, MeOThDP.

The BCM biosynthetic pathway was recently characterized in *Clostridium botulinum A* ATCC 19397 (Begley Lab- *manuscript in preparation*). In this pathway, a cytosine ribonucleotide serves as the precursor for BCM (Figure 4.2). In contrast, bacterial thiamin biosynthesis utilizes a complex chemical reaction to convert 5-aminoimidazole ribotide to HMP-P.[6] HMP-P is then further phosphorylated by a

HMPP kinase to HMP-PP.[1, 7] BcmD is annotated as the gene *thiD-2* in *C. botulinum.* BcmD converts BCM to BCM-pyrophosphate (BCM-PP) through two sequential phosphorylations. This reaction primes BCM to be shuttled into the thiamin biosynthetic pathway for production of MeOThDP. BcmD is highly similar in sequence to known HMPP kinases, which are members of the ribokinase structural family. In previous studies, the structure of HMPP kinase from *Salmonella typhimurium* (*St*HMPP kinase) has been solved with HMP bound in the active site (PDB ID: IJXI).[8] Structural comparison of *St*HMPP kinase with other members of the ribokinases superfamily led to a model for ATP binding and subsequent HMP phosphorylation. BcmD is expected to act in a similar manner and the structure of BcmD is of interest in understanding the substrate specificity of BCM biosynthetic enzymes.

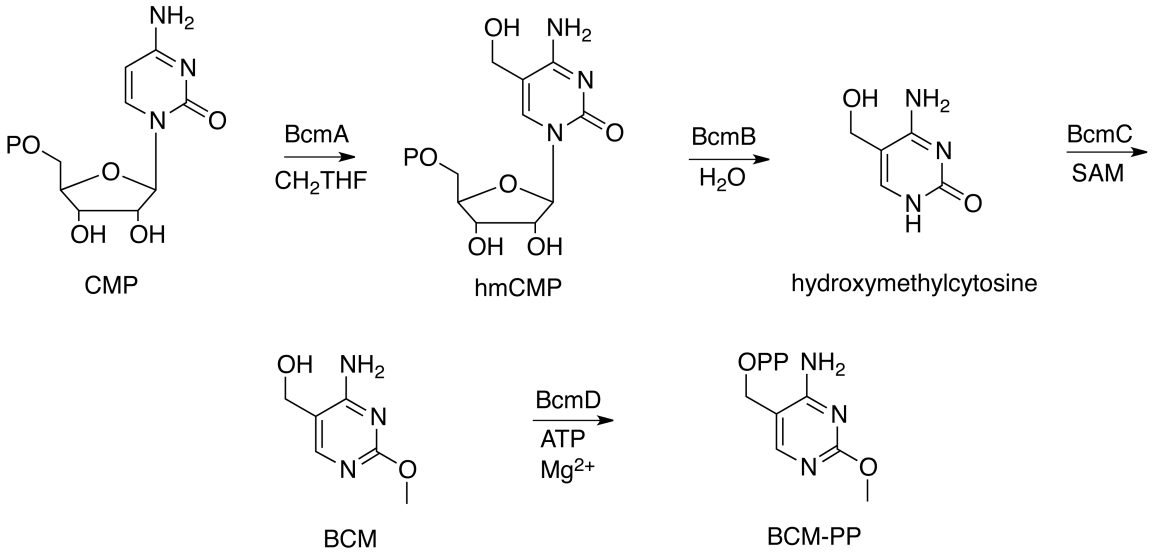


Figure 4.2. Bacimethrin biosynthesis pathway. A 5-hydroxymethyl is added to cytidine 5′-phosphate (CMP) by BcmA to create hydroxymethylCMP (hmCMP). The hmCMP *N*-glycosidic bond is cleaved by BcmB and the hydroxymethylcytosine product is methylated at O2 by BcmC and SAM to produce bacimethrin (BCM). In the final step, BCM is phosphorylated twice by BcmD to produce BCM-PP.

Functional studies show enzymatic synthesis of BCM when BcmA, BcmB, and BcmC are incubated with the necessary precursor molecules (Begley Lab- *manuscript in preparation*). Incubation of BCM with BcmD and ATP and Mg^{2+} cofactors results in the depletion of ATP and BCM, with two new peaks observed when the products were analyzed by HPLC. These two peaks represented molecules that are more polar than BCM, consistent with the formation of BCM-P and BCM-PP. Treatment of the reaction products with alkaline phosphatase caused the products to revert back to BCM. When ATP or BcmD was omitted from the reaction mixture, no consumption of BCM was observed. BcmD showed similar activity toward HMP (Begley Lab- in preparation). This activity reveals that BcmD is a kinase for both BCM and HMP. In this study, the structure of BcmD was analyzed and compared to *St*HMPP kinase to determine if there is a structural basis for differentiating between these very similar pyrimidine substrates. Here we report the structure of BcmD at 2.79 Å, compare it with structures of related ribokinases family members, and discuss the functional implications for its role in BCM-PP and MeOThDP biosynthesis.

Section 4.2. Materials and Methods

Cloning, Overexpression and Purification of BcmD. The gene encoding BcmD, *thiD-2*, was cloned from the genomic DNA of *C. botulinum* A str. ATCC 19397. The gene was ligated into the pTHT vector, a modified pET-28 with an N-terminal His_6 tag followed by a TEV protease cleavage site. Standard DNA manipulation methods were followed during all of the cloning procedures and the DNA was sequenced to verify plasmid accuracy.[9] The plasmids were transformed into *Escherichia coli* BL21(DE3) cells and grown overnight at 37 °C on selective

kanamycin (30 μg/mL) containing agar. For each starter culture, a single colony was selected and transferred to 10 mL sterile LB Lennox media with 30 μg/mL kanamycin at 37 °C with shaking overnight. Approximately 18 h later, the overnight starter cultures were added directly to 1.5 L volumes of sterile LB media. The cells were grown with shaking at 200 RPM in a 37 °C incubator until reaching an OD600 of 0.6. At this point the incubator temperature was reduced to 15 °C. After reaching the induction temperature (15 °C) and an OD600 of 0.8, solid isopropyl-β-D-1-thiogalactopyranoside (IPTG) was added to reach a final concentration of 0.5 mM IPTG in each 1.5 L flask. The cells were shaken for approximately 18 h after addition of IPTG and then harvested by centrifugation at 2000*g* for 20 min. The cell pellet was collected in 3 L aliquots and frozen at -20 °C for storage.

The cell pellet was later thawed and resuspended in 30 mL of lysis buffer (50 mM Tris-HCl, 300 mM NaCl, and 20 mM imidazole at pH 8.0). The resuspension was subjected to three rounds of sonication for 3 min on and 10 min off. The lysed cell extract was next centrifuged at 40,000*g* for 30 min at 4 °C. The supernatant was collected and loaded on to a preequilibrated 2 mL NiNTA column (Qiagen). The column was washed with 50 mL of lysis buffer with a flow rate of 1 ml/min to remove any nonspecifically bound contaminants. The protein was eluted from the column by passing 20 mL of elution buffer (50 mM Tris-HCl, 300 mM NaCl, and 250 mM imidazole at a final pH of 8.0) through the column. The elution volume containing BcmD was collected.

BcmD purity was assessed by SDS-PAGE analysis. The sample was buffer exchanged by overnight dialysis into 10 mM Tris-HCl at pH 7.5 and 150 mM KCl,

then concentrated to 14 mg/mL using a centrifugal filter device with a molecular weight cutoff of 10,000 Da. The purification process yielded approximately 80 mg protein per L of cell culture. The purified protein was aliquoted and flash frozen in liquid nitrogen for storage at -80 °C.

Crystallization of BcmD. The sitting drop vapor diffusion method was used for crystallization and the initial crystallization condition was determined using the commercially available Wizard II sparse matrix screen (Emerald Biosystems). The protein solution was thawed on ice and equal volumes of protein and reservoir solutions were mixed and equilibrated at 22 °C against a total volume of 700 μL well solution. The optimized crystallization conditions were 6-12% (w/v) polyethylene glycol 8,000, 0.1 M sodium/potassium phosphate pH 5.6, and 0.2 M NaCl. Plate-like BcmD crystals grew approximately 150 μm by 100 μm in about two to three days. Crystals were cryoprotected for data collection in a solution composed of the mother liquor supplemented with 15% (v/v) glycerol.

X-ray Data Collection and Processing. X-ray diffraction experiments were conducted at APS NE-CAT beamline 24-ID-E (Argonne National Laboratory) using a Quantum 315 detector (Area Detector Systems Corp.) at a distance of 400 mm. The oscillation method was used with 1.0º rotation per frame for 180º with cryocooling. Diffraction data was collected at wavelength of 0.97918 Å using 1 s exposures. HKL2000 was used for indexing, integration, and scaling of the data.[10] Data collection and processing statistics are listed in Table 4.1.

Structure Determination, Model Building, and Refinement. The BcmD crystal belongs to space group $P2_1$ and contains two molecules per asymmetric unit. It has a

Table 4.1. Summary of data collection and refinement statistics

	BcmD
beamline	APS 24-ID-E
resolution (Å)	2.79
wavelength (Å)	0.97918
space group	$P2_1$
a (Å)	47.3
b (Å)	109.5
c (Å)	53.6
β (°)	106.1
Matthews coefficient	2.22
% solvent	44.7
mol/asu	2
Measured reflections	48,203
Unique reflections	13,017
Average I/σ	22.0 (3.9)[a]
Redundancy	3.7 (3.6)
Completeness (%)	100 (100)
R_{merge} (%)[b]	9.3 (38.6)
Reflections in working set	37537
Reflections in test set	1877
R-factor (%)	21.5
R_{free} (%)	26.4
rms deviation from ideals	
bonds (Å)	0.01
angles (deg)	1.385
average *B* factor for protein (Å^3)	52.6
Ramachandram plot	
most favored (%)	95.0
allowed (%)	5.0
disallowed (%)	0.0

[a]Values in parentheses are for the highest-resolution shell.
[b]$R_{merge} = \Sigma\Sigma_i | I_i - \langle I \rangle | / \Sigma \langle I \rangle$, where $\langle I \rangle$ is the mean intensity of the N reflections with intensities I_i and common indices h,k,l.
[c]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, calculated over all reflections used in the refinement. R_{free}, is similar to R_{work} but calculated over a subset of reflections (5%) excluded from all stages of refinement.

