2-1-2012

Evolution of structure and function among hotdog-fold thioesterases and HAD family phosphatases

Wang Min

Min Wang
Candidate

Chemistry and Chemical Biology
Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Debra Dunaway-Mariano, Chairperson

Patrick S Mariano

Wei Wang

Karen N Allen

EVOLUTION OF STRUCTURE AND FUNCTION AMONG HOTDOG-FOLD THIOESTERASES AND HAD FAMILY PHOSPHATASES

BY

MIN WANG

B.S., Applied Chemistry, Sichuan University, China, 2002
M.S., Chemistry, Sichuan University, China, 2005

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy

Chemistry

The University of New Mexico
Albuquerque, New Mexico

December, 2011

DEDICATION

To

My parents, Shaoming Wang and Guorong Ru

To

My wife, Li Zheng

ACKNOWLEDGMENTS

First of all, I would like to express my sincere appreciation and gratitude to my advisor Dr. Debra Dunaway-Mariano for her guidance, advice, support and encouragement. She has taught me how to think, work and write in a scientific way and help me finding the career path. Her invaluable advice and everyday encouragement make all these years enjoyable and memorable. I'm also very grateful for the high degree of confidence she has in me. She is the best teacher and scientist I met in my life. Without her I could have never gone so far. She has the respect from the bottom of my heart.

I would like to express my sincere appreciation to Dr. Patrick S. Mariano for his detailed guidance on the organic synthesis and revising my dissertation. I would like to extend my sincere appreciation to my committee members: Dr. Mary Anne Nelson and Dr. Karen Anne for their insightful suggestions and assistance on my project and my dissertation manuscript.

I would like to express special thanks to Dr. Kelly Daughtery, Dr. Rui Wu and Dr. Karen Allen at Department of Chemistry, Boston University for her dedicated work in solving the BT3352 crystal structure, and their efforts in solving the structures of *Bacteriodes fragilis* and *Porphyromonas gingivalis* fusion and truncated proteins and *Pseudomonas* thioesterase.

I would like to thank all of the past and present members of Dr. Dunaway-Mariano's research group for their help, support and friendship. They made my graduate school a happy and fulfilling experience. Many thanks to Dr. Ling Li, Dr. Liangbing Wang and Dr. Jiwen Zou for their helping me in the beginning of my work and teaching me the basis techniques.

Finally, my warmest thanks to my dearest family for their love and support, my parents, Shaoming Wang and Guorong Ru. A special thanks to my wife Li Zheng for her love and support. It was her belief in me that encouraged me to achieve my best.

EVOLUTION OF STRUCTURE AND FUNCTION AMONG HOTDOG-FOLD THIOESTERASES AND HAD FAMILY PHOSPHATASES

BY

MIN WANG

ABSTRACT OF DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

Chemistry

The University of New Mexico

Albuquerque, New Mexico

December, 2011

EVOLUTION OF STRUCTURE AND FUNCTION AMONG HOTDOG-FOLD THIOESTERASES AND HAD PHOSPHATASES

Min Wang

B.S., Applied Chemistry, Sichuan University, China, 2002

M.S., Chemistry, Sichuan University, China, 2005

Ph.D., Chemistry, University of New Mexico, 2011

ABSTRACT

My doctoral research begun with the function assignment to two domains of fusion protein BT4699, from *Bacteriodes thetaiotaomicron,* BF1314 from *Bacteriodes fragilis* and PG1653 from *Porphyromonas gingivalis*. BT4699, however, is insoluble, so its orthologue BF1314 and PG1653 were prepared and studied. The fusion proteins consist of a domain that belong to a large family of phosphatases known as Haloacid Dehalogenase (HAD) superfamily, and a domain belong to the hotdog-fold thioesterase superfamily. The genes encoding the fusion proteins located at an operon that encoding enzymes involved in the menaquinone biosynthetic pathway. So we assumed that the thioesterase domain is connected to the biosynthesis of menaquinone, which sever as a crucial role in the anaerobic bacteria respiratory chain. While the HAD domain is involved in hydrolysis of pyrophosphate produced by MenA in the lower segment of the pathway. The thioesterase domain was predicted to catalyze the hydrolysis of 1,4-dihydroxyl-2-napthoyl-CoA (DHNA-CoA) to 1,4-dihydroxyl-2-napthoate (DHNA) as the precursor of menaquinone synthesis.

To study the function of thioesterase domain, the putative substrate 1,4-dihydroxy naphthoyl-CoA was prepared by chemienzymatic strategy. The structure was confirmed by H NMR and mass spectrum. HPLC analysis indicated that DHNP-CoA is very unstable, which

could be self-hydrolyzed within 16 hrs at pH=7.5. By using a combination of techniques like steady-state kinetics, HPLC, and NMR, the function of thioesterase domain of the fusion protein were investigated. Steady-state kinetics studies indicated that PG1653 has high activity over DHNP-CoA. HPLC and H-NMR proved the formation of hydrolyzed products. The high specificity of BF1314 and PG1653 over DHNP-CoA testified the function of thioesterase domain.

While the function of HAD domain is not that as we assumed. It didn't even show any activity towards pyrophosphate. PG1653 fusion protein did not have any phosphatase activity due to the mutation of Asp to Phe, which is a key catalytic residue of HAD phosphatase. The other fusion protein BF1314 just showed very weak activities towards a few phosphate substrates. Both truncated HAD domain and thioesterase domain were also prepared to investigate the function of HAD domain. Size exclusion chromatography was used to examine the effect of native structure on the functions of both HAD domain and thioesterase domain. The results indicated that the HAD domain may serve as an oligmerization domain that help the tetramerization of the thioesterase domain as an active thioesterase tetramer. The structure supportive role of the HAD domain disabled its phosphatase activity.

The HAD domain homologue, BT3352, shares 47% identity with fusion protein BT4699. This high level of identity suggests that these two HADs are related by gene duplication. One gene copy must have fused with the thioesterase domain and took on a structural function that supports the thioesterase function while losing its catalytic function. The other copy may carry on the function of the original phosphatase. BT3352 exhibited good sugar phosphatase activities, and erythrose 4-phosphate turned out to be its physiological substrate. By comparing with its

homologue yidA from *E. coli*, which may play a house keeping role in the galactonate degradation pathway, we hypothesized that BT3352 may also play house keeping role in likewise pathway.

Bioinformatic analysis suggests that the 4 gene cluster might encode a partial β-oxidation pathway in *Pseudomonas aeruginosa*, which supported the pathway with seven other gene clusters. The target gene cluster contain two putative β-oxidation pathway enzymes, acyl-CoA dehydrogenase (PA5187), 3-hydroxy-acyl-CoA dehydrogenase (PA5188), a putative iron-containing alcohol dehydrogenase (PA5186) and a putative hotdog-fold thioesterase (PA5185). Substrate screen results indicated that PA5187 target C4-C15 fatty acyl-CoA, while PA5188 showed good activity towards C4-C10 and peak at 3-hydroxyhexanoyl-CoA. The hotdog-fold thioesterase exhibited narrow substrate specificities, which is good for C2-C4 acyl-CoA thioeters, and propionyl-CoA is the best substrate. Based on these results we conclude that This is a fatty acid oxidation pathway which augments the central fatty acid degradation pathway found in all bacteria. While the structure of the liganded PA5185 is needed to understand how it selects for the small substrates.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SCHEMES

LIST OF ABBREVIATIONS

ACP	acyl-carrier protein
ADE	Adenosine
ADP	adenosine 5’-diphosphate
Ala or A	alanine
AMP	adenosine 5’-monophosphate
Arg or R	arginine
Asp or D	aspartate
Asn or N	Asparagine
AT	acyl transferase
ATCC	American Type Culture Collection
ATP	adenosine 5’-triphosphate
BF1314	HAD-thioesterase fusion protein from *Bacteroides fragilis*
BF-HAD	HAD domain of BF1314
BF-TE	Thioesterase domain of BF1314
BLAST	basic local alignment search tool
bp	base pair(s)
BT1666	sugar phosphatase from *Bacteroides thetaiotaomicron*
BT3352	standalone HAD domain homolog of BF1314 from *Bacteroides thetaiotaomicron*
BT4131	sugar phosphatase from *Bacteroides thetaiotaomicron*
BT4699	HAD-thioesterase fusion protein from *Bacteroides thetaiotaomicron*
BVU_2420	standalone thioesterase domain homolog of BF1314 from *Bacteroides vulgatus*
4-CBA	4-chlorobenzoic acid

CoASH	coenzyme A
Cys or C	cysteine
DCPIP	2,6-dichlorophenolindophenol
DDT	dichlorophenyltrichloroethene
DEBS	6-eoxyerythronolide B synthase
DH	dehydratase
DHN-CoA	1,4 dihydroxynapthoyl-CoA
DMF	Dimethylformamide
DNA	deoxyribonucleic acid
DTNB	5’,5’-dithio-bis-(2-nitrobenzoic acid)
DTT	Dithiothreitol
E4P	Erythrose 4-phosphate
Echs1	enoyl-CoA hydratase *Rattus norvegicus*
E. coli	*Escherichia coli*
ER	enoyl-CoA reductase
EXPASY	expert protein analysis system
FAD	flavin adenine dinucleotide
FAS	Fatty acid synthase
FPLC	fast protein liquid chromatography
g	gram or gravity (depending on the context)
Gln or Q	glutamine
Glu or E	glutamate
Gly or G	glycine

h	hour
HADSF	haloacid dehalogenase superfamily
4-HBA	4-hydroxybenzoate
His or H	histidine
HPLC	high performance (pressure) liquid chromatography
4-IBA	4-Iodobenzoate
Ile or I	Isoleucine
IPTG	isopropylthio-β-galactoside
k	rate constant
k_{cat}	enzyme turnover rate
kDa	kilo-Dalton
KDPG	2-keto-3-deoxy-6-phosphogluconate
KDPGal	2-keto-3-deoxy-6-phosphogalaconate
K^+HEPES	(N-(2-hydroxyethyl)piperazine-N'-2-ethanesulfonic acid)potassium salt
K_i	inhibition constant
K_M	Michaelis-Menten constant
KR	ketoreductases
KS	ketosynthase
L	liter
LB	Luria-Bertani
Leu or L	Leucine
Luci	luciferase
Lys or K	Lysine

M	molar
4-MBA	4-methylbenzoate
mCi	milli-Curie (10^{-3} Ci)
MenA	1,4-dihydroxy-2-naphthoate octaprenyltransferase
MenB	1,4-dihydroxynaphthoyl-CoA synthase
MenC	O-succinylbenzoate-CoA synthase
MenD	2-succinyl-6-hydroxy-2,4- cyclohexadiene-1-carboxylate synthase
MenE	O-succinylbenzoyl-CoA ligase
MenF	isochorismate synthase
MenG	Demethylmenaquinone methyltransferase
MenH	2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase
MESG	2-amino-6-mercapto-7-methylpurine riboside
Met or M	Methionine
mg	milligram (10^{-3} gram)
min	minute
MK	menaquinone
mM	millimolar
mL	milliliter (10^{-3} L)
MS	mass spectrum
$M_{w, calc}$	Calculated molecular weight
$M_{w, obs}$	Observed molecular weight by size exclusion chromatography
NAD^+	nicotinamide adenine dinucleotide
β-NADH	β-nicotinamide adenine dinucleotide, reduced form

NCBI	National Center for Biotechnology Information
NMR	nuclear magnetic resonance
NP	Naphthoate
NRPS	non-ribosomal peptide synthetase
OD	optical density
ORF	open reading frame
OSB	O-Succinylbenzoate
PA3302	R-specific Enoyl-CoA Hydratase from *Pseudomonas aeruginosa PAO1*
PA5185	Hot dog-fold thioesterase from *Pseudomonas aeruginosa PAO1*
PA5186	iron-containing alcohol dehydrogenase
PA5187	acyl-CoA dehydrogenase from *Pseudomonas aeruginosa PAO1*
PA5188	3-hydroxyacyl-CoA dehydrogenase from *Pseudomonas aeruginosa PAO1*
PaaI	Phenylacetyl-CoA thioesterase
PAGE	polyacrylamide gel electrophoresis
PCB	polychlorinated biphenyls
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
PEP	phosphoenolpyruvate
PG1653	HAD-thioesterase fusion protein from *Porphyromonas gingivalis W83*
PG-HAD	Truncated HAD domain of PG1653
PG-TE	Truncated thioesterase domain of PG1653
PheA	phenylalanine adenylation

Phe or F	phenylalanine
Pi	inorganic phosphate
pI	isoelectric point
PKS	polyketide
PMS	phenazine methosulphate
PMSF	phenylmethylsulfonyl fluoride
pNPP	para-Nitrophenylphosphate
PPi	pyrophosphate
PPTase	Phosphopantetheinyl transferase
Pro or P	proline
PSP	phosphoserine phosphatase
RNA	ribonucleic acid
RMSD	Root mean square deviation
rpm	rotation per minute
SDS	sodium dodecyl sulfate
S	substrate
SAM	S-adenosylmethionine
SEC	Size exclusion chromatography
sec or s	second
Ser or S	serine
SHCHC	2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate
TE	Thioesterase
Thr or T	threonine

Tris	tris[hydroxymethyl]aminomethane
Trp or W	tryptophan
Tyr or Y	tyrosine
μL	microliter
μM	micromolar
Val or V	Valine
v	volume
V_{max}	maximum velocity
WT	wild type
yidA	Phosphatase from *Escherichia coli*

CHAPTER 1: GENERAL BACKGROUND INFORMATION

1.1 Hotdog-fold Thioesterase superfamily

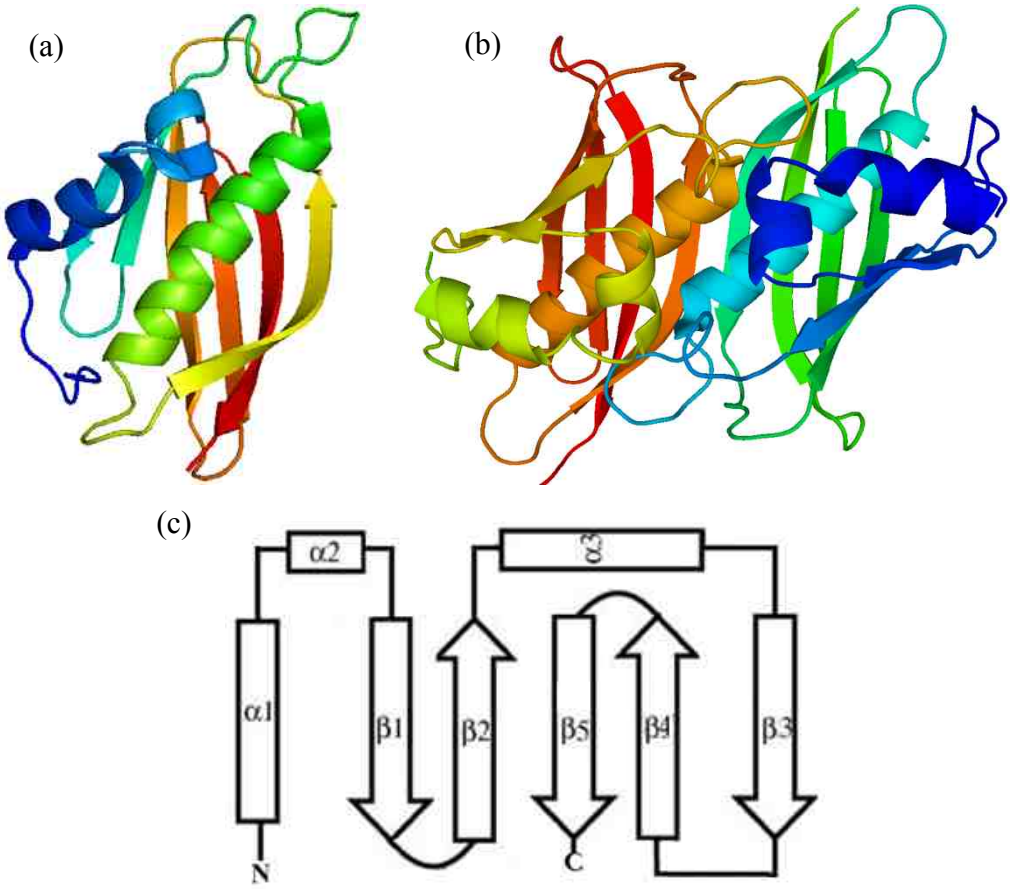


Figure 1.1 a). Ribbon presentation of *E coli.* ydii crystal structure, b) The hotdog dimmer structure of ydii and c) Schematic diagram of ydii secondary structure topology.

Thioesters play a central role in the cells where they participate in metabolism, membrane synthesis, signal transduction, and gene regulation (1). The carboxylic acid components of biological thioesters are metabolites of varied size, shape, and polarity.

They are converted to thioesters by ligases for the purpose of solubility, transport, signaling, or activation for reaction in biosynthetic or biodegradation pathways. The naturally occurring thiols include coenzyme A (CoA), glutathione, the pantetheine unit of the holo-acyl carrier protein (ACP), and the cysteine residue of a protein. Thioesters are reverted to the thiol and carboxylic acid components by hydrolysis catalyzed by thioesterases. Thioesterases have primarily evolved within the α/*β*-fold hydrolase enzyme superfamily (2) and the hotdog-fold enzyme superfamily (3). The conserved fold of the hotdog fold superfamily proteins consists of a five-stranded antiparallel *β*-sheet wrapped around an elongated α-helix. This topology is reminiscent of a hotdog "bun" wrapping around a "sausage", hence the name hotdog fold (4).

The "hot-dog" fold superfamily of enzymes is a large group of structurally related enzymes with diverse biochemical functions. Family members are found in eukaryotes, bacteria and archaea wherein there are involved in a range of cellular processes (3). The "hot-dog" fold was first observed in the structure of E. coli. *β*-hydroxydecanoyl thioester dehydratase (4). The hotdog domain consists of a five-stranded anti-parallel β-sheet wrapped about an elongated α-helix, reminiscent of a hotdog in a bun (Figure 1.1a and 1.1c). Two hotdog subunits associate to form the functional unit (Figure 1.1b), an extended β-sheet structure with two active-site clefts located at the respective ends of the subunit-subunit interface (Figure 1.1b).

Two backbone segments frame the catalytic scaffold of the active site, contributing catalytic and substrate-binding groups. One segment is the N-terminus of the elongated α-helix and its connecting loop from one subunit (subunit A), and the other is the center segment of the elongated α-helix on the opposing subunit (subunit B) (colored

green and blue, respectively in Figure 1.2). The backbone amide NH of the α-helix N-terminus residue of subunit A engages in hydrogen-bond formation with the substrate thioester C=O, thereby orienting and polarizing the C=O for nucleophilic attack. A carboxylate residue is used as a general base or nucleophilic catalyst for the hydrolysis reaction (Figure 1.2). It is positioned at either of two locations in the active site scaffold. The location divides the hotdog thioesterase family into two "catalytic motif" subfamilies. For one group (type AA, Figure 1.2), the catalytic carboxylate residue (Asp 17) is located on the same subunit as the polarizing α-helix N-terminus, where it is positioned on the loop that enters the α-helix N-terminus (Tyr 24). For group type AB, the polarizing α-helix N-terminus is located on one subunit (Figure 1.2, His 65) and the catalytic carboxylate residue Glu73 (Figure 1.2) is found on the α-helix of the opposing subunit. The thioesterase domain of the HAD-thioesterase fusion proteins is type AB (5-9).

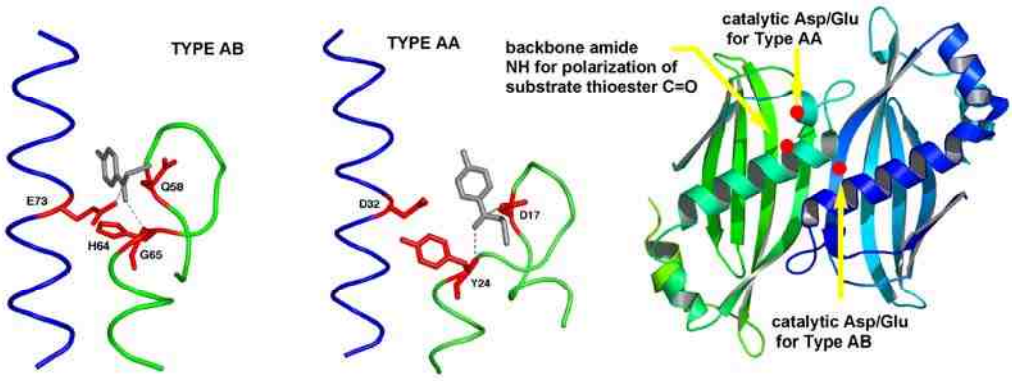


Figure 1.2 (Left) the catalytic scaffold type AB and type AA represented by the *Arthrobacter* and *Pseudomonas* 4-hydroxbenzoyl-CoA thioesterases, respectively (4-hydroxybenzoylthioester unit of the truncated 4-hydydroxybenzoyl-CoA substrate ligand

is colored grey. (Right) The hot dog thioesterase dimer with the locations of the catalytic residues indicated by red dots.

1.2 Catalytic Mechanism of Hotdog-fold Thioesterases

The catalytic carboxylate residue can function in base or nucleophilic catalysis (Figure 1.3 A). In the former, a water molecule is deprotonated by the catalytic carboxylate as it attacks the thioester C=O. If a tetrahedral intermediate is formed, the proton may hop from oxygen to sulfur prior to expulsion of the CoA. Alternatively, a general acid or electrophilic residue might assist in thiolate departure, although the need for thiolate protonation is not indicated by the results of earlier studies of solution hydrolysis of thioesters.

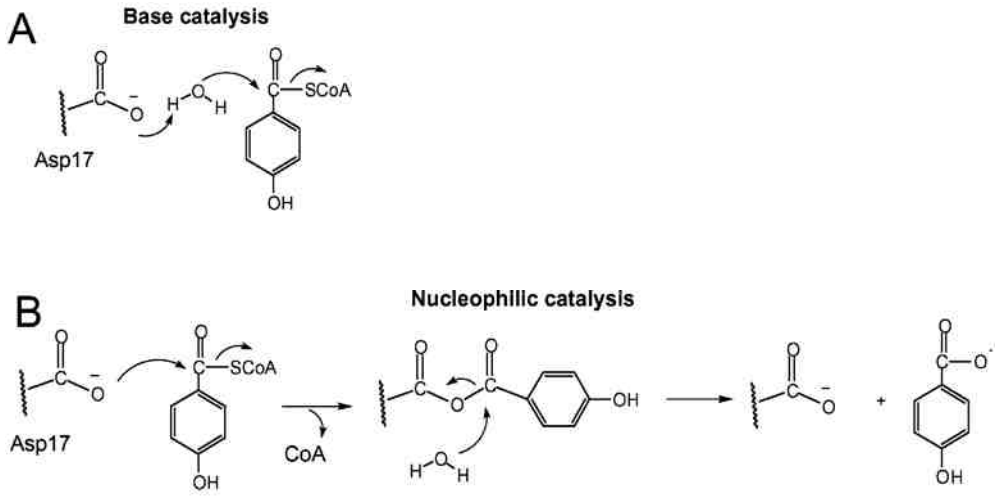


Figure 1.3 The A) nucleophilic catalysis and, B) general base catalysis in the hotdog thioesterase.

In nucleophilic catalysis, the carboxylate attacks the thioester C=O, resulting in the transient formation of a tetrahedral oxyanion that subsequently generates an unsymmetrical anhydride intermediate. The anhydride undergoes nucleophilic attack at the C=O by a water molecule, from a position that conforms to the stereoelectronic requirements for carbonyl addition (10). An active site residue may engage in hydrogen bond formation with the water molecule for positioning, and possibly deprotonation, required for its nucleophilic attack at one of the two C=Os of the unsymmetrical anhydride.

1.3 Biological Functions of Thioesterases

1.3.1 Aromatic Compounds Degradation Pathway

Some bacterial thioesterases function in the degradation pathways of certain aromatic compounds. The 4-chlorobenzoate (4-CBA) dehalogenation pathway (Figure 1.4) that converts the common 4-CBA to the metabolite 4-hydoxybenzoate (4-HBA). Several bacterial strains (viz. *Pseudomonas* sp. CBS3, *Alcaligenes* sp. NTP-1 and *Arthrobacter* sp. strain SU) harboring pathway have been isolated from soil and sludge, based on their ability to utilize 4-CBA as a sole source of carbon. In each of these bacteria, 4-CBA is first converted to 4-HBA, and 4-HBA can be further degraded via the

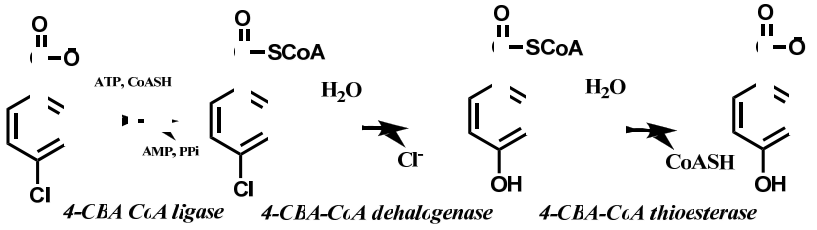


Figure 1.4 4-Chlorobenzoate Dehalogenation Pathway

the *ortho* or *meta*-aromatic ring oxidation pathways, which funnel into the β-ketoadipate pathway. The acetyl-CoA and succinyl-CoA products enter the TCA cycle for energy generation.

The 4-CBA dehalogenation pathway of *Pseudomonas* sp. strain CBS3 is the best characterized. The genes encoding this pathway in *Pseudomonas* sp. strain CBS3 are chromosome borne. In 1986, a 9.5 kb chromosomal DNA fragment from this strain that encoded the dehalogenation pathway enzymes was cloned into the cosmid vector pPSA843 (11). Selective expression of the dehalogenation pathway genes revealed three polypeptides of molecular weight of 57, 30 and 16 kDa as possible components of the dehalogenation system. Recombinant protein production of the pathway components showed that three enzymes, i.e., 4-CBA CoA ligase, 4-CBA-CoA dehalogenase and 4-HBA-CoA thioesterase function in a sequential three-step reaction pathway (12). The 4-CBA:CoA ligase utilizes Mg (II), ATP and coenzyme A (CoA) to form the 4-CBA-CoA thioester. The 4-CBA-CoA dehalogenase catalyzes a novel aromatic substitution reaction involving the replacement of the chloride with a hydroxyl group (13). The thioesterase of this dehalogenation pathway is responsible for the ensuing hydrolysis of 4-HBA-CoA to generate 4-HBA and CoA (12).

1.3.2 Menaquinone Biosynthesis Pathway.

In prokaryotes, ubiquinone (*aka* coenzyme Q) and menaquinone (MK) (*aka* vitamin K) are lipid-soluble molecules that shuttle electrons between membrane-bound protein complexes in the electron transport chain (14, 15). For example, *Escherichia coli*, a facultative anaerobe, utilizes ubiquinone under aerobic conditions, yet uses MK (MK-8) when grown anaerobically (15). In contrast, vertebrates use MK as a cofactor for certain carboxylases involved in blood coagulation, bone metabolism, tissue calcification and cell cycle regulation (16). In mammals, the electron transport chain is located in the inner mitochondrial membrane, where ubiquinone is used instead of menaquinone. Menaquione can also act as a signaling molecule that regulates the transcription of genes involved in its own metabolism (17). The principal structural variations within the two main classes of quinones occur from the length of the isoprenoid chain. In mitochondria the ubiquinone has a side chain of 10 isoprene units (CoQ-10), while in *E. coli* the ubiquinone has a side chain of 8 isoprene units (CoQ-8) (18, 19).

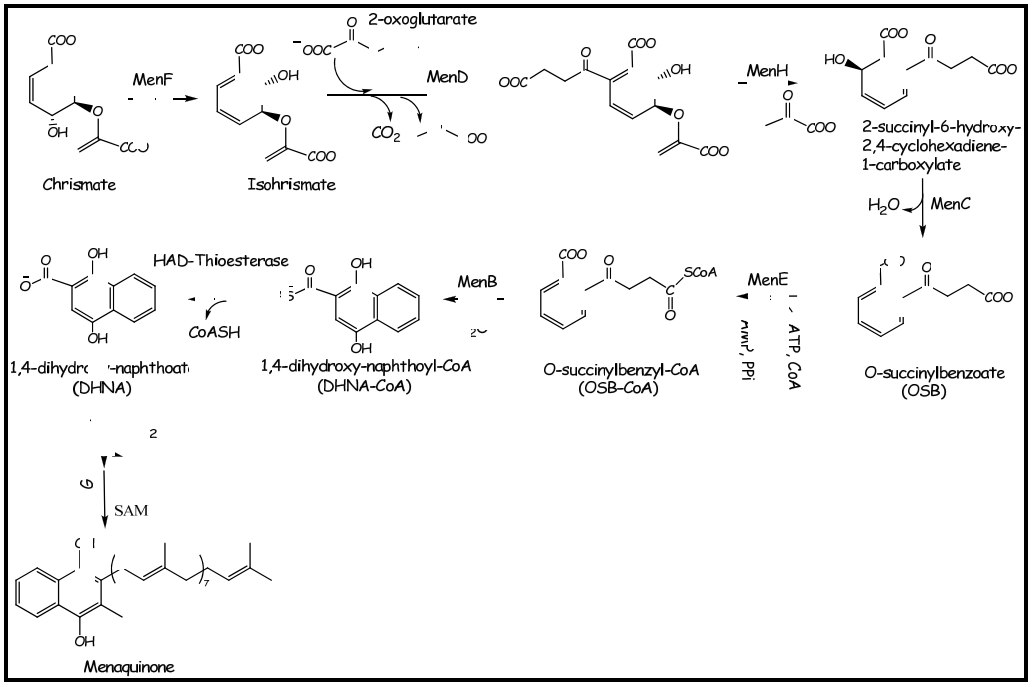


Figure 1.5 The bacterial menaquinone biosynthesis pathway

The biosynthesis of menaquinone has mainly been studied in *E. coli* (15, 16). Chorismate (Figure 1.5), derived from the shikimate pathway, is initially converted into isochorismate by MenF, MenD, MenH, and MenC to O-succinylbenzoate (OSB), the succinyl chain is activated by ligation with CoA catalyzed by MenE, and then cyclized, producing 1,4 dihydroxy 2-naphthoyl-CoA (DHNA-CoA). Subsequently, a specific CoA thioesterase removes CoA and frees the carboxyl group of 1,4 dihydroxy 2-naphthoate (DHNA) and allows its conjugation to the polyisoprenyl chain for anchoring into biological membranes.

1.3.3 Fatty Acid Biosynthesis Pathway.

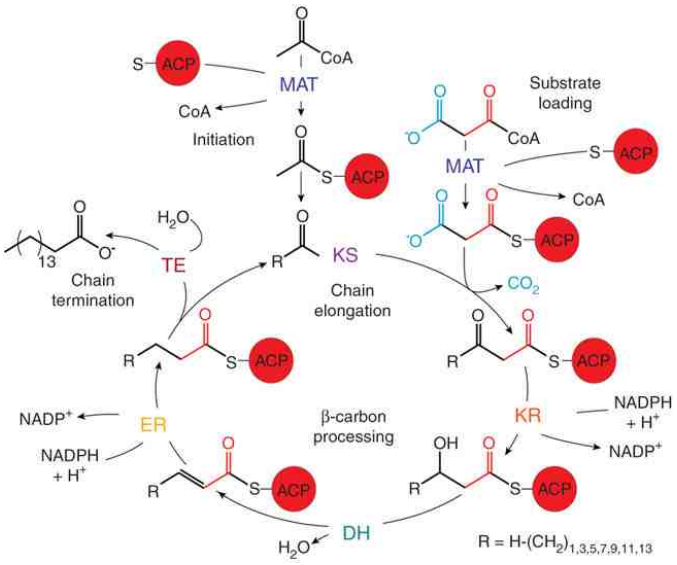


Figure 1.6 The fatty acid biosynthesis reaction cycle initiates with transfer of the acetyl moiety to the KS via an ACP-bound intermediate (initiation). The malonyl thioester is

similarly transferred to an ACP (substrate loading) and then condensed with the KS-bound acyl chain (chain extension). The resulting β-ketone is then reduced and dehydrated, yielding a saturated acyl chain (β-carbon processing) that is delivered to the KS, initiating the next cycle. After seven cycles, the 16-carbon acyl chain, palmitate, is released by thioesterase (chain termination) (20).