Matthews coefficient[11] of 2.22 $Å^3$/Da and corresponding estimated solvent content of 44.7%. The structure of BcmD was determined by the molecular replacement method using *St*HMPP kinase with 42% sequence identity, as the search model (PDB ID: 1JXH). Two chains were positioned in the asymmetric unit by the automated search program MolRep.[12] Iterative rounds of model building were conducted using COOT,[13] followed by refinement with RefMac5[14] in initial rounds and PHENIX.refine[15] in final rounds. The quality of the final structure was analyzed using MolProbity[16] and COOT.[13] Structure refinement converged with a final R-factor value of 21.5% and R_{free} of 26.4%. Complete refinement statistics are available in Table 1.

Section 4.3. Results

Overall Structure of BcmD. The structure of BcmD was solved to 2.79 Å in the $P2_1$ space group, with two protein chains in the asymmetric unit. The two chains, A and B, are related by 2-fold noncrystallographic symmetry and interact with an interface area from each chain of about 1390 $Å^2$ (Figure 4.3A). The homodimer spans 75 Å and is 41 Å wide by 33 Å deep. At the dimer interface there are nine hydrogen bonds and 148 hydrophobic contacts. The final model includes residues 2-239, with residues 117, 126, 178-180, and 197-208 in chain A and 109-11, 175-181, and 196-208 in chain B disordered in the crystal structure. The R-factor and R-free of the final BcmD model are 21.5% and 26.4%, respectively.

Tertiary structure. Each monomer forms an α-β-α three layer sandwich (Figure 4.3B). The central β-sheet, with topology β2β1β5β6β7β8β9β10, is parallel, with the exception of β9 that runs anti-parallel to β8 and β10 (Figure 4.3C). The α-helices flank both sides of the β–sheet, with α1, α7, and α8 on one side and α2, α3,

α4, α5, and α6, on the other. The remaining β-strands are part of an active-site flap in which residues 41 to 53 create the anti-parallel β3 and β4 strands that form a protective flap over the active site. A second active site flap is created by a flexible loop composed of residues 106 to 114.

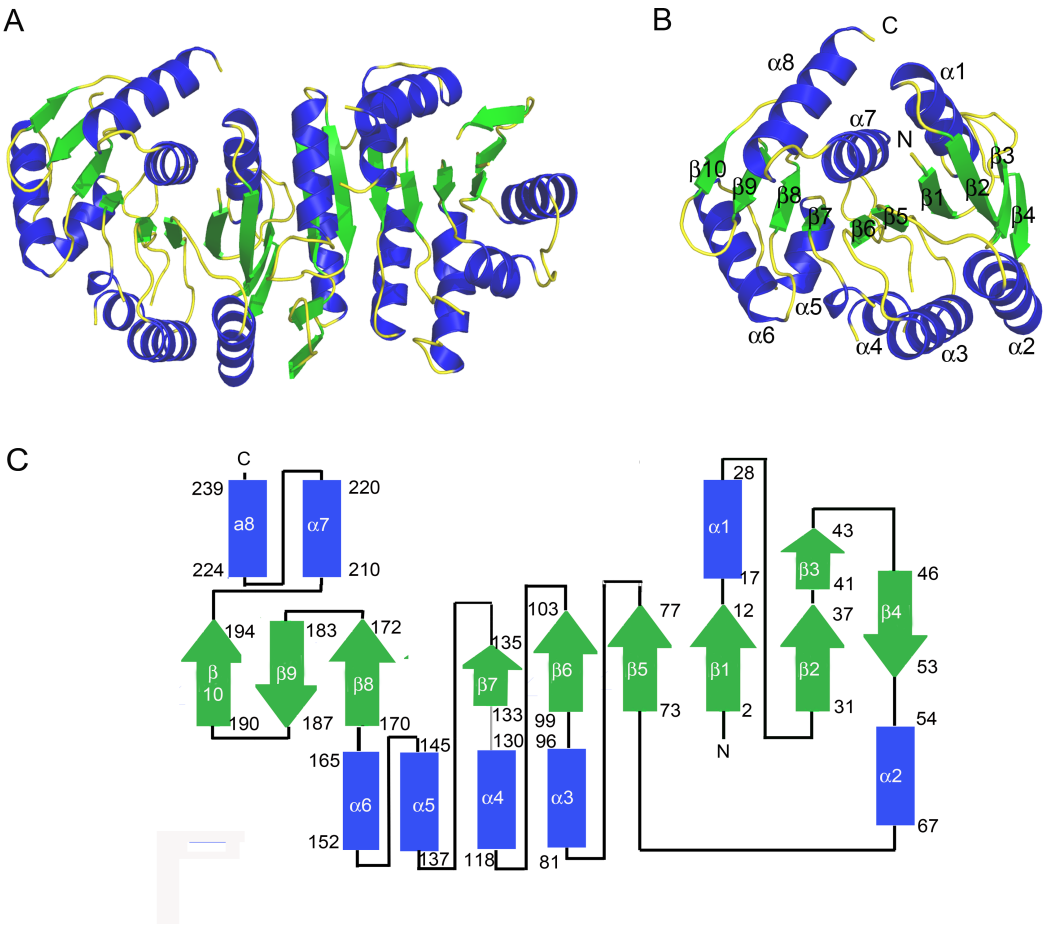


Figure 4.3. Structure of BcmD. A) BcmD is a homodimer with α-helices depicted in blue and β-strands depicted in green. B) The BcmD monomer α-β-α three layer sandwich is shown with the same color scheme as in A. A core β-sheet is surround on either side by α-helices. β-strands 3 and 4 form the active site binding flap. C) A topology diagram of BcmD displays the arrangement of the secondary structural elements

BcmD Active Site. Each BcmD molecule contains two active sites, each self-contained within the individual protomers. Co-crystallization and soaking experiments were unsuccessful in obtaining a liganded BcmD structure. The amino acids involved in substrate binding and catalysis are predicted from the superposition of BcmD with *St*HMPP kinase with bound HMP (PDB ID 1JXI) (Figure 4.4). In this model, BcmD Glu42 makes a hydrogen bond with N4 of the pyrimidine substrate. The BcmD Met78 and Gly9 amino backbone groups bind a water molecule that hydrogen bonds with the pyrimidine N1. The pyrimidine is also stabilized by water mediated hydrogen bonding by Asp21. The ATP binding site is highly conserved in ribokinases. While disordered in BcmD, the predictive anion binding GTGC motif common to this family is conserved in BcmD (residues 206-209) (Figure 4.5). BcmD Asp103, Glu140, and backbone residues create a phosphate-binding pocket, with Lys173, Asp183, and Lys233 also positioned for favorable ATP binding interactions.

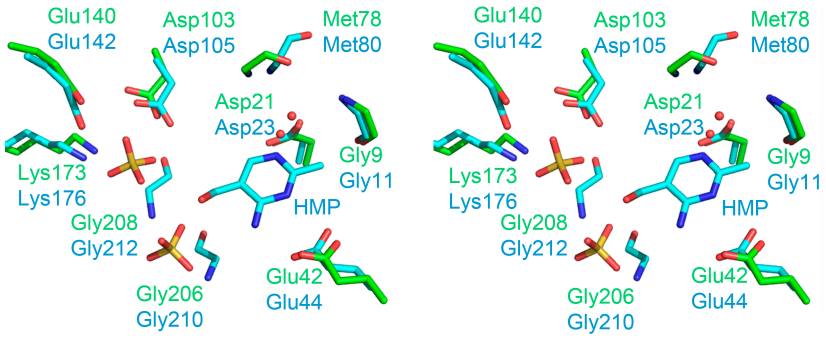


Figure 4.4. Stereoview of BcmD active site aligned with StHMPP kinase/HMP and two sulfates. The BcmD active site was identified by alignment with StHMPP kinase. All active site residues are conserved. Gly208 and Gly206 are disordered in BcmD but conserved in sequence.

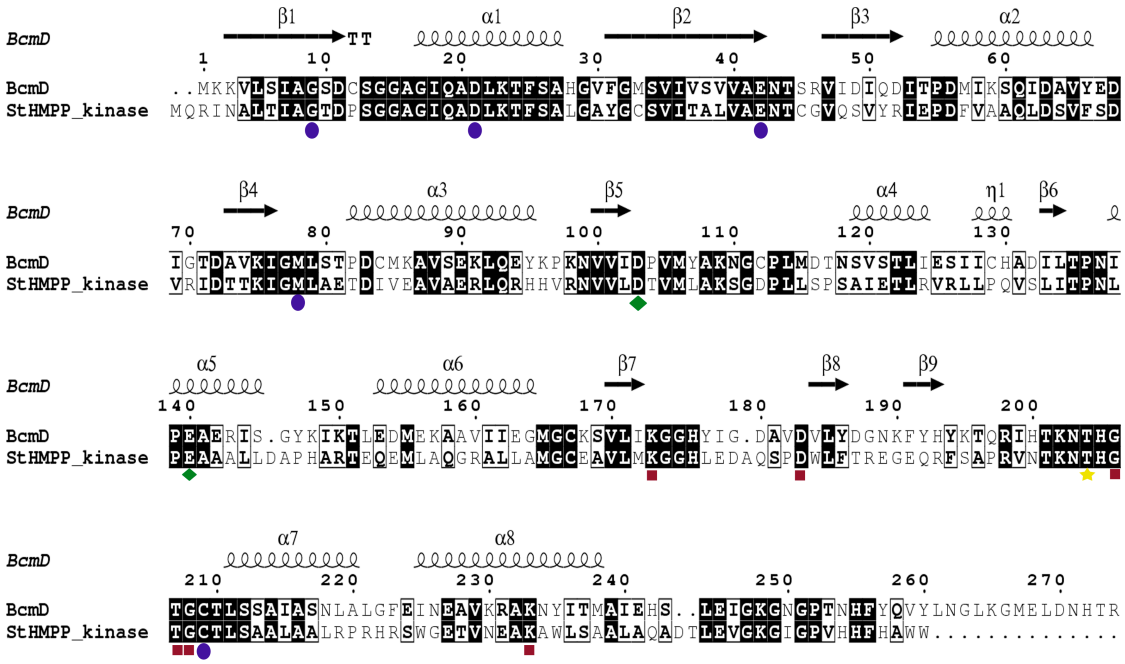


Figure 4.5. BcmD/StHMPP kinase sequence alignment. The residues of importance in StHMPP kinase activity are highlighted using the following scheme: purple circles (HMP binding), red squares (ATP binding), yellow star (Mg^{2+} coordination), green diamond (conserved active site residues). All of the residues important for StHMPP kinase activity are conserved in BcmD.