Fatty acids are essential components of all biological membranes and also represent an important form of energy storage in both animals and plants. Their biosynthesis occurs universally. The synthesis of fatty acids *de novo* is achieved by sequential condensation of two-carbon units derived initially from acetyl-CoA. In eukaryotes, this process is mediated by fatty acid synthase. This large enzyme complex is comprised of two multifunctional polypeptide chains, each containing seven discrete functional domains, juxtaposed head-to-tail such that two separate centers for fatty acid assembly are formed at the subunit interface. In prokaryotes, the fatty acid synthetic pathway is the same, however the enzymes that catalyze the reaction steps are separate entities (e.g., ketoacyl-ACP synthase, ketoreductase, dehydrase and reductase) (Figure 1.6).

The FAS thioesterase (thioesterase I) is the chain-terminating enzyme in both bacterial and human fatty acid syntheses. It releases the fatty acids by hydrolyzing the thioester bond in newly synthesized fatty acyl-ACP. The thioesterase in FAS has a marked specificity for long-chain substrates, exhibiting a 10-fold preference for C16 over C14 straight-chain acyl-thioesters (21). This specificity dictates that C-16 straight-chain saturated fatty acid is the major product of *de novo* fatty acid biogenesis in most biological systems.

In nature, fatty acids having shorter chain lengths (C12 and C14) are also abundant. The modification of the chain length of the FAS product requires the participation of a second chain-terminating thioesterase, thioesterase II, which is not associated with the FAS and which has preference for acyl thioesters of medium chain length. Thioesterase II is a discrete monofunctional, monomeric protein of 29 kDa that exhibits hydrolase activity toward medium chain-length acyl thioesters (22). Thioesterase II can access the growing acyl chain on the 4'-phosphopantetheine of the FAS to hydrolyze the thioester bond, thus releasing the medium chain-length saturated fatty acids that are typically 8 to 14 carbon atoms in length. It is believed that the interactions between the multi-component FAS and thioesterase II is dynamic rather that static and that substantial conformation changes accompany catalysis (23).

1.3.4 Thioesterases in mediation of intracellular lipid metabolism.

The abilities of acyl-CoA thioesterases to catalyze the hydrolysis of acyl-CoAs to free fatty acid and coenzyme A (CoASH) provide the potential to regulate intracellular levels of acyl-CoAs, free fatty acids and CoASH. These thioesterases have been observed in almost all cellular compartments, including the endoplasmic reticulum, cytosol, mitochondria and peroxisome. Recently, several thioesterases functioning in the different cellular compartments were purified, and their encoding genes cloned. These thioesterases are varied in their molecular weight and substrate specificity.

In the cytosol of eukaryote cells, acyl-CoAs and free fatty acids may regulate gene transcription via nuclear receptors. One such example is nuclear receptor HNF-4α that activates or inhibits transcription depending on the presence or absence of acyl-CoA

(24). Cytosolic thioesterase I (CTE-I) has a suggested role in controlling levels of acyl-CoA/fatty acids in the cytosol by hydrolyzing acyl-CoA. Thus, the activity of this thioesterase has a direct impact on gene transcription.

In mitochondria, where the β-oxidation occurs, high levels of CoA are required to maintain high metabolic capacity. The mitochondria acyl-CoA thioesterases may play important roles in maintaining sufficient CoA levels for optimal mitochondrial function.

1.3.5 Microbial polyketide natural product biosynthesis pathways

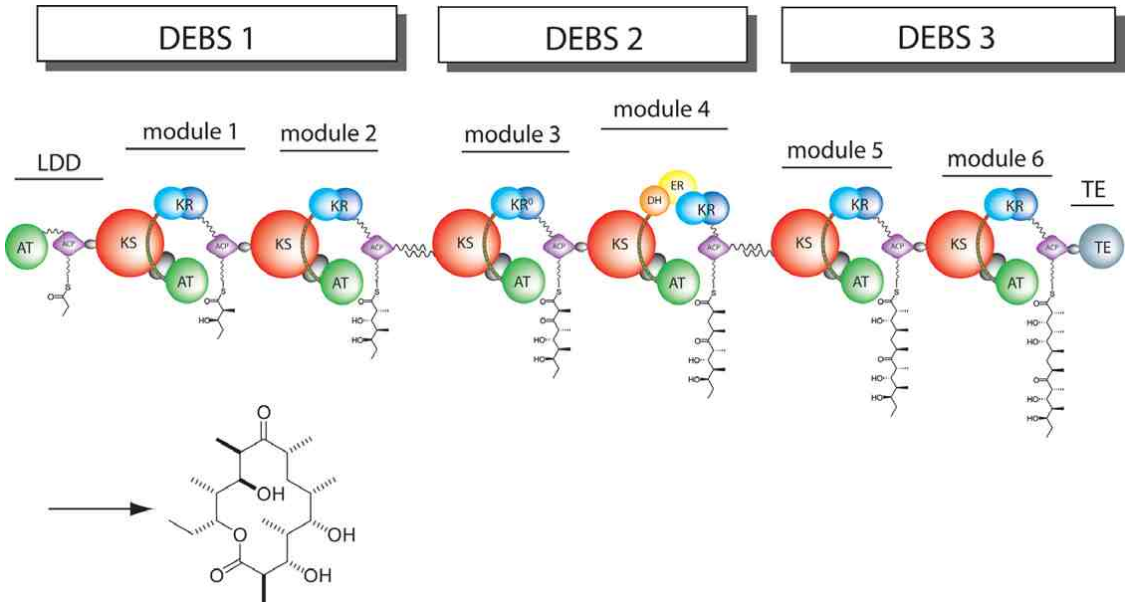


Figure 1.7 Biosynthesis of 6-deoxyerythronolide B

Thioesterase domains are also found in polyketide (PKS) and nonribosomal peptide synthetases (NRPS) which produce the polyketide and the cyclic peptide core of pharmacologically important compounds, including the antibiotics erythromycin, rifamycin, vancomycin, the immunosuppressive agent cyclosporin A and the anti-cancer

agent epithilone. The NRPS and PKS consist of modules of linked protein domains. Natural product acyl chains grow from an N-terminal initiation enzyme module through a series of elongation modules, one for each monomer added; then a C-terminal termination module releases the full acyl chain from its covalent thioester linkage (Figure 1.6). Each module has at least three protein domains: 1) a catalytic domain for selection and loading of the monomer, 2) the pantetheinylated acyl carrier domain, and 3) a catalytic domain that mediates chain extension, consisting of C-C bond formation to a β-keto thioester (a thioclaisen condensation) for PKS and C-N amide bond formation for NRPS. Based on the domain content of the PKS or NRPS, we can predict the backbone structures of the encoded natural products. Release of the full length acyl/peptidyl chains is mediated by thioesterases, some of which catalyze hydrolysis while others catalyze regiospecific macrocyclization. Tailoring of PK and NRP chains, by acylation, alkylation, glycosylation and oxidoreduction, occurs both during tethered chain growth and after thioesterase-mediated release (25).

The biosynthesis of the polyketide core of erythromycin A, 6-deoxyerythronolide B (6-dEB), has provided the paradigm for understanding the structure and function of the PKSs that are responsible for assembling complex polyketides (Figure 1.7). The 6-eoxyerythronolide B synthase (DEBS), which catalyzes the formation of 6-dEB, consists of three large subunits, DEBS1, DEBS2 and DEBS3, each containing two modules and above 300 kD in size. There are 2 domains in the N-terminal loading module, responsible for priming the synthase with a proprionate starter unit, and 26 domains in the six extender modules. Each extender module contains at least three essential domains: a ketosynthase (KS), an acyl transferase (AT) and an acyl carrier protein (ACP). The AT

domain selects the appropriate carbon extender unit and transfers the units from acyl-CoA onto the phosphopantetheine arm of ACP. The KS domain accepts the polyketide chain from the previous module and catalyzes the chain elongation reaction by adding an ACP-bound extender unit through decarboxylative condensation. After the extender unit is added, it can be further processed by optional tailoring domains, including ketoreductases (KRs), dehydratases (DHs), and enoyl reductases (ERs), to yield a hydroxyl, enoyl, or methylene group at the beta-position. Finally, the thioesterase (TE) domain that located at the C-terminus of DEBS module 6 promotes the macrocyclization event, which releases the final product, 6-dEB (26).

1.4 Haloalkanoate Dehalogenase (HAD) Enzyme Superfamily

The haloalkanoate dehalogenase (HAD) enzyme superfamily is a large, ubiquitous enzyme superfamily. It contains more than 19,000 unique sequences which are distributed among all three superkingdoms of life. Numerous members are found in each organism; there are 28 in *Escherichia coli*, 30 in *Mycobacterium tuberculosis*, 169 in *Arabidopsis thaliana*, and 183 in *Homo sapiens*. The HAD superfamily, named after the archetypal enzyme haloacid dehalogenase (27), includes enzymes catalyzing carbon or phosphoryl group transfer reactions on a diverse range of substrates, using an active site aspartate in nucleophilic catalysis. The majority of the enzymes in this superfamily are involved in phosphoryl transfer. These are the phosphate monoester hydrolases (phosphatases) or phosphoanhydride hydrolases P-type (ATPases), phosphonoacetaldehyde hydrolase (phosphonatase) and phosphomutases.

All cellular organisms synthesize and degrade nucleic acids, phospholipids, and phosphorylated proteins in addition to a vast number of primary and secondary organophosphate metabolites. Consequently a large number of phosphohydrolases and phosphotransferases have evolved and they have done so within a relatively small set of superfamilies (28, 29).

Members of the HAD superfamily have assumed numerous disparate biological functions, which vary in their degree of essentiality to the cell. A central question regarding this family is how the catalytic platform has been adapted through evolution to act on a wide range of substrates, a process which has termed the "evolutionary exploration of substrate space" (30). The availability of numerous crystal structures of representatives belonging to diverse branches of the HAD superfamily has facilitated the study of the diversification of structure and function within the HAD superfamily.

1.4.1 Structural and Functional Aspects of the HAD Superfamily

Except the four highly conserved catalytic motifs, the level of sequence divergence among members of the HAD enzyme superfamily is quite high (sequence identity <15%) (31). With the development of sequencing techniques, the number of sequenced genes grows geometrically. More and more genes are found to be misannotated especially in HAD superfamily. The question of conservation of backbone fold and active-site diversification within the family is best addressed through comparison of members catalyzing different chemistries, by comparing the folds, active-site scaffolds, and catalytic mechanisms of HAD members, with the ultimate goal of gaining insight into the evolution of catalysis within the HAD enzyme superfamily.

1.4.2 Structural of the HAD Superfamily Catalytic Domain

The catalytic domain of the HAD superfamily is a three layered α/β sandwich comprised of repeating β-α units which adopt the topology typical of the Rossmannoid class of α/β folds. The central sheet is parallel and is typically comprised of at least five strands in a β5, β4, β1, β2, β3 order (Figure 1.8).

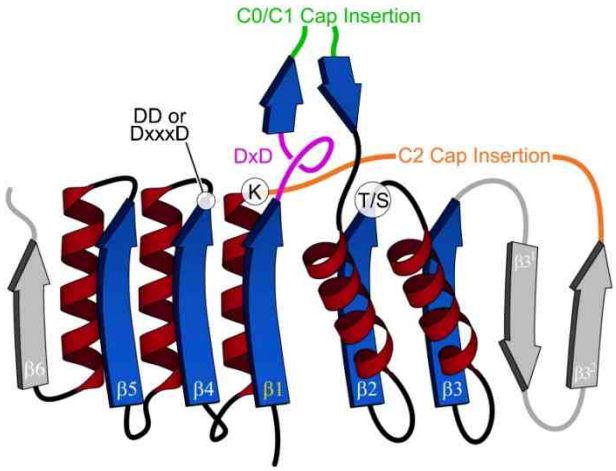


Figure 1.8 Topology diagrams of domains representative of the major divisions of Rossmann-like folds with catalytic residues.

Sequence comparisons have shown that practically all members of the HAD superfamily contain four highly conserved sequence motifs (32) (Figure 1.9a and Figure1.10). Sequence motif I corresponds to strand β1 and the DxD signature is present at the end of this strand. The carboxylate group of the first Asp and the backbone C=O of the second Asp coordinate the Mg^{2+} cofactor (Figure 1.9b). Additionally, the first Asp in motif I acts as the nucleophile that forms an aspartyl-phosphate intermediate during

catalysis (33). In phosphatase and phosphomutase members of the superfamily the second acidic residue acts as a general acid-base. It binds and protonates the substrate leaving group in the first step and deprotonates the nucleophile of the second step (33). In the ATPases, the occurrence of a threonine at this position allows for a reduced rate of aspartyl phosphate hydrolysis, which may allow for the time lag necessary for the consequent conformational change. In the phosphonatases, there is an alanine instead of the second aspartate, which is consistent with the unique role played by the enamine intermediate (formed with the insert domain) as a general base catalyst in aspartyl phosphate hydrolysis.

Motif II corresponds to the β2 strand, which is characterized by a highly conserved threonine or a serine at its end. Motif III contains a conserved lysine that occurs near the N terminus of the helix located upstream of β4. Motif II and motif III contribute to the stability of the reaction intermediates of the hydrolysis reaction. The lysine in motif III is reminiscent of the basic residues termed arginine fingers that stabilize the negative charge on reaction intermediates in many other phosphohydrolases, particularly those of the P-loop NTPase fold (34). It is likely that the Lys residues play a similar role even in the HAD hydrolases. An analysis of the available structures shows that the lysine in motif III may occur in either of two structural contexts in different HAD hydrolases. In the P-type ATPases, acid phosphatases, phosphoserine phosphatases and the Cof phosphatases the lysine is incorporated into the helix immediately preceding strand S4. However, in all other HAD hydrolases it emerges from the loop immediately prior to the helix. On account of this difference in the secondary structure context of the lysine, motif III is

poorly conserved relative to the other motifs. Motif IV maps to strand β4 and the conserved acidic residues located at its end.

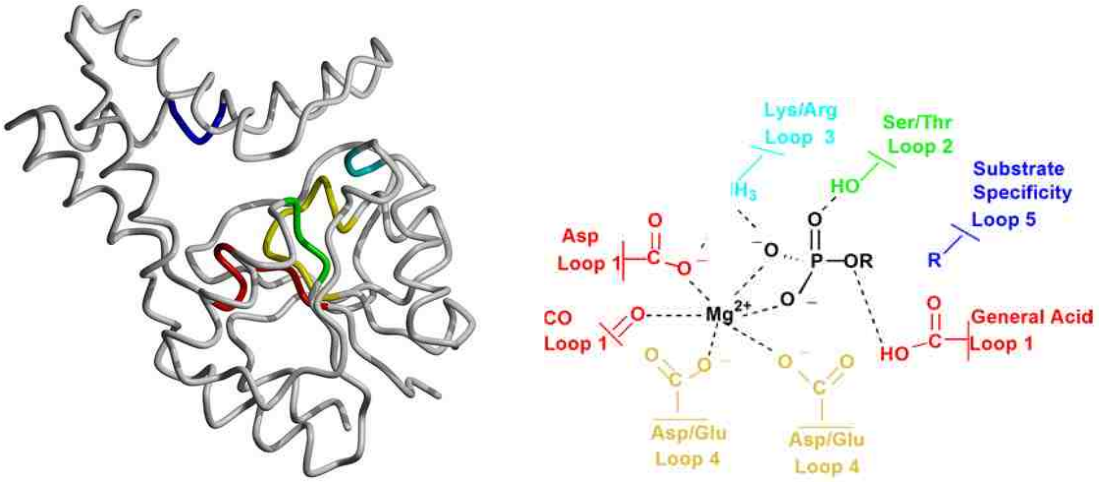


Figure 1.9. a). The conserved core domain loops shown in the typical superfamily members, the Haloalkanoic Acid Dehalogenase from *Pseudomonas sp.* Loop 1 is colored in red, Loop 2 is colored in green, Loop 3 is colored in blue and Loop 4 is colored in yellow. b). Chemdraw depiction of the active site.

The terminal acidic residues of motif IV typically exhibit one of three basic signatures: DD, GDxxxD, or GDxxxxD. These acidic residues, along with those in motif I, are required for coordinating the Mg ion in the active site (31, 35, and 36). Motif I to motif IV are spatially arranged around a single "binding cleft" at the C-terminal end of the strands of the central sheet that forms the active site of the HAD superfamily. This binding cleft is partly covered by the β-hairpin flap occurring after β1. An additional insert occurring between the two strands of the flap or in the region immediately after β3,

provides extensive shielding of the active site from solvent as is required for catalytic activity. These inserts, termed caps, often contain residues which contribute to substrate specificity or assume auxiliary catalytic functions. The cap domains play a central role in the reactions catalyzed by most HAD hydrolases (37-39).

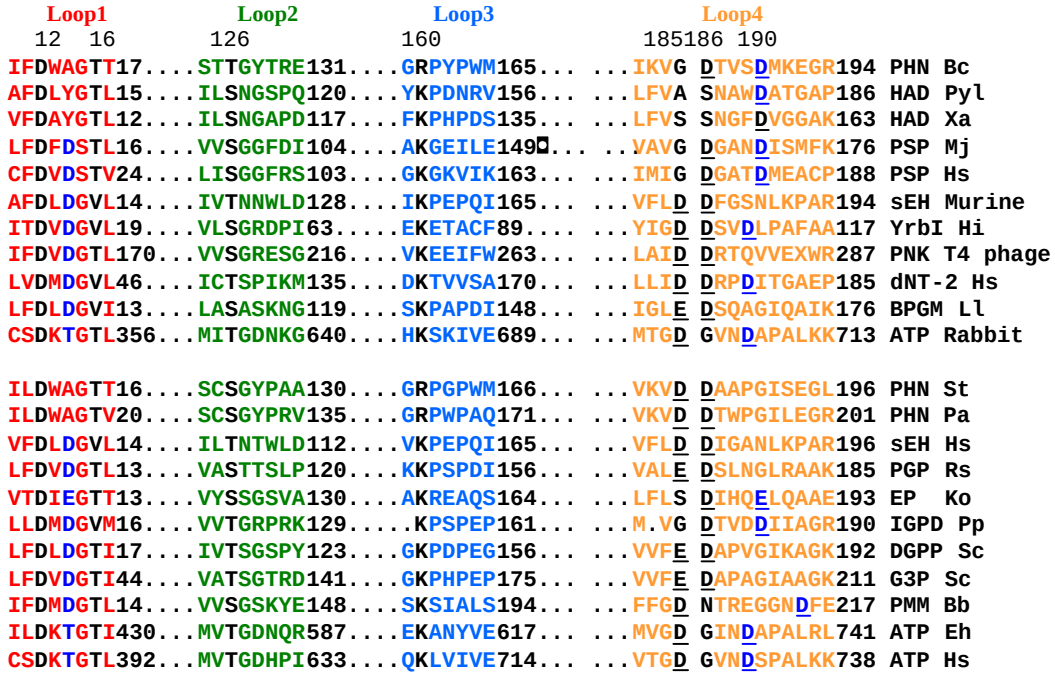


Figure 1.10 Sequence alignment of HAD core domain. In each loop, the residues identified by computer analysis of the primary sequence are colored black and the residues identified by structure-function analysis are colored dark blue (loops 1 and 4, will be discussed next), in loop 4, the metal ion pair is underlined. The upper part of the figure is for the sequences with known X-ray structures.

Despite the low overall sequence identity between the dehalogenase, phosphonatase and phosphatase, their folds are remarkably similar, particularly within the core domain (Figure 1.9a). The similarity in the overall folds together with the high sequence identity

within the conserved catalytic motifs (37) supports the hypothesis that they evolved by divergent evolution from a common progenitor. In the case of convergent evolution to a common protein structure and common chemistry, it would be expected that there would be no significant sequence identity and that the catalytic residues would be provided by different loops and therefore by different regions of the polypeptide chain. This is not the case in the HAD members where the conserved core domain supports a common active-site scaffold that consists of four loops located at topological switch points of the core domain.

1.4.3 Role of the Cap Domain in the Catalytic Mechanism of the HAD Superfamily

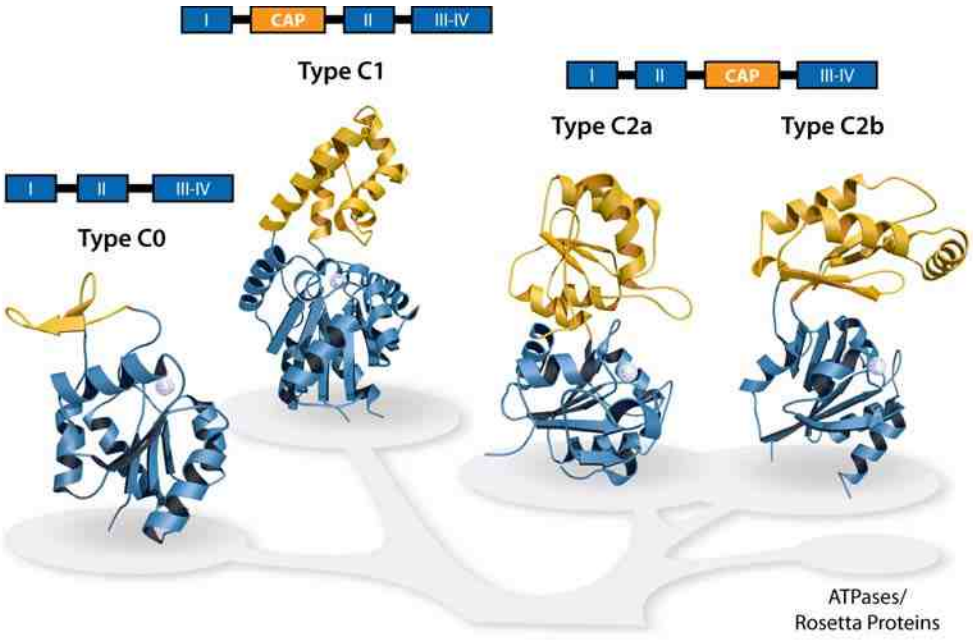


Figure 1.11 The HADSF can be divided into subfamilies by the topology and location of the cap domain. C1 members have a helical bundle inserted between motif I and II, C2a and C2b members have one of two different α/β domains inserted between motif II and

III and C0 have no cap (insertion site of cap between motifs shown as primary sequence diagram).

The most notable sequence inserts observed in the HAD superfamily are the caps, which, despite their diversity, can be classified into three categories: 1) type C0 members are structurally the simplest representatives of the HAD superfamily. They have only small inserts in either of the two points of cap insertion. 2) type C1 members are defined as HADs that contain inserts occurring in the middle of the β-hairpin of the flap motif, and fold into a structural unit distinct from the core domain. 3) The type C2 members are defined as HADs that contain inserts occurring in the linker immediately after strand S3 (C2a or C2b) (Figure 1.10). Most members of the HAD superfamily have either a C1 cap or a C2 type cap, although in few cases proteins may simultaneously possess both C1 and C2 caps (40).

Key aspects of the HAD catalytic mechanism that emerged from the reported family analysis are (40): 1) the alternation between open and close states and 2) a preliminary reaction favored by solvent exclusion and a subsequent step favored by extensive solvent contact. The principal features of the core domain responsible for this process are the squiggle and the flap. The squiggle, being close to a helical conformation, appears to be a structure that can be alternatively tightly or loosely wound. This differential winding in turn induces a movement in the flap immediately juxtaposed to the active site and alternatively results in the closed and open states (Figure 1.11). The strict conservation of the squiggle and the flap across the HAD superfamily suggests that they are likely to be part of a universal essential functional feature of this superfamily (37, 39).

The conformational changes in the squiggle and flap comprise the minimal apparatus for solvent exclusion and access at the active site of these enzymes. Given this minimal state, natural selection appears to have favored the emergence of cap modules as they made the process of solvent exclusion and acyl phosphate formation more efficient. In addition to aiding the basic catalytic mechanism the emergence of diverse caps also provided a means of substrate recognition by supplying new surfaces for interaction with substrates, which was not afforded by the ancestral active site alone (37, 38).

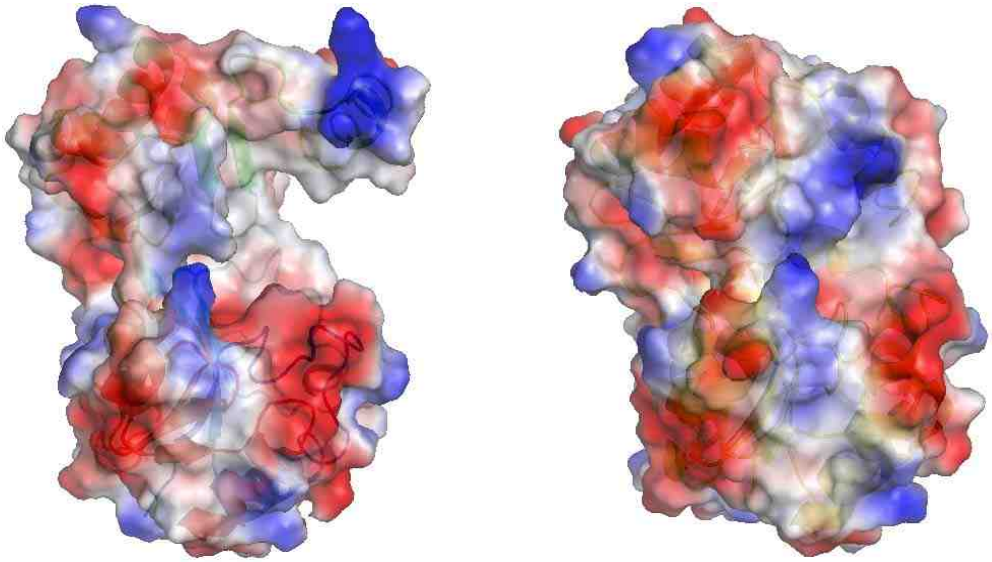


Figure 1.12 Space filling model of BT4131 in the open conformation (left), and close conformation (right) (41).

1.4.4 Evolution of the HAD Superfamily

The higher-order structural relationships of the HAD fold suggest that it first emerged as a part of the expansion of the phosphoesterase or Mg^{2+} chelating class of

Rossmannoid folds (40). The emergence of the squiggle and flap motifs might have allowed for a rudimentary solvent exclusion mechanism that allowed the HAD superfamily to acquire a catalytic mechanism based on the concomitant formation of an acyl phosphate intermediate (40). HAD enzymes do not participate directly in DNA replication and protein/RNA synthesis. The ancestral HAD phosphatase, like the ancestral version of the Rossmannoid folds, might have acted on nucleotides (40). This is suggested by the observation that nucleotide substrates are encountered in all the major branches of HAD superfamily including members of the earliest branching C0 assemblage, specifically the bifunctional polynucleotide kinase/phosphatase. The early branching C0 members appears to have specialized in large substrates such as proteins and nucleic acids, which precluded the need for large solvent-excluding caps (42). The emergence of various caps appears to have provided an additional structurally variable interaction module that allowed different representatives of HAD superfamily to accept a diverse range of substrates, typically small molecules.

1.4.5 Function and Structure in PSP Families of HAD C1 Subfamily

The PSP family is characterized is actually a hybrid of the C1 and C2 cap domain families. The family prototype phosphoserine phosphatase (PSP) catalyzes the dephosphorylation of L-3-phosphoserine in the biosynthetic pathway of serine (43, 44). PSP active-site residues are located at the interface between the two domains. The active site can be divided into two sub-sites: a phosphate-binding site and a serine-binding site. Lys 144, Asn 170, Gly 100, and a magnesium ion in the phosphate-binding site form a cationic cavity that facilitates the transfer of the phosphate group. All of the phosphate-

binding residues are from three conserved sequence motifs in the core domain. The specificity and orientation of the serine group of L-phosphoserine are achieved through interactions with residues from the helical cap domain: Glu 20 and Arg 56 form H-bonded salt-bridges with the amino group and the carboxyl group of L-phosphoserine, respectively, while Met 43 and Phe 49 form van der Waals contacts with the serine group (45).

1.4.6 Function and structure of C2b subfamily protein BT3352 and yidA

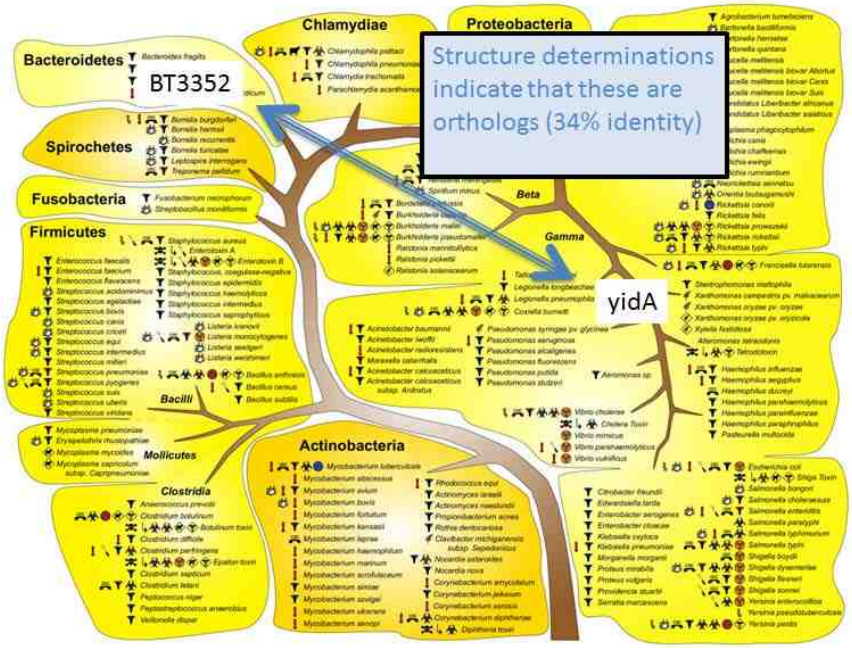


Figure 1.12 Bacterial pathogens (46)

The HAD superfamily type C2 phosphatases are prevalent in secondary metabolic pathways, and thus provide an excellent test case for *in vivo* function assignments and they provide an opportunity to explore the unusual but apparent functional redundancy

within the HAD superfamily. That is, in members of this subfamily (e.g., the *NagD* gene product and *Bacteriodes thetaiotaomicron* BT4131) there appears to be low level (k_{cat} = 1-26 s^{-1}; k_{cat}/K_m = 1 x 10^3 M^{-1} s^{-1}) of overlapping activities with a number of common sugar substrates, including D-glucose 6-phosphate, D-fructose 6-phosphate, D-gluconate 6-phosphate, D-arabinose 5-phosphate, and N-acetyl-glucoseamine-6-phosphate (41). In order to understand this phenomenon, the *substrate range* of each target enzyme must be identified. Also it is likely that the combined application of the substrate docking and library screening plus gene knock-out phenotyping will be required for *in vivo* function determination.

1.5 Goals of Doctoral Research

My doctoral research initially focused on determining the physiological substrates and structural determinants of substrate recognition in two HAD superfamily members and two hotdog-fold superfamily members. The first target was selected because it is a Rosetta protein comprised of a HAD superfamily domain and a hotdog-fold superfamily domain. A possible function for the thioesterase domain was suggested by the inclusion of the encoding gene in a gene cluster that encodes the menaquinone pathway in certain Bacteriodetes species. My initial goal was to demonstrate the function of the thioesterase in this pathway and then develop and test hypotheses for the function of the tethered HAD superfamily domain. In order to define the structural determinants for the recognition of the respective physiological substrates, I would supply purified protein to our collaborator Professor Karen Allen (of Boston University) for crystallization and X-ray structure determination. The structure obtained would be used to direct my site-

directed mutagenesis based analysis of the structural determinants of substrate recognition. Because the fusion protein did not crystallize, I identified a pair of single protein orthologues for structural analysis. The structure of the HAD superfamily member BT3352 was determined, however the thioesterase did not crystallize. Nevertheless, I demonstrated that it, and the thioesterase domain of the fusion protein, function in the menaquinone pathway by catalyzing the hydrolysis of the pathway intermediate 1,4-dihydroxynapthoyl-CoA. Substrate specificity analysis of the HAD member BT3352 revealed that it is a promiscuous phosphatase which prefers small (C4 and C5) sugar phosphates. The structural basis for substrate recognition was determined through the kinetic analysis of site-directed mutants.

The second part of my doctoral research focused on identifying the chemical steps of a pathway in *Pseudomonas aeruginosa* which employs a hotdog-fold family member (PA5185) to catalyze one of the steps. The pathway enzymes are encoded by an operon that contains genes of unknown function. The objective was to use bioinformations to identify possible chemical functions of the encoded enzymes and then test these functions using a focused substrate activity screen. By linking the chemical reaction catalyzed and the substrate specificities of each of the pathway enzymes I hoped to define a new metabolic pathway.

1.6 References

1. Hunt, M. C., and Alexson, S. E. (2002) The role Acyl-CoA thioesterases play in mediating intracellular lipid metabolism, *Prog Lipid Res.,* 41, 99-130.

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

3. Shane CD, Alex Bateman, (2004) The Hotdog fold: wrapping up a superfamily of thioesterases and dehydratases, *BMC Bioinformatics*, 5, 109.

4. Leesong, M., Henderson, B. S., Gillig, J. R., Schwab, J. M., and Smith, J. L. (1996) Structure of a dehydratase-isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: two catalytic activities in one active site, *Structure*, 4, 253-264.

5. Benning, M. M., Wesenberg, G., Liu, R., Taylor, K. L., Dunaway-Mariano, D., and Holden, H. M. (1998) The three-dimensional structure of 4-hydroxybenzoyl-CoA thioesterase from Pseudomonas sp. Strain CBS-3, *J. Biol. Chem.,* 273, 33, 572-579.

6. Scholten, J.D., Chang, K. H., Babbitt, P. C., Charest, H., Sylvestre, M., and Dunaway-Mariano, D, (1991) Novel enzymic hydrolytic dehalogenation of a chlorinated aromatic, *Science*, 253, 182-185.

7. Thoden, J. B., Zhang, Z., Dunaway-Mariano, D., Holden, H. M., (2003) The structure of 4-hydroxybenzoyl-CoA thioesterase from Arthrobacter sp. strain SU, *J. Biol. Chem.,* 278, 43709-43716.

8. Cheng Z, Song F, Shan X, Wei Z, Wang Y, Dunaway-Mariano D, Gong W., (2006) Crystal structure of human thioesterase superfamily member 2, *Biochem. Biophys. Res. Commun.*, 349, 172-177.

9. Zhuang, Z., Song, F., Zhang, W., Taylor, K., Archambault, A., Dunaway-Mariano, D., (2002) Kinetic, Raman, NMR, and site-directed mutagenesis studies of the Pseudomonas sp. strain BS3 4-hydroxybenzoyl-CoA thioesterase active site, *Biochemistry*, 41, 11152-11160.

10. Bruice, T. C. (2002) A view at the millennium: the efficiency of enzymatic catalysis, *Acc. Chem. Res.,* 35, 139-148.

11. Savard, P., L. Peloquin, and M. Sylvestre. (1986) Cloning of Pseudomonas sp. strain CBS3 genes specifying dehalogenation of 4-chlorobenzoate. J Bacteriol, 168, 81-85.

12. Chang, K. H., P. H. Liang, W. Beck, J. D. Scholten, and D. Dunaway-Mariano. (1992) Isolation and characterization of the three polypeptide components of 4-chlorobenzoate dehalogenase from Pseudomonas sp. strain CBS-3, *Biochemistry*, 31, 5605-5610.

13. Liang, P. H., G. Yang, and D. Dunaway-Mariano. (1993) Specificity of 4-chlorobenzoyl coenzyme A dehalogenase catalyzed dehalogenation of halogenated aromatics, *Biochemistry*, 32, 12245-12250.

14. Bentley, R. Maganathan, R. (1982) Biosynthesis of vitamin K (menaquinone) in bacteria, *Microbiol. Rev.,* 46, 241–280.

15. Meganathan, R. (2001) Biosynthesis of menaquinone (vitamin K2) and ubiquinone (coenzyme Q): a perspective on enzymatic mechanisms. *Vitam. Horm*., 61, 173–218.

16. Vermeer, C and Shearer, M and Zittermann, A and Bolton-Smith, C and Szulc, P and Hodges, S and Walter, P and Rambeck, W and Stocklin, E and Weber, P. (2004) Beyond deficiency: Potential benefits of increased intakes of vitamin K for bone and vascular health, *Eur. J. Nutr.*, 43, 325–335.

17. Ichikawa T, Horie-Inoue K, Ikeda K, Blumberg B, Inoue S (2006) Steroid and xenobiotic receptor SXR mediates vitamin K2-activated transcription of extracellular matrixrelated genes and collagen accumulation in osteoblastic cells, *J Biol. Chem.,* 281, 16927–16934.

18. Lester, R. L., and Crane, F. L. (1959) The natural occurrence of coenzyme Q and related compounds, *J. Biol. Chem.,* 234, 2169–2175

19. Bishop, D. H. L., Pandya, K. P., and King, H. K. (1962) Ubiquinone and vitamin K in bacteria, *Biochem. Biochem. J.,* 83, 606-614

20. Brignole EJ, Smith S, Asturias FJ. (2009) Conformational flexibility of metazoan fatty acid synthase enables catalysis, *Nat. Struct. Mol. Biol.,* 16(2), 190-197.

21. Lin, C. Y., and S. Smith. (1978) Properties of the thioesterase component obtained by limited trypsinization of the fatty acid synthetase multienzyme complex*, J. Biol. Chem.,* 253,1954-1962.

22. Libertini, L. J., S. Smith. (1978) Purification and properties of a thioesterase from lactating rat mammary gland which modifies the product specificity of fatty acid synthetase, *J Biol. Chem*., 253, 1393-1401.

23. Mikkelsen, J., A. Witkowski, and S. Smith. (1987) Interaction of rat mammary gland thioesterase II with fatty acid synthetase is dependent on the presence of acyl chains on the synthetase, *J. Biol. Chem.,* 262, 1570-1574.

24. Hertz, R., J. Magenheim, I. Berman, and J. Bar-Tana. (1998) Fatty acyl-CoA thioesters are ligands of hepatic nuclear factor-4alpha, *Nature*, 392, 512-516.

25. Walsh, C.T., (2008) The chemical versatility of natural-product assembly lines, *Acc. Chem. Res.,* 41, 1, 4-10.

26. Khosla C, Tang Y, Chen AY, Schnarr NA, Cane DE, (2007) Structure and mechanism of the 6-deoxyerythronolide B synthase, *Annu Rev Biochem.,* 76, 195-221.

27. Koonin, E. V., and Tatusov, R. L. (1994) Computer analysis of bacterial haloacid dehalogenases defines a large superfamily of hydrolases with diverse specificity. Application of an iterative approach to database search, *J. Mol. Biol.,* 244,125-244,132.

28. Vincent, J. B., Crowder, M. W., and Averill, B. A. (1992) Hydrolysis of phosphate monoesters: a biological problem with multiple chemical solutions, *Trends Biochem. Sci.*, 17, 105-110.

29. Vetter, I. R., and Wittinghofer, A. (1999) Nucleoside triphosphate-binding proteins: different scaffolds to achieve phosphoryl transfer, *Q. Rev. Biophys.,* 32, 1-56.

30. Anantharaman, V., Aravind, L., and Koonin, E. V. (2003) Emergence of diverse biochemical activities in evolutionarily conserved structural scaffolds of proteins, *Curr. Opin. Chem. Biol.,* 7, 12-20.

31. Morais, M. C., Zhang, W., Baker, A. S., Zhang, G., Dunaway-Mariano, D., and Allen, K. N. (2000) The crystal structure of bacillus cereus phosphonoacetaldehyde hydrolase: insight into catalysis of phosphorus bond cleavage and catalytic diversification within the HAD enzyme superfamily, *Biochemistry*, 39, 10385-10396.

32. Aravind, L., Galperin, M. Y., and Koonin, E. V. (1998) The catalytic domain of the P-type ATPase has the haloacid dehalogenase fold. Trends Biochem Sci 23, 127-129.

33. Lahiri, S. D., Zhang, G., Dunaway-Mariano, D., and Allen, K. N. (2002) Caught in the act: the structure of phosphorylated beta-phosphoglucomutase from Lactococcus lactis, *Biochemistry*, 41, 8351-8359.

34. Ahmadian, M. R., Stege, P., Scheffzek, K., and Wittinghofer, A. (1997) Confirmation of the arginine-finger hypothesis for the GAP-stimulated GTP-hydrolysis reaction of Ras, *Nat. Struct. Biol.*, 4, 686-689.

35. Lahiri, S. D., Zhang, G., Radstrom, P., Dunaway-Mariano, D., and Allen, K. N. (2002) Crystallization and preliminary X-ray diffraction studies of beta-phosphoglucomutase from Lactococcus lactus, *Acta. Crystallogr. D. Biol. Crystallogr.*, 58, 324-326.

36. Peisach, E., Selengut, J. D., Dunaway-Mariano, D., and Allen, K. N. (2004) X-ray crystal structure of the hypothetical phosphotyrosine phosphatase MDP-1 of the haloacid dehalogenase superfamily, *Biochemistry*, 43, 12, 770-779.

37. Baker, A. S., Ciocci, M. J., Metcalf, W. W., Kim, J., Babbitt, P. C., Wanner, B. L., Martin, B. M., and Dunaway-Mariano, D. (1998) Insights into the mechanism of catalysis by the P-C bond-cleaving enzyme phosphonoacetaldehyde hydrolase derived from gene sequence analysis and mutagenesis*, Biochemistry*, 37,9305-9315.

38. Olsen, D. B., Hepburn, T. W., Moos, M., Mariano, P. S., and Dunaway-Mariano, D. (1988) Investigation of the Bacillus cereus phosphonoacetaldehyde hydrolase. Evidence for a Schiff base mechanism and sequence analysis of an active-site peptide containing the catalytic lysine residue, *Biochemistry*, 27, 2229-2234.

39. Kurihara, T., Liu, J. Q., Nardi-Dei, V., Koshikawa, H., Esaki, N., and Soda, K. (1995) Comprehensive site-directed mutagenesis of L-2-halo acid dehalogenase to probe catalytic amino acid residues, *J. Biochem.,* 117, 1317-1322.

40. A. Maxwell Burroughs, K. N. A., Debra Dunaway-Mariano, L. Aravind. (2006) Evolutionary Genomics of the HAD Superfamily: Understanding the Structural

Adaptations and Catalytic Diversity in a Superfamily of Phosphoesterases and Allied Enzymes, *J. Mol. Biol.,* 361.

41. Lu Z, Dunaway-Mariano D, Allen KN. (2005) HAD superfamily phosphotransferase substrate diversification: structure and function analysis of HAD subclass IIB sugar phosphatase BT4131, *Biochemistry,* 44(24):8684-**86**96.

42. Eara Peisach, L. W., A. Maxwell Burroughs, L. Aravind, Debra Dunaway-Mariano, Karen N. Allen. (2008) The X-ray crystallographic structure and activity of the HAD enzyme superfamily evidences a novel biochemical function. *Proteins,* 70, 197-207.

43. Borkenhagen, L. F., and Kennedy, E. P. (1958) The enzymic equilibration of L-serine with O-phospho-L-serine, *Biochim. Biophys. Acta., 28*, 222-223.

44. Neuhaus, F. C., and Byrne, W. L. (1958) O-Phosphoserine phosphatase. *Biochim. Biophys. Acta.,* 28, 223-224.

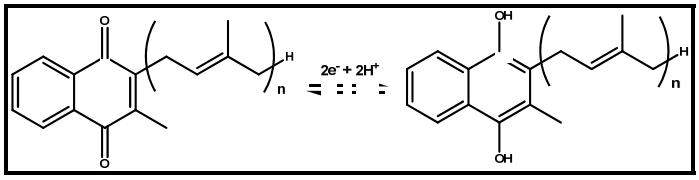
45. Wang, W., Cho, H. S., Kim, R., Jancarik, J., Yokota, H., Nguyen, H. H., Grigoriev, I. V., Wemmer, D. E., and Kim, S. H. (2002) Structural characterization of the reaction pathway in phosphoserine phosphatase: crystallographic "snapshots" of intermediate states, *J. Mol. Biol.,* 319, 421-431.

46. Hugenholtz P, Goebel BM, Pace NR, (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity, *J. Bacteriol.*, **180,** 4765-4774.

Chapter 2 The HAD Phosphatase: Hotdog-fold Thioesterase Rossetta Protein of the Menaquinone (Vitamin K2) Pathway

2.1 Menaquinone Biosynthesis

Scheme 2.1 Oxidized and reduced forms of Menaquinone.



Menaquinone is a lipid-soluble molecule that serves important biological functions. In humans and animals, it is a vitamin (K2) that has to be acquired from the diet or bacteria in the gut *(1)*. This substance is crucial in blood clotting through its involvement in the *γ*-carboxylation of glutamate residues *(2)*, bone metabolism, and *(3)* calcification of arteries and other soft tissues *(4)*. In bacteria, menaquinone is an electron-transporting molecule where it serves as the only electron carrier in the respiratory chain of aerobic Gram-positive bacteria and most anaerobic bacteria (Scheme 2.1) *(1)*. It is also needed in the respiration of facultative bacteria, such as *Escherichia coli*, during anaerobic growth *(1)*. It is essential for the survival of bacteria and, as a consequence of its absence in humans and animals, the menaquinone biosynthesis pathway has been a potential target for the development of new antibiotics.

The biosynthesis of menaquinone in *E. coli (5,) Bacillus subtilis (6)* and *Mycobacterium phlei (7)* shown in Fig. 1 has been carefully investigated. The *E. coli* menaquinone contains a 40-carbon isoprene chain (MK-8). In the pathway, chorismate,

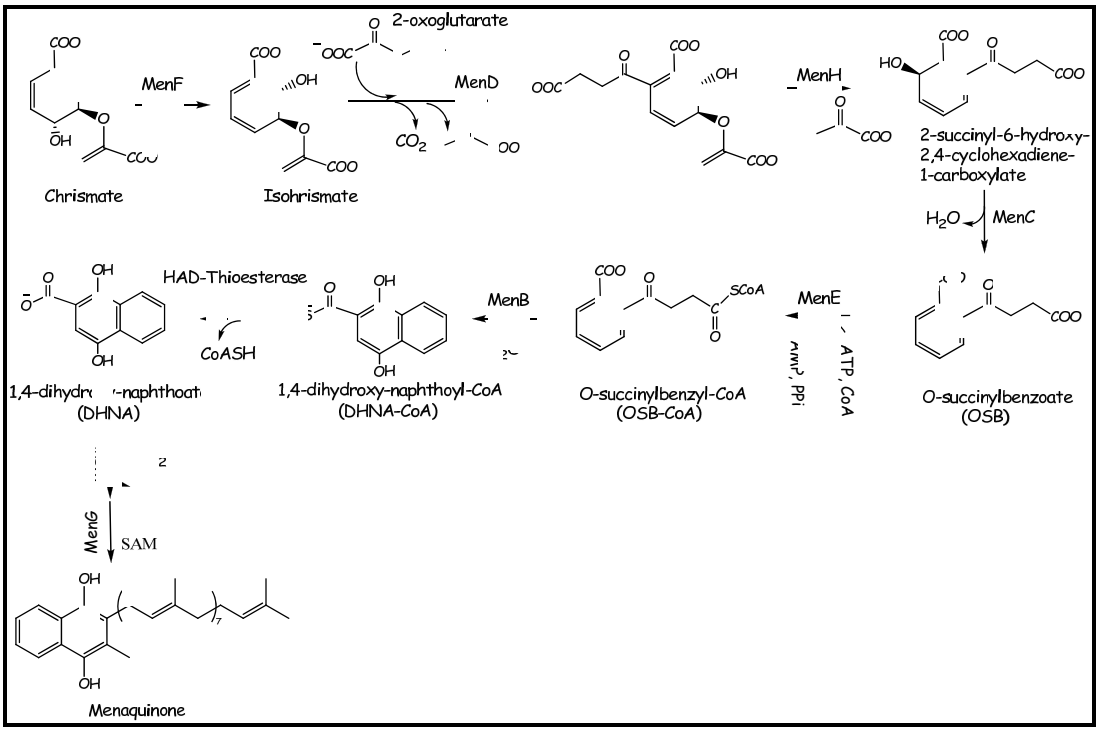


Figure 2.1 The bacterial menaquinone biosynthesis pathway

derived from the shikimate pathway, is initially converted to isochorismate by MenF, an isochorismate synthase *(8),* and then to 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate (SHCHC) by the action of the thiamin-dependent enzyme MenD *(9, 10).* SHCHC is then dehydrated by MenC (a member of the enolase superfamily) to give the aromatic compound O-succinylbenzoate *(11).* O-Succinylbenzoate (OSB) is converted to OSB-CoA by MenE *(12),* which is a member of acyl-CoA ligase superfamily *(13).* The naphthoquinol skeleton of DHN is generated via a cyclization reaction catalyzed by 1,4-dihydroxynaphthoyl-CoA synthase, MenB *(14, 15).* The 1,4-dihydroxynaphthoyl-CoA is hydrolyzed to DHN by an uncharacterized thioesterase. Finally, in the last two steps of the pathway, menaquinone is generated from DHN by the action of MenA, which

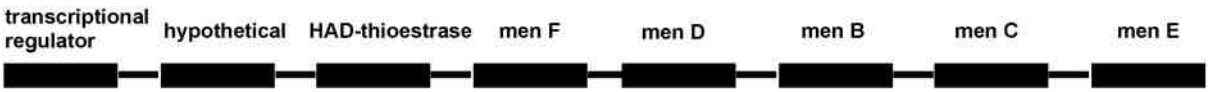
catalyzes a prenylation process, and by MenG, an S-adenosylmethionine (SAM)-dependent methyl transferase. Polyprenylpyrophosphate required by MenA is synthesized via a non-mevalonate isoprenoid pathway *(16)*.

The human internal cavity is colonized by a variety of bacteria that establish long-term symbiotic relationships with the host. Most predominant are the *Bacteriodes* species *thetaiotamicron, vulgatus, distasonis* and *fragilis,* which inhabit the gut and *Porphyromonas gingivalis*, which inhabits the mouth. The menaquinone operon of *Bacteriodes* (depicted in Figure 2.1) that encodes enzymes that catalyze the conversion of chorismate to DHN-CoA also includes a gene that encodes a two-domain protein of unknown function (*P. gingivalis* ExPasy accession Q7MU91, *B. caccae* ExPasy accession A5ZGV9, *B. thetaiotamicron* ExPasy accession Q89YN2, *B. ovatus* ExPasy accession A7M602, *B. uniformis* ExPasy accession A7VAE9, *B. fragilis* ExPasy accession Q64WR2 and Q5LFS7, *B. stercoris* ExPasy accession B0NN98, *Parabacteriodes merdae* ExPasy accession A7AG7, *P. distasonis* ExPasy accession A6LB33). An alignment of the two-domain proteins of the *Bacteriodes* species is shown in Figure 2.2B. Presumably, the two domain protein arose by gene fusion. In bacteria the genes that encode proteins that function together in a common biological process are often organized in a gene cluster. A mutation event that connects the two genes results in a "Rosetta protein", which typically possesses the functions of the two ancestral proteins.

The N–terminal domain (residue 1-265 using the numbering from the *P. gingivalis* protein sequence) of the two domain protein belongs to the haloacid dehalogenase superfamily (HADSF) and the C-terminal domain (residues 290-407) belongs to the hotdog fold superfamily.

A

B




```
Q7MU91 MKYKIIVIDVFNTLLNNKYELTEENRNALLYIQKELGIRVVLTSGRTVAALKPLAEELNLREYSGYLIPYHGGQIVNCRSGKVIASHRMATSFVEVLDAR
A5ZGV9 MKYKLLVLDVDGTLLNDAKEISKRTLAALLKVQQMG-VRIVLASGRPTYGLMPLAKSLELGNYGGFILSYNGCQIINAQNGEILFERRINPEMLPYLEKK
Q89YN2 MKYKLLVLDVDGTLLNDAKEISKRTLAALLKVQQMG-VRIVLASGRPTYGLMPLAKSLELGNYGGYILSYNGCQIINAQNGEILFERRINPEMLPYLEKK
A7M602 MKYKLLVLDVDGTLLNDAKEISKRTLAALLKVQQMG-VRIVLASGRPTYGLMPLAKMLELGNYGGFILSYNGCQIINAQNGEILFERRINPEMLPYLEKK
A7VAE9 MKYKLLVLDVDGTLLNDEREISKRTLAALLKVQQMG-VRIVLASGRPTYGLMPLAKTLELGNYGGFVLSYNGCQIIKAQNGEILFERRINPEMLPYLEKK
Q64WR2 MKYKLLVLDVDGTLLNDEKEITPRTLATLLKVQQMG-VHIVLASGRPTYGILPLAKKLELGNYGGYILSYNGAQVINAKNGEVLLERRINPEMLPYLEKK
Q5LFS7 MKYKLLVLDVDGTLLNDEKEITPRTLATLLKVQQMG-VHIVLASGRPTYGILPLAKKLELGNYGGYILSYNGAQVINAKNGEVLLERRINPEMLPYLEKK
B0NN98 MKYKLLVLDVDGTLLNDAKEISKRTLASLLKVQQMG-IRVALASGRPTYGLMPLAKTLELGNYGGFIISYNGGQIINAQNGEILFERRINPEMLPYLEKK
A7AG77 MKYKLLVLDVDGTLLNNQKEISPRTLAALLKVQQMG-VHIVLASGRPTNGVMPIAEKLELNHYGGYILSYNGGQIINVQTGELLFEKRIDPEWIPYFEKK
A6LB33 MKYKLLVLDVDGTLLNDKKEITPRTHAALLKAQQMG-VHVVLASGRPTNGVQPLAEALELNHYGGFILSYNGGQIINAQTGELMFEKRIDPAMIPYLNRK
       ****::*:** .****:  *:: ..  :**  *:   :::.*:***.. .: *:*: *:* .*.*:::.*:* *::: :.*::: .:*: .  :  :: :

Q7MU91 ARAEGVTLLAFGKDSIFSTDPHDPLAVKESDICGQPLVSKHPR----SSEVIKCSFVG-EAEKIAELEVVMRRDYGDKLNIFRSSETLLECLPSEVQKIK
A5ZGV9 ARKNGFALFTYHDDTIITDSPENEHIQNEARLNNLRIIEETEFSAAVDFAPCKCMLVSDNEEALIGLEDHWKRRLNGALDVFRSEPYFLEVVPCAIDKAN
Q89YN2 ARKNNFALFTYHDDTIITDTPENEHIQNEARLNNLKVIKEEEFSVAIDFAPCKCMLVSDDEEALVSLEGHWKRRLNGALDVFRSEPYFLEVVPCAIDKAN
A7M602 ARKNGFALFTYHDDTIITDSPENEHIQNEARLNDLQIIKEEEFSAAVDFAPCKCMLVSDDEEALLGLEDHWKRRLNGALDVFRSEPYFLEVVPCAIDKAN
A7VAE9 ARKNGFAIFTYHDDTLITDSPDNEYIKNEALLNNLKIIKEDEFSTAIDFAPCKCMLVSDKEKALIGLEQHWEKRLAGTLDAFRSEPYFLEVVPCGVNKAN
Q64WR2 ARKNGFAIFTYTEDRMIADQADNEHILQEAFLNRMELIEEPEFSVAVDFAPSKCMLVSDDEEALIGLEEHWKKRLNGALDVFRSEPYFLEVLPCGIDKST
Q5LFS7 ARKNGFAIFTYTEDRMIADQADNEHILQEAFLNRMELIEEPEFSVAVDFAPSKCMLVSDDEEALIGLEEHWKKRLNGALDVFRSEPYFLEVLPCGIDKST
B0NN98 ARKNNFAIFTYHDDTILTDSSDNEHVRAEANLNNLKIIQEEEFSTAIDFAPCKCILVSNDEEALKDLEEHWKKRLDGTLDVFCSEPYFLEVVPCGIDKAN
A7AG77 AKKNDFAIFTYHKDFILTDKPDNKYVIQEANLNKMQIVGVDNFAEAVDFPPCKCMLASDDEEALIGLENHWKKRLDGVLDAFRSEPYFLEVVPQFIDKGN
A6LB33 AKENGFAIFTYHKDYILTDSPENKHVQEEAELNKMRIIGVENFPEAVDFAPCKCILTSDDENNLVGLENHWKKRLDGVLEAFRSEDYFLEVAPHFINKGN
       *: :..:::::: .* :::  ..:     *: :    ::            .** :.. . :    :**   .:   . *: * *.  :**  *  ::* .

Q7MU91 AISDLLDKLDLSREELVSVGDSYCDVEMIQLAGLGVAVANAREAVKACADYITTSNEENGVAHLVDKYIRHEYEAVPFSVEDVNSIVPGTLMESLGIRCT
A5ZGV9 SLGALLEELDVKREEVIAIGDGVCDVTMIQLAGLGVAMGHSQDSVKICADYVTASNEEDGVAIAVEKAIIAEVRAAEIPLDQLNAQARHALMGNLGIQYT
Q89YN2 TLGALLEELDVKREEVIAIGDGVCDVTMIQLAGLGVAMGHSQDSVKVCADYVTASNEEDGVALAVEKAIIAEVRAAEIPLDQLNAQARHALMGNLGIQYT
A7M602 SLGALLEVLGMKREEVIAVGDGVCDVTMIQLAGLGIAMGHSQDSVKACADYVTASNEEDGVAVAVEKAIISEVRVAEIPLDQLNAQARHALMGNLGIQYT
A7VAE9 TLGALLEHLGVTREEVIAVGDGVCDVTMLQLAGMGVAMGHSQDSVKVCADYVTASNEEDGVALAVEKLILAEVRAAEVPLDLLNERARHALMGNLGIQYT
Q64WR2 SLGALLSHLDITPEEIIVIGDGVCDVSMIQFAGLGIAMGNAQDSVKVCADVVTASNEEDGVALAVEKAILSEIRPAEIPLDQLNERARHALMGNLGIQYT
Q5LFS7 SLGALLSHLDITPEEIIVIGDGVCDVSMIQFAGLGIAMGNAQDSVKVCADVVTASNEEDGVALAVEKAILSEIRPAEIPLDQLNERARHALMGNLGIQYT
B0NN98 TLGVLLSYLNIAREEVIAIGDGVCDVNMLQVAGLGIAMGHAQDSVKVCADYVTASNEEDGVAQSVEKLILAEVHAAEIPLDLLNERARHALMGNLGIQYT
A7AG77 TLGVLMTKLSVSPEEVIAIGDGVCDVTMLQLAGVGIAMGNAEDSVKACVDKTTLTNEEEGVAVAVERTILAEIRPTEVPLDQLNMRARHALMGNLGIQYT
A6LB33 TLAVLMEMLNITTEEVVAIGDGVADVSMLQLAGTGVAMGNARDSVKACADFTTLSNNMDGVAVAIEKAILATIKPTEVPLDQLNARAKHALMGNLGIQYT
       ::. *:  *.:  **:: :**. .** *:*.** *:*:.::.::** *.**:*: :***  ::: *    . . ..:: :*  .  :** .***: *

Q7MU91 KIARGYVEATMPVDIRTRQPMGILHGGASLAFAETLAGFGSVALCNPGEIQVGLQVSGNHVSSALEGDVLRGEASIMHQGRSTHVWSINIYSTKSGKLIC
A5ZGV9 YAAEDRVEATMPVDHRTRQPFGILHGGASLALAETVAGLGSMILCQPDEIVVGMQVSGNHISSAHEGDTVRAVATIVHKGRSSHVWNVDVFTS-TNKLVS
Q89YN2 YADEDRVEATMPVDHRTRQPFGILHGGATLALGETVAGLGSMILCQPDEIVVGMQVSGNHISSAHEGDTVRAVATIVHKGRSSHVWNVDVFTS-TNKLVS
A7M602 YADEDRVEATMPVDHRTRQPFGILHGGATLALAETVAGLGSMILCQPDEIVVGMQVSGNHISSAHEGDTVRAVGTVVHKGRSSHVWNVDVFTS-TNKLVS
A7VAE9 YASDERVEATMPVDYRTRQPFGILHGGATLALAETVAGLGSMIICEPDEIVVGMQVSGNHISSAHEGDTVRAVATIVHKGRSSHVWNVDVFTS-TNKLVS
Q64WR2 YASEDRVEATMPVDERTRQPFGILHGGATLALAETVAGLGSMILCQPDEIVVGMQVSGNHMSSAHEGDTVRAVGTIIHKGRSSHVWNVDVFTS-TDKLVS
Q5LFS7 YASEDRVEATMPVDERTRQPFGILHGGATLALAETVAGLGSMILCQPDEIVVGMQVSGNHMSSAHEGDTVRAVGTIIHKGRSSHVWNVDVFTS-TDKLVS
B0NN98 YASEERIEATMPVDHRTRQPFGILHGGATLALAETVAGLGSMITCQPDEIVVGMQVSGNHISSAHEGDTVRAVATIVHKGRSSHVWNVDVFTS-TNKLVS
A7AG77 YASEDRVEATMPVDERTRQPFGILHGGATLALAETVAGLGSMILAKPDEIVVGMQVSGNHMSSAHEGDTVRAVGTIIHKGRSSHVWNVDVFTS-TDKLVS
A6LB33 YASEDRVEATMPVDERTRQPFGILHGGATLALAETVAGLGSMILAKPDEMVVGMQVSGSHMSSAHEGDTVRAVGTIIHKGRSSHVWNVDIFTS-TDKLVS
         :******* *****:*******:**:.**:**:**:  .:*.*:     **:****.*:*** ***.:*. .:::*:***:***.:::::: :.**:.

Q7MU91 TCRVLNSILKQR
A5ZGV9 SIRVVNSVMKKR
Q89YN2 SIRVVNSVMKKR
A7M602 SIRVVNSVMKKR
A7VAE9 SIRVVNSVIKKR
Q64WR2 SIRVVNSILKKR
Q5LFS7 SIRVVNSILKKR
B0NN98 SVRVVNSVLKKR
A7AG77 SIRVVNSILKKK
A6LB33 SVRVINSILKKK
       : **:**::*::
```

Figure 2.2 The gene map of the menaquinone biosynthetic pathway operon. B. An alignment of sequences of HAD-thioesterase fusion proteins.

The intervening sequence (residues 266-289) might serve as a solvated linker that connects the HADSF domain to the hotdog fold domain. The HADSF is comprised of phosphotransferases, a majority of which are phosphatases. The hotdog fold superfamily is comprised primarily of acyl-CoA and acyl-carrier protein thioesterases.