Section 4.4. Discussion

Structure Comparison. BcmD adopts an α-β-a sandwich fold with high structural similarity to known HMPP kinases and is classified as a member of the ribokinase superfamily based on its function and tertiary structure. A DALI search of BcmD against the PDB reveals proteins with high structural homology (Table 4.2). The HMPP kinase from *Thermus thermophilus* (PDB ID 1UB0) has the highest structural homology, with a Z-score of 31.1 and rmsd of 1.6. It shares 41% sequence identity with BcmD and also forms a dimer. The DALI structural search also reveals the HMPP kinases from *S. typhimurium*, *Bacillus subtilis,* and yeast have similarly

high Z-scores and rmsd values as compared to BcmD. As expected from previous structural studies of *St*HMPP kinase, BcmD also shares high structural homology to many known pyridoxine and ribokinases. Although the sequence similarity decreases between BcmD and the pyridoxine and ribokinases as compared to the HMPP kinases, the overall structure remains conserved. The structural homologs listed in Table 2 are all dimers in their biologically relevant states, further supporting the close relationship within this family of enzymes.

BCM Binding Site. The BcmD active site was identified by alignment with the *St*HMPP kinase structure containing HMP and sulfate in the active site. The residues necessary for HMP binding in *St*HMPP kinase are conserved in BcmD such that BCM is presumed to bind BcmD similarly to HMP binding in *St*HMPP kinase (Figure 4). Residues that make hydrophobic contacts with HMP in StHMPP kinase are conserved in BcmD. The BcmD binding pocket has room to accommodate the addition of the 2-methoxy group in BCM, as BCM only differs from HMP at this position. Preliminary studies suggest that BcmD may actually favor HMP as a substrate over BCM (Begley Lab- in preparation). Any interactions that favor HMP over BCM are not clear from analysis of the active site binding pocket residues alone.

ATP Binding and Kinase Reaction. Members of the ribokinase family phosphorylate the hydroxymethyl group of a small molecule substrate. In the absence of a *St*HMPP kinase/ATP structure, the binding and reaction mechanisms were deduced by comparison to other ribokinases structures. In this reaction, an inline displacement mechanism is utilized whereby an aspartate or cysteine activated water acts a base to deprotonates the HMP hydroxymethyl group. The ATP co-substrate is

Table 4.2. Summary of proteins structurally similar to BcmD

Protein	PDB ID	Z score	rmsd	% identical	no. of aligned residues
T. thermophilus HMP Kinase	1UB0	31.1	1.6	41	212
S. typhimurium HMP Kinase	1JXI	29.7	1.5	43	211
B. subtilis HMP kinase	2I5B	29.1	2.1	43	218
Yeast Thi20	3RM5	27.7	1.8	38	212
E. coli pyridoxine kinase	2DDO	27.1	2.0	17	217
Y. pestis pyridoxamine kinase	3PZS	25.5	2.1	17	220
Human pyridoxal kinase	2YXU	25.0	2.0	19	218
E. coli ribokinase	1RKA	19.3	2.5	18	204

positioned in an ATP binding pocket near the small molecule hydroxymethyl acceptor group. The ATP is highly coordinated by Mg^{2+} and Thr204. Its γ-phosphate is transferred to HMP to produce HMP-P. During the reaction, the bound Mg^{2+} and the enzyme's anion hole stabilize the transition state. After the first phosphorylation, the ADP dissociates, then the HMP-P phosphate moves to an alternate binding location where it is stabilized by lysine, and a new ATP molecule binds similarly to the first. Finally, the second phosphorylation to produce HMP-PP occurs in a similar manner. Superposition of *St*HMPP kinase and BcmD shows conservation of the ATP and substrate binding sites. Based on functional and structural similarity, this mechanism appears conserved in BcmD and by analogy, BcmD acts on BCM to produce BCM-PP.

Relationship to Ribokinases. BcmD is classified as a member of the ribokinase family based on a conserved fold and active site architecture. Ribokinases diverge in

the identity of the substrate, their substrate binding residues, and method of shielding the active site from the solvent. Like *St*HMPP kinase, BcmD has two active site flaps formed by β-strands 3 and 4 and the loop region of residues 106-114 that close upon substrate binding. In the unliganded BcmD model, this flap is in the "open" conformation (Figure 4.6). The residues equivalent to this active site flap are ordered in the liganded *St*HMPP kinase revealing the "closed" state (PDB ID: 1JXI), but were disordered in its apo-structure (PDB ID: 1JXH). The "open" structure of BcmD reveals a 10 Å movement of this flap upon substrate binding. Members of the ribokinases family shield their bound substrates using structural features of varying complexity.[8] The BcmD active site shielding mechanism is consistent with *St*HMPP kinase. This method is distinct among members of this family, as the movement of the β-flap and loop region upon substrate binding appears to be one of the simplest shielding mechanisms. Relatives such as ribokinases, adenosine kinase, and glucokinase evolved to contain insertions of additional secondary structural features that protect their active site.[8]

Role in Bacimethrin Biosynthesis. *C. botulinum* A str. ATCC 19397 contains two *thiD* genes, with 52% identity to each other. In this *C. botulinum* strain, the *thiD-1* gene is clustered with a thiazole kinase and thiamin phosphate pyrophosphorylase, suggesting it is the HMPP kinase utilized for thiamin biosynthesis. BcmD, encoded by *thiD-2,* is clustered in the newly established BCM biosynthesis pathway. With high sequence similarity, overall structural homology, and nearly identical active site architecture, BcmD is very closely related to the thiamin biosynthesis HMPP kinase. BcmD has demonstrated substrate flexibility for either HMP or BCM, but it is not yet

clear whether BCM is a substrate for the HMPP kinase encoded by *thiD-1*. In order for BCM to act as a toxin, it would be advantageous to the BCM producer if HMPP kinases phosphorylate BCM, allowing either BCM or BCM-PP to act as a toxin. BcmD may have evolved from a duplication event of the *thiD* gene, as its kinase activity prepares BCM to be shuttled into the thiamin biosynthesis pathway for production of the toxic ThDP analog, MeOThDP. While the substrate preferences of BcmD for HMP and BCM is still being established, this preliminary finding and structural similarity establishes a close relationship between BcmD and the thiamin HMPP kinase.

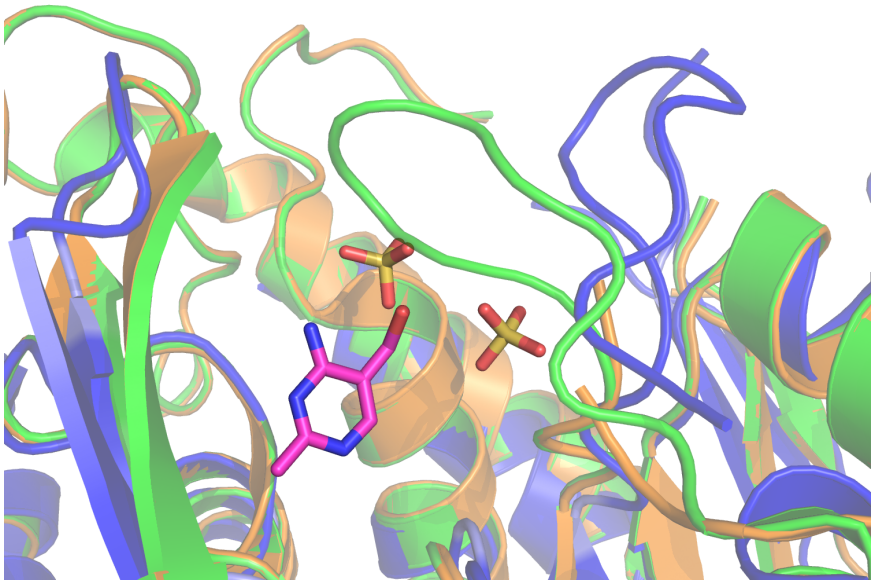


Figure 4.6. HMPP kinase/BcmD active site binding flap movement upon substrate binding. Purple represents apo-BcmD and is superimposed on the structure of *St*HMPP kinase with HMP and sulfate bound (green). The binding flap in apo-*St*HMPP kinase is disordered (orange).

REFERENCES

1. 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.
2. Drautz, H., Messerer, W., Zahner, H., Breidingmack, S., and Zeeck, A. (1987) Metabolic Products of Microorganisms .239. Bacimethrin Isolated from Streptomyces-Albus Identification, Derivatives, Synthesis and Biological Properties, *J Antibiot 40*, 1431-1439.
3. Tanaka, F., Umezawa, H., and Sumiki, Y. (1961) Bacimethrin, a New Antibiotic Produced by B Megatherium, *J Antibiot 14*, 161-&.
4. Zilles, J. L., Croal, L. R., and Downs, D. M. (2000) Action of the thiamine antagonist bacimethrin on thiamine biosynthesis, *J Bacteriol 182*, 5606-5610.
5. Reddick, J. J., Saha, S., Lee, J. M., Melnick, J. S., Perkins, J., and Begley, T. P. (2001) The mechanism of action of bacimethrin, a naturally occurring thiamin antimetabolite, *Bioorg Med Chem Lett 11*, 2245-2248.
6. Estramareix, B., and David, S. (1990) Conversion of 5-Aminoimidazole Ribotide to the Pyrimidine of Thiamin in Enterobacteria - Study of the Pathway with Specifically Labeled Samples of Riboside, *Biochim Biophys Acta 1035*, 154-160.
7. Petersen, L. A., and Downs, D. M. (1997) Identification and characterization of an operon in Salmonella typhimurium involved in thiamine biosynthesis, *J Bacteriol 179*, 4894-4900.