Although the menaquinone pathway in *E. coli.* has been investigated thoroughly, the enzyme(s) acting to remove CoA from DHN-CoA has not been clearly identified yet. Initially, CoA was thought to be cleaved non-catalytically following the cyclization catalyzed by MenB *(17).* The observation that the product of MenB catalyzed reaction was found to be DHN-CoA instead of DHN invalidated this mechanistic proposal *(1).* An α/β-fold enzyme termed MenH, which can also act as a thioesterase, was identified through inspection of the menaquinone biosynthetic operon of *E. coli (10).* In spite of the fact that no direct evidence shows that MenH or its orthologs displays activity against DHN-CoA, this protein was assumed to be the missing CoA thioesterase *(7).* In 2008, Ming et al. demonstrated that MenH actually catalyzes the conversion of 2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylate to (1R,6R)-2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate, which is an earlier step in the menaquinone biosynthetic pathway (18). In early 2009, Joshua et al. reported a hot-dog thioesterase, Slr0204 from *Synechocystis,* which participates in the biosynthesis of phylloquinone, as the enzyme responsible for the hydrolysis of DHN-CoA *(19)*

In the study described below, we demonstrated that the C-terminal hotdog fold domain of the *Bacteroides* fusion protein catalyzes the hydrolysis of DHN-CoA and thus functions in the menaquinone biosynthesis pathway.

2.2 Experiments and Materials

2.2.1 Materials

All chemicals were obtained from Sigma-Aldrich. Primers, T4 DNA ligase, and restriction enzymes were purchased from Invitrogen, *Pfu*, *Pfu* Turbo polymerases were from Stratagene, the pET23a and pET-3A vector kits were from Novagen, the Geneclean Spin and the Qiaprep Spin Miniprep kits were from Qiagen, Genomic DNA from *Bacteroides thetaiotaomicron VPI-5482* (ATCC-29148D) was from ATCC, and DEAE Sepharose was from Amersham Biosciences. All aqueous solutions were prepared using deionized water. Biosynthesis products were analyzed and separated by using a Shimadzu LC-20AB HPLC system equipped with Beckman Coulter ultrasphere C18 analytical column and Restek Ultra II C18 column.

^{1}H and ^{13}C spectra were recorded using a Bruker Avance 500 spectrometer. ^{1}H NMR data are reported as follows: chemical shifts (ppm), and multiplicities (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet). ^{13}C NMR chemical shifts are reported in ppm relative to NH_4HCO_3 (160 ppm). Mass spectrometric data were obtained at the University of the New Mexico mass spectrometry facility.

2.2.2 Preparation of the recombinant HAD-thioesterase fusion proteins PG1653 and BF1314.

The genes encoding the HAD-thioesterase fusion protein PG1653 (NCBI accession NP905773, Swiss Pro accession: Q7MU91) and BF1314 (NCBI accession YP098598, Swiss Pro accession: Q5LFS7) were amplified from *Porphyromonas gingivalis W83* and *Bacteroides fragilis* genomic DNAs by using PCR and then subcloned into a pET-28a vector, which adds a His_6 tag to the N-terminus of each protein. The oligonucleotide primers were 5'-CAA ATT TTA CTG CAT ATG AAG TAT AAA ATC-3' (PG1653 forward), 5'-GGA AAA AAC GTC CTC GAG CCG CTG TTT CAG GAT-3' (PG1653 reverse), 5'-GAG CAT ATG AAG TAT AAC TGT TAG TCC-3' (BF1314 forward) and 5'-TTG AAG CTT CTA ACG TCA TCT TTT CTT-3' (BF1314 reverse), which contain restriction the endo-nuclease cleavage sites *Nde I* and *Xho I*. The integrities of the subcloned genes were demonstrated by using full length DNA sequencing.

Gene expressions were performed by using *E. coli* BL21 (DE3) competent cells. Following growth in 3 L of LB media containing 50 μg/mL of Kanamycin at 37 °C for 6 h with 220 rpm mixing, induction was initiated by the addition of 0.2 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG). After incubation for 12 h at 20 °C and 170 rpm mixing, the cells were harvested by using centrifugation (6,500 rpm for 15 min at 4 °C), resuspended in 300 mL of lysis buffer (50 mM HEPES, 300 mM NaCl, 10 mM imidazole, pH 8.0) and lysed by 2 passages through a French Press cell (12,000 psi). The cell lysate was centrifuged at 20,000 rpm for 60 min at 4 °C. The supernatant containing the fusion proteins was eluted through a Ni^{+2} affinity column. The column was washed with 200 mL of lysis buffer and 200 mL of wash buffer (50 mM HEPES, 300 mM NaCl, 20 mM imidazole, pH 8.0) before introduction of the fusion proteins and subsequently

eluted with 100 mL of elution buffer (50 mM HEPES, 300 mM NaCl, 250 mM imidazole, pH 8.0). The protein containing fractions were concentrated by using a 10 kDa Macrosep centricon (Pall Filtron). Imidazole was removed by dialysis in 50 mM HEPES, 300 mM NaCl, pH 7.5. The purities of the proteins were determined by using SDS-PAGE gel electrophoresis and their concentrations were determined by using the Bradford method (absorbance at 280 nm, PG1653 ε = 21478 $M^{-1}cm^{-1}$ and BF1314 ε = 24579 $M^{-1}cm^{-1}$).

2.2.3 Preparation of recombinant O-succinylbenzoyl-coenzyme A ligase

The genes encoding the O-succinylbenzoyl-coenzyme A ligase (MenE) from *E. coli* (K12), were amplified from genomic DNA by using PCR and cloned into the pET-23a vector, which adds a His_6 tag to the C-terminus of the protein for later purification. The oligodeoxynucleotide primers used in gene amplification were 5'-GGA GCG GTT CAT ATG ATC TTC TCT GAC-3' (MenE forward), 5'-GAA GTG AAA CTC GAG GGT TAA ATA TTG CTG ACG TTG CAC-3' (MenE reverse), which contain the restriction endonuclease cleavage sites *NdeI* and *XhoI*. Plasmid DNAs were purified using a Qiaprep Spin Miniprep Kit. Gene sequencing was carried out at the Center for Genetics in Medicine, University of New Mexico School of Medicine.

Gene expression was carried out in transformed *E. coli* BL21(DE3) cells. Following growth in 3 L of LB media containing 50 μg/mL carbarnicillin at 37 °C for 12 h with mixing at 220 rpm, induction was initiated by the addition of 1 mM isopropyl-1-thio-β-D-galactopyranoside. After incubation for 10 h at 20 °C and 170 rpm, the cells were harvested by centrifugation (6500 rpm for 15 min at 4 °C), resuspended in 300 mL of lysis

buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, 1 mM DTT, pH 8.0) and lysed by 2 passages through a French Press cell (12,000 psi). The lysate was centrifuged at 20,000 rpm for 90 min at 4 °C, and the supernatant containing MenE was passed through a Ni^{+1} affinity column for His_6-tagged protein binding. The column was washed with 100 mL of lysis buffer and 100 mL of wash buffer (50 mM NaH_2PO_4, 600 mM NaCl, 50 mM imidazole, 1 mM DTT, pH 8.0) before MenE was subsequently eluted using 100 mL of elution buffer (50 mM NaH_2PO_4, 600 mM NaCl, 250 mM imidazole, 1 mM DTT, pH 8.0). The eluant was concentrated by using a 10 kDa Macrosep centricon (Pall Filtron) device. Imidazole was removed by chromatography on a S-200 resin using 50 mM NaH_2PO_4, 600 mM NaCl, pH 7.5 as the elution buffer. The purity of the protein was determined by using SDS-PAGE and its concentration was determined by using the Bradford method (absorbance at 280 nm, $\varepsilon = 104{,}770\ M^{-1}\ cm^{-1}$).

2.2.4 Preparation of recombinant 1,4-dihydroxynaphthoyl-CoA synthase

The genes encoding 1,4-dihydroxynaphthoyl-CoA synthase (MenB) from *E. coli* (K12), were amplified from genomic DNA by using PCR and cloned into a pET-14b vector, which adds a His_6 tag to the N-terminus of the protein for later purification. The oligodeoxynucleotide primers used in gene amplification were 5'-CAC CAT ATG ATT TAT CCT GAT GAA GCA-3' (MenB forward), 5'-ATA CCT GCG CGC CTC GAG TTA CGG ATT CCG-3' (MenB reverse), which contain the restriction endonuclease cleavage sites *NdeI* and *XhoI*. Plasmid DNAs were purified using a Qiaprep Spin Miniprep Kit. Gene sequencing was carried out at the Center for Genetics in Medicine, University of New Mexico School of Medicine.

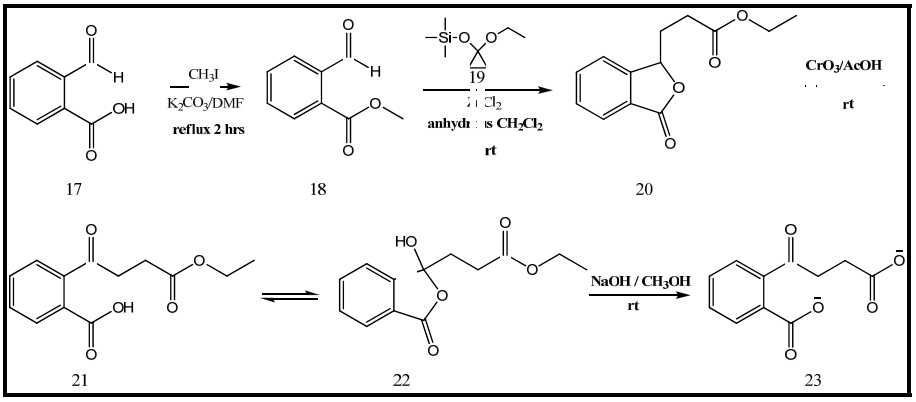
Gene expression was carried-out in transformed *E. coli* BL21(DE3) cells. Following growth in 3 L of LB media containing 50 μg/mL carbarnicillin at 37 °C for 12 h with mixing at 220 rpm, induction was initiated with the addition of 1 mM isopropyl-1-thio-β-D-galactopyranoside. After incubation for 10 h at 20 °C 170 rpm, the cells were harvested by centrifugation (6500 rpm for 15 min at 4 °C), resuspended in 300 mL of lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, 1 mM DTT, pH 8.0) and lysed by 2 passages through a French Press cell (12,000 psi). The cell lysate was centrifuged at 20,000 rpm for 90 min at 4 °C, and the supernatant containing MenB was passed through a Ni^{+1} affinity column. The column was washed with 100 mL of lysis buffer and 100 mL of wash buffer (50 mM NaH_2PO_4, 600 mM NaCl, 50 mM imidazole, 1 mM DTT, pH 8.0) before MenB was subsequently eluted using 100 mL of elution buffer (50 mM NaH_2PO_4, 600 mM NaCl, 250 mM imidazole, 1 mM DTT, pH 8.0). The protein was concentrated by using a 10 kDa Macrosep centricon (Pall Filtron) device. Imidazole was removed by chromatography on S-200 resin using 50 mM NaH_2PO_4, 600 mM NaCl, pH 7.5 as the elution buffer. The purity of the proteing was determined by using SDS-PAGE and its concentration was determined by using the Bradford method (absorption at 280 nm (MenB extinction coefficient $\varepsilon = 41{,}370\ M^{-1}\ cm^{-1}$).

2.2.5 Preparation of *O*-succinoylbenzoate 23

The route used for preparation of this substance is outlined in Scheme 2.2. A solution of 2-formylbenzoic acid **20** (27 g, 0.18 mol), iodomethane (49 g, 0.34 mol), and K_2CO_3 (14 g, 0.10 mol) in 200 mL of DMF was stirred at reflux for 1 h. The mixture was poured into 150 mL of water and then extracted with CH_2Cl_2. The organic layer was

washed successively with water, aqueous HCl, aqueous $NaHCO_3$, and brine, dried over anhydrous Na_2SO_4, and then concentrated *in vacuo* to give a colorless oil, which slowly crystallized to give pure methyl 2-formylbenzoate **18** (30.2 g, 100%). ^{1}H NMR (500 MHz, d_4-$CDCl_3$): δ 10.57 (s, 1H), 7.92(t, 2H), 7.60 (m, 2H), 3.93 (s, 3H). This spectrum matched that previously reported for **18** *(20)*.

Scheme 2.2 Organic synthesis of O-succinoylbenzoate



A solution of **18** (2 g, 0.011 mol) in dry dichloromethane (100 mL) containing suspended anhydrous zinc chloride (8.16 g, 0.060 mol) was stirred at 0 °C for 30 min. To this solution was added 1-ethoxy-1-trimethylsiloxycyclopropane **19** (4.2 g, 0.024 mol) dropwise at 0 °C. The mixture was warmed to room temperature and stirred over night. After addition of dilute HCl, the organic layer was separated and concentrated *in vacuo*, giving a residue that was subjected to flash chromatography on silica gel using 5:1 hexane:acetone as the eluant to yield the lactone **20**. ^{1}H NMR (500 MHz, d_4-$CDCl_3$): δ 7.81 (d, 1H), 7.63 (s, 1H), 7.48(s, 1H), 7.42 (d, 1H), 5.50 (t, 1H), 4.05 (t, 2H), 2.39 (m, 3H), 1.94 (m, 1H), 1.18 (s, 3H). ^{13}C NMR (125 MHz, d_4- $CDCl_3$): δ172.31, 169.98,

149.06, 133.99, 129.16, 125.53, 121.76, 79.85, 60.47, 29.66, 29.24, 13.97. These spectra matched those previously reported for **20** *(20)*.

A solution of **20** (1 g, 4.27 mmol) in glacial acetic acid (20 mL) containing chromium trioxide (2 g, 0.020 mol) was stirred at room temperature for 12 h. The mixture was diluted with water and extracted with ethyl acetate. Concentration of the extract *in vacuo* gave a residue that was subjected to silica gel chromatography using 2:1 hexane:acetone as the eluant to yield ethyl *o*-succinoylbenzoate **21**. ^{1}H NMR (500 MHz, d_4-$CDCl_3$): δ 7.78 (s, 1H), 7.66 (d, 1H), 7.53 (t, 2H), 4.10 (d, 2H), 2.70 (d, 1H), 2.52 (s, 1H), 2.43 (s, 1H), 2.24 (s, 1H), 1.22 (s, 3H). ^{13}C NMR (125 MHz, d_4- $CDCl_3$): δ 204.48, 174.12, 168.32, 148.91, 134.66, 130.56, 126.36, 125.40, 122.23, 106.49, 61.23, 33.90, 28.98, 14.00. These spectra matched those previously reported for **21** *(20)*.

A solution of **21** (1 g, 4.0 mmol) in 50 mL methanol containing 3 mL of 50% (w/v) aqueous NaOH was stirred at room temperature overnight. Concentration *in vacuo* gave compound O-succinylbenzoate **23**. ^{1}H NMR (500 MHz, d_4-H_2O): δ 7.67 (d, 1H), 7.61 (s, 2H), 7.57 (q, 1H), 3.23 (t, 2H), 2.57 (t, 2H), ^{13}C NMR (125 MHz, d_4- H_2O): δ 201.61, 174.64, 169.56, 130.74, 124.31, 122.32, 120.59, 120.22, 31.06, 27.18, 24.88. MS: *m/z* calculated for $C_{11}H_{18}O_3$ $[M+H\}^+$, 221.03; found 221.04.

2.2.6 Enzymatic synthesis of 1,4-dihydroxynaphthoyl-CoA

1,4-Dihydroxynaphthoyl-CoA was biosynthesized by using the menaquinone pathway proteins MenE and MenB. To a solution of *o*-succinylbenzoate (26.1 mg, 100 mmol), coenzyme A (44 mg, 56 mmol), and ATP (30.9 mg, 56 mmol) in 470 μL of reaction buffer (20 mM Tris, 1 mM $MgCl_2$, 50 mM $NaHCO_3$, pH 7.8) was added 10 μL

MenB (2 mM) and 20 μL MenE (300 uM). The resulting mixture was incubated at 37 °C for 30 min and filtered by using a Microcon ultracel centrifuge filter (pore size 30 kDa). The filtrate was subjected to Shimadzu LC-20 HPLC using a Beckman Coulter ultrasphere C18 analytical column and a linear gradient from (100% solvent A) to (60% solvent A plus 40% solvent B) over 45 min at a flow rate of 1 mL/min. Solvent A was a 50 mM ammonium acetate (pH 5.9) and solvent B methanol. Fractions containing DHN-CoA were combined and lyophilized to give solid 1,4-dihydroxynaphthoyl-CoA. ^{1}H NMR (500 MHz, D_2O): δ 8.69 (s, 1H), 8.63 (s, 1H), 8.47 (s, 1H), 8.41 (s, 1H), 7.91 (s, 1H), 7.78 (s, 1H), 6.24 (d, 1H), 6.14 (d, 1H), 4.66 (s, 1H), 4.32 (s, 2H), 4.06 (s, 1H), 3.91 (s, 1H), 3.74 (s, 1H), 3.68 (s, 2H), 3.51 (m, 4H), 2.79 (s, 1H), 2.50 (s, 1H), 2.33 (s, 1H), 0.98 (s, 3H), 0.87 (s, 3H). Mass spectrometric analysis: *observed* 954.13 Da; *calculated* (M^+H for $C_{32}H_{42}N_7O_{19}P_3S$) 954.15.

2.2.7 Preparation of the truncated HAD domain of PG1653

The HAD domain of the fusion protein PG1653 was prepared by using site-directed mutagenesis. The codon for Val271 was mutated to a stop codon TAA through quick change techniques. Oligonucleotide primers 5'- CGA ATA TGA AGC CTA ACC TTT CTC CGT AGA -3' (forward), and 5'- TCT ACG GAG AAA GGT TAG GCT TCA TAT TCG -3' (reverse) were used in the reaction. Plasmid DNA was purified using a Qiaprep Spin Miniprep Kit. Gene sequencing was carried out at the Center for Genetics in Medicine, University of New Mexico School of Medicine.

Gene expression was carried out in transformed BL21 (DE3) cells. Following growth in 3 L of LB media containing 50 μg/mL Kanamycin at 37 °C for 6 h with mixing

at 220 rpm, induction was initiated by the addition of 0.2 mM isopropyl-1-thio-β-D-galactopyranoside. After incubation for 12 h at 20 °C 170 rpm, the cells were harvested by centrifugation (6500 rpm for 15 min at 4 °C), resuspended in 300 mL of lysis buffer (50 mM HEPES, 300 mM NaCl, 10 mM imidazole, pH 8.0) and lysed by 2 passages through a French Press cell (12,000 psi). The lysate was centrifuged at 20,000 rpm for 60 min at 4 °C and the supernatant containing PG-HAD was passed through a Ni^{+2} affinity column. The column was washed with 200 mL of lysis buffer and 200 mL of wash buffer (50 mM HEPES, 300 mM NaCl, 20 mM imidazole, pH 8.0) before elution with 100 mL of a mixture of 50 mM HEPES, 300 mM NaCl, and 250 mM imidazole (pH 8.0). The protein solution was concentrated by using a 10 kDa Macrosep centricon (Pall Filtron) device. Imidazole was removed by dialysis in 50 mM HEPES, 300 mM NaCl (pH 7.5). The protein purity was verified by using SDS-PAGE and the protein concentration was determined by using the Bradford method and by measuring the absorption at 280 nm (ε =16,765 M^{-1} cm^{-1}).

2.2.8 Preparation of the truncated thioesterase domain of PG1653

The genes encoding the thioesterase domain of fusion protein PG1653 gene (NCBI accession NP905773, Swiss Pro accession: Q7MU91) was amplified from *Porphyromonas gingivalis W83* and P*fu* DNA polymerase and P*fu* DNA polymerase. The PCR product was cloned into the PET-28b vector, which was used to transform competent *E. coli.* BL21 (DE3) cells. Oligonucleotide primers 5'-TAT GAA GCC CAT ATG TTC TCC GTA GAA-3' (forward) and 5'-ATG GGA AAA AAG CTT TTA CCG CTG TTC AGG-

3'(reverse), containing the restriction endonuclease cleavage sites *NdeI* and *HindIII,* were used in the PCR reactions. The Pro272 codon was mutated to a start codon.

Gene expression was performed using transformed BL21 (DE3) cells. Following growth in 3 L of LB media containing 50 μg/mL Kanamycin at 37 °C for 6 h with mixing at 220 rpm, induction was initiated by the addition of 0.2 mM isopropyl-1-thio-β-D-galactopyranoside. After incubation for 12 h at 20 °C 170 rpm, the cells were harvested by centrifugation (6500 rpm for 15 min at 4 °C), resuspended in 300 mL of lysis buffer (50 mM HEPES, 300 mM NaCl, 10 mM imidazole, pH 8.0) and lysed by 2 passages through a French Press cell (12,000 psi). The lysate was centrifuged at 20,000 rpm for 60 min at 4 °C, and the supernatant was passed through a Ni^{+2} affinity column. The column was washed with 200 mL of lysis buffer and 200 mL of wash buffer (50 mM HEPES, 300 mM NaCl, 20 mM imidazole, pH 8.0) before using 100 mL of elution buffer (50 mM HEPES, 300 mM NaCl, 250 mM imidazole, pH 8.0). SDS-PAGE gel analysis shows that all the PG-TE was present in the pellet.

2.2.9 Mutagenesis of the PG1653 fusion protein and the truncated HAD domain

The F11D mutants were prepared by using quick change mutagenesis. Oligonucleotide primers 5'- TA GTC ATA GAT GTA GAC AAC ACC CTG CTC A -3' (forward), 5'- T GAG CAG GGT GTT GTC TAC ATC TAT GAC TA -3' (reverse). The gene sequences were confirmed at the Center for Genetics in Medicine, University of New Mexico School of Medicine.

Gene expression was carried out in BL21 (DE3) cells. Following growth in 3 L of LB media containing 50 μg/mL Kanamycin at 37 °C for 6 h with mixing at 220 rpm,

induction was initiated by the addition of 0.2 mM isopropyl-1-thio-β-D-galactopyranoside. After incubation for 12 h at 20 °C 170 rpm, the cells were harvested by centrifugation (6500 rpm for 15 min at 4 °C), resuspended in 300 mL of lysis buffer (50 mM HEPES, 300 mM NaCl, 10 mM imidazole, pH 8.0) and lysed by 2 passages through a French Press cell (12,000 psi). The lysate was centrifuged at 20,000 rpm for 60 min at 4 °C, and the supernatant was passed through a Ni^{+2} affinity column. The column was washed with 200 mL of lysis buffer and 200 mL of wash buffer (50 mM HEPES, 300 mM NaCl, 20 mM imidazole, pH 8.0) before using 100 mL of elution buffer (50 mM HEPES, 300 mM NaCl, 250 mM imidazole, pH 8.0). The proteins solution were concentrated by using 10 kDa Macrosep centricon (Pall Filtron) devices. Imidazole was removed by dialysis in 50 mM HEPES, 300 mM NaCl (pH 7.5). The proteins' purities were verified by using SDS-PAGE and their concentrations were determined by using the Bradford method and by measuring the absorption at 280 nm (ε =16500 M^{-1} cm^{-1} for HAD domain mutant and 25500 M^{-1} cm^{-1} for fusion protein mutant)

2.2.10. Preparation of the truncated HAD domain of BF1314

The HAD domain of the fusion protein PG1653 was prepared by using site-directed mutagenesis. The Arg273 codon was mutated to a stop codon TAA using quick change techniques. Oligonucleotide primers 5'- CTT TCT GAG ATT TAA CCG GCA GAG ATTCCT-3' (forward) and 5'- AGG AAT CTC TGC CGG TTA AAT CTC AGA AAG-3' (reverse). Plasmid DNA was purified using a Qiaprep Spin Miniprep Kit. Gene sequencing was carried out at Mclab.

Gene expression was performed using BL21 (DE3) competent cells. Following growth in 3 L of LB media containing 50 μg/mL Kanamycin at 37 °C for 6 h with mixing at 220 rpm, induction was initiated by the addition of 0.2 mM isopropyl-1-thio-β-D-galactopyranoside. After incubation for 12 h at 20 °C 170 rpm, the cells were harvested by centrifugation (6500 rpm for 15 min at 4 °C), resuspended in 300 mL of lysis buffer (50 mM HEPES, 300 mM NaCl, 10 mM imidazole, pH 8.0) and lysed by 2 passages through a French Press cell (12,000 psi). The lysate was centrifuged at 20,000 rpm for 60 min at 4 °C and the supernatant was passed through a Ni^{+2} affinity column. The column was washed with 200 mL of lysis buffer and 200 mL of wash buffer (50 mM HEPES, 300 mM NaCl, 20 mM imidazole, pH 8.0) before elution with 100 mL of a mixture of 50 mM HEPES, 300 mM NaCl, and 250 mM imidazole (pH 8.0). The protein solution was concentrated by using a 10 kDa Macrosep centricon (Pall Filtron) device. Imidazole was removed by dialysis in 50 mM HEPES, 300 mM NaCl (pH 7.5). The protein purity was verified by using SDS-PAGE and its concentration was determined by using the Bradford method and by measuring the absorption at 280 nm (ε =19335 M^{-1} cm^{-1}).

2.2.11. Preparation of truncated thioesterase domain of BF1314

The gene encoding the HAD domain of the fusion protein BF1314 gene (NCBI accession YP098598, Swiss Pro accession: Q5LFS7) from *Bacteroides fragilis* was amplified by using PCR with genomic DNA from *Bacteroides fragilis* and P*fu* DNA polymerase. The PCR product was cloned into the PET-28b vector, which was used to transform using *E. coli.* BL21 (DE3). Oligonucleotide primers 5'- CTT TCT GAG ATT CAT ATG GCA GAG ATT CCT-3' (forward) and 5'-ATG GGA AAA AAG CTT TTA

CCG CTG TTC AGG-3'(reverse), containing the restriction endonuclease cleavage sites *NdeI* and *HindIII,* were used in the PCR reactions. The Pro274 codon was mutated to a start codon. Gene sequencing was carried out at Mclab.

Gene expression was performed using BL21 (DE3) competent cells. Following growth in 3 L of LB media containing 50 μg/mL Kanamycin at 37 °C for 6 h with mixing at 220 rpm, induction was initiated by the addition of 0.2 mM isopropyl-1-thio-β-D-galactopyranoside. After incubation for 12 h at 20 °C 170 rpm, the cells were harvested by centrifugation (6500 rpm for 15 min at 4 °C), resuspended in 300 mL of lysis buffer (50 mM HEPES, 300 mM NaCl, 10 mM imidazole, pH 8.0) and lysed by 2 passages through a French Press cell (12,000 psi). The lysate was centrifuged at 20,000 rpm for 60 min at 4 °C, and the supernatant was passed through a Ni^{+2} affinity column. The column was washed with 200 mL of lysis buffer and 200 mL of wash buffer (50 mM HEPES, 300 mM NaCl, 20 mM imidazole, pH 8.0) before using 100 mL of elution buffer (50 mM HEPES, 300 mM NaCl, 250 mM imidazole, pH 8.0). The protein solution was concentrated by using a 10 kDa Macrosep centricon (Pall Filtron) device. Imidazole was removed by dialysis in 50 mM HEPES, 300 mM NaCl (pH 7.5). The protein purity was verified by using SDS-PAGE and its concentration was determined by using the Bradford method and by measuring the absorption at 280 nm (ε =8480 M^{-1} cm^{-1}).

2.2.12. Steady-state kinetic measurements

Thioesterase Activity Assay. For most substrates tested, thioesterase activities were determined by measuring the rate of formation of CoA using the 5,5'-dithio-bis-(2-nitrobenzoic acid) (DTNB) coupling assay. Reaction solutions (1 mL) initially containing

specified concentrations of the substrates, 1 mM DTNB, 0.15 M KCl and 50 mM K^+HEPES (pH 7.5), were monitored at 412 nm.

Phosphatase Activity Assay.The phosphatase activities of the HAD domains were determined by measuring the production of phosphate using an EnzChek phosphate assay kit from Invitrogen. Reaction solutions (1 mL) initially containing specified concentrations of substrate, purine nucleoside phosphorylase, MESG, 20 mM Tris-HCl, 2 mM $MgCl_2$, and 2 mM sodium azide (pH 7.5), were monitored at 360 nm.

MenE Activity Assay. The activities of substrates toward MenE were determined by using a coupled assay that detects the formation of AMP through the sequential reactions catalyzed by adenylate kinase, pyruvate kinase, and lactate dehydrogenase. Reaction solutions (1 mL) which initially contained specified concentrations of substrates, 15 mM $MgCl_2$, 200 μM NADH, 3 mM PEP, 5 mM KCl, 11 units adenylate kinase, 9 units pyruvate kinase and 9 units lactate dehydrogenase in 50 mM K^+HEPES (pH 7.5, 25 °C) were monitored 340 nm.

MenB Activity Assay. Reaction solutions which initially contained 50 mM K^+HEPES (pH 7.5), 200 μM ATP, 200 μM CoA, 1 mM $MgCl_2$, 1 μM MenE and specified concentrations of substrate were incubated for 5 min at 25 °C to covert the OSB to OSB-CoA. Subsequently MenB was added to the reaction mixture. The increase in absorbance at 392 nm resulting from the formation of DHN-CoA was monitored.

Steady-state Kinetic Constant Determination.

The initial velocities of the reactions were measured as a function of the concentration of the substrates, which varied in the range of 0.5 to 10-fold K_M. The data were fitted using an

enzyme kinetic module in sigmaplot to define V_{max} and K_M. The k_{cat} was calculated from the ratio of V_{max} and enzyme concentration.

$$V = V_{max}[S]/([S] + K_M)$$

where V is the initial velocity, V_{max} the maximum velocity, [S] the substrate concentration, K_M the Michaelis constant.

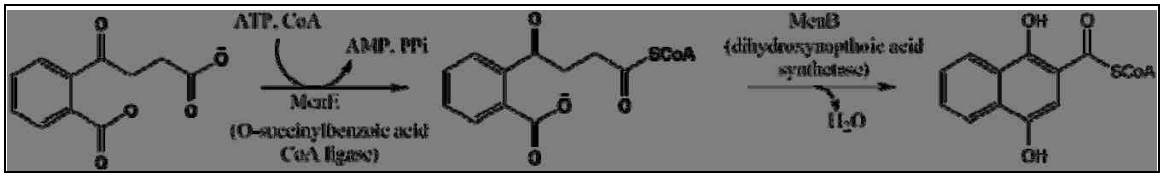
2.2.13 Protein native mass determination by size exclusion chromatography

Chromatographies were carried out using an ÄKTA P-920 FPLC system and Hiprep 16/60 Sephacryl S-200 HR column (GE Healthcare). The size exclusion column was equilibrated with 50 mM HEPES (pH 7.5), 500 mM NaCl, and 1 mM DTT. The column was calibrated using the gel-filtration low molecular weight and high molecular weight gel filtration calibration kit from Amersham Biosciences. Purified protein (2 mL, ca. 200 mM) was applied to the column.