8. 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 angstrom resolution, *Structure 10*, 225-235.

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

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

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

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

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

14. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures, *Acta Crystallogr. D Biol. Crystallogr. 67*, 355-367.

15. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive

Python-based system for macromolecular structure solution, *Acta Crystallogr D Biol Crystallogr 66*, 213-221.

16. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Cryst. D 66*, 12-21.

CHAPTER 5

STRUCTURAL CHARACTERIZATION OF SsAziG, A *STREPTOMYCES SAHACHIROI* THIOESTERASE INVOLVED IN AZINOMYCIN B BIOSYNTHESIS

Section 5.1. Introduction

Azinomycin B (Figure 5.1A) is a complex natural product with antitumor activity isolated from *Streptomyces sahachiroi* and *Steptomyces griseofuscus*.[1, 2] It is of interest to biochemists because of its unique structure, antitumor potential, and novel DNA crosslinking mechanism. Azinomycin B and its analog, azinomycin A (Figure 5.1A), contain unusual and densely assembled functional groups. Along with a naphthoate ring, epoxide, and 2-amino-1, 3-dicarbonyl groups, the azinomycins also contain an unprecedented aziridino[1,2a]pyrrolidine ring. This unique architecture allows azinoymcin B to bind in the major groove of DNA and covalently crosslink to the N7 position of purine bases by electrophilic attack at its C10 and C21 positions.[3] Using this mechanism, azinomycin B binds in the major groove and creates DNA interstrand crosslinks with apparent sequence selectivity.[4-7] Initial clinical studies on mice with leukemia show azinomycin B is effective at nanomolar concentrations and treatment with azinomycin B increases lifespan as compared to other DNA interstrand crosslinking drugs.[2] These preliminary findings have led to drug discovery efforts to develop and evaluate azinomycin.[8-10]

The gene cluster responsible for azinomycin B biosynthesis in *S. sahachiroi* was recently identified.[11] Sequence analysis revealed that the gene annotated as *aziB* likely encodes a polyketide synthase (PKS). PKS megaenzymes generate polyketide

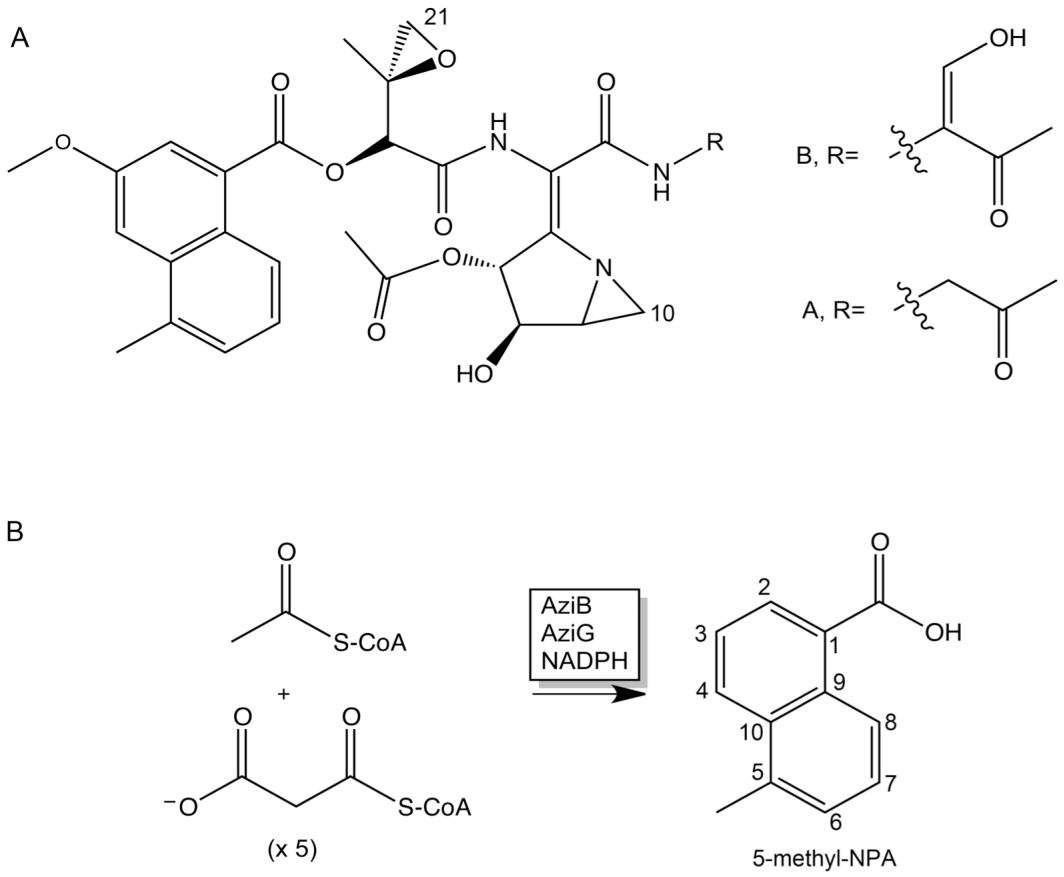


Figure 5.1. A) Structure of the azinomycins B) The 5-methyl-NPA moiety of azinomycin B is synthesized by AziB and AziG, by ligation and condensation of five malonyl-CoAs and one acetyl-CoA

chains from acetyl-CoA and malonyl-CoA precursors. These polyketide chains may then be modified and further condensed, either by the PKS itself or other partner enzymes. Upon identification of the *S. sahachiroi azi* gene cluster, it was hypothesized the protein encoded by the *aziB* gene, SsAziB, is responsible for production of the naphthoic acid (NPA) moiety. SsAziB is a 187 kDa protein and the only typical PKS gene in the *azi* gene cluster. Sequence analysis reveals SsAziB contains ketosynthase, acyltransferase, dehydratase, ketoreductase, and acyl carrier

protein domains. It was initially predicted that AziB iteratively condenses one acetyl-CoA and five malonyl CoA molecules to form a hexaketide. Subsequent intramolecular aldol condensations and dehydrations between C1/C9 and C5/C10 form the 5-methyl-NPA product later incorporated into azinomycin B (Figure 5.1B).

Initial studies reported that SsAziB is responsible and sufficient for the biosynthesis of 5-methyl-NPA.[11, 12] More recent findings from the Watanabe group conflict with these reports. Only 2-methylbenzoic acid forms when SsAziB is incubated with acetyl-CoA, malonyl-CoA, and NADPH (Watanabe Lab- *manuscript in preparation*).[13] PKSs frequently contain thioesterase domains[14] or have closely associate thioesterase partner proteins that catalyze the hydrolysis of a thioester metabolite in order to release an organic acid or CoA from intermediates in a stalled biosynthesis pathway. When SsAziG, a thioesterase found in the *S. sahachiroi azi* gene cluster, is added to the reaction mixture, the mass of the reaction product is consistent with the production of 5-methyl-NPA. The synthesis of a standard to confirm this product is underway. To aid in functional characterization, we solved the X-ray crystal structure of SsAziG. Further details about potential substrate binding interactions were revealed by the crystal structure of SsAziG complexed with the product analog, 3-hydroxy-2-NPA, and by structural comparison to known thioesterases. These studies support the role of SsAziG in 5-methyl-NPA biosynthesis and lead to a proposed mechanistic and functional role for SsAziG in the azinomycin B pathway.

Section 5.2. Materials and Methods

SsAziG Cloning, Overexpression, and Purification. The gene encoding SsAziG, *aziG,* was cloned into a pET-24a vector (Novagen) from genomic DNA from an azinomycin B producing strain of *S. sahachiroi* using standard DNA manipulation techniques.[15] The DNA was then sequenced to verify plasmid accuracy. The plasmid was transformed into chemically competent *Escherichia coli* BL21(DE3) cells. The transformed cells were plated on selective kanamycin (30 μg/mL) cotaining agar and incubated at 37 °C overnight. For each 1.5 L volume to be grown, a single colony was selected and transferred to 10 mL sterile LB Lennox media supplemented with 30 μg/mL of kanamycin and shaken overnight at 37 °C. After approximately 18 h, 10 mL of overnight starter culture was added directly to 1.5 L volumes of sterile LB. The flasks were shaken at 200 RPM and 37 °C until the media reached an OD600 of approximately 0.6. Upon reaching this OD600, the incubator temperature was turned down to 15 °C. After the incubator reached the induction temperature of 15 °C and the media OD600 was approximately 0.8, solid isopropyl-β-D-1-thiogalactopyranoside (IPTG) was added to reach a final concentration of 0.5 mM IPTG. After induction of protein expression by addition of IPTG, the cells were incubated with shaking for approximately 18 h before harvesting. The cells were harvested by centrifugation at 2000*g* for 15 min and stored in 3L aliquots frozen at -20 °C until purification.

To purify the SsAziG protein, the 3 L cell pellet was thawed and resuspended in 45 mL of lysis buffer (50 mM Tris-HCl, 300 mM NaCl, and 20 mM imidazole at pH 7.5). The resuspended cells were lysed on ice with three rounds of sonication for 3

min, with 10 min off after each round. The lysed cell suspension was centrifuged at 40,000*g* for 30 min at 4 °C. The cleared lysate was decanted and loaded onto a 2 mL NiNTA column (Qiagen) preequilibrated with lysis buffer. Following loading of the cleared lysate, the column was washed with 50 mL of lysis buffer at a flow rate of 1 ml/min to remove non-specifically bound contaminants.

To elute SsAziG, 20 mL of elution buffer (50 mM Tris-HCl, 300 mM NaCl, and 250 mM imidazole at a final pH of 7.5) was loaded on to the column and the volume containing SsAziG was collected. SsAziG purity was assessed by SDS-PAGE gel electrophoresis. The sample was buffer exchanged into 10 mM Tris-HCl, 50 mM NaCl at a pH of 7.5 by overnight dialysis and concentrated to 10 mg/ml using a centrifugal filter device with a molecular weight cut off of 3,000 Da. The purified protein was aliquoted and flash frozen into liquid nitrogen for storage at -80 °C. This purification process yielded approximately 25 mg of protein per L of purified cell culture.

Crystallization of SsAziG. The hanging drop vapor diffusion method was used for crystallization. In this method, the frozen protein aliquots were thawed on ice and equal volumes of protein and reservoir solution were mixed and equilibrated at 22 °C against a total volume of 500 μL well solution. Using the commercially available sparse matrix Wizard screens (Emerald Biosystems), three crystallization conditions were identified. Crystals were observed in Wizard I condition 13 (1.32 M ammonium sulfate and 0.1 M cacodylate pH 6.5), Wizard II condition 39 (0.1 M CAPS pH 10.5, 0.2 M NaCl, and 20% PEG 8,000), and Wizard III condition 33 (1.6 M magnesium sulfate and 0.1 M MES pH 6.5). Optimized crystals grew in 1.25 M ammonium

sulfate and 0.1 M cacodylate pH 6.0, 1.6 M magnesium sulfate and 0.1 M MES pH 5.5, or 0.1 M CAPS pH 10.8, 0.2 M NaCl, and 14% PEG 8000. In all three conditions, crystals with similar morphologies appeared overnight. The crystals were square bipyramidal, 60 μm long, and about 40 μm in width at the center. The crystals were cryprotected for data collection in a solution composed of their mother liquor supplemented with 15% (v/v) glycerol before flash freezing in liquid nitrogen. For soaking experiments, the crystals were soaked with 10 mM 3-hydroxy-2-NPA for 5 min before cryoprotection and flash freezing.

X-ray Data Collection and Processing. X-ray diffraction experiments were conducted at CHESS (Cornell High Energy Synchrotron Source) beamline F1 using a Q270 detector (Area Detector Systems Corp.) at a distance of 304.7 mm and wavelength of 0.91830 Å. The oscillation method was used with 1º rotation per frame for 90º with cryocooling and 10 s exposures. Crystals grown in all three conditions were tested for diffraction quality. The data sets reported are for crystals grown in the ammonium sulfate containing condition. Additional diffraction experiments on the ligand soaked crystals were conducted at APS NE-CAT beamline 24-ID-E (Argonne National Laboratory) using a Quantum 315 detector (Area Detector Systems Corp.) at a distance of 325 mm and wavelength of 0.97919 Å. The oscillation method was also used, with 1º rotation per frame for 70º with cryocooling and 1 s exposure time. HKL2000 was used for indexing, integration, and data scaling.[16] Complete data collection and refinement statistics are listed in Table 1.

Table 5.1. AziG data collection and refinement statistics

	AziG	AziG/3-hydroxy-2-NPA
beamline	CHESS F1	NE-CAT ID-E
resolution (Å)	2.15	2.4
wavelength (Å)	0.9183	0.97919
space group	$I4_122$	$I4_122$
a (Å)	68.5	68.2
c (Å)	139.6	139.7
Matthews coefficient	2.73	2.71
% solvent	54.9	54.6
mol/asu	1	1
Measured reflections	62888	36636
Unique reflections	9373	6783
Average I/σ	37.3(5.8)	22.2(3.6)
Redundancy	6.7(5.7)	5.4(5.5)
Completeness (%)	99.5(100)	99.6(100)
R_{merge} (%)[b]	9.1(38.6)	9.4(35.6)
No. of protein atoms	920	927
No. of ligand atoms	0	14
No. of water atoms	34	28
Reflections in working set	8883	6448
Reflections in test set	446	323
R-factor (%)	22.2	20.2
R_{free} (%)	26.6	25.0
rms deviation from ideals		
bonds (Å)	0.007	0.008
angles (deg)	1.17	1.101
average *B* factor for protein (Å^3)	45.2	26.4
Ramachandram plot		
most favored (%)	95.1	96.7
allowed (%)	4.9	3.3
disallowed (%)	0	0

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

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

[c]$R_{work} = \Sigma_{hkl} | |F_{obs}| - k |F_{cal}| | / \Sigma_{hkl} |F_{obs}|$ where F_{obs} and F_{cal} are observed and calculated structure factors, respectively, calculated over all reflections used in the refinement. R_{free}, is similar to R_{work} but calculated over a subset of reflections (5%) excluded from all stages of refinement.

Structure Determination, Model Building, and Refinement. The SsAziG crystal belongs to space group $I4_122$ and contains one promoter per asymmetric unit. It has a Matthew's coefficient of 2.73 $Å^3$/Da with corresponding solvent content of 54.9%.[17] The structure of SsAziG was determined using the molecular replacement method, with *Mycobacterium tuberculosis* thioesterase (PDB ID: 3S4K) as the model. The program Chainsaw was used to create a search model from 3S4K.[18] Molecular replacement using this search model was conducted with MolRep.[19] Iterative rounds of model building were conducting with COOT,[20] with refinement using RefMac5[21] in initial rounds, and PHENIX.refine[22] in final rounds. The quality of the final structure was assessed using MolProbity.[23] The final structure refinement was completed with an R-factor value of 22.2% and R_{free} of 26.6%. This structure was used as the model for refinement of the SsAziG/3-hydroxy-2-NPA structure. PHENIX.elbow was used for ligand generation before placement into $F_o - F_c$ electron density. Model improvement of the liganded structure was performed as described above. Structure refinement for SsAziG/3-hydroxy-2-NPA converged with an R-factor of 20.2% and R_{free} of 25.0%. Complete refinement statistics for both structures are listed in Table 1.

Section 5.3. Results

Overall Quality of SsAziG Model. The structure of SsAziG was solved to 2.15 Å in the $I4_122$ space group, with one protein chain in the asymmetric unit. The final model includes residues 11 to 134 and is approximately 40 Å long, and 20 Å by 25 Å wide. There are 34 ordered water molecules in the final model. The R-factor and R_{free} of the apo structure are 22.2% and 26.6%, respectively. The SsAziG/3-hydroxy-2-NPA crystal diffracted to 2.4 Å and also belonged to $I4_122$ space group with one

protein chain in the asymmetric unit. The complexed structure has 28 water molecules in the final structure and 14 ligand atoms. Refinement converged with a R-factor of 20.2% and R_{free} of 25.0% for the SsAziG/3-hydroxy-2-NPA model.

SsAziG Tertiary Structure. The SsAziG protomer displays the characteristic hotdog fold in which an antiparallel β-sheet surrounds a long α-helix on three sides (Figure 5.2). The SsAziG antiparallel β-sheet, with strand order β1↓β2↑β4↓β5↑β6↓β3↑, twists around the 22 residues of α-helix 3. Three additional α-helices 1, 2, and 4, are short secondary structural elements of only three to five amino acids found within the long loop regions of SsAziG.

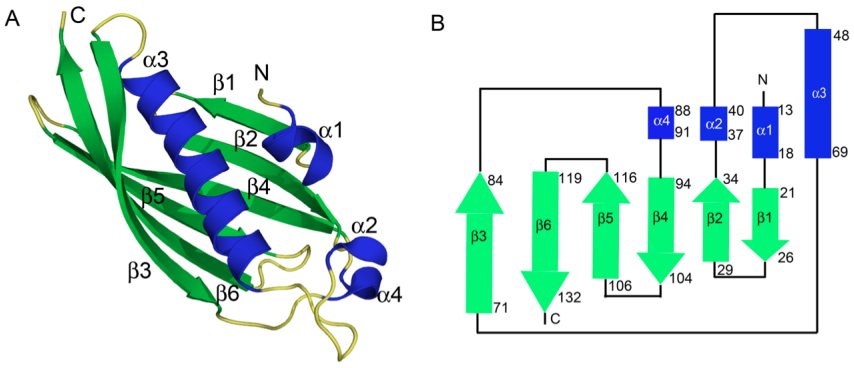


Figure 5.2. Monomeric structure of AziG. A) Ribbon drawing depicts the hotdog fold that AziG adopts. β-strands are in green and α-helices in blue. B) A topology diagram of the AziG fold, following the same coloring scheme as in A.

SsAziG Quaternary Structure. The SsAziG tetramer is formed via crystallographic symmetry (Figure 5.3). The tetramer is composed of four equivalent chains, with a total of four active sites. Each active site requires residues contributed from two chains. The largest interface area occurs between chains A and B. Both

chains contribute 1,180 Å^2 of surface area to the interaction, with 26 interface residues from each. Within this area there are 10 hydrogen bonds and 135 non-bonding interactions. Chains C and D form an interface area equivalent to the AB interface. The AC interface and BD interface areas are equivalent. They are 447 Å^2 in area and include six hydrogen bonds and 51 hydrophobic interactions, with 10 interface residues from each chain participating. The AD and CB interfaces are 153 Å^2 in area and three interface residues from each chain form a total of seven hydrophobic interactions.

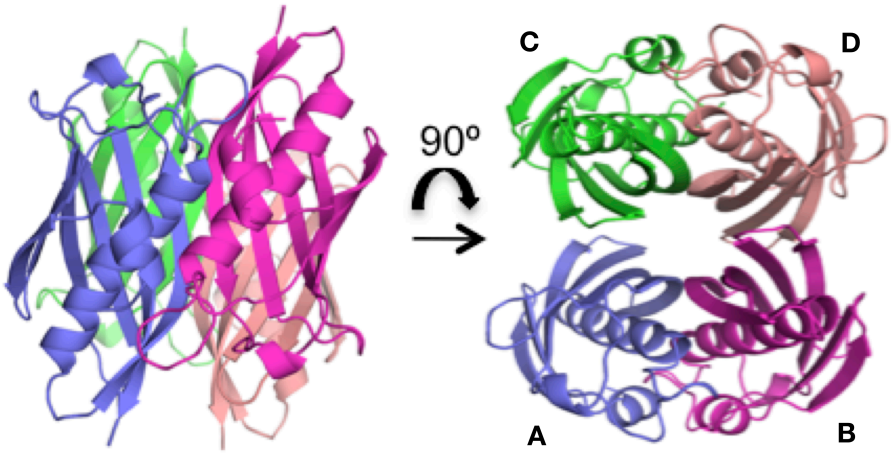


Figure 5.3. The biological tetramer formed by AziG. The β-sheet of each monomer faces the core and the main α-helices face outward.