2.3 Results and Discussion

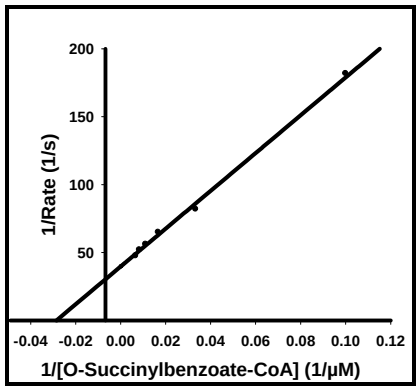
2.3.1 Biosynthesis of 1,4 dihydroxy Naphthoyl-CoA

Scheme 2.3 Biosynthesis of 1,4 dihydroxy naphthoyl-CoA

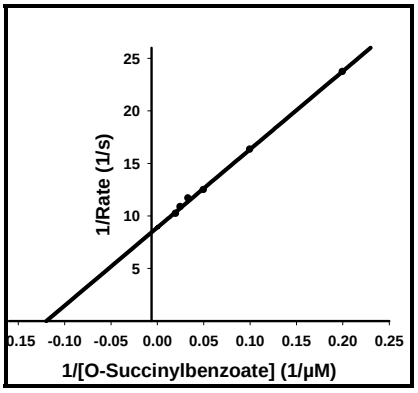


Synthetic O-succinylbenzoate (OSB) was converted to the corresponding CoA thioester using recombinant *E. coli* menE as the catalyst and (3 mM) ATP and (1 mM)

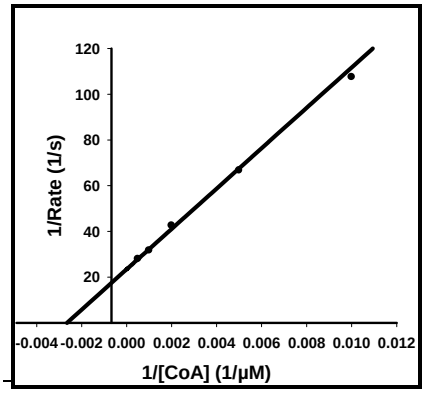
CoA as co-substrates (at pH 7.5 and 25 ^{o}C k_{cat} = 11 ± 0.2 s^{-1}, O-succinylbenzoate K_m = 8 ± 0.5 μM and O-succinylbenzoate k_{cat}/K_m = 1.4 x 10^6 M^{-1} s^{-1}). At a fixed OSB concentration of 50 μM and CoA concentration of 300 μM the k_{cat} = 9 ± 0.2 s^{-1} and ATP K_m = 380 ± 30 μM. At a fixed OSB concentration of 50 μM and ATP concentration of 400 μM k_{cat} = 10 ± 0.2 s^{-1} and K_m = 227 ± 16 μM (see Lineweaver-Burk plots in Figure. 2.3). The OSB-CoA generated *in situ* was converted to the DHN-CoA using recombinant MenB as catalyst (Scheme 2.3) (k_{cat} = 5 ± 0.2 s^{-1} and K_m = 35 ± 3 μM).



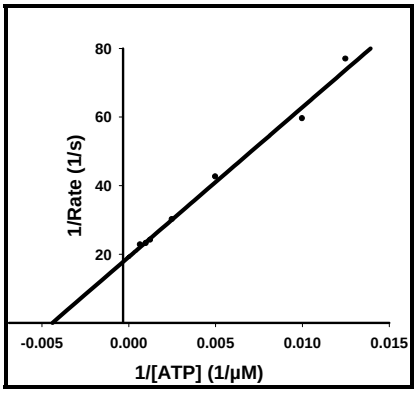
(a)



(b)



(c)



(d)

Figure 2.3 Lineweaver-Burk plots of the initial velocity data derived from (a)

MenB catalyzed reaction of OSB-CoA. (b) MenE catalyzed reaction of OSB (concentration varied), ATP (3mM) and CoA (1mM). (c) MenE catalyzed reaction of OSB (50 μM), ATP (400 μM) and CoA (concentrated varied). (d) MenE catalyzed reaction of OSB (50 μM), ATP (concentration varied) and CoA (300 μM).

The DNH-CoA was separated from the crude reaction mixture by HPLC (see Figure 2.4). A typical preparative reaction provided DNH-CoA in 60% yield.

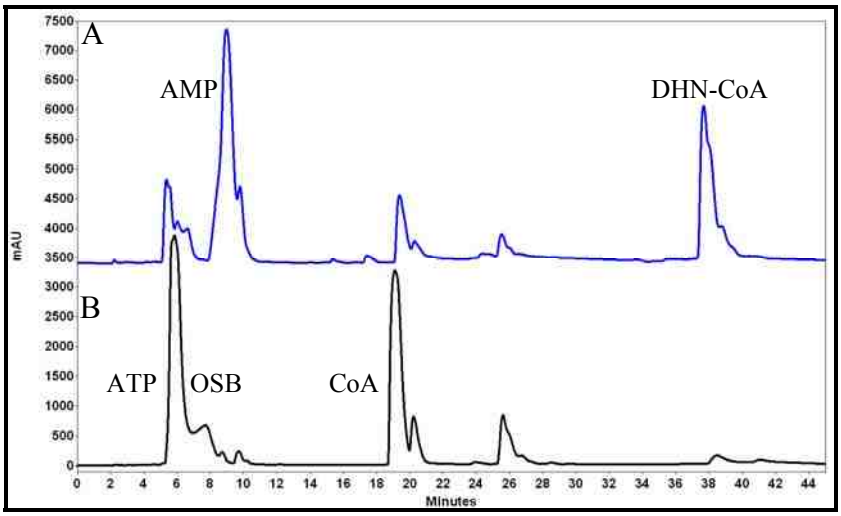


Figure 2.4. HPLC traces of MenB and MenE mixtures from catalyzed reactions of OSB to form DHN-CoA synthesis monitored at 292 nm. (Trace B is the mixture of 2.5 mM OSB, ATP and CoA without enzymes; Trace B is the MenE and MenB assisted biosynthesis of DHN-CoA. Reactions were carried out at 37 °C with 50 mM K^+HEPES, 2.5 mM ATP, 2.5 mM CoA, 1 mM $MgCl_2$, pH 7.5 containing 0.1 mM MenE and 0.2 mM MenB and 10 mM of the MenB activator sodium bicarbonate. Retention times: DNH-CoA 37.9 min, OSB 5.8 min, ATP 7.7 min and CoA 19.1 min.

The stability of DNH-CoA in 25 °C in 50 mM HEPES pH 7.5 buffer was examined. The HPLC trace of the reaction solution measured at 8 and 20 h of incubation (Figure 2.5) indicates a half life ~ 8 h.

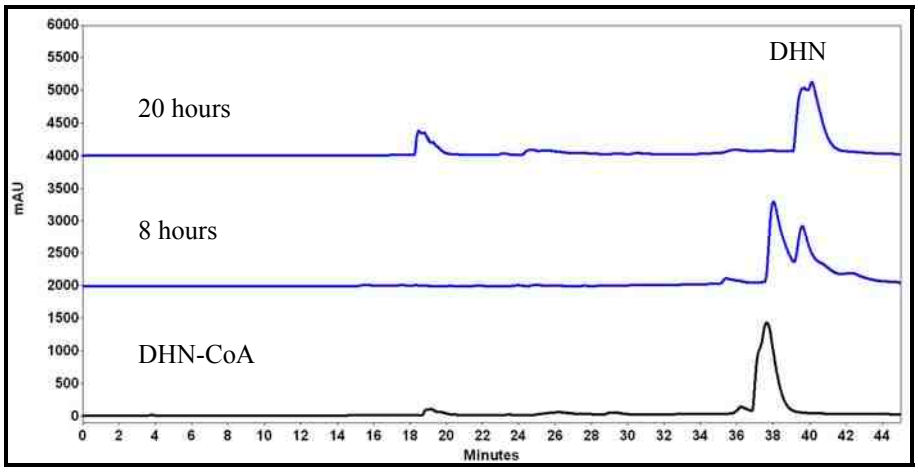
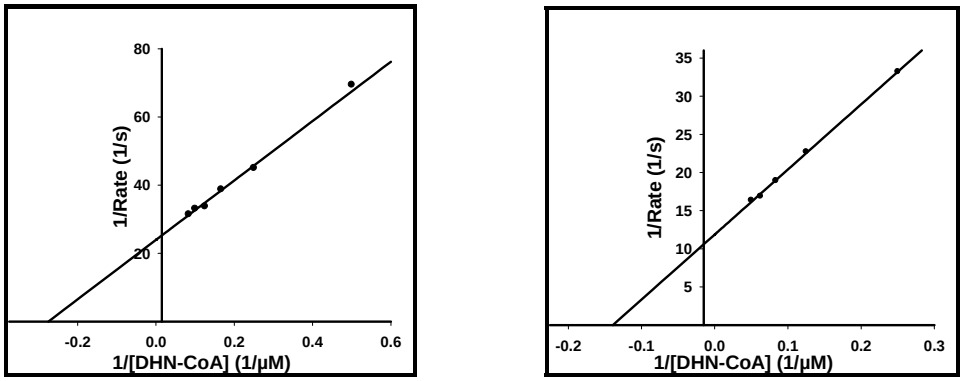


Figure 2.5. HPLC trace of the DHN-CoA decomposition reaction in 50 mM HEPES buffer (pH=7.5) at 25 °C.

2.3.2 Function of the thioesterase domain of the fusion protein determined by steady-state kinetic substrate screening

The steady-state kinetic constants (Table 2.1) of possible substrates for the *P. gingivalis* and *B. fragilis* HAD-thioesterase fusion protein (PG1653 and BF1314) were determined at pH 7.5 and 25 °C by carrying out initial velocity measurements, in which the enzyme is held at a fixed concentration while the concentration of the thioester substrate was varied in the range of 0.5-10 fold K_m.



(a) (b)

Figure 2.6 Lineweaver-Burk plots of a) PG1653 and b) BF1314 catalyzed DHN-CoA hydrolysis reactions in 50 mM K^+HEPES (pH 7.5, 25 ^oC)

Table 2.1 Steady-state kinetic constants (k_{cat}, K_m and k_{cat}/K_m) for substrates for hydrolysis reactions catalyzed by the HAD-thioesterase fusion protein PG1653 and BF1314 (50 mM K^+HEPES, pH 7.5, 25 ^oC) using the DTNB assay unless otherwise noted.

Substrate	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1}. s^{-1}$)
butyl-CoA	$< 1 \times 10^{-4}$	--------	--------
Benzoyl-CoA	$(3.0 \pm 0.1) \times 10^{-2}$	$(5.8 \pm 0.5) \times 10^{1}$	5.5×10^{2}
4-Hydroxybenzoyl-CoA	4.6 ± 0.1	$(8.0 \pm 0.5) \times 10^{2}$	5.8×10^{3}
4-Chlorobenzoyl-CoA	2.8 ± 0.1	$(1.2 \pm 0.1) \times 10^{2}$	2.3×10^{5}
4-Iodobenzoyl-CoA	5.4 ± 0.1	$(6.5 \pm 0.5) \times 10^{1}$	8.3×10^{5}
4-Methoxybenzoyl-CoA	$(2.5 \pm 0.2) \times 10^{-1}$	$(2.2 \pm 0.1) \times 10^{2}$	1.1×10^{3}
Naphthoyl-CoA	6.9 ± 0.2	$(1.7 \pm 0.2) \times 10^{2}$	4×10^{5}
1,4-Dihydroxynaphthoyl-CoA (PG1653)	$(1.0 \pm 0.1) \times 10^{1}$	4.0 ± 0.3	2.5×10^{6}
1,4-Dihydroxynaphthoyl-CoA (BF1314)	$(1.2 \pm 0.1) \times 10^{1}$	7.0 ± 0.5	1.7×10^{6}
3,4-Dihydroxy phenylacetyl-CoA	$(4.0 \pm 0.1) \times 10^{-2}$	$(2.2 \pm 0.2) \times 10^{2}$	1.6×10^{2}

4-Hydroxyphenylacetyl-CoA	$(1.7 \pm 0.1) \times 10^{-1}$	$(1.4 \pm 0.1) \times 10^{2}$	1.2×10^{3}
4-Hydroxyphenylacetyl-CoA[a]	$(1.5 \pm 0.1) \times 10^{-1}$	$(3.9 \pm 0.3) \times 10^{1}$	3.8×10^{3}
3-Hydroxyphenylacetyl-CoA	$(3.1 \pm 0.2) \times 10^{-1}$	$(1.7 \pm 0.1) \times 10^{2}$	1.8×10^{3}

[a] The direct assay was used

The UV spectrum of PG1653 and BF1314 showed a 280 nm λ_{max} with no additional absorbance at 260 nm. Likewise neither enzyme reacted with DTNB. Thus, CoA did not co-purify with either of the two proteins. PG1653 and BF1314 had k_{cat} and K_m values of 10.0 ± 0.3 s^{-1}, 4.0 ± 0.3 μM, and 12.0 ± 0.3 s^{-1}, 7.0 ± 0.5 μM, respectively, for reaction of DHN-CoA. The large respective k_{cat} /K_m values of 2.5×10^{6} and 1.7×10^{6} M^{-1} s^{-1} (Fig. 2.6a, 2.6b and Table 2.1) indicate that DNH-CoA is the physiological substrate. The substrate specificity profile determined for PG1653 supports this conclusion. Namely, catalysis is limited to aromatic CoA thioesters and among these the napthyl > benzoyl > phenyacetyl. Even removal of the two hydroxyl groups from the DNH-CoA napthyl ring significantly reduced the k_{cat}/K_m value (Table 2.1).

The BF1314 gene was genetically engineered to express the thioesterase domain as a standalone protein (residues 274-410) (BF-TE). This protein was purified and tested for catalytic activity towards the DHN-CoA substrate. No activity was observed.

2.3.3 Function of HAD domain of fusion protein determined by steady-state kinetic substrate screen

The function of the HAD domain of the HAD-thioesterase fusion proteins was investigated by screening catalytic activity towards the hydrolysis of a variety of

phosphate esters. BF1314 displayed a very low level of activity towards substrates that are most susceptible to spontaneous hydrolysis, namely carbamoyl phosphate, imidodiphosphate and acetyl phosphate (respective k_{cat}/K_m values of 54 M^{-1} s^{-1}, 37 M^{-1} s^{-1} and 16 M^{-1} s^{-1}). The organophosphate metabolites that were found to have no detectable substrate activity when tested at a concentration of 1 mM in solutions containing 10 μM BF1314, 1 mM $MgCl_2$ and 20 mM Tris-HCl (pH 7.5 and 25 ^oC) include 100 μM sodium azide.

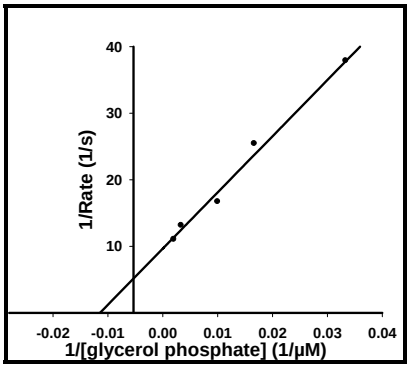
The BF1314 gene was genetically engineered to express the HAD domain as a standalone protein (residues 1 to 273). This protein (BF-HAD) displayed higher activities as compared to the fusion protein. For example, the k_{cat}/K_m value of 1.9 x 10^2 M^{-1} s^{-1} was observed for glycerol-phosphate at pH 7.5 (Table 2.2). Replacing the Mg^{2+} cofactor with Mn^{2+} did not significantly change the activity, whereas replacing it with Co^{2+} precluded the activity. The pH rate profile (Figure 2.7b) showed that the BF-HAD is maximally active at pH 5.5 (k_{cat}/K_m = 1.2 x 10^3 M^{-1} s^{-1}; Figure 2.7a and Table 2.2). Thus, the HAD domain has a low level of activity towards a phosphate ester metabolite which is somehow suppressed by the presence of the thioesterase domain.

Table 2.2 Steady-state kinetic data of HAD-thioesterase fusion protein BF1314 and BF-HAD (20 mM Tris, 1 mM metal ions, pH 7.5, 25 ^oC).

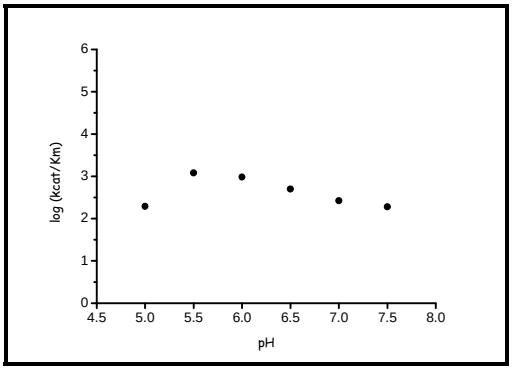
Substrate	BF1314			BF-HAD		
	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (M^{-1} s^{-1})	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m (M^{-1} s^{-1})
para-Nitrophenyl phosphate	< 1 x 10^{-4}	--------	--------	< 1 x 10^{-4}	--------	--------
Carbamoyl phosphate	(2.2 ± 0.1) x 10^{-2}	(4.4 ± 0.3) x 10^2	5.4 x 10^1	(4.1 ± 0.3) x 10^{-2}	(9.7 ± 0.8) x 10^2	4.2 x 10^1

Imidodi-phosphate	$(2.1 \pm 0.1) \times 10^{-1}$	$(5.7 \pm 0.5) \times 10^{3}$	3.7×10^{1}	$(3.8 \pm 0.2) \times 10^{-1}$	$(7.4 \pm 0.5) \times 10^{3}$	5.1×10^{1}
Acetyl-phosphate	$(8.0 \pm 0.2) \times 10^{-3}$	$(5.1 \pm 0.4) \times 10^{2}$	1.6×10^{1}	$(2.0 \pm 0.1) \times 10^{-2}$	$(8.7 \pm 0.6) \times 10^{2}$	2.3×10^{1}
Glycerol-phosphate	$< 1 \times 10^{-4}$	--------	--------	$(1.1 \pm 0.1) \times 10^{-2}$	$(6.3 \pm 0.2) \times 10^{1}$	1.7×10^{2}
Glycerol-phosphate [a]				$(1.0 \pm 0.1) \times 10^{-1}$	$(8.7 \pm 0.7) \times 10^{1}$	1.1×10^{3}

[a] Measured at pH 5.5.



(a)



(b)

Figure 2.7 (a) Lineweaver-Burk plots of the BF-HAD catalyzed glycerol-phosphate hydrolysis reaction in 20 mM MES (pH 5.5, 25 ^{o}C), k_{cat}/K_m = 1.1 x 10^3 M^{-1} s^{-1}; (b) pH profile of the BF-HAD catalyzed glycerol-phosphate hydrolysis reaction.

Except for the listed substrate in table 2.2, several other organophosphate esters are also tested, and both the fusion protein BF1314 the truncated HAD domain of BF1314 are inactive towards these substrates. These substrates are: pyrophosphate, erythrose 4-phosphate, fructose 1-phosphate, glucose 6-phosphate, glucose 1-phosphate, ATP, AMP, ADP, UMP, farnesyl monophosphate, farnesyl-pyrophosphate, fructose 1,6-bisphosphate, thiamine phosphate, threonine phosphate, serine phosphate, ribose 5-phosphate, arabinose 5-phosphate, 2-keto, 3-deoxy, 6-phosphogluconate, Glyceraldehye

3-phosphate, 6-phosphogluconic acid, 3-phospho glycerate, 2-phosphoglycolate, phosphoenolpyruvate.

Whereas the BF1314 active site contains a full set of stringently conserved catalytic residues, the PG1653 does not. Specifically, the key Asp acid/base is replaced by a Phe residue. Neither the wild-type enzyme of the F11D mutant displaced activity against the panel of organophosphate substrates tested with the BF1314.

2.3.4 Native molecular weight of HAD-thioesterase fusion protein BF1314 and truncated BF-HAD and BF-TE

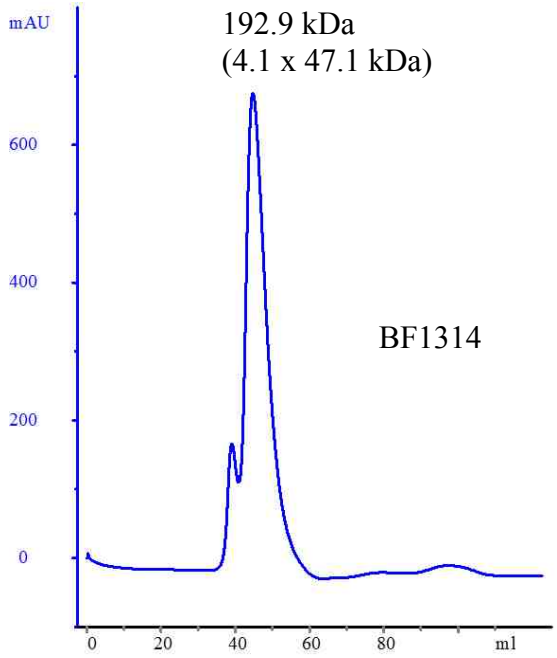
To better understand the function of the purpose of the HAD domain in the BF1314 fusion protein, the quaternary structures of *holo* enzyme and the two genetically engineered single domain proteins (BF-HAD and BF-TE) were determined by size exclusion chromatography (SEC). The calculated masses of the monomeric proteins ($M_{w,\ calc}$) and the observed ($M_{w,\ obs}$) masses of the native proteins in solution are listed in Table 2.3.

The *holo* BF1314 (residues 1-410) exclusively formed tetramers in solution (Figure 2.8a), and tetramerization was also observed for the purified BF-HAD (residues 1-273; Figure 2.8b). The BF-TE (residues 274-410; Figure 2.8c), on the other hand is a monomer. Notably, the minimal catalytic unit of a hotdog-fold thioesterase is a dimer because the active site is located at the subunit-subunit interface. Bacterial hotdog-fold thioesterases are dimer of dimer or trimer of dimer. HADs that possess cap domains are monomers or dimers. The closest homolog to the BF1314 HAD domain is the type C2B capped HAD phosphatase BT3352, which is described in Chapter 3. BT3352 is a

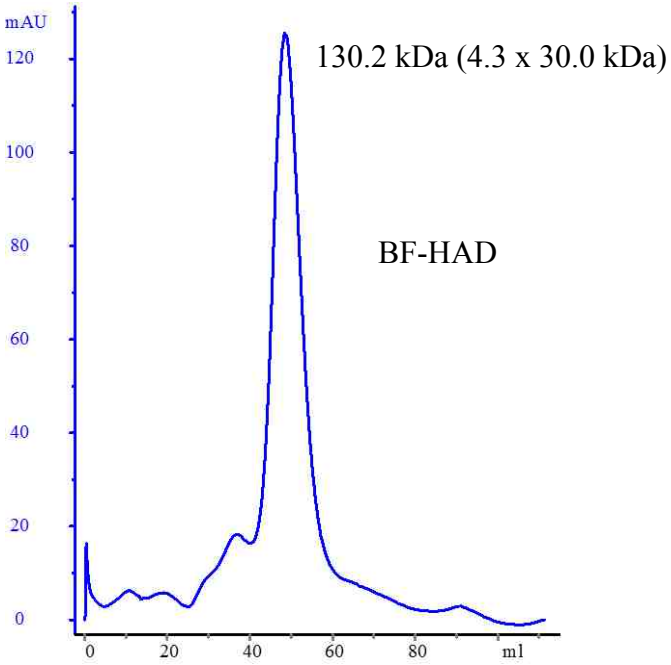
monomer. HADs dimerize either using the cap domain or the catalytic domain. To form a tetramer both the cap and catalytic domain would have participate. *A priori*, this would preclude normal cap function. This might explain why the HAD domain has little catalytic activity despite having the requisite catalytic residues (Table 2.2).

If the HAD domain functions to drive the dimerization of the thioesterase domain then the removal of the HAD domain would explain why the thioesterase exists as a monomer and as the monomer has an incomplete active site why the engineered thioesterase domain is catalytically inactive (Table 2.4). The ability of the HAD domain to drive dimerization of the thioesterase domain requires that the two domains are tethered. Specifically, the size exculsion chromatography of a mixture of BF-HAD and BF-TE separated the two proteins which indicates that the HAD and thioesterase domains do not associate with one another.

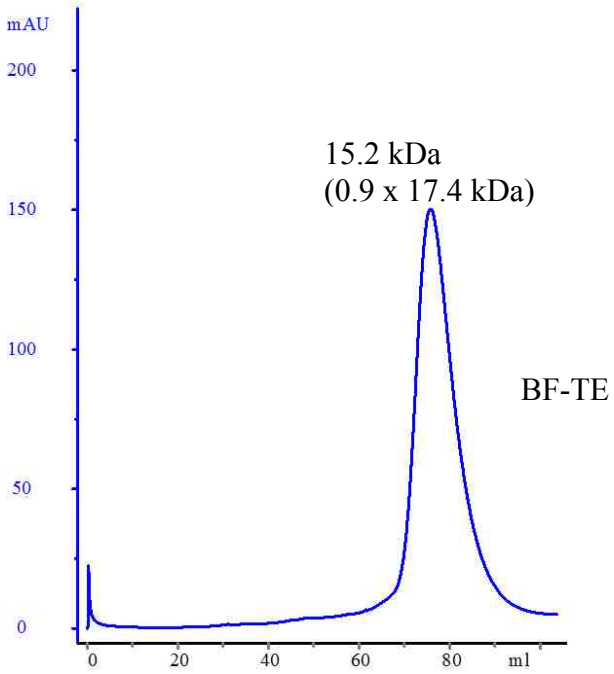
(a)



(b)


mAU
120
100
80
60
40
20
0
0
20
40
60
80
ml
130.2 kDa (4.3 x 30.0 kDa)
BF-HAD

(c)


mAU
200
150
100
50
0
0
20
40
60
80
ml
15.2 kDa
(0.9 x 17.4 kDa)
BF-TE

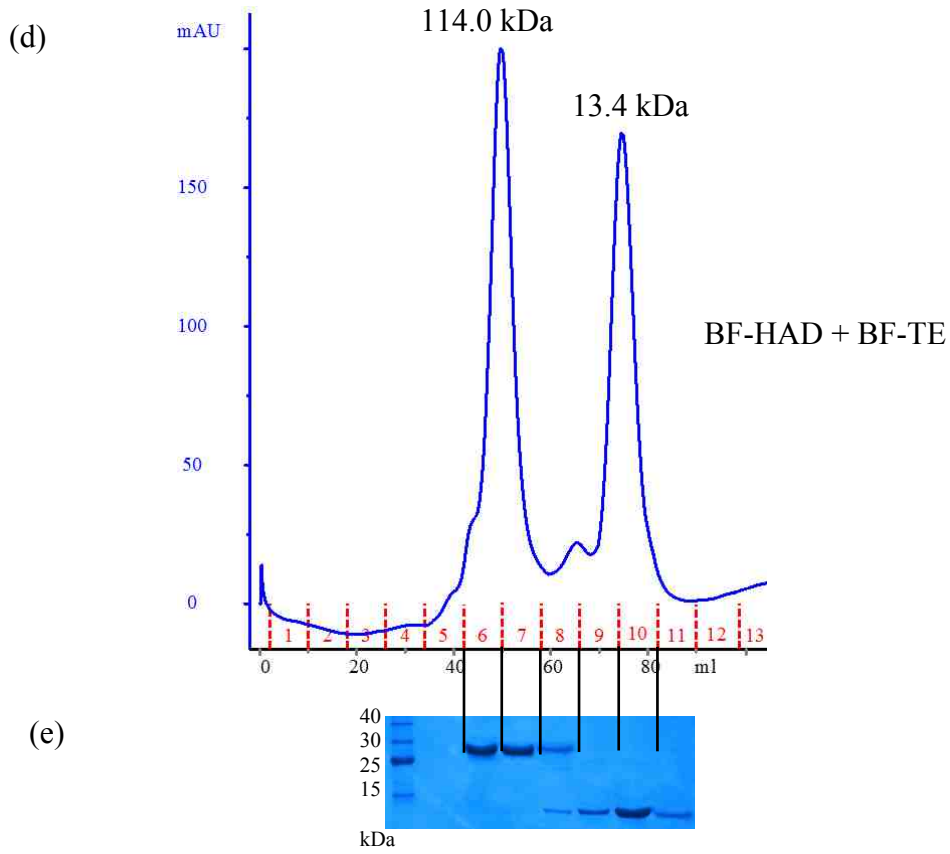


Figure 2.8 Size exclusion chromatography of (a) *holo* BF1314 (residues 1-410); (b) BF-HAD; (c) BF1314 BF-TE, residues 274-410); (d) mixture of BF-HAD and BF-TE; (e) 16% SDS-PAGE and Coomassie staining of (d).

Table 2.3. Molecular Masses holo BF1314, BF-HAD and BF-TE determined by size exculsion chromatography.

Enzymes	Monomer		Sec	Oligomer
	$M_{w,\ calc}$ (kDa)	$M_{w,\ obs}$ (kDa)	x-mer_{obs}	
BF1314	47.1	192.9	4.1	Tetramer
BF-HAD	30.0	130.2	4.3	Tetramer
BF-TE	17.4	15.2	0.9	Monomer

2.3.5 Effects between N-terminal HAD domain and C-terminal thioesterase domain on the kinetic activities of fusion protein and truncated proteins

To further investigate the interaction between the truncated N-terminal HAD domain and the C-terminal thioesterase domain, steady-state kinetic analysis was carried out BF-HAD phosphates activity in the presence and absence of BF-TE, and on the BF-TE thioesterase activity in the presence and absence of BF-HAD. Kinetic constants were determined using equimolar amounts of the two single domain constructs. The k_{cat}/K_m values reported in Tables 2.4 and 2.5 show that the presence of domain partner had essentially no effect on the phosphatase or thioesterase activities. This indicates that the two domains do not bind to one another, which is in agreement with the results from the size exclusion chromatography.

Table 2.4 Steady state kinetic values measured for the HAD-thioesterase fusion protein BF1314, BF-TE and a 1:1 mixture of BF-HAD and BF-TE for catalysis of 1,4-DHN-CoA thioester hydrolysis (50 mM K^+ HEPES, pH 7.5, 25 ^oC).

Enzymes	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1}\ s^{-1}$)
BF1314	$(1.2 \pm 0.1) \times 10^{1}$	7.0 ± 0.5	1.7×10^{6}
BF-TE	$< 1 \times 10^{-4}$	--------	--------
BF-TE + BF-HAD	$< 1 \times 10^{-4}$	--------	--------

Table 2.5 Steady state kinetic values measured for BF-HAD and a 1:1 mixture of BF-HAD and BF-TE phosphatase activity (50 mM K^+ HEPES, pH 7.5, 25 ^oC).

Substrate	BF-HAD	BF-HAD + BF-TE

	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1}\ s^{-1}$)	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1}\ s^{-1}$)
Carbamoyl phosphate	$(4.1 \pm 0.3) \times 10^{-2}$	$(9.7 \pm 0.8) \times 10^{2}$	4.2×10^{1}	$(4.3 \pm 0.1) \times 10^{-2}$	$(9.1 \pm 0.8) \times 10^{2}$	4.7×10^{1}
Imido-diphosphate	$(3.8 \pm 0.2) \times 10^{-1}$	$(7.4 \pm 0.5) \times 10^{3}$	5.1×10^{1}	$(4.0 \pm 0.3) \times 10^{-1}$	$(8.2 \pm 0.5) \times 10^{3}$	4.9×10^{1}
Acetyl-phosphate	$(2.0 \pm 0.1) \times 10^{-2}$	$(8.7 \pm 0.6) \times 10^{2}$	2.3×10^{1}	$(2.2 \pm 0.1) \times 10^{-2}$	$(9.2 \pm 0.6) \times 10^{2}$	2.4×10^{1}
Glycerol-phosphate	$(1.1 \pm 0.1) \times 10^{-2}$	$(6.3 \pm 0.2) \times 10^{1}$	1.7×10^{2}	$(1.1 \pm 0.1) \times 10^{-2}$	$(5.9 \pm 0.2) \times 10^{1}$	1.9×10^{2}

2.4 Conclusion

In the study described above the fusion proteins, PG1653 and BF1314 were shown to catalyze the hydrolysis of DHN-CoA to CoASH and DHN as a step in the menaquinone (vitamin K12). The DHN-CoA was biosythnesized by from synthetic OSB by the coupled reactions of OSB: CoA ligase and DHN-CoA synthase. Both PG1653 and BF1314 exhibited high catalytic efficiency in the hydrolysis of DHN-CoA. The BF1314 was found to have very weak phosphatse activity and the N-terminal HAD domain of PG1653 had none. The truncated HAD domain of BF1314 is able to dephosphorylate glycerol phosphate with $k_{cat}/K_m = 1.2 \times 10^3\ M^{-1}\ s^{-1}$ at pH 5.5. The truncated thiosterase domain of PG1653 was prone to precipatation wherease the truncated BF1314 thioesterase domain is stable and soluble in aqueous buffer. Nieghter truncate was catalytically active, which might be explained by the fact that they are monomeric and thus do not have a complate active site. Because the fusion protein is a tetramer we believe that the HAD domain functions as the oligomerization domain. This would require that it use both the cap and catalytic domain for oligerimization and this in turn might explain the low phosphatase activity. Noteably, the standalone DHN-CoA thioesterase from *Bacteroides vulgatus* (see Chapter 3) was found to be an octomer. The fusion protein is a product of a mutation

which fused adjacent HAD and thioesterase domians. The fact that the HAD domain is not required as a phosphatase is indicated by the loss of the essential acid/base in PG1653.