Substrate Binding Site. The product analog, 3-hydroxy-2-NPA, is modeled into the final F_o-F_c electron density (Figure 5.4A). The bound 3-hydroxy-2-NPA molecule reveals the active site locations at the AB and CD interfaces. There are two

binding sites in each interface (Figure 5.4B), for four total active sites per molecule. In the SsAziG/3-hydroxy-2-NPA structure, the 3-hydroxy group forms a hydrogen bonding interaction with His48 and the NPA carboxy group forms a hydrogen bond with the amine group of Gly49, positioned at the positive dipole end of the long α-helix 3 (Figure 5.4A). 3-hydroxy-2-NPA is found in a binding pocket formed at the AB or CD interface (Figure 5.4C). This cavity is approximately 7 Å in diameter and 20 Å deep.

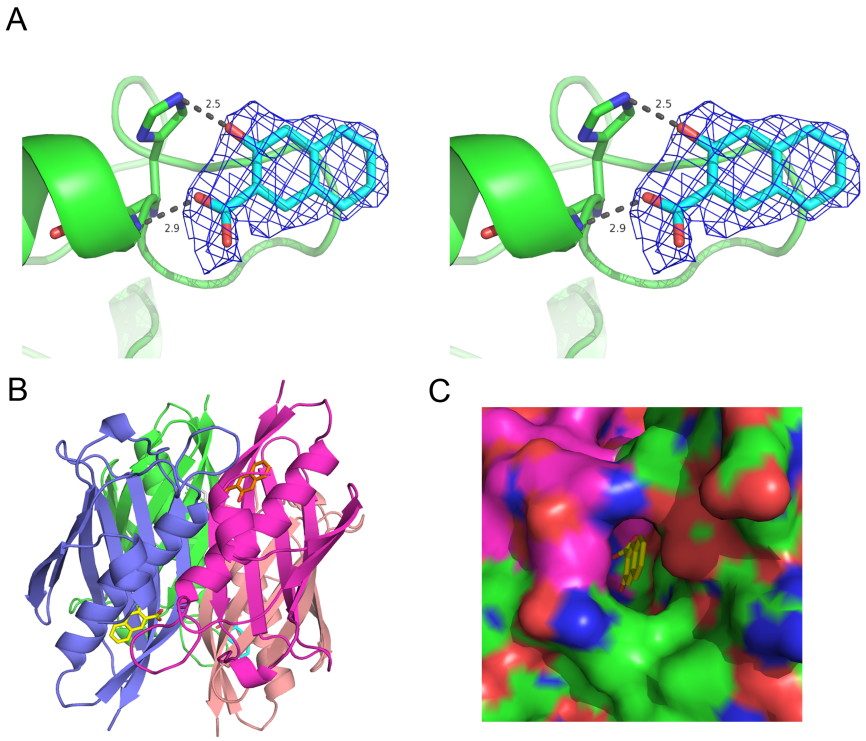


Figure 5.4. AziG active site. A) Stereoview shows electron density for 3-hydroxy-2-NPA and its interactions with the N-terminus of the central α-helix. B) Each tetramer has four active site. The location of two of the 3-hydroxy-2-NPA (yellow and orange) binding sites are drawn. C) The four equivalent 3-hydroxy-2-NPA binding pockets are created by the interface between chains A and B or C and D.

5-methyl-NPA Modeling. Using manual modeling in COOT, the product of the SsAziB/SsAziG reaction, 5-methyl-NPA, was placed into the active site based on the location of 3-hydroxy-2-NPA. The orientation and placement of the NPA ring remains nearly the same. The model with 5-methyl-NPA with the most favorable binding interactions and no clashes was chosen (Figure 5.5). In this model, the 5-methyl-NPA carboxy group interacts with the amide group of Gly49, Ser61, and Gln42. 5-methyl-NPA is within 3.0 Å of the expected catalytic residue, Glu57.

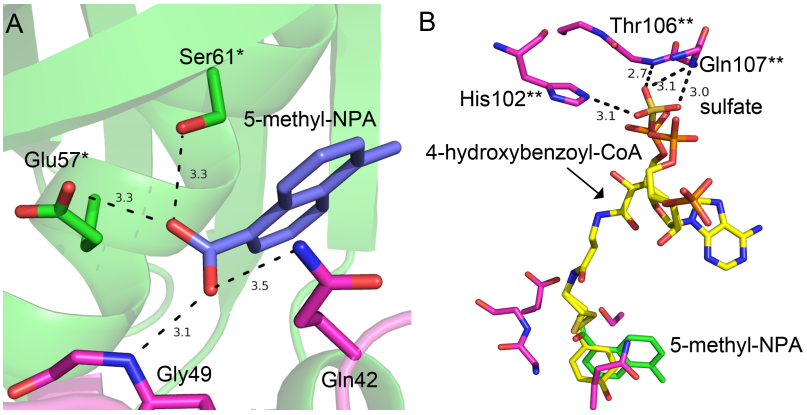


Figure 5.5. A) Model for 5-methyl-NPA binding and probable interactions. B) Superposition of AziG and 5-methyl-NPA with *Anthrobacter* thioesterase with 4-hydroxybenzoyl-CoA (PDB 1Q4U) reveals potential CoA binding mode.

Section 5.4. Discussion

Structural Comparison. SsAziG forms a long central α-helix surrounded by a six-stranded anti-parallel β-sheet and three small α-helices located within the connecting loop regions. This tertiary arrangement is characterized as a hotdog fold, one of the two main structural families to which thioesterases belong.[24] Thioesterases found in the hotdog fold superfamily are minimally a dimer, as residues from both subunits are necessary for catalysis. While the quaternary structure varies among

hotdog fold thioesterases, SsAziG is a tetramer and most closely related to other thioesterases that also form tetramers. A DALI structural search alignment against the PDB reveals proteins with tertiary structure closely related to SsAziG (Table 2). SsAziG is most highly structurally related to a putative esterase, Rv1847 from *Mycobacterium tuberculosis* (PDB ID: 3S4K), as revealed by a rmsd value of 0.9 between the two structures. SsAziG and Rv1847 share 50% sequence identity and both form tetramers. The DALI search also reveals two structurally related thioesterases whose functions have been extensively studied. Alignment with PaaI, a phenylacetic acid degradation protein from *Thermus thermophilus* (PDB ID: 1WLV), and 4-hydroxybenzoyl-CoA thioesterase, from *Anthrobacter* sp. SU (PDB ID: 1Q4S), results in RMSD values of 1.1 and 1.3, respectively. These thioesterases both share 31% sequence identity with SsAziG. Despite their diverse physiological functions, the details of PaaI and 4-hydroxybenzoyl-CoA thioesterase mechanisms are informative in investigating the function and mechanism of SsAziG. Finally, SsAziG shares 24% sequence identity with the human thioesterase superfamily member 2 (Them2, PDB ID: 2H4U) and their structure alignment has a rmsd value of 1.7. Functional studies suggest Them2 is a fatty acyl-CoA thioesterase, important for fat metabolism. While serving different physiological functions, the structural similarity of Them2 to SsAziG demonstrates the diversity of organisms that utilize hotdog fold thioesterases for various metabolic functions.

Table 5.2. Summary of proteins structurally similar to SsAziG, as analyzed by DALI

Protein	PDB ID	Z score	rmsd	% identical	no. of aligned residues
M. tuberculosis putative esterase, Rv1847	3S4K	22.8	0.9	50	121
T. thermophilus PaaI, Phenylacetic acid degradation protein	1WLV	19.8	1.1	31	116
Anthrobacter 4-hydroxybenzoyl-CoA thioesterase	1Q4S	18.5	1.3	31	120
H. sapiens thioesterase, them2	2H4U	17.9	1.7	24	117

Active Site. SsAziG was co-crystallized with the product analog, 3-hydroxy-2-NPA (Figure 5.4). The α-helix dipole at the N-terminus of SsAziG α-helix 3 creates a positively charged environment at the bottom of the binding cavity. This results in an electrostatically favorable binding environment for the negatively charged carboxy group of 5-methyl-NPA. Based on the orientation of 3-hydroxy-2-NPA, the expected product, 5-methyl-NPA, was modeled into the active site binding cavity (Figure 5.5A). The location and orientation of 5-methyl-NPA is supported by alignment with 4-hydroxybenzoyl-CoA thioesterase co-crystallized with 4-hydroxybenzyl CoA (PDB ID 1Q4U) (Figure 5.5B). These structurally similar thioesterase enzymes have evolved in different organisms and function in specialized cellular roles, but their similarities reveal a fairly conserved mechanism. As a result, the mechanism of SsAziG can be inferred based on structural comparison to these enzymes. The residues important for catalysis in 4-hydroxybenzoyl-CoA thioesterase (Glu73*, Gly65, Gln58, Thr77*) are mostly conserved in the SsAziG active site. The corresponding residues (Glu57*, Gly49, Gln42, and Ser61*) form the predicted

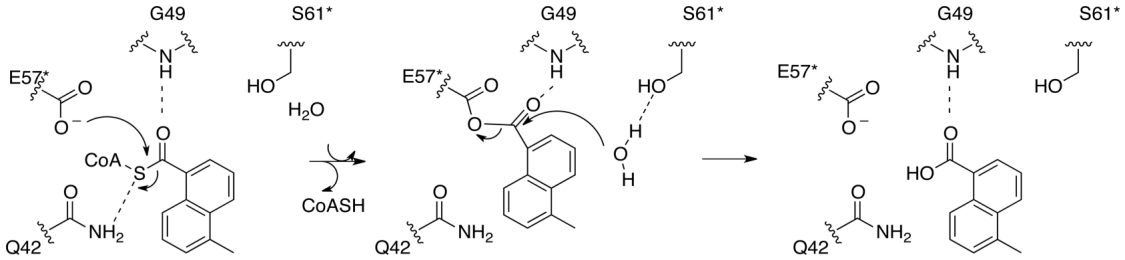


Figure 5.6. Proposed mechanism for 5-methyl-NPA-CoA hydrolysis by AziG

SsAziG catalytic machinery. Glu57* (from chain B) acts as the nucleophile, while the amide groups from Gly49* and Gln42 aid in transition state stabilization via hydrogen bond formation with the thioester moiety (Figure 5.6). Whereas Thr77* orients a water molecule for nucleophilic attack in 4-hydroxybenzyl-CoA thioesterase, SsAziG Ser61* is positioned to act similarly.