2.5 References

1. James J. Truglio, Karsten Theis, Yuguo Feng, Ramona Gajda, Carl Machutta, Peter J. Tonge, Caroline Kisker, (2003) Crystal structure of Mycobacterium tuberculosis MenB, a key enzyme in vitamin K2 biosynthesis, *J. Biol. Chem.,* **278, 43,** 42352-42360.

2. Dowd, P., Ham, S.W., Naganathan, S., and Hershline, R. (1995) Vitamin K and energy transduction: a base strength amplification mechanism, *Annu. Rev. Nutr.,* **15**, 419-440.

3. Bugel, S. (2008) Vitamin K, pp 393–416, Elsevier Academic Press Inc, San Diego.

4. Shanahan, C. M., Proudfoot, D., Farzaneh-Far, A., and Weissberg, P. L. (1998) The role of Gla proteins in vascular calcification, Crit. Rev. Eukaryot. Gene Exp., 8, 357–375.

5. Meganathan, R. (2001) Biosynthesis of menaquinone (vitamin K2) and ubiquinone (coenzyme Q): a perspective on enzymatic mechanisms, *Vita. Horm.,* **61**, 173-218.

6. Rowland, B., Hill, K., Miller, P., Driscoll, J., Taber, H. (1995) Structural organization of a Bacillus subtilis operon encoding menaquinone biosynthetic enzymes, *gene (Amst.),* **167,** 105-109.

7. Heide, L., Arendt, S., and Leistner, E. (1982) Enzymatic synthesis, characterization, and metabolism of the coenzyme A ester of o-succinylbenzoic acid, an intermediate in menaquinone (vitamin K2) biosynthesis, *J. Biol. Chem.,* **257,** 7396-7400.

8. Daruwala, R., Kwon, O., Meganathan, R., Hudspeth, M. E. (1996) A new isochorismate synthase specifically involved in menaquinone (vitamin K2) biosynthesis encoded by the menF gene, *FEMS Microbiol. Lett.,* **140**, 159-163.

9. Meganathan, R., Bentley, R. (1983) Thiamine pyrophosphate requirement for o-succinylbenzoic acid synthesis in Escherichia coli and evidence for an intermediate, *J. Bacteriol.,* **153,** 739-746.

10. Palaniappan, C., Sharma, V., Hudspeth, M. E., Meganathan, R. (1992) Menaquinone (vitamin K2) biosynthesis: evidence that the Escherichia coli menD gene encodes both 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylic acid synthase and alpha-ketoglutarate decarboxylase activities, *J. Bacteriol.,***174**, 8111-8118.

11. Sharma, V., Meganathan, R., Hudspeth, M.E. (1993) Menaquinone (vitamin K2) biosynthesis: cloning, nucleotide sequence, and expression of the menC gene from Escherichia coli., *J. Bacteriol.,* **175**, 4917-4921

12. Sharma, V., Hudspeth, M. E., and Meganathan, R. (1996) Menaquinone (vitamin K2) biosynthesis: localization and characterization of the menE gene from Escherichia coli *Gene* (*Amst.*), **168,** 43–48.

13. Chang, K. H., Xiang, H., and Dunaway-Mariano, D. (1997) Acyl-Adenylate Motif of the Acyl-Adenylate/Thioester-Forming Enzyme Superfamily : A Site-Directed Mutagenesis Study with the Pseudomonas Sp. Strain Cbs3 4-Chlorobenzoate-Coenzyme-A Ligase, *Biochemistry,* **36,** 15650–15659.

14. Meganathan, R., and Bentley, R. (1981) Identification of Bacillus subtilis men mutants which lack O-succinylbenzoyl-coenzyme A synthetase and

dihydroxynaphthoate synthase. *Biochemistry,* **20,** 5336–5340.

15. Sharma, V., Suvarna, K., Meganathan, R., and Hudspeth, M. E. (1992) Menaquinone (vitamin K2) biosynthesis: nucleotide sequence and expression of the menB gene from Escherichia coli, *J.Bacteriol.,* **174,** 5057–5062.
16. Kemp, L. E., Bond, C. S., and Hunter, W. N. (2002) Structure of 2C-methyl-D-erythritol 2,4- cyclodiphosphate synthase: an essential enzyme for isoprenoid biosynthesis and target for antimicrobial drug development, *Proc. Natl. Acad. Sci. U. S. A.,* **99,** 6591–6596.
17. Ming Jiang, Minjiao Chen, Zu-Feng Guo and Zhihong Guo (2010) A bicarbonate cofactor modulates 1,4-dihydroxy-2-naphthoyl-coenzyme a synthase in menaquinone biosynthesis of Escherichia coli, *J. Biol. Chem.,* **285**,30159-30169
18. Joshua R. Widhalm, Chloe van Oostende, Fabienne Furt, and Gilles J. C. Basset (2009) A dedicated thioesterase of the Hotdog-fold family is required for the biosynthesis of the naphthoquinone ring of vitamin K1, *Proc Natl Acad Sci USA*, 106, (14), 5599-5603
19. Gore V. G. ; Chordia M. D. ; Narasimhan N. S. (1990) Improved Syntheses of Shihunine, the Spiro Phthalide Pyrrolidine Alkaloid, *Tetrahedron*, 46 (7): 2483-2494.

Chapter 3. Structure-function Analyses of HAD Phosphatase BT3352 and Hotdog-fold Thioesterase BVU_2420

3.1 Introduction

In Chapter 2 I described the 2-domain fusion protein that functions as a 1,4-dihydroxynapthoyl-CoA thioesterase (DHN-CoA thioesterase) in the menaquinone biosynthetic pathway in *Bacteriodes thetaiotamicron* (BF1413) and the HAD domain-thioesterase domain fusion protein. The N-terminal domains of these two fusion proteins belong to the haloalkanoate dehalogenase (HAD) enzyme superfamily whereas the C-terminal domains which are responsible for the thioesterase activity belong to the hotdog-fold enzyme superfamily. The two fusion proteins proved to be poor subjects for structure determination. Moreover, the HAD domain displayed little to no phosphatase activity. So that I might carry out structure function analysis of the respective domains their closest sequence homologues were prepared for study. In this Chapter I report on the structure and function of the HAD phosphatase BT3352 from *B. thetaiotamicron* and on the oligomerization state and kinetic properties of the hotdog-fold thioesterase BVU2420 from *Bacteroides vulgatus.*

3.2 Materials and Methods

3.2.1 Gene Cloning, Expression, and Purification of Wild Type and Mutant BT3352 HAD Proteins

The gene encoding the BT3352 was amplified from the *B. thetaiotaomicron* genomic DNA by PCR and then cloned into the pET-28a vector, which adds a His_6 tag to

the protein N-terminus. Oligonucleotide primers containing the restriction endo-nuclease cleavage sites *NdeI* and *HindIII,* were used in the PCR reactions. Recombinant plasmids (pET28/BT3352) were isolated by using a Qiaprep Spin Miniprep Kit. Gene sequencing was carried out at the Center for Genetics in Medicine, University of New Mexico School of Medicine.

Gene expression was induced in transformed BL21(DE3) competent cells grown for 6 h in 3 L of LB media containing 50 μg/mL Kanamycin a (37 °C with mixing at 220 rpm) using 1 mM isopropyl-1-thio-β-D-galactopyranoside. Following incubation for 12 h at 20 °C and 170 rpm, the cells were harvested by centrifugation (6500 rpm for 15 min at 4 °C), resuspended in 300 mL of lysis buffer (50 mM HEPES, 500 mM NaCl, 10 mM imidazole, 1 mM DTT, pH 8.0) and lysed by two passages through a French Press cell (12,000 psi). The cell debris was removed by centrifugation at 20,000 rpm for 60 min at 4 °C, and the supernatant containing BT3352 was passed through a Ni^{+2} affinity column. The column was washed with 200 mL of lysis buffer and 200 mL of wash buffer (50 mM HEPES, 500 mM NaCl, 20 mM imidazole, 1 mM DTT, pH 8.0) before BT3352 was subsequently eluted using 100 mL of the elution buffer (50 mM NaH_2PO_4, 600 mM NaCl, 250 mM imidazole, 1 mM DTT, pH 8.0). The protein solution was concentrated by using a 10 kDa Macrosep centricon (Pall Filtron) device. The imidazole was removed by dialysis in 50 mM HEPES, 300 mM NaCl, pH 7.5. The purity of the protein was verified by using SDS-PAGE. The protein concentration was determined by using the Bradford method and by measuring the absorbance at 280 nm ($\varepsilon = 20359\ M^{-1}\ cm^{-1}$).

The selenomethionine substituted protein was expressed in *E. coli* B834(DE3) cells transformed with pET28/BT3352. The protein was purified at 4 ^{o}C using Ni-NTA column

chromatography followed by size exclusion S-200 column chromatography. The protein was concentrated by utilizing a 10 kDa Macrosep centricon (Pall Filtron) device to give a final concentration of 21 mg/mL.

Site-directed mutants were prepared by employing a Quickchange Site-Directed mutagenesis kit from Stratagene with the pET28/BT3352 clone serving as the template in conjunction with custom-designed commercial primers in the PCR reaction. The mutated plasmid was used to transform competent BL21(DE3) cells which were then grown in LB media. The mutant proteins were purified using the same procedure described for the wild type protein.

3.2.2 Gene Cloning, Expression, and Purification Wild-Type and Mutant BVU2420

The gene encoding BVU2420 was amplified by PCR using *Bacteroides vulgatus* genomic DNA and cloned into the pET-28a vector, which adds a His_6 tag to the protein N-terminus. Oligonucleotide primers containing *NdeI* and *HindIII* cleavage sites were used in the PCR reactions. Plasmid DNAs were purified using a Qiaprep Spin Miniprep Kit. The gene sequencing was carried out at the Center for Genetics in Medicine, University of New Mexico School of Medicine.

Gene expression was performed using transformed *E. coli* BL21(DE3) cells. Following growth in 3 L of LB media (1x) containing 50 μg/mL Kanamycin at 37 °C for 6 h with mixing at 220 rpm, induction was initiated with the addition of 0.2 mM isopropyl-1-thio-β-D-galactopyranoside. After incubation for 12 h at 20 °C 170 rpm, the cells were harvested by centrifugation (6500 rpm for 15 min at 4 °C), resuspended in 90 mL of lysis buffer (50 mM HEPES, 300 mM NaCl, 10 mM imidazole, pH=8.0) and lysed

by 2 passages through a French Press cell (12,000 psi). The cell debris was removed by centrifugation at 20,000 rpm for 60 min at 4 °C, and the supernatant containing BVU2420 was passed through a Ni^{+2} affinity column for His_6-tagged protein binding. The column was washed with 200 mL of lysis buffer and 100 mL of wash buffer (50 mM HEPES, 300 mM NaCl, 20 mM imidazole, pH 8.0) before the BVU2420 was subsequently eluted using 100 mL of elution buffer (50 mM HEPES, 300 mM NaCl, 250 mM imidazole, pH 8.0). The protein solution was concentrated to ~30 mL by using a 10 kDa Macrosep centricon (Pall Filtron) device. To remove the His_6-tag, 30 μL of thrombin (500U/mL) were added to the concentrated protein solution at room temperature (25 °C) and the resulting solution was incubated overnight and then passed through a Ni^{+2} affinity column. The eluate was dialyzed at 4 ^oC in 50 mM HEPES and 100 mM NaCl (pH 7.5). The protein purity was verified by SDS-PAGE. The protein concentration was determined by the Bradford method and by measuring the absorption at 280 nm.

3.2.3 Steady-State Kinetic Measurements

The rate of enzyme catalyzed hydrolysis of *p*-nitrophenyl phosphate (PNPP) hydrolysis was determined by monitoring the increase in absorbance at 410 nm ($\Delta\varepsilon$ = 18.4 $mM^{-1}cm^{-1}$) at 25 °C or 37 °C. The 0.5 mL assay mixtures contained 50 mM HEPES (pH 7.0) and 5 mM $MgCl_2$. Hydrolysis reactions of all other phosphate esters were monitored using the Enzcheck phosphate assay kit (Invitrogen).The phosphate release in solutions containing 20 mM Tris-HCl, 2 mM $MgCl_2$, and 2 mM sodium azide (pH 7.5) was monitored at 25°C by measuring the increase in absorbance at 360 nm. Initial velocities of the reactions were measured as a function of substrate concentrations, which were varied in

the range of 0.5 to 10-fold K_M. The initial velocity data were fitted using equation 1 and the computer program KinetAsyst to define V_{max} and K_M. The k_{cat} values were calculated from the ratios of V_{max} and enzyme concentrations.

$$V = V_{max}[S]/([S] + K_M)$$

where V is the initial velocity, V_{max} the maximum velocity, [S] the substrate concentration, K_m the Michaelis constant.

Thioesterase activity was measured using the DTNP coupling assay. The reaction solutions (1 mL) contained specified concentrations of substrate and 1 mM DTNB in 0.15 M KCl and 50 mM K^+HEPES (pH 7.5). The progress of the reaction was monitored at 412 nm. The initial velocity of the reaction was measured as a function of the concentration of the substrates, varied in a range of 0.5 to 10-fold K_M. The initial velocity data were fitted with the enzyme kinetic module in Sigmaplot to define V_{max} and K_M. The k_{cat} was calculated from the ratio of V_{max} and enzyme concentration.

$$V = V_{max}[S]/([S] + K_M)$$

where V is the initial velocity, V_{max} the maximum velocity, [S] the substrate concentration, K_m the Michaelis constant.

3.2.4 Size exclusion chromatography

SEC was performed on ÄKTA P-920 FPLC system and Hiprep 16/60 Sephacryl S-200 HR column (GE Healthcare). The size exclusion column was equilibrated with 50 mM HEPES (pH 7.5), 500 mM NaCl, and 1 mM DTT. The column was calibrated using the gel-filtration low molecular weight and high molecular weight gel filtration

calibration kit from Amersham Biosciences. Purified protein (2 mL, ca. 200 mM) was applied to the column.

3.2.5 BT3352 Crystallization

The crystallization and X-ray structure determination of BT3352 were carried out in Dr. Karen Allen's laboratory at Boston University. Crystals of BT3352 were initially obtained using the Index HT screen (Hampton Research) in 200 mM NaCl, 100 mM Bis-Tris buffer (pH 6.5) and 25% PEG 3,350. The crystallization condition was optimized by employing the vapor diffusion method with hanging drop geometry and using the Additive Screen (Hampton Research). Final conditions, using 2 µl drops comprised of equal amounts (1 µL each) of 25 mg/mL protein equilibrated against a reservoir solution containing 100 mM Bis-Tris buffer, pH 6.5, 200 mM NaCl, 28% PEG 3350 and 30 mM Gly-Gly-Gly, yielded crystals in *ca.* 2-3 d (figure 3.2). A solution containing 5 mM sodium tungstate and 5 mM magnesium chloride was added directly to the hanging drop containing the crystals and incubated for 5 min, before the crystals were dragged through Paratone (which caused clusters to break into single crystals). A single crystal was abstracted using a loop and directly frozen in the LN2 stream. The crystal belonged to the space group $P2_12_12_1$ with unit cell dimensions of a = 34.913 Å, b = 66.744 Å, c = 97.615 Å. Data were collected at a wavelength of 1.2134 Å optimized for tungsten anomalous edge at BeamLine X12C at NSLS of BNL, Upton, NY, which is equipped with a Brandeis 2 x 2 CCD detector (Figure 3.2).

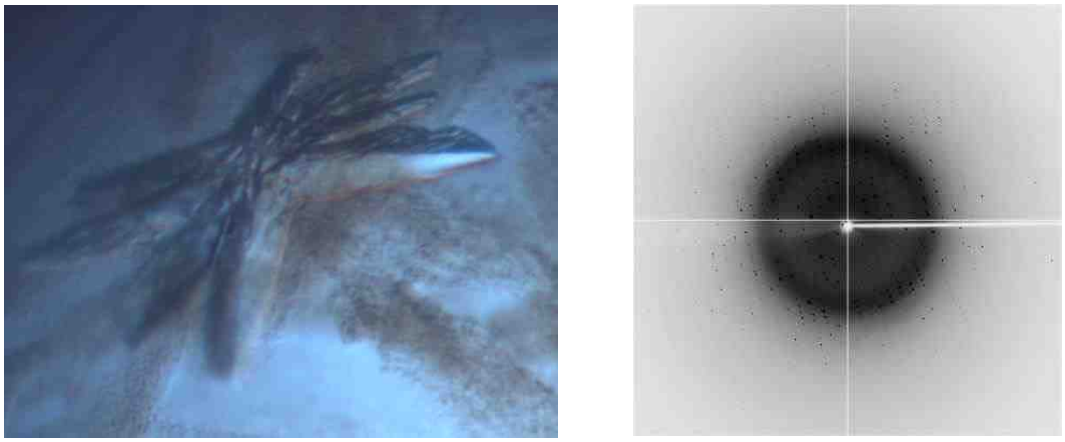


Figure 3.1. BT3352 crystals. Optimized crystals of BT3352 (left panel), oscillation photo of tungstate bound BT3352 at λ = 1.2134 Å (right panel).

3.2.6 BT3352 X-ray Structure Determination

The structure of BT3352 was determined at 1.99 Å using phases calculated by using the SAD method from a single tungsten atom (as tungstate (WO_4)). Phenix AutoSol was used to determine the heavy atom site (Hyss), calculate phases (Phaser), perform density modification (Resolve) and build a partial model (130 of 270 residues) *(10)*. Phenix AutoBuild was used to build the remaining residues of the model of BT3352 (267 of 270 residues). Iterative rounds of manual refinement in COOT *(11)* and of atomic position, atomic displacement factor and simulated annealing refinement were carried out in Phenix Refine *(12)* until an acceptable R factor and R_{free} were obtained.

Table 3.1. Data Collection, Phasing and Refinement Statistics for the BT3352 apo structure with Mg^{2+} as a cofactor and tungstate.

Data Statistics	
Space group	$P2_12_12_1$
Cell dimension (Å)	a = 34.913 b = 63.744 c =

	97.615
Molecules / ASU	1
Wavelength (Å)	1.2134
Resolution (Å)	1.99
Observed reflections	273,888
Unique reflections	27,423
Completeness (%)	94.8 (97.9)
R_{merge} (%)	11.8 (31.8)
$<I/\sigma>$	22.5(13.6)
Redundancy	5.1 (4.1)
Figure of Merit	0.843
Phase Error	22.59
Refinement	
R / R_{free}	18.42 / 22.38
Average B Factor (Å^2)	
protein	21.4
water	25.0
Mg	20.9
WO_4	20.4
Rms deviation from ideal Bonds	0.0312
Rms deviation from ideal Angles	1.4215
Ramachandran Plot Statistics favored region	262 (98.2%)
Ramachandran Plot Statistics allowed region	5 (1.8%)

3.3 Results and Discussion

3.3.1 BT3352 Structure Analysis

Mass spectrometric analysis of the recombinant BT3352 shows that the M_W of the protein is 32086.3 Da, which matches with the theoretical M_W of 32085.5 Da computed

by using ExPASy. The native molecular mass determined by size exclusion chromatography is 21 kDa which indicates that BT3352 is a monomer.

X-ray structure determination and analysis were carried out by Dr. Kelly Daughtery in Professor Karen Allen's laboratory at Boston University. The structure of BT3352, complexed with tungstate (WO_4) and magnesium, was solved at 1.99 Å using tungsten (as WO_4) to phase the structure (Table 1). The overall three-dimensional structure of the protein is similar to that of other HADSF members (Figure 3.3). The Rossmann fold core domain is conserved and the four conserved motif residues are found to form the active-site (figure 3.3). Tungstate is observed to be bound in the phosphate transfer site. BT3352 has a type C2B cap (100 residues), which has a αββαβαβαββ topology inserted between HAD motifs two and three (Figure 3.4) of the Rossmann fold.

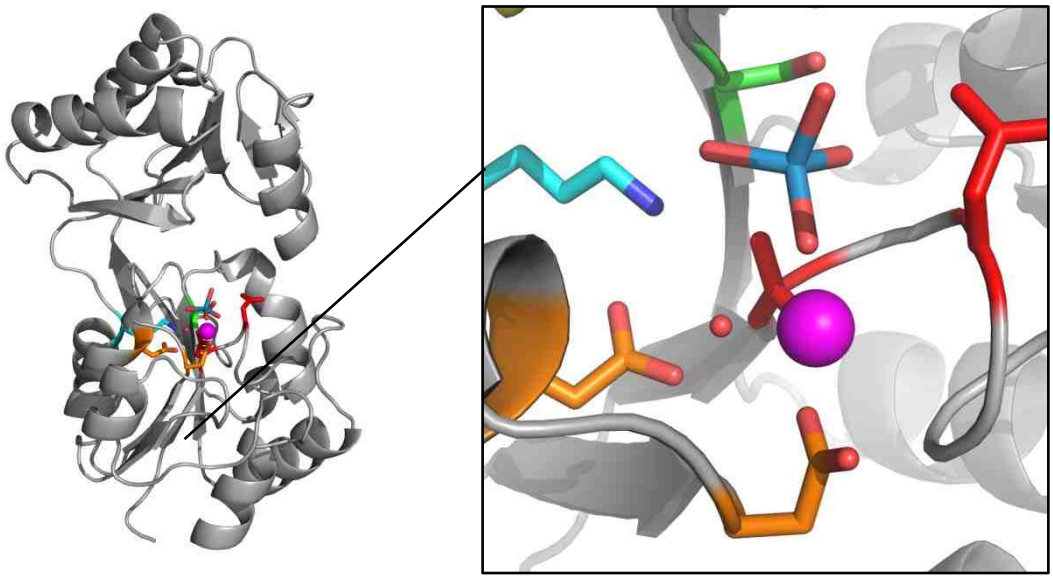


Figure 3.2. BT3352 X-ray crystallographic structure. BT3352 bound with Mg^{2+} and tungstate is shown in Pymol as a cartoon diagram (left panel) with HAD motif residues shown in stick (motif 1 red, motif 2 green, motif 3 cyan, motif 4 orange) and magnesium as magenta sphere. Zoom in of BT3352 active site with HAD motifs (right panel) oxygen and water (sphere) are shown in red, nitrogen is shown in blue, and tungsten in light blue.

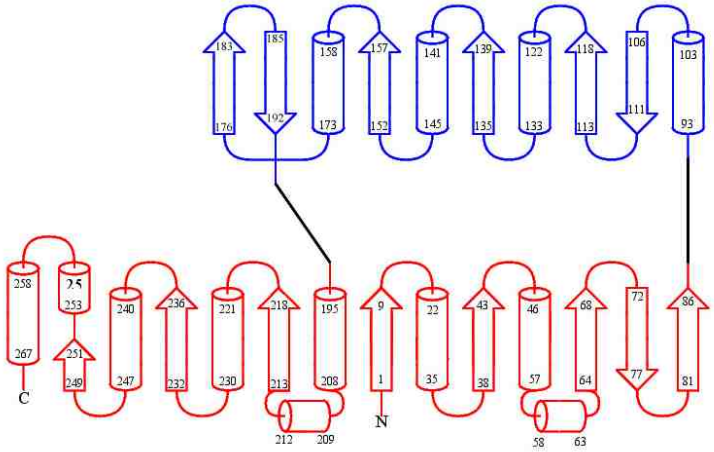


Figure 3.3. Topological diagram of BT3352. Generated by using TopDraw *(13)* in CCP4 *(14)*. Core domain shown in red and cap domain shown in blue. Figure compliments of Dr. Daughtery.

To determine if the BT3352 structure determined by using crystallographic methods is in the cap-open or cap-closed conformation, the catalytic domain of BT3352 was overlayed (by Dr. Kelly Daughtery) with cap-open and cap-closed structures of C2B family members (closed PDB ID 1RKQ, 1YMQ, open 3MPO, 1NF2) using the Dali structural alignment server. The structures 3MPO, 1RKQ, 1NF2 and 1YMQ share 31%, 29%, 29% and 25% sequence identity with BT3352 and have an RMSD of 2.4, 2.6, 2.0 Å and 2.3 Å, respectively. The cap-open structures of 3MPO and 1NF2 overlay well with BT3352, while the cap closed 1RKQ and 1YMQ structures display cap position

differences. Thus, the structure BT3352 is considered to be in the cap-open conformation (Figure 3.5). Because of the comparatively higher sequence identity with BT3352, the structure of 1RKQ (yidA from *E. coli*) was used to generate a cap-closed model of BT3352. The model depicts a solvent excluded active site.

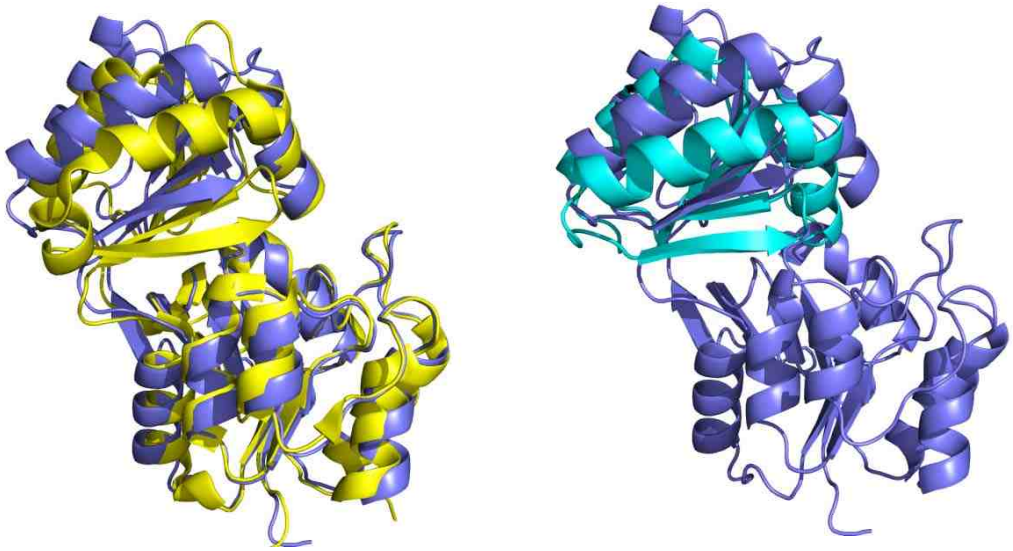


Figure 3.4. BT3352 Cap Analysis. BT3352 (purple) overlayed with 1RKQ (yellow) (left panel), and the cap-closed model of BT3352 (cyan) overlayed with the BT3352 cap-open structure (purple) (right panel). Figure prepared by Dr. Daughtery.

```
BT3352          MKYKLIVLDLDGTLTN-SKKEISSRNRETLIRIQEQGIRLVLASGRPTYGIVPLANELRM 59
3MPO            MTIKLIAIDIDGTLLN-EKNELAQATIDAVQAAKAQGIKVVLCTGRPLTGVQPYLDAMDI 59
1RKQ            MAIKLIAIDMDGTLLL-PDHTISPAVKNAIAAARARGVNVVLTTGRPYAGVHNYLKELHM 60
1NF2            -MYRVFVFDLDGTLLN-DNLEISEKDRRNIEKLS-RKCYVVFASGRMLVSTLNVEKKYFK 57
1YMQ            -MTKALFFDIDGTLVSFETHRIPSSTIEALEAAHAKGLKIFIATGRP-KAIINNLSELQD 58
                   : : :*:****       :.      :     :   :.: :**   .     .

BT3352          NEFGGFILSYNGGEIINWESKEMMYENVLPNEVVPVLYECARTNHLSILTYDGAEIVTEN 119
3MPO            DGDDQYAITFNGSVAQTIS-GKVLTNHSLTYEDYIDLEAWARKVRAHFQIETPDYIYTAN 118
1RKQ            EQPGDYCITYNGALVQKAADGSTVAQTALSYDDYRFLEKLSREVGSHFHALDRTTLYTAN 120
1NF2            RTFP--TIAYNG-AIVYLPEEGVILNEKIPPEVAKDIIEYIKPLNVHWQAYIDDVLYSEK 114
1YMQ            RNLIDGYITMNG--AYCFVGEEVIYKSAIPQEEVKAMAAFCEKKGVPCIFVEEHNISVCQ 116
                       :: **          : :  :. :    :    .           :   :
```

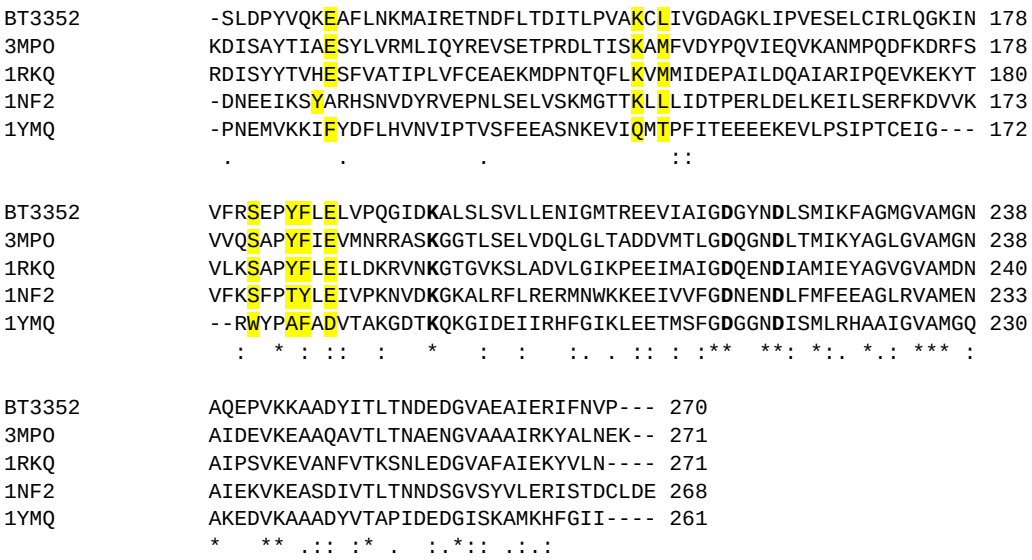


Figure 3.5. Primary Sequence Alignment of C2B members. Sequence alignments of BT3352, 3MPO, 1RKQ, 1NF2and 1YMQ (Clustal @ EBI), in which highlighted residues are proposed specificity residues in C2B members. Bold residues indicate HAD motifs, (*) indicates absolutely conserved; (**:**) indicates mostly conserved, (.) indicates somewhat conserved.