Substrate Binding. In addition to the catalytic residues, several other residues are predicted to aid in 5-methyl-NPA-CoA substrate binding. Many residues important for CoA binding in 4-hydroxbenzyl-CoA thioesterase are conserved in SsAziG. In the structure of SsAziG, which was crystallized in ammonium sulfate, a sulfate ion occupies the predicted CoA phosphate-binding site. Additional CoA binding interactions come from SsAziG His102** (from chain C); positioned to interact with the negatively charged CoA nucleotide phosphate group (Figure 5.5B). The amide backbone groups of Thr106** and Glu107** also create favorable CoA binding interactions. Two arginine residues aid in binding the phosphate group in the 4-hydroxybenzyl-CoA thioesterase structure. Comparably, SsAziG Arg133 is structurally conserved in the C-terminal loop and could also aid in stabilizing and binding the negatively charged phosphate portion of 5-methyl-NPA-CoA.

Proposed Mechanism. Through structural comparisons with known thioesterases, a probable mechanism is hypothesized (Figure 5.6). Throughout the reaction, Gly49 forms a hydrogen bond to the NPA carboxy group to anchor it during catalysis. In this model, Glu57* acts as the active site nucleophile and transfers the intermediate NPA moiety from CoA to the enzyme. Subsequently, a water molecule oriented by Ser61* for nucleophilic attack at the carbon of the carboxy-enzyme intermediate. In concert, the bond between the catalytic Glu57* and the substrate is broken when the hydroxyl group is added. The thioesterase reaction is then complete and the proposed product, 5-methyl-NPA can then be released.

Functional Implications. The structure of SsAziG crystallized with 3-hydroxy-2-NPA supports its function as the thioesterase involved in synthesis of the azinomycin B NPA moiety. PKSs are able to catalyze the polymerization of polyketides by tethering the growing chain to a CoA arm and transferring the substrate to various catalytic domains. The final step in the biosynthesis of the polyketide product is to release it from the CoA anchor, by thioester bond cleavage. In the case of azinomycin B 5-methyl-NPA biosynthesis, AziB catalyzes the growth and condensation of 5-methyl-NPA moiety, but a thioesterase is necessary for its release. Lacking a thioesterase domain, AziB utilizes SsAziG for this final step. Previous studies that attribute AziB to complete biosynthesis of 5-methyl-NPA were conducted *in vivo* in a heterologous host. This may have led to nonspecific activity from an unidentified host thioesterase that released the NTA moiety from AziG. Alternatively, non-enzymatic hydrolysis is also a possible explaination for this observation.

Efforts to capture an AziB-AziG complex using crosslinking studies, size exclusion chromatography, and crystallography were so far unsuccessful. The interaction between these two enzymes may be transient or require additional components. Nevertheless, the role of SsAziG in 5-methyl-NPA biosynthesis is supported by both these structural studies and recent biochemical evidence. Future efforts to crystallize and characterize AziB will shed more light on its mechanism for polyketide chain growth, shuttling of intermediates between different domains, and interactions with AziG.

REFERENCES

1. Nagaoka, K., Matsumoto, M., Oono, J., Yokoi, K., Ishizeki, S., and Nakashima, T. (1986) Azinomycin-a and Azinomycin-B, New Antitumor Antibiotics .1. Producing Organism, Fermentation, Isolation, and Characterization, *J Antibiot 39*, 1527-1532.
2. Ishizeki, S., Ohtsuka, M., Irinoda, K., Kukita, K. I., Nagaoka, K., and Nakashima, T. (1987) Azinomycin-a and Azinomycin-B, New Antitumor Antibiotics .3. Antitumor-Activity, *J Antibiot 40*, 60-65.
3. Foulke-Abel, J., Agbo, H., Zhang, H., Mori, S., and Watanabe, C. M. (2011) Mode of action and biosynthesis of the azabicycle-containing natural products azinomycin and ficellomycin, *Nat Prod Rep 28*, 693-704.
4. Armstrong, R. W., Salvati, M. E., and Nguyen, M. (1992) Novel Interstrand Cross-Links Induced by the Antitumor Antibiotic Carzinophilin Azinomycin-B, *Journal of the American Chemical Society 114*, 3144-3145.
5. Coleman, R. S., Perez, R. J., Burk, C. H., and Navarro, A. (2002) Studies on the mechanism of action of azinomycin B: definition of regioselectivity and sequence selectivity of DNA cross-link formation and clarification of the role of the naphthoate, *J Am Chem Soc 124*, 13008-13017.
6. David-Cordonnier, M. H., Casely-Hayford, M., Kouach, M., Briand, G., Patterson, L. H., Bailly, C., and Searcey, M. (2006) Stereoselectivity, sequence specificity and mechanism of action of the azinomycin epoxide, *Chembiochem 7*, 1658-1661.

7. Coleman, R. S., Woodward, R. L., Hayes, A. M., Crane, E. A., Artese, A., Ortuso, F., and Alcaro, S. (2007) Dependence of DNA sequence selectivity and cell cytotoxicity on azinomycin A and B epoxyamide stereochemistry, *Org Lett 9*, 1891-1894.

8. Hartley, J. A., Hazrati, A., Kelland, L. R., Khanim, R., Shipman, M., Suzenet, F., and Walker, L. F. (2000) A Synthetic Azinomycin Analogue with Demonstrated DNA Cross-Linking Activity: Insights into the Mechanism of Action of this Class of Antitumor Agent The authors gratefully acknowledge the financial support provided by the CRC and the EPSRC. We are indebted to the EPSRC National Mass Spectrometry Centre for performing mass spectral measurements, and the EPSRC Chemical Database Service at Daresbury.1, *Angew Chem Int Ed Engl 39*, 3467-3470.

9. Casely-Hayford, M. A., Pors, K., James, C. H., Patterson, L. H., Hartley, J. A., and Searcey, M. (2005) Design and synthesis of a DNA-crosslinking azinomycin analogue, *Org Biomol Chem 3*, 3585-3589.

10. Casely-Hayford, M. A., Pors, K., Patterson, L. H., Gerner, C., Neidle, S., and Searcey, M. (2005) Truncated azinomycin analogues intercalate into DNA, *Bioorg Med Chem Lett 15*, 653-656.

11. Zhao, Q., He, Q., Ding, W., Tang, M., Kang, Q., Yu, Y., Deng, W., Zhang, Q., Fang, J., Tang, G., and Liu, W. (2008) Characterization of the azinomycin B biosynthetic gene cluster revealing a different iterative type I polyketide synthase for naphthoate biosynthesis, *Chem Biol 15*, 693-705.

12. Ding, W., Deng, W., Tang, M., Zhang, Q., Tang, G., Bi, Y., and Liu, W. (2010) Biosynthesis of 3-methoxy-5-methyl naphthoic acid and its incorporation into the antitumor antibiotic azinomycin B, *Mol Biosyst 6*, 1071-1081.

13. Walsh, C. T. (2008) The chemical versatility of natural-product assembly lines, *Acc Chem Res 41*, 4-10.

14. Keatinge-Clay, A. T. (2012) The structures of type I polyketide synthases, *Nat Prod Rep 29*, 1050-1073.

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

16. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode, In *Methods Enzymol.* (Carter, C. W. J., and Sweet, R. M., Eds.), pp 307-326, Academic Press, New York.

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

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

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

20. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot, *Acta Crystallogr D Biol Crystallogr 66*, 486-501.

21. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures, *Acta Crystallogr. D Biol. Crystallogr. 67*, 355-367.

22. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr D Biol Crystallogr 66*, 213-221.

23. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Cryst. D 66*, 12-21.

24. Cantu, D. C., Chen, Y., and Reilly, P. J. (2010) Thioesterases: a new perspective based on their primary and tertiary structures, *Protein Sci 19*, 1281-1295.

CHAPTER 6

SUMMARY

Protein X-ray crystallography is a powerful tool for determining the structural details of macromolecules at atomic resolution. Obtaining an enzyme crystal structure can provide a multitude of information. From a mechanistic enzymology prospective for example, the structure is valuable for identifying the enzyme active site. Structures determined in the presence of substrates, products, or inhibitors are often critical in obtaining a mechanistic understanding of the enzymatic reaction. Once the active site residues are identified, mutant kinetic studies are frequently conducted to confirm the role of individual residues in catalysis. These biochemical studies compliment the structural findings and can lead to a proposed mechanism by which the enzyme catalyses its reaction. This structure-function approach is especially useful when there is more than one plausible mechanism, or the mechanism is completely unknown. Other potential applications for X-ray crystallographic studies are immense. In addition to mechanistic studies, X-ray crystallography is frequently used to understand protein-protein interactions, discover conformational changes, and help explain enzyme function.[1]

The preceding studies used X-ray crystallography to study enzymes that catalyze reactions in bacterial metabolic pathways. The metabolites of interest are either necessary for cell viability, as introduced in Chapters Two and Three on nucleotide and thiamin metabolism, or are used by the cell for some biochemical purpose, as discussed in Chapters Four and Five on bacimethrin and azinomycin B synthesis. Understanding the basic biochemistry and selectivity of the processes by

which these metabolites are formed and degraded enriches our understanding of cell biology and chemistry. Additionally, knowing the exact position of each atom in an enzyme enables exploration of mechanistic possibilities.