Owing to the lack of liganded structures of C2B family members, it is not possible to assign specificity residues. However, structural and primary sequence alignments suggest variable residues in the cap of each enzyme (
Figure 3.6). Variability is seen in a loop between β10 and β11 of the cap across HAD type 2B family members (BT3352 – Glu 183, 3MPO - Ala, 1RKQ – Ala, 1NF2 – Phe, 1YMQ – Tyr), suggesting their possible importance in specificity. The beta-strands β10 and β11 are part of a structurally conserved beta-sheet in C2B cap domains. Loop residues are found within non-bonded interaction ranges of the active site of C2B

members. Lu et. al. proposed that this loop contains specificity residues (termed the primary specificity loop) *(9),* a hypothesis which our structural analysis supports.

3.3.2 Substrate Specificity and Docking Models

BT3352 contains the conserved catalytic HAD motifs that are found in all other HADSF members. Based on the initial broad substrate specificity screen alone, the physiological substrate of this protein was unclear owing to the low K_m and k_{cat} values

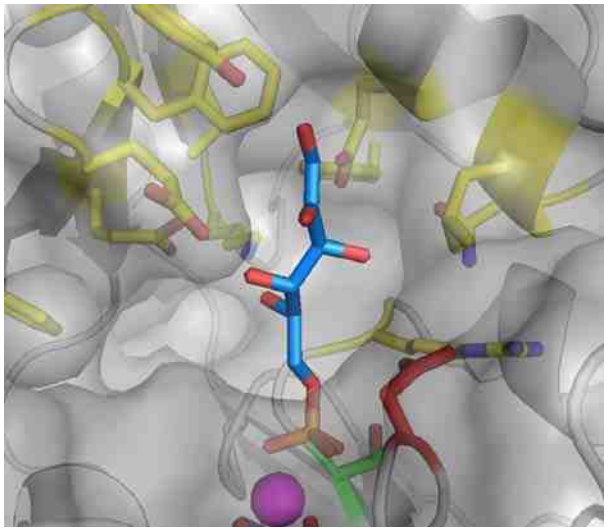


Figure 3.6. Substrate docking in BT3352 active site. BT3352 (gray cartoon) with HAD motifs (according to standard coloring) and possible substrate binding residues in the cap (yellow). Fructose 6-phosphate (teal) with BT3352 van der Waal surface shown in gray.

measured for a wide range of substrates. In this series, fructose 6-phosphate (k_{cat} / K_m = 3.3 x 10^2 $M^{-1}s^{-1}$), imidodiphosphate (k_{cat} / K_m = 3.1 x 10^3 $M^{-1}s^{-1}$) and threonine phosphate (k_{cat} / K_m = 8.2 x 10^2 $M^{-1}s^{-1}$) have the highest k_{cat} / K_m values, but still are relatively poor substrates compared to those of other C2B phosphatases (e.g. k_{cat} / K_m on the order of 10^5

$M^{-1}s^{-1}$ *(4)*). This finding indicated that the true substrate probably had not been identified by using the initial screen. Employing the cap open structure, a van der Waals surface was generated and a cavity surface of the BT3352 active site was made using PyMOL *(15).* These two surfaces were used by Dr. Kelly Daughtery in Professor Karen Allen's laboratory at Boston University to manually dock several possible substrates into the active site by overlaying the substrate phosphoryl moiety with the tungstate ion bound in the active site. Fructose 6-phosphate fits well into the cap open structure, but it appears to be too large to be accommodated in the cap closed model (Figure 3.5). This finding correlates well with the large K_m value (K_m = 1.4 mM) of fructose 6-phosphate. Imidodiphosphate and threonine phosphate are both relatively small molecules that do not reside within hydrogen bonding distance of cap residues when docked in the active site of cap-closed BT3352. Both compounds are low avtivity substrates for BT3352. Based on this structural analysis, it appeared that a feasible substrate for BT3352 could be a 4 or 5 carbon phosphorylated sugar. As a result, kinetic studies were carried out with 4 and 5 carbon phosphorylated sugars (see below).

Table 3.2. Steady-state kinetic contants measured for BF1314, PG1653, BFHAD and BT3352 catalyzed hydrolysis of phosphate esters in 20 mM Tris, 1 mM $MgCl_2$, 0.1 mM sodium azide at pH 7.5 and 25 ^oC.

Substrate	Variable Proteins	k_{cat} (s^{-1})	K_m (μM)	K_m/k_{cat} ($M^{-1}s^{-1}$)
para-Nitrophenyl phosphate	BF1314	$< 1 \times 10^{-4}$	--------	--------
	BFHAD	$< 1 \times 10^{-4}$	--------	--------
	PG1653	$< 1 \times 10^{-4}$	--------	--------
	BT3352	$(1.5 \pm 0.1) \times 10^{-3}$	$(4.4 \pm 0.5) \times 10^{2}$	3.3
Carbamoyl phosphate	BF1314	$(2.4 \pm 0.1) \times 10^{-2}$	$4.4 \pm 0.3) \times 10^{2}$	5.4×10^{1}
	BFHAD	$(3.5 \pm 0.3) \times 10^{-2}$	$(9.8 \pm 0.8) \times 10^{2}$	3.6×10^{1}
	PG1653	$< 1 \times 10^{-4}$	--------	--------
	BT3352	$(4.6 \pm 0.4) \times 10^{-2}$	$(1.5 \pm 0.1) \times 10^{3}$	3.1×10^{1}

Imidodi phosphate	BF1314	$(2.1 \pm 0.1) \times 10^{-1}$	$(5.7 \pm 0.7) \times 10^{3}$	3.7
	BFHAD	$(2.0 \pm 0.2) \times 10^{-1}$	$(7.4 \pm 0.8) \times 10^{3}$	2.7
	PG1653	$< 1 \times 10^{-4}$	--------	--------
	BT3352	$(2.0 \pm 0.1) \times 10^{-1}$	$(6.2 \pm 0.2) \times 10^{1}$	3.2×10^{2}
Uridine mono phosphate (UMP)	BF1314	$< 1 \times 10^{-4}$	--------	--------
	BFHAD	$< 1 \times 10^{-4}$	--------	--------
	PG1653	$< 1 \times 10^{-4}$	--------	--------
	BT3352	$< 1 \times 10^{-4}$		--------
Adenosine mono phosphate (AMP)	BF1314	$< 1 \times 10^{-4}$	--------	--------
	BFHAD	$< 1 \times 10^{-4}$	--------	--------
	PG1653	$< 1 \times 10^{-4}$	--------	--------
	BT3352	$< 1 \times 10^{-4}$	--------	--------
Glycerol phosphate	BF1314	$< 1 \times 10^{-4}$	--------	--------
	BFHAD	$(1.1 \pm 0.1) \times 10^{-2}$	$(6.3 \pm 0.2) \times 10^{1}$	1.7×10^{2}
	PG1653	$< 1 \times 10^{-4}$	--------	--------
	BT3352	$(8.5 \pm 0.1) \times 10^{-1}$	$(2.7 \pm 0.2) \times 10^{3}$	3.1×10^{2}
Thiamine phosphate	BF1314	$< 1 \times 10^{-4}$	--------	--------
	BFHAD	$< 1 \times 10^{-4}$	--------	--------
	PG1653	$< 1 \times 10^{-4}$	--------	--------
	BT3352	$< 1 \times 10^{-4}$	--------	--------
Threonine phosphate	BF1314	$< 1 \times 10^{-4}$	--------	--------
	BFHAD	$< 1 \times 10^{-4}$	--------	--------
	PG1653	$< 1 \times 10^{-4}$	--------	--------
	BT3352	$(8.8 \pm 0.3) \times 10^{-1}$	$(1.1 \pm 0.1) \times 10^{3}$	8.2×10^{2}
Farnesyl monophophate	BF1314	$< 1 \times 10^{-4}$	--------	--------
	BF-HAD	$< 1 \times 10^{-4}$	--------	--------
	PG1653	$< 1 \times 10^{-4}$	--------	--------
	PGF11D	$< 1 \times 10^{-4}$	--------	--------
Farnesyl pyrophophate	BF1314	$< 1 \times 10^{-4}$	--------	--------
	BF-HAD	$< 1 \times 10^{-4}$	--------	--------
	PG1653	$< 1 \times 10^{-4}$	--------	--------
	PGF11D	$< 1 \times 10^{-4}$	--------	--------
Serinol phosphate	BF1314	$< 1 \times 10^{-4}$	--------	--------
	BFHAD	$< 1 \times 10^{-4}$	--------	--------
	PG1653	$< 1 \times 10^{-4}$	--------	--------
	BT3352	$(7.9 \pm 0.2) \times 10^{-1}$	$(1.3 \pm 0.1) \times 10^{3}$	6.2×10^{2}
Erythrose 4-phosphate	BF1314	$< 1 \times 10^{-4}$	--------	--------
	BFHAD	$< 1 \times 10^{-4}$	--------	--------
	PG1653	$< 1 \times 10^{-4}$	--------	--------
	BT3352	$(1.8 \pm 0.1) \times 10^{1}$	$(1.4 \pm 0.1) \times 10^{2}$	1.3×10^{5}
Glyceraldehye 3-phosphate	BF1314	$< 1 \times 10^{-4}$	--------	--------
	BFHAD	$< 1 \times 10^{-4}$	--------	--------
	PG1653	$< 1 \times 10^{-4}$	--------	--------
	BT3352	$(3.3 \pm 0.1) \times 10^{-1}$	$(2.7 \pm 0.2) \times 10^{3}$	1.2×10^{2}
6-phospho gluconic acid	BF1314	$< 1 \times 10^{-4}$	--------	--------
	BFHAD	$< 1 \times 10^{-4}$	--------	--------
	PG1653	$< 1 \times 10^{-4}$	--------	--------
	BT3352	$(1.1 \pm 0.1) \times 10^{-2}$	$(9.4 \pm 0.9) \times 10^{3}$	1.1
3-phospho glycerate	BF1314	$< 1 \times 10^{-4}$	--------	--------
	BFHAD	$< 1 \times 10^{-4}$	--------	--------
	PG1653	$< 1 \times 10^{-4}$	--------	--------
	BT3352	$< 1 \times 10^{-4}$	--------	--------
2-phospho glycolate	BF1314	$< 1 \times 10^{-4}$	--------	--------
	BFHAD	$< 1 \times 10^{-4}$	--------	--------
	PG1653	$< 1 \times 10^{-4}$	--------	--------

PEP	BT3352	$(9.2 \pm 0.5) \times 10^{-1}$	$(1.1 \pm 0.1) \times 10^{3}$	8.2×10^{2}
	BF1314	$< 1 \times 10^{-4}$	--------	--------
	BFHAD	$< 1 \times 10^{-4}$	--------	--------
	PG1653	$< 1 \times 10^{-4}$	--------	--------
Fructose 6-phosphate	BT3352	$< 1 \times 10^{-4}$	--------	--------
	BF1314	$< 1 \times 10^{-4}$	--------	--------
	BFHAD	$< 1 \times 10^{-4}$	--------	--------
	PG1653	$< 1 \times 10^{-4}$	--------	--------
Glucose 6-phosphate	BT3352	$(4.6 \pm 0.1) \times 10^{-1}$	$(1.4 \pm 0.1) \times 10^{3}$	3.3×10^{2}
	BF1314	$< 1 \times 10^{-4}$	--------	--------
	BFHAD	$< 1 \times 10^{-4}$	--------	--------
	PG1653	$< 1 \times 10^{-4}$	--------	--------
2-keto-3-deoxy-6-phospho galactonate	BT3352	$< 1 \times 10^{-4}$	--------	--------
	BF1314	$< 1 \times 10^{-4}$	--------	--------
	BFHAD	$< 1 \times 10^{-4}$	--------	--------
	PG1653	$< 1 \times 10^{-4}$	--------	--------
	BT3352	$< 1 \times 10^{-4}$	--------	--------

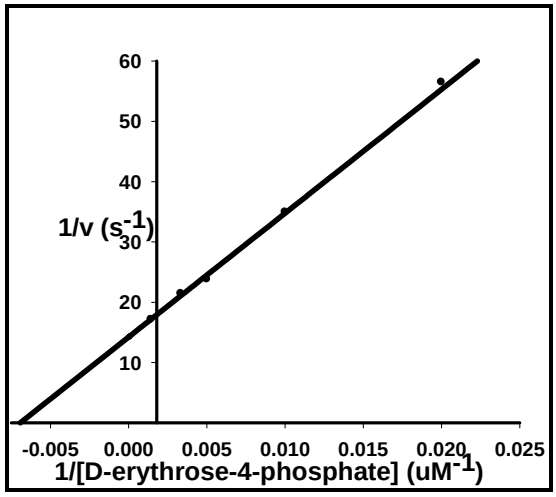


Figure 3.7. Lineweaver-Burk plot of the BT3352 catalyzed erythrose 4-phosphate hydrolysis reaction (20 mM Tris, 1 mM $MgCl_2$, 0.1mM sodium azide, pH 7.5, 25 ^{o}C)

The results of a four and five carbon phosphorylated sugar substrate screen indicate that erythrose 4-phosphate is the best substrate of BT3352 with a $k_{cat} / K_m = 1.3 \times 10^5$ (figure 3.8). The low K_m (K_m = 140 μM) of this substance indicates that it could be the physiological substrate of BT3352. Moreover, the substrate range of BT3352 is narrow and the k_{cat}/K_m values of alternate substrates are not physiologically relevant.

The BT3352 is not located within an operon, however there are two notable genes in the neighborhood; BT3361, a putative erythronate-4-phosphate dehydrogenase and BT3356 a predicted 6-phosphofructokinase.

3.3.3 Evolution of the C2B HAD Phosphatases in *B. thetatiotaomicron*

In total, four C2B HAD phosphatases were identified by using a BLAST search of the *B. thetaiotaomicron* genome, including BT1666, BT3352, BT4131 and BT4699 (the HAD domain-thioesterase domain fusion protein described in Chapter 2). An alignment of their sequences is shown in Figure 3.9. BT1666 and BT4131 share a relatively high sequence identity (37%) and, consequently, are likely to be related via a gene duplication event. BT3352 and the HAD domain of BT4699 share 48% sequence identity, indicating that they too are likely to be related by a gene duplication event. Following gene duplication one of the two paralogous gene fused with the gene encoding the DHN-CoA thioesterase. The redundancy of the fused HAD domain allowed it to evolve a new function which we have proposed in Chapter 2 is that of a dimerization domain.

BT1666 (21) and BT4131 *(9)* have been shown be broad specificity, low catalytic efficiency phosphatases, with k_{cat}/K_m values for sugar phosphate substrates in the range of 1×10^2-1×10^3 M^{-1} s^{-1}. These two HAD members are likely to function in limiting the accidental accumulation of sugar phosphate metabolites and thereby protect the cell from their inhbitory effects on off target enzymes. In contrast, BT3352 functions as a highly specific erythrose 4-phosphate phosphatase with comparatively high catalytic efficency. It might function to hydrolyze erythrose 4-phosphate or its oxidation product erythronate-4-phosphatase ($k_{cat}/K_m = 2 \times 10^3$ $M^{-1}s^{-1}$).

```
BT3352          MKYKLIVLDLDGTLTN-SKKEISSRNRETLIRIQEQGIRLVLASGRPTYGIVPLANELRM 59
BT1666          -MIKVLLLDVDGTLLSFETHKVSQSSIDALKKVHDSGIKIVIATGRAASDLHEIDAVP-- 57
BT4131          -MTKALFFDIDGTLVSFETHRIPSSTIEALEAAHAKGLKIFIATGRPKAIINNLSELQDR 59
BT4699          MKYKLLVLDVDGTLLN-DAKEISKRTLAALLKVQQMGVRIVLASGRPTYGLMPLAKSLEL 59
                   * :.:*:**** . . :.:.. .  :*   :  *:::.:*:**.   :  :

BT3352          NEFGGFILSYNGGEIINWESKEMMYENVLPNEVVPVLYECARTNHLSILTYDGAEIVTEN 119
BT1666          --YDGVIALNGAECVLRD--GSVIRKVAIPAQDFRKSMELAREFDFAVALELNEGVFVNR 113
BT4131          NLIDGYITMNGAYCFVG---EEVIYKSAIPQEEVKAMAAFCEKKGVPCIFVEEHNISVCQ 116
BT4699          GNYGGYILSYNGCQIINAQNGEILFERRINPEMLPYLEKKARKNNFALFTYHDDTIITDT 119
                   .* *   ..  .:     .:: :  :  : .      ..   ..        : .

BT3352          SLDPYVQKEAFLNKMAIRETNDFLTDITLPVAKCLIVG-DAGKLIPVESELCIRLQGKIN 178
BT1666          LTPTVEQIAGIVEHPVPPVVDIEEMFERKECCQLCFYFDEEAEQKVMPLLSG-----LS 167
BT4131          PNEMVKKIFYDFLHVNVIPTVSFEEASNKEVIQMTPFITEEEEKEVLPSIPT-----CE 170
BT4699          PENEHIQNEARLNNLKVIKEEEFSVAIDFAPCKCMLVSDDEEALVSLEGHWKRRLNGALD 179
                      :    . :  :    .:           :           :           .

BT3352          VFRSEPYFLELVPQGIDKALSLSVLLENIGMTREEVIAIGDGYNDLSMIKFAGMGVAMGN 238
BT1666          ATRWHPLFADVNVAGTSKATGLSLFADYYRVKVSEIMACGDGGNDIPMLKAAGIGVAMGN 227
BT4131          IGRWYPAFADVTAKGDTKQKGIDEIIRHFGIKLEETMSFGDGGNDISMLRHAAIGVAMGQ 230
BT4699          VFRSEPYFLEVVPCAIDKANTLGALLEELDVKREEVIAIGDGVCDVTMIQLAGLGVAMGH 239
                  *  * * ::   .  *   :. :     :. .* :: ***  *:.*:: *.:*****:

BT3352          AQEPVKKAADYITLTNDEDGVAEAIERIFNVP---------------------------- 270
BT1666          ASEKVQSVADFVTDTVDNSGLYKALKHFGVI---------------------------- 258
BT4131          AKEDVKAAADYVTAPIDEDGISKAMKHFGII---------------------------- 261
BT4699          SQDSVKVCADYVTASNEEDGVALAVEKAIIAEVRAAEIPLDQLNAQARHALMGNLGIQYT 299
                :.: *:  **::* . ::.*:  *:::

BT3352          ------------------------------------------------------------
BT1666          ------------------------------------------------------------
BT4131          ------------------------------------------------------------
BT4699          YADEDRVEATMPVDHRTRQPFGILHGGATLALGETVAGLGSMILCQPDEIVVGMQVSGNH 359


BT3352          -------------------------------------------------
BT1666          -------------------------------------------------
BT4131          -------------------------------------------------
BT4699          ISSAHEGDTVRAVATIVHKGRSSHVWNVDVFTSTNKLVSSIRVVNSVMKKR 410
```

Figure 3.8. Primary sequence alignment of *B. thetaiotaomicron* C2B members. The coloring and labeling patterns are the same as those used in Figure 3.6.

3.3.4 Divergence in Structure and Function in Homologues *B. thetaiotaomicron* BT3352 and *E. coli* yidA

Table 3.3. Steady-state kinetic constants k_{cat} and K_m for wild-type and mutant BT3352 catalyzed hydrolysis of erythrose 4-phosphate in 20 mM Tris, 1 mM $MgCl_2$, 0.1 mM sodium azide, pH 7.5, 25 ^oC.

BT3352	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1}s^{-1}$)	yidA	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1}s^{-1}$)
WT	$(1.8 \pm 0.1) \times 10^{1}$	$(1.4 \pm 0.1) \times 10^{2}$	1.2×10^{5}	WT	8 ± 1	$(4.2 \pm 0.1) \times 10^{2}$	1.9×10^{4}
R46A	$<10^{-4}$	--------	--------	R46A	$(1.6 \pm 0.08) \times 10^{-2}$	$(4.0 \pm 0.4) \times 10^{2}$	4.0×10^{1}
E128A	$(4.4 \pm 0.2) \times 10^{-2}$	$(1.9 \pm 0.2) \times 10^{3}$	2.3×10^{1}	E130A	$(8.0 \pm 0.3) \times 10^{-2}$	$(2.6 \pm 0.2) \times 10^{2}$	3.1×10^{2}
K152A	$(2.6 \pm 0.1) \times 10^{-2}$	$(3.9 \pm 0.6) \times 10^{2}$	6.6×10^{1}	K154A	$(7.8 \pm 0.2) \times 10^{-2}$	$(4.2 \pm 0.3) \times 10^{2}$	1.9×10^{2}
L154A	2.3 ± 0.1	$(1.2 \pm 1.3) \times 10^{3}$	1.9×10^{3}	M156A	$(1.3 \pm 0.6) \times 10^{1}$	$(1.5 \pm 0.2) \times 10^{3}$	8.2×10^{3}
S182A	$(2.3 \pm 0.1) \times 10^{-1}$	$(7.1 \pm 0.2) \times 10^{2}$	3.2×10^{2}	S184A	1.60 ± 0.04	$(3.6 \pm 0.2) \times 10^{2}$	4.4×10^{3}
Y185A	$(4.8 \pm 0.2) \times 10^{-1}$	$(3.6 \pm 0.4) \times 10^{2}$	1.3×10^{3}	Y187A	$(2.2 \pm 0.02) \times 10^{-1}$	$(8.0 \pm 0.2) \times 10^{1}$	2.8×10^{3}
F186A	$(9.1 \pm 0.3) \times 10^{-2}$	$(3.8 \pm 0.3) \times 10^{2}$	2.4×10^{2}	F188A	5.0 ± 0.2	$(4.8 \pm 0.3) \times 10^{2}$	1.8×10^{4}
E188A	$(3.7 \pm 0.1) \times 10^{-2}$	$(7.6 \pm 0.6) \times 10^{2}$	4.8×10^{1}	E190A	$(4.6 \pm 0.1) \times 10^{-1}$	$(1.4 \pm 0.1) \times 10^{3}$	3.4×10^{2}

The active site residues of special interest in these enzymes include Arg45, Glu128, Lys152, Leu154, Ser182, Tyr185, Phe186 and Glu188. Arg45 is conserved among HADSF members. Also, Ala replacement of the Asp11 acid/base partner Arg46 leads to a reduction of the activity of BT3352 to below the detection limit (Table 3). Lu *et al*. proposed that as general base, Asp11 is held in position by a hydrogen bond formed with the side chain of Arg45 *(20)*. The other important residues may play important roles in substrate recognition.

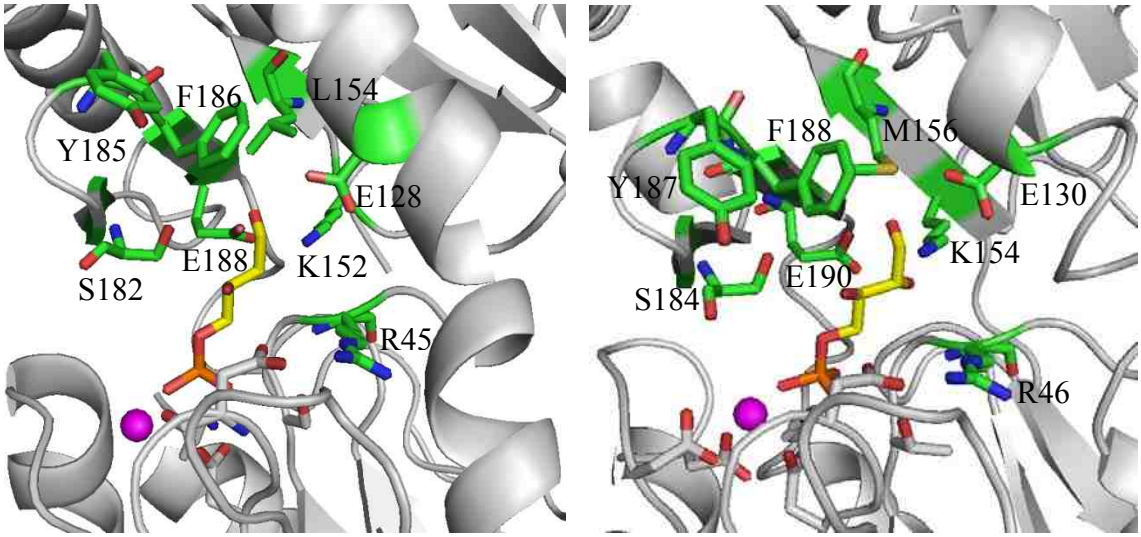


Figure 3.9. Pymol generated stereoview of the zoom-in structures of BT3352 (left) and yidA (right) with active sites showing the amino acid residues that immediately surround docked erythrose 4-phosphate.

Table 3.4. Comparisons of the phosphatase activities of BT3352 and yidA towards a variety of substrates in 20 mM Tris, 1 mM $MgCl_2$, 0.1 mM sodium azide, pH 7.5, 25 ^oC.

Substrate	BT3352			yidA		
	k_{cat} (s^{-1})	K_m (μM)	K_m/k_{cat} ($M^{-1}s^{-1}$)	k_{cat} (s^{-1})	K_m (μM)	K_m/k_{cat} ($M^{-1}s^{-1}$)
para-Nitrophenyl phosphate	$(1.5 \pm 0.1) \times 10^{-3}$	$(4.4 \pm 0.5) \times 10^{2}$	3.3	$(3.5 \pm 0.1) \times 10^{-3}$	$(2.2 \pm 0.1) \times 10^{3}$	1.7×10^{1}
Glycerol phosphate	$(8.5 \pm 0.1) \times 10^{-1}$	$(2.7 \pm 0.2) \times 10^{3}$	3.1×10^{2}	1.0 ± 0.1	$(3.7 \pm 0.5) \times 10^{3}$	2.7×10^{-1}
Erythrose 4-phosphate	$(1.8 \pm 0.1) \times 10^{1}$	$(1.4 \pm 0.1) \times 10^{2}$	1.3×10^{5}	8.0 ± 1.0	$(4.2 \pm 0.1) \times 10^{2}$	1.9×10^{4}
erythronate 4-phosphate	1.8 ± 0.1	$(2.0 \pm 0.2) \times 10^{2}$	9×10^{3}	1.0 ± 0.1	$(1.0 \pm 0.1) \times 10^{2}$	1×10^{4}
Ribose 5-phosphate	1.1 ± 0.1	$(1.5 \pm 0.2) \times 10^{3}$	7.3×10^{2}	8.4 ± 0.5	$(1.1 \pm 0.1) \times 10^{4}$	8.1×10^{2}
Ribulose 5-phosphate	$(6.3 \pm 0.1) \times 10^{-1}$	$(2.1 \pm 0.1) \times 10^{3}$	3.0×10^{2}	7.0 ± 0.5	$(5.8 \pm 0.5) \times 10^{3}$	1.2×10^{3}
Arabinose 5-phosphate	$(6.3 \pm 0.1) \times 10^{-1}$	$(1.0 \pm 0.1) \times 10^{3}$	6.3×10^{2}	$(1.4 \pm 0.4) \times 10^{1}$	$(2.2 \pm 0.2) \times 10^{3}$	6.5×10^{3}
Fructose 6-phosphate	$(4.6 \pm 0.1) \times 10^{-1}$	$(1.4 \pm 0.1) \times 10^{3}$	3.3×10^{2}	3.8 ± 0.2	$(6.6 \pm 0.7) \times 10^{3}$	5.8×10^{2}
Fructose 1-phosphate	$< 1 \times 10^{-4}$	--------	--------	1.5 ± 0.1	$(2.7 \pm 0.2) \times 10^{4}$	5.4×10^{2}

Mannose 6-phosphate	$< 1 \times 10^{-4}$	--------	--------	6.3 ± 0.8	$(3.7 \pm 0.3) \times 10^{4}$	1.7×10^{2}
Galactose 1-phosphate	$< 1 \times 10^{-4}$	--------	--------	$(3.6 \pm 0.3) \times 10^{-1}$	$(2.5 \pm 0.2) \times 10^{4}$	1.4×10^{1}
Glucose 6-phosphate	$< 1 \times 10^{-4}$	--------	--------	6 ± 0.6	$(2.1 \pm 0.3) \times 10^{4}$	2.9×10^{-1}
Glucose 1-phosphate	$< 1 \times 10^{-4}$	--------	--------	4.2 ± 0.1	$(2.7 \pm 0.2) \times 10^{3}$	1.5×10^{3}
2-keto-3-deoxy-6-phospho galactonate	$< 1 \times 10^{-4}$	--------	--------	1.6 ± 0.1	$(2.5 \pm 0.4) \times 10^{2}$	6.4×10^{3}
6-phospho gluconic acid	$(1.1 \pm 0.1) \times 10^{-2}$	$(9.4 \pm 0.9) \times 10^{3}$	1.1	$(1.4 \pm 0.1) \times 10^{-1}$	$(9.2 \pm 0.8) \times 10^{3}$	1.6

In order to obtain a "virtual" picture of the catalytic site in the enzyme-substrate complex, erythrose 4-phosphate was docked into BT3352. The resulting model shows that one of the substrate's hydroxyl groups is located within hydrogen bonding distance of the side chains of E128 (3.0 Å) and E188 (3.3 Å) (Figure 3.11). In comparison, E130 (3.0 Å) and E190 (4.3 Å) of yidA participate in long hydrogen bonding interactions with erythrose 4-phosphate, an observation, which indicates the existence of a larger active site that can accommodate larger substrates. The substrate screen results (Table 3) also demonstrated that yidA has a more roomy active site that can accommodate a wide substrate group ranging from tetrose to hexose phosphate. In contrast, the active site of BT3352 appears to be more compatible with substrates like tetrose and pentose phosphate.

Wild type BT3352 is 10-fold more active in the hydrolysis of erythrose 4-phosphate than is yidA and it displays a greater sensitivity to the mutations. A zoom-in snapshot of the catalytic site reveals that BT3352 provides more residues for interaction with the hydroxyl groups of erythrose 4-phosphate, an observation which may account for the 3-fold lower K_m of BT3352 in contrast to that of yidA (Figure 3.11).

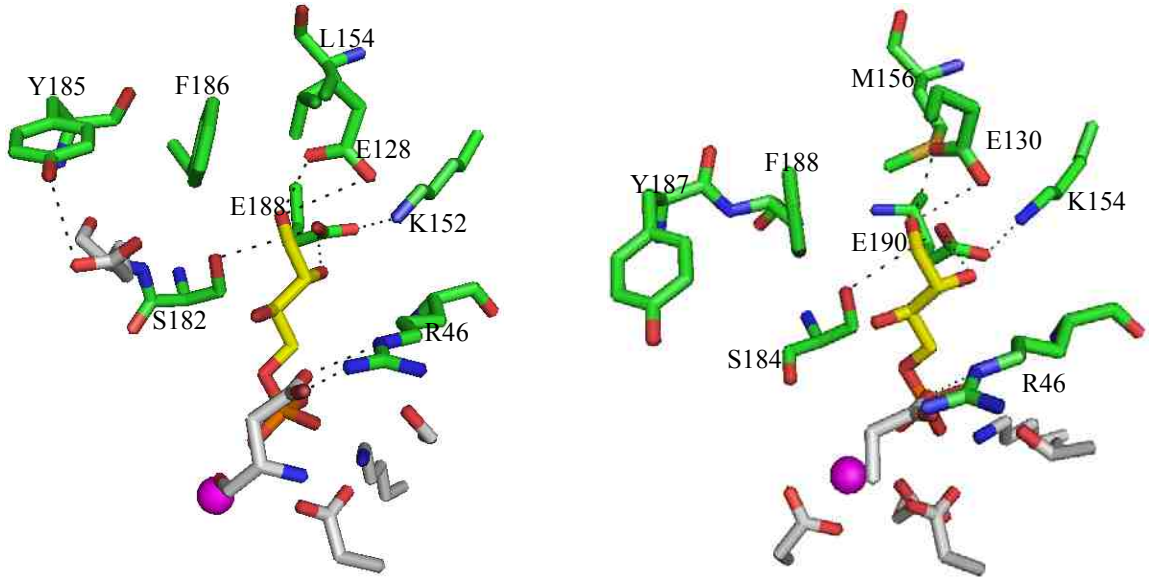


Figure 3.10. A Pymol generated stereoview of the hydrogen bonding patterns between the active sites of BT3352 (A), and yidA (B) and erythrose 4-phosphate.