Novel structural details revealed the basis of substrate preference by NDT family members at the substrate 2′-position. Introduced in Chapter Two, MilB is the CMP hydrolase involved in the early steps of biosynthesis of the commercially available antifungal compound, mildiomycin.[2] Also discussed in Chapter Two is BcmB from the BCM biosynthetic pathway. BcmB is closely related in sequence, function, and structure to MilB. These two enzymes are classified in the NDT superfamily, whose members catalyze *N*-glycosidic bond cleavage of 2′-deoxynucleosides via a covalent 2-deoxyribosyl-enzyme intermediate.[3] Conservation of key active site residues suggested that members of the NDT superfamily share a common mechanism; however, the enzymes noticeably differ in their substrate preferences. Substrates vary in the type of nucleobase, the presence or absence of a 2′-hydroxyl group, and the presence or absence of a 5′-phosphate group. We determined the structures of MilB and BcmB and compared them to previously determined structures of the NDT superfamily members: NDT, PTD, and Rcl.[3-6] The comparisons reveal how these enzymes differentiate between ribosyl and deoxyribosyl nucleotides or nucleosides, and among different nucleobases. The structures of BcmB and the MilB-CMP complex reveal the presence of an active site phenylalanine near the substrate 2′-binding position. Whereas MilB and BcmB prefer substrates containing 2′-ribosyl groups and have a phenylalanine positioned in the active site, NDT family members with preference for 2′-deoxyribosyl groups have a tyrosine

residue equivalently positioned. This difference was not obvious from sequence comparisons alone. Further studies show that this phenylalanine is critical for MilB and BcmB specificity towards CMP, and mutation of this phenylalanine residue to tyrosine results in a 1000-fold reversal of substrate specificity from its 2′-ribosyl containing substrate, CMP, to the 2′-deoxy containing analog, dCMP. This novel finding is especially important in understanding the substrate specificity of NDT family members. Commercially, NDT is exploited for enzymatic generation of nucleosides and this finding provides a potential basis for engineering NDT specificity at the 2′-position.

Thiamin metabolism is another area of biochemistry that presents interesting biochemical and physiological questions. The recent identification of the *C. botulinum* BCM biosynthesis pathway added to the interest in this area. The genes encoding the enzymes responsible for biosynthesis of BCM, a thiamin antimetabolite, are clustered with a thiaminase I. The physiological role of thiaminase I is a long standing mystery in thiamin biochemistry. Thiaminase Is degrade thiamin into its pyrimidine and thiazole derived moieties. The 2.2 Å resolution X-ray crystal structure of the catalytically inactive Cb-thiaminase I C143S mutant cocrystallized with thiamin provides the first structure of a thiaminase I with bound substrate. The C143S/thiamin complex reveals the atomic level details of thiamin orientation upon binding to Cb-thiaminase I. This structure was critical for the identification of active site residues involved in substrate binding and catalysis. The specific roles of each active site residue was elucidated through site directed mutagenesis and kinetic studies, leading to a detailed mechanism for Cb-thiaminase I. In this mechanism, a

catalytic cysteine is deprotonated by a nearby glutamate. The cysteine initiates thiaminase activity by nucleophilic attack at the thiamin C6′. The C143S/thiamin structure also brought attention to the role of the active site Asp94 and Tyr80 in substrate binding. Kinetic assays established the role of these two active site amino acids in the expulsion of the thiazole moiety. Cb-thiaminase I is also a rare example of an enzyme belonging to the periplasmic binding protein (PBP) superfamily. Structural analyses reveal similarities and differences between Cb-thiaminase I and PBPs. Thiaminase I lacks the characteristic PBP "open" and "closed" conformational change upon substrate binding. The structure of Cb-thiaminase I was also compared to the biochemically similar, but structurally and functionally unique, thiaminase II. This comparison revealed the two classes of thiaminases utilize distinct thiamin binding modes and are likely not closely related. These structural and mechanistic studies support and expand upon the previous mechanistic and evolutionary understanding of thiaminase I.

Natural product biosynthesis pathways have evolved to create intricate chemical structures that intrigue scientists due to their novel chemistry and biochemical mechanisms. While BCM is a fairly chemically simple, it is a physiologically puzzling natural product that acts as a thiamin antimetabolite. It has been isolated from several prokaryotic organisms[7, 8] and the BCM biosynthetic gene cluster in *C. botulinum* was recently identified. First, BcmA adds a hydroxymethyl group to the CMP cytosine subsequent *N*-glycosidic bond cleavage by BcmB then frees hydroxymethylcytosine. In Chapter Two the structure of the hmCMP hydrolase, BcmB, is discussed in context of its relationship to other *N*-glycosidase family

members. The structure of BcmB supports the proposed mechanism for hydrolysis and the substrate specificity for CMP or hmCMP.

BcmD is another enzyme of interest in the BCM biosynthetic pathway. In Chapter Four, the structure of BcmD is discussed and compared to similar enzymes. In order to act as a substrate for the thiamin metabolic pathway, BCM must be diphosphorylated to BCM-PP. BcmD is the HMPP kinase responsible for two sequential phosphorylations of BCM to create BCM-PP. This reaction is analogous to the diphosphorylation of HMP before ligation with THZ-P to create ThMP (Figure 1.3). Before the structure of BcmD was solved, it was not known whether it had structural features that allowed it to distinguish between, and possibly favor, BCM over HMP. Structural comparison of BcmD to known HMPP kinases did not reveal any structural differences that may favor BCM binding. Subsequent biochemical studies showed that BcmD did not favor BCM over HMP and raises further questions about the evolution and regulation of the BCM biosynthetic genes.

Azinomycin B is another interesting natural product whose biosynthesis presents novel biochemistry. The genes responsible for its biosynthesis were recently identified in *Streptomyces sahachiroi*. Azinomycin B contains densely assembled and uniquely tailored peptide and naphthoate precursors. Previous reports suggested that the *S. sahachiroi* polyketide synthase SsAziB, encoded by the *aziB* gene, was the only enzyme in the pathway necessary for production of the naphthoate moiety.[9] Recent findings suggest that SsAziB acts in concert with a thioesterase, SsAziG, to produce the naphthoate (Watanabe Lab- *manuscript in preparation*). SsAziB polymerizes and condenses acetyl-CoA and malonyl-CoA precursor molecules by shuttling the growing

polyketide to its various catalytic domains via a covalent CoA linker. Then SsAziG hydrolyzes the 5-methyl-NPA-CoA thioester bond to release the product for later incorporation into azinomycin B. Crystallographic studies of SsAziG with a bound product analog, 3-hydroxy-NPA, allowed for modeling of the proposed product, 5-methyl-NPA. Structural comparison between SsAziG and related thioesterases leads to the proposed SsAziG function and mechanism. In the proposed mechanism, the azinomycin B naphthoate moiety is transferred by from the SsAziB CoA to the SsAziG active site glutamate. The NPA fragment is then released for incorporation into azinomycin B by hydrolysis of the covalent enzyme-substrate bond. This mechanism is very similar to other hotdog-fold thioesterases that hydrolyze the covalent bond between a metabolite and coenzyme A.[10] Because of the complexity and antitumor properties of the azinomycins, their biosynthesis is of continued interest to chemists and biologists. Future structural studies involving enzymes in this pathway will provide a greater understanding of their functions and mechanisms. Armed with this information, a detailed picture of azinomycin B biosynthesis will emerge, enabling synthetic chemists and biochemists to exploit its chemical properties for future studies. The role SsAziG plays in NPA synthesis presents a previously unexplored component necessary for azinomycin B biosynthesis. Its relationship to AziB adds to the understanding of the highly complex PKS enzymes that are just now beginning to be biochemically and structurally analyzed.

The work done on the bacimethrin and azinomycin B biosynthesis pathways, as well as the NDT superfamily members, are examples where X-ray crystallography provided valuable biochemical information. The structures elucidated in the preceding

studies and supporting biochemical characterization revealed interesting functional and mechanistic possibilities. Each of the unique structures discussed- MilB, BcmB, Cb-thiaminase I, BcmD and SsAziG- provided an opportunity to better understand how these biologically relevant enzymes catalyze complex chemical reactions.

REFERENCES

1. Campbell, I. D. (2002) Timeline: the march of structural biology, *Nat Rev Mol Cell Biol 3*, 377-381.
2. Li, L., Xu, Z., Xu, X., Wu, J., Zhang, Y., He, X., Zabriskie, T. M., and Deng, Z. (2008) The mildiomycin biosynthesis: initial steps for sequential generation of 5-hydroxymethylcytidine 5'-monophosphate and 5-hydroxymethylcytosine in Streptoverticillium rimofaciens ZJU5119, *Chembiochem 9*, 1286-1294.
3. Short, S. A., Armstrong, S. R., Ealick, S. E., and Porter, D. J. (1996) Active site amino acids that participate in the catalytic mechanism of nucleoside 2'-deoxyribosyltransferase, *J Biol Chem 271*, 4978-4987.
4. Anand, R., Kaminski, P. A., and Ealick, S. E. (2004) Structures of purine 2'-deoxyribosyltransferase, substrate complexes, and the ribosylated enzyme intermediate at 2.0 A resolution, *Biochemistry 43*, 2384-2393.
5. Doddapaneni, K., Mahler, B., Pavlovicz, R., Haushalter, A., Yuan, C., and Wu, Z. (2009) Solution structure of RCL, a novel 2'-deoxyribonucleoside 5'-monophosphate N-glycosidase, *J Mol Biol 394*, 423-434.
6. Yang, Y., Padilla, A., Zhang, C., Labesse, G., and Kaminski, P. A. (2009) Structural characterization of the mammalian deoxynucleotide N-hydrolase Rcl and its stabilizing interactions with two inhibitors, *J Mol Biol 394*, 435-447.
7. Drautz, H., Messerer, W., Zahner, H., Breidingmack, S., and Zeeck, A. (1987) Metabolic Products of Microorganisms .239. Bacimethrin Isolated from

Streptomyces-Albus Identification, Derivatives, Synthesis and Biological Properties, *J Antibiot 40*, 1431-1439.

8. Tanaka, F., Umezawa, H., and Sumiki, Y. (1961) Bacimethrin, a New Antibiotic Produced by B Megatherium, *J Antibiot 14*, 161-&.

9. Zhao, Q., He, Q., Ding, W., Tang, M., Kang, Q., Yu, Y., Deng, W., Zhang, Q., Fang, J., Tang, G., and Liu, W. (2008) Characterization of the azinomycin B biosynthetic gene cluster revealing a different iterative type I polyketide synthase for naphthoate biosynthesis, *Chem Biol 15*, 693-705.

10. Cantu, D. C., Chen, Y., and Reilly, P. J. (2010) Thioesterases: a new perspective based on their primary and tertiary structures, *Protein Sci 19*, 1281-1295.