Mutations in the caps of both BT3352 and yidA have a greater impact on k_{cat} than on K_m, with the exception of the L154A/M156A double mutation which leads in BT3352 to a 10-fold increase in K_m and an 8-fold decrease in the k_{cat}. In contrast there is little change in the yidA k_{cat}. The biggest effects on catalytic activity are promoted by the following mutations of the three charged residues: E188A/E190A, E128A/E130A, K152A /K154A. The S182A/S184A mutation causes a large effect on the activity of BT3352 (300-fold decrease k_{cat}/K_m) and only a small effect on that of yidA (4-fold decrease). Similarly, the F186A/F188A mutation reduces the k_{cat}/K_m of the BT3352 500-fold but does not affect the k_{cat}/K_m of yidA. The Y185A/Y187A mutation has a 10-fold greater effect on the catalytic efficiency of BT3352 than that of yidA. In addition to the effects on catalytic efficiency, replacement of S182, Y185, and F186 with Ala leads to a great destabilization of BT3352, while the corresponding mutations S184A, Y187A, and F188A have no noticable effect on the stability of yidA. Overall, yidA is less efficient and less responsive

to mutations. Nevertheless, the sensitivity of k_{cat} to the cap mutations suggests that the interaction between the cap and catalytic domains is important to productive binding and/or transition state stabilization.

Table 3.5. Steady-state kinetic constants k_{cat} and K_m for BT3352 mutants L154A and L154M for hydrolysis of erythrose-4-phosphate in 20 mM Tris, 1 mM $MgCl_2$, 0.1mM sodium azide, pH 7.5, 25 ^oC.

Substrate	L154A			L154M		
	k_{cat} (s^{-1})	K_m (μM)	K_m/k_{cat} ($M^{-1}s^{-1}$)	k_{cat} (s^{-1})	K_m (μM)	K_m/k_{cat} ($M^{-1}s^{-1}$)
erythrose 4-phosphate	2.3 ± 0.1	(1.2 ± 1.3) x 10^3	1.9 x 10^3	(6.8 ± 0.2) x10^{-1}	(7.1 ± 0.5) x10^2	9.6 x 10^2
α glucose 1-phosphate	< 1 x 10^{-4}	--------	--------	< 1 x 10^{-4}	--------	--------
2-keto-3-deoxy-6-phospho galactonate	< 1 x 10^{-4}	--------	--------	< 1 x 10^{-4}	--------	--------

The active site residues L154 in BT3352 and M156 in yidA may play roles in the determining the active site environment. Mutagenesis of both L154 and M156 to alanine results in significant decreases in K_m values (10-fold for BT3352 and 4-fold for yidA). Mutation of L154 to Met does not greatly change the specificity or activity of BT3352 (K_m is decreased 5-fold and the k_{cat} decreased 30-fold; Table 5). Mutation of Phe186, Glu128 and Glu188 to Ala results in significant changes in BT3352's specificity as demonstrated by the observation that 2-keto-3-deoxy-6-phosphogluconate (KDPG) and 2-keto-3-deoxy-6-phosphogalaconate (KDPGal) are good substrates for the mutants but not for the wild-type enzyme. In comparison to the activity of wild type BT3352 for hydrolysis of KDPG, the K_m of E128A decreased 3-fold and the k_{cat} is increased 400-fold (k_{cat}/K_m value of 2.3 x

10^4 $M^{-1}s^{-1}$). The KDPG K_m of E188A is decreased 30-fold but the k_{cat} is decreased 5-fold. Finally, the E128A mutant also shows modest phosphatase activity for KDPGAL (Table 3.6).

Table 3.6. Steady-state kinetic constants k_{cat} and K_m for wild-type and mutant BT3352 catalyzed hydrolysis of 2-keto-3-deoxy-6-phosphogluconate (KDPG) and 2-keto-3-deoxy-6-phosphogalaconate (KDPGal) in 20 mM Tris, 1 mM $MgCl_2$, 0.1 mM sodium azide, pH 7.5, 25 ^oC)

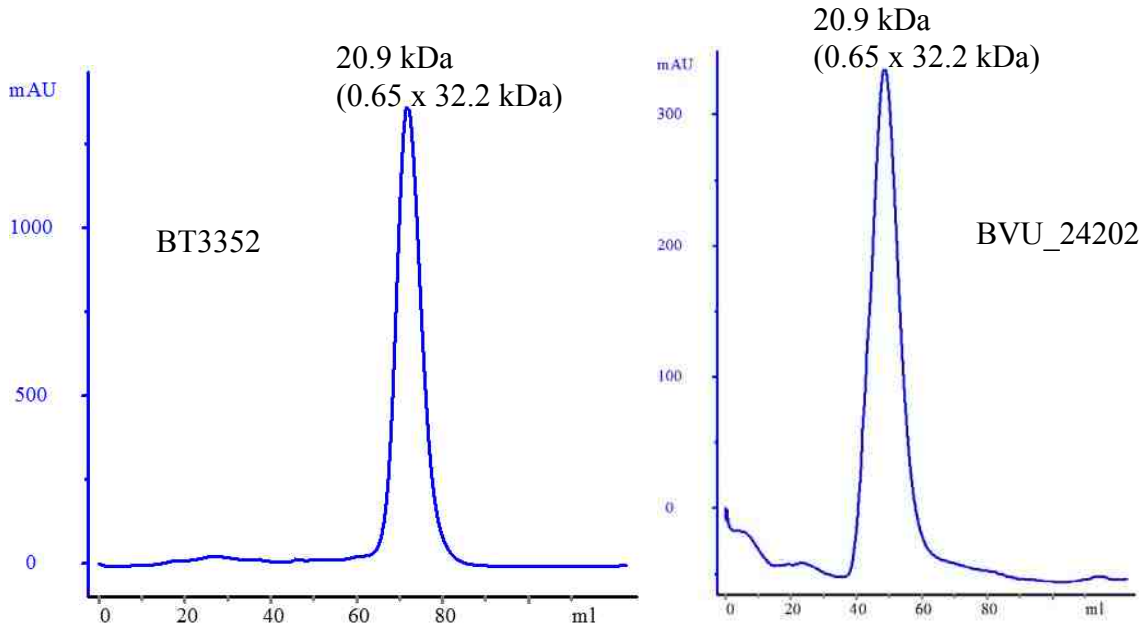
	KDPG			KDPGal		
	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1}s^{-1}$)	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1}s^{-1}$)
WT	$(1.3 \pm 0.1) \times 10^{-2}$	$(5.9 \pm 0.4) \times 10^2$	2.2×10^1	$<10^{-4}$	--------	--------
R46A	$<10^{-4}$	--------	--------	$<10^{-4}$	--------	--------
E128A	4.8 ± 0.1	$(2.1 \pm 0.1) \times 10^2$	2.3×10^4	2.1 ± 0.1	$(1.0 \pm 0.1) \times 10^3$	2.0×10^3
K152A	$<10^{-4}$	--------	--------	$<10^{-4}$	--------	--------
L154A	$<10^{-4}$	--------	--------	$<10^{-4}$	--------	--------
L154M	$<10^{-4}$	--------	--------	NA		
S182A	$<10^{-4}$	--------	--------	$<10^{-4}$	--------	--------
Y185A	$<10^{-4}$	--------	--------	$<10^{-4}$	--------	--------
F186A	$(2.1 \pm 0.1) \times 10^{-1}$	$(5.9 \pm 0.3) \times 10^3$	3.6×10^1	$(3.7 \pm 0.1) \times 10^{-2}$	$(7.5 \pm 0.4) \times 10^2$	4.9×10^1
E188A	$(6.9 \pm 0.3) \times 10^{-2}$	$(2.2 \pm 0.2) \times 10^1$	3.1×10^3	$(5.8 \pm 0.1) \times 10^{-2}$	$(1.0 \pm 0.1) \times 10^3$	5.8×10^1

The HAD subfamily IIB cap domains can be subclassified into small (*αββαββ*) or large (*αββ(αβαβ)αββ*) domain groups. The large domain family contain an *αβαβ* insertion of a small domain by using two loops to cover the core domain. The *αββαββ* loop is

utilized for determining primary substrate specificity whereas the *αβαβ* loop is the secondary substrate specificity domain. Both BT3352 and yidA contain primary (residues 180-190 in BT3352 numbering and residues 182-192 in yidA numbering) and secondary (residues 122-133 in BT3352, and 124-135 in yidA) substrate specificity loops. The results of mutagenisis experiments demonstrate that these loops act to position the polar residues Glu188/Glu190 on their primary substrate specificity loops and Glu128/Glu130 on their secondary loops to accommodate the polar leaving groups in hydrolysis reactions.

3.3.5 **Native molecular weight of BT3352 and BVU_2420**

Size exclusion chromatography was carried out to determine the native molecular weight BT3352 and BVU_2420 (Figure 3.12; Table 3.6)). BT3352 is a monomer in solution whereas BVU_2420 is an octomer.



(a) (b)

Figure 3.11 SEC of recombinantly expressed and purified proteins (a)BT3352; (b) BVU2420.

Table 3.7. Size exclusion chromatography evaluation table

Enzymes	Monomer		Sec	Oligomer
	$M_{w,\,calc}$ (kDa)	$M_{w,\,obs}$ (kDa)	x-mer_{obs}	
BT3352	32.2	20.9	0.65	monomer
BVU_2420	16.7	130.2	8.3	octomer

3.3.6 Determination of the substrate specificity profile for BVU_2420 and comparison to those of the HAD-thioesterase fusion proteins BF1314 and PG1653.

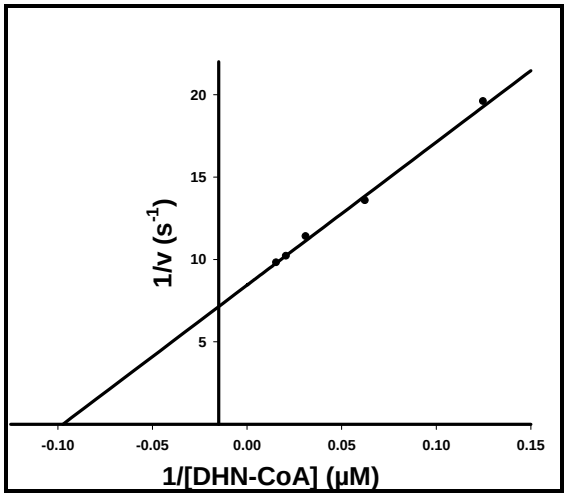


Figure 3.12. Lineweaver-Burk plot of the BVU_2420 catalyzed DHN-CoA hydrolysis in 50 mM HEPES (pH 7.5, 25 ^{o}C).

Table 3.8. Comparison of the substrate thioesterase activity of BF1314, PG1653, and BVU_2420. The steady-state kinetic constants k_{cat} and K_m were measured in 50 mM K^+HEPES (pH 7.5, 25 ^oC) using the DTNB assay.

Substrate	Variable Proteins	k_{cat} (s^{-1})	Km (uM)	K_m/k_{cat} ($M^{-1}s^{-1}$)
Butyryl-CoA	PG1653	$<10^{-4}$	--------	--------
	BF1314	$<10^{-4}$	--------	--------
	BVU_2420	$(3.5 \pm 0.1) \times 10^{-2}$	$(6.9 \pm 0.2) \times 10^{2}$	5.1×10^{1}
Isobutyryl-CoA	PG1653	$<10^{-4}$	--------	--------
	BF1314	$<10^{-4}$	--------	--------
	BVU_2420	$(2.2 \pm 0.1) \times 10^{-2}$	$(6.1 \pm 0.4) \times 10^{2}$	3.6×10^{1}
Hexanoyl-CoA	PG1653	$<10^{-4}$	--------	--------
	BF1314	$<10^{-4}$	--------	--------
	BVU_2420	$(4.2 \pm 0.2) \times 10^{-2}$	$(3.3 \pm 0.3) \times 10^{3}$	1.3×10^{1}
Octanoyl-CoA	PG1653	$<10^{-4}$	--------	--------
	BF1314	$<10^{-4}$	--------	--------
	BVU_2420	$(3.2 \pm 0.3) \times 10^{-3}$	$(1.7 \pm 0.1) \times 10^{4}$	1.9×10^{-1}
Decanoyl-CoA	PG1653	$<10^{-4}$	--------	--------
	BF1314	$<10^{-4}$	--------	--------
	BVU_2420	$<10^{-4}$	--------	--------
Myristoryl-CoA	PG1653	$<10^{-4}$	--------	--------
	BF1314	$<10^{-4}$	--------	--------
	BVU_2420	$<10^{-4}$	--------	--------
DHN-CoA	PG1653	$(1.0 \pm 0.1) \times 10^{1}$	4.0 ± 0.3	2.5×10^{6}
	BF1314	$(1.2 \pm 0.1) \times 10^{1}$	7.0 ± 0.5	1.7×10^{6}
	BVU_2420	$(1.3 \pm 0.1) \times 10^{1}$	$(1.5 \pm 0.1) \times 10^{1}$	8.7×10^{5}
NP-CoA	PG1653	6.9 ± 0.2	$(1.7 \pm 0.2) \times 10^{2}$	4.0×10^{4}
	BF1314	6.4 ± 0.3	$(2.3 \pm 0.4) \times 10^{2}$	2.9×10^{4}
	BVU_2420	4.6 ± 0.2	$(1.8 \pm 0.2) \times 10^{2}$	2.5×10^{4}
4HBA-CoA	PG1653	4.6 ± 0.1	$(8.0 \pm 0.6) \times 10^{2}$	5.8×10^{3}
	BF1314	1.1 ± 0.1	$(2.3 \pm 0.2) \times 10^{2}$	4.7×10^{3}
	BVU_2420	1.2 ± 0.1	$(3.3 \pm 0.2) \times 10^{1}$	3.6×10^{4}
Benzoyl-CoA	PG1653	$(3.0 \pm 0.1) \times 10^{-2}$	$(5.8 \pm 0.5) \times 10^{1}$	5.2×10^{2}
	BF1314	$(2.4 \pm 0.2) \times 10^{-2}$	$(8.4 \pm 0.8) \times 10^{1}$	2.8×10^{2}
	BVU_2420	$(2.8 \pm 0.1) \times 10^{-2}$	$(3.1 \pm 0.4) \times 10^{1}$	9.0×10^{2}
3HPA-CoA	PG1653	$(3.1 \pm 0.2) \times 10^{-1}$	$(1.7 \pm 0.3) \times 10^{2}$	2.0×10^{3}
	BF1314	$(1.9 \pm 0.1) \times 10^{-1}$	$(4.6 \pm 0.4) \times 10^{1}$	4.1×10^{3}
	BVU_2420	$(3.4 \pm 0.1) \times 10^{-2}$	$(8.1 \pm 0.6) \times 10^{2}$	4.2×10^{1}

Like BF1314 and PG1653, BVU2420 displayed the highest activity towards the catalyzed the hydrolysis of DHN-CoA: k_{cat}/K_m value of 9×10^5 $M^{-1}s^{-1}$ (Table 3.7 and Figure 3.12) and a preference for aromatic thioester substrates. In contrast to BF1314 and

PG1653, BVU2420 showed some activity with short chain fatty acyl-CoA thioester metabolites.

3.4 Conclusion

The investigation of BT3352 was initiated to gain further insight into why the HAD domain of the HAD-thioesterase fusion proteins display little to no phosphatase activity and to obtain a crystal structure of a close homolog. Based on its substrate activity profile we conclude that BT3352 is a promiscuous phosphatase which based on its X-ray structure is designed to target small (C4 and C5) sugar substrates. It's (type C2b) monomeric structure is in contrast to the tetrameric structure of the HAD-thioesterase fusion protein. The similarity between the structures of BT3352 and the *E. coli* enzyme ydiA suggest common ancestory despite the ability of the ydiA to act on larger sugar phosphate substrates. Active site residues Asp128/130, and Asp188/190 in BT3352/yidA play important roles in recognizing and binding substrates. In addition, residues S182/S184, K152/K154 in BT3352/yidA may be used to orient E188/E190 in appropriate substrate binding positions. Replacement of Asp128 and Asp188 with Ala enabled BT3352 to catalyze the dephosphorylation of KDPG and KDPGal with modest efficiency, and thus behave like BT3352.

The HAD-thioesterase fusion protein thioesterase domain homolog proved to be a DHN-CoA thioesterase but disspointingly, to also be a poor subject for X-ray stucture determination. Its high oligeromization state might be the cause. In order to obtain the structural information that will provide insight into substrate recognition other close homologues should be pursued in future work.

3.5 References

1. Finn, R.D., et al., *The Pfam protein families database.* Nucleic Acids Res, 2010. **38**(Database issue): p. D211-22.
2. Dai, J., et al., *Conformational cycling in beta-phosphoglucomutase catalysis: reorientation of the beta-D-glucose 1,6-(Bis)phosphate intermediate.* Biochemistry, 2006. **45**(25): p. 7818-24.
3. Kim, Y., et al., *A conserved phosphatase cascade that regulates nuclear membrane biogenesis.* Proc Natl Acad Sci U S A, 2007. **104**(16): p. 6596-601.
4. Silvaggi, N.R., et al., *The X-ray crystal structures of human alpha-phosphomannomutase 1 reveal the structural basis of congenital disorder of glycosylation type 1a.* J Biol Chem, 2006. **281**(21): p. 14918-26.
5. Zhang, G., et al., *Investigation of metal ion binding in phosphonoacetaldehyde hydrolase identifies sequence markers for metal-activated enzymes of the HAD enzyme superfamily.* Biochemistry, 2004. **43**(17): p. 4990-7.
6. Burroughs, A.M., et al., *Evolutionary genomics of the HAD superfamily: understanding the structural adaptations and catalytic diversity in a superfamily of phosphoesterases and allied enzymes.* J Mol Biol, 2006. **361**(5): p. 1003-34.
7. Allen, K.N. and D. Dunaway-Mariano, *Markers of fitness in a successful enzyme superfamily.* Current Opinion in Structural Biology, 2009. **19**(6): p. 658-665.
8. Zhang, G., et al., *Catalytic cycling in beta-phosphoglucomutase: a kinetic and structural analysis.* Biochemistry, 2005. **44**(27): p. 9404-16.
9. Lu, Z., D. Dunaway-Mariano, and K.N. Allen, *HAD superfamily phosphotransferase substrate diversification: structure and function analysis of*

HAD subclass IIB sugar phosphatase BT4131. Biochemistry, 2005. **44**(24): p. 8684-96.

10. Terwilliger, T.C., et al., *Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard.* Acta Crystallogr D Biol Crystallogr, 2008. **64**(Pt 1): p. 61-9.
11. Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.
12. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.
13. Bond, C.S., *TopDraw: a sketchpad for protein structure topology cartoons.* Bioinformatics, 2003. **19**(2): p. 311-2.
14. Potterton, E., et al., *The CCP4 molecular-graphics project.* Acta Crystallogr D Biol Crystallogr, 2002. **58**(Pt 11): p. 1955-7.
15. Schrödinger, L., *The PyMOL Molecular Graphics System, Version 1.2r3pre*.
16. Peisach, E., et al., *The X-ray crystallographic structure and activity analysis of a Pseudomonas-specific subfamily of the HAD enzyme superfamily evidences a novel biochemical function.* Proteins, 2008. **70**(1): p. 197-207.
17. Oschlies, M., et al., *A C-Terminal Phosphatase Module Conserved in Vertebrate CMP-Sialic Acid Synthetases Provides a Tetramerization Interface for the Physiologically Active Enzyme.* Journal of Molecular Biology, 2009. **393**(1): p. 83-97.

18. Zhang, J., *Evolution by gene duplication: an update.* Trends in Ecology & Evolution, 2003. **18**(6): p. 292-298.

19. McGinnis, S. and T.L. Madden, *BLAST: at the core of a powerful and diverse set of sequence analysis tools.* Nucleic Acids Res, 2004. **32**(Web Server issue): p. W20-5.

20. Lu, Z., D. Dunaway-Mariano, and K.N. Allen, *The catalytic scaffold of the haloalkanoic acid dehalogenase enzyme superfamily acts as a mold for the trigonal bipyramidal transition state.* Proc Natl Acad Sci, 2008. 105(15):p.5687-5692.

21. Lu, Z., D. Dunaway-Mariano, and K.N. Allen, The X-ray crystallographic structure and specificity profile of HAD superfamily phosphohydrolase BT1666: Comparison of paralogous functions in *B. thetaiotaomicron*, Proteins**,** 2011. 79(11):p.3099-3107.

Chapter 4. A Genomic Enzymology Approach to the Discovery of a Novel Fatty-Acid β-Oxidation Pathway in *Pseudomonas aeruginosa PAO1*

4.1 Introduction

Enzymes, which function in a common constitutive metabolic pathway in bacteria are often encoded by gene operons or if not operons, then by gene clusters. In this manner, the synthesis of the individual enzymes can be coordinated so that all are present when needed. Through the process of systematic function determination of neighboring genes it is often possible to construct a novel metabolic pathway. The first step to be taken is to use a BLAST search of protein sequence databanks to identify the closest characterized homolog to each enzyme and hence the ranges of chemistries catalyzed by known members of the respective enzyme families. Like pieces of a puzzle the investigator tries to place the profiled enzymes in order to generate what might be a feasible sequence of reaction types which would could ultimately be linked to the correct starting materials/products. To test the model *in vitro* substrate activity testing can be carried out using the recombinant enzymes and synthetic substrates.

Because a focus of my doctoral work was to discover novel hotdog-fold thioesterase functions in bacterial systems I was most interested in identifying a novel chemical pathway in which a hotdog thioesterase participated. Thus my first step was to identify an uncharacterized hotdog-fold thioesterase whose gene context was suggestive of a partnership with the protein products of neighboring genes. This was accomplished using the web server *Superfamily* (http://supfam.cs.bris.ac.uk/SUPERFAMILY/

index.html) to identify hotdog-fold thioesterase paralogues in genome sequenced bacteria. I selected the putative thioesterase PA5185 of the human pathogen *Pseudomonas aeruginosa PAO1* as a promising candidate.

In this chapter I report the results from a combined bioinformatic and experimental analysis of PA51585 and it's the protein products of its neighboring genes that support the assignment of novel metabolic pathway in the pathogen *Pseudomonas aeruginosa* which is catalyzed by four the encoded enzymes: PA5185 (the putative hotdog fold thioesterase), PA186 (a putative iron containing alcohol dehydrogenase), PA1857 (a putative acyl-CoA dehydrogenase) and PA188 (a putative 3-hydroxyacyl-CoA dehydrogenase). The nature of the types of reactions possible catalyzed by homologues of these four enzymes suggested that they are functioning in an organic acid β-oxidation pathway. My goal was to define the substrate, intermediates and the product of the pathway through defining the substrate specificity profiles of the pathway enzymes.

4.2 Materials and Methods

4.2.1 PA5185 Gene Cloning and Expression in *E. coli* and Recombinant Protein Purification

The gene encoding the PA5185 (NCBI accession NP253872, Swiss Pro accession: Q9HU04) was amplified by PCR from *Pseudomonas aeruginosa PAO1* genomic DNA and then subcloned into the pET-28a vector, which adds a His_6 tag to the protein N-terminus. Oligonucleotide primers were 5'- A GGA CCG CAT ATG GCG ACC GCC CCG CGA (PA5185 forward), 5'- G CGA AAG CTT TCA CTG GGC CGA GCT CTG

AA (PA5185 reverse), which contained the restriction endo-nuclease cleavage sites *Nde I* and *Hind III*. The cloned gene was confirmed by full length DNA sequencing.

Gene expression was performed using BL21(DE3) competent cells. Following growth in 3 L of LB media (1x) containing 50 μg/mL Kanamycin at 37 °C for 6 h with mixing at 220 rpm, induction was initiated with the addition of 0.2 mM isopropyl-1-thio-β-D-galactopyranoside. After incubation for 12 h at 20 °C and 170 rpm, the cells were harvested by centrifugation (6500 rpm for 15 min at 4 °C), resuspended in 90 mL of lysis buffer (50 mM HEPES, 300 mM NaCl, 10 mM imidazole, pH 8.0) and lysed by 2 passages through a French Press cell (12,000 psi). The cell debri was removed by centrifugation at 20,000 rpm for 60 min at 4 °C, and the supernatant containing PA5185 was passed through a Ni^{+2} affinity column for His_6-tagged protein binding. The column was washed with 200 mL of lysis buffer and 100 mL of wash buffer (50 mM HEPES, 300 mM NaCl, 20 mM imidazole, pH 8.0) before PA5185 was subsequently eluted using 100 mL of elution buffer (50 mM HEPES, 300 mM NaCl, 250 mM imidazole, pH 8.0). The protein solution was concentrated by using 10 kDa Macrosep centricon (Pall Filtron) device.

The His_6-tagged PA5185 (5 mL, 3.3 mg/mL and pH=7.4) was digested by incubating with 30 uL of Thrombin (500 units/mL DDI H_2O) at ambient temperature for 12 h. The removal of the His_6-tag protein was verified by MS spectroscopy. The protein solution was dialyzed against 6 L 50 mM HEPES and 300 mM NaCl (pH 7.5) at 4 ^oC. The protein purity was verified by SDS-PAGE analysis. The protein concentration was determined by the Bradford method and by measuring the absorption of the protein solution at 280 nm (ε = 15540 M^{-1} cm^{-1}).

4.2.2 Preparation of PA5185 site directed mutants.

The D26A, D26N and D41A PA5185 mutant genes were prepared by using the quick change mutagenesis kit (Stratgene) in conjunction with the pET-28a/PA5185 clone as template. Plasmid DNA was purified using the Qiaprep Spin Miniprep Kit. Gene sequencing was carried out at the Center for Genetics in Medicine, University of New Mexico School of Medicine.

Gene expression was performed using BL21(DE3) competent cells. Following growth in 3 mL of LB media (1x) containing 50 μg/mL Kanamycin at 37 °C for 6 h with mixing at 220 rpm, induction was initiated with the addition of 0.2 mM isopropyl-1-thio-β-D-galactopyranoside. After incubation for 12 h at 20 °C and 170 rpm, the cells were harvested by centrifugation (6500 rpm for 15 min at 4 °C), resuspended in 90 mL of lysis buffer (50 mM HEPES, 300 mM NaCl, 10 mM imidazole, pH 8.0) and lysed by 2 passages through a French Press cell (12,000 psi). The cell debri was removed by centrifugation at 20,000 rpm for 60 min at 4 °C, and the supernatant containing D26A, D26N or D41A PA5158 was passed through a Ni^{+2} affinity column. The column was washed with 200 mL of lysis buffer and 100 mL of wash buffer (50 mM HEPES, 300 mM NaCl, 20 mM imidazole, pH 8.0) before elution with 100 mL of elution buffer (50 mM HEPES, 300 mM NaCl, 250 mM imidazole, pH 8.0). The protein solution was concentrated by using a 10 kDa Macrosep centricon (Pall Filtron) device and then dialyzed at 4 ^{o}C against 50 mM HEPES and 300 mM NaCl (pH 7.5). The protein purity was verified by SDS-PAGE analysis. The protein concentration was determined by the Bradford method and by measuring the absorption at 280 nm.

4.2.3 Synthesis of 2-keto-3-propanyl-CoA.

A 15 uL (0.6 mmol) aliquoate of acetone bromide was dissolved in 2 mL of anhydrous THF before adding dropwise to a stirred solution of 50 mg (0.063 mmol) of CoA (Lithium salt) in 1 mL of deionized water at room temperature. The reaction solution was maintained at pH 9.0 by adding 0.1 M LiOH or 0.1 M HCl. The mixture was stirred for 2 h and then was extracted 3 times with 4 mL of ethyl acetate. The water layer was frozen and lyophilized. The solid was dissolved in 300 μL of deionized water and then loaded onto a Sephadex G-10 column (95 x 1.5 cm).The column was eluted with deionized water at room temperature. The 3 mL fractions were analyzed by injecting 50 μL aliquots on to a Rainin Dynamax HPLC system equipted with a reversed-phased C-18 column (Beckmen Ultrasphere), pre-equilibrated with 20% solution B (20 mM $(NH_4)_2HPO_4$ in 64% acetonitrile DI water solution, pH 6.7) and 80% solution A (20 mM $(NH_4)_2HPO_4$ in DI water, pH 6.7). A linear gradient (20% B 2 minutes, 20~65% solution B14 min, 65~80% solution B 2 minutes and 20% solution B 2 min) was employed to elute the column at a flow rate of 1.0 mL/min. The eluant was monitored at 260 nm. The retention time of the product is 3.3 min. The fractions 19-29 containing pure 2-keto-3-propanyl-CoA were pooled and concentrated by lyophilization, the yield is *ca.*20%, and the structure was confirmed by ^{1}H NMR and Mass analysis. (500 MHz, D_2O): δ 8.63 (s, 1H), 8.36 (s, 1H), 8.41 (s, 1H), 6.24 (d, 1H), 6.22 (d, 1H), 4.92 (d, 2H), 4.91 (d, 2H), 4.64 (s, 1H), 4.28 (s, 2H), 4.06 (s, 1H), 3.87 (t, 1H), 3.86 (t, 1H), 3.85 (t, 1H), 3.61 (d, 1H), 3.60 (d, 1H), 3.58 (s, 3H), 3.51 (s, 2H), 3.36 (s, 2H), 2.64 (s, 2H), 2.51 (s, 2H), 2.26 (s, 3H), 0.94 (s, 3H), 0.81 (s, 3H). MS: *m/z* calculated for $C_{24}H_{39}N_7O_{17}P_3S$ 823.13 compared to the observed mass of 822.14.

4.2.4 PA5187 and PA5188 Gene Cloning and Expression in *E. coli* and Recombinant Protein Purification

The genes encoding the PA5187 and PA5188 (NCBI accession NP253874, NP253875, Uni Prot accession: Q9HU02, Q9HU01 respectively) were amplified by PCR from *Pseudomonas aeruginosa PAO1* genomic DNA and then subcloned into the pET-28a vectors, which add a His_6 tag to the protein N-terminus. Oligonucleotide primers were 5'- AGG TGC GCT CAT ATG GAT TAC CAG GCC CCC (PA5187 forward), 5'- GCG AGA AAG CTT TCA GAA CGA GAA ATG CTC (PA5187 reverse), 5'- GAG GAC AAG CAT ATG AGC CAG AAC TTT TCC (PA5188 forward) and 5'- TGG TAA AAG CTT TCA GCG CAC CTC CTG CAG (PA5188 reverse), which contained the restriction endo-nuclease cleavage sites *Nde I* and *Hind III*. Subcloned genes were confirmed by full length DNA sequencing.

Both PA5187 and PA5188 were purified by FPLC. Gene expression was performed using BL21(DE3) competent cells. Following growth in 3 L of LB media containing 50 μg/mL ampicillin at 37 °C for 6 h with mixing at 220 rpm, induction was initiated with 0.2 mM isopropyl-1-thio-β-D-galactopyranoside. After incubation for 12 h at 20 °C 170 rpm, the cells were harvested by centrifugation (6500 rpm for 15 min at 4 °C), resuspended in 50 mL of binding buffer (20 mM HEPES, 500 mM NaCl, 40 mM imidazole, pH 7.4) and lysed by a single pass through a French Press cell (12,000 psi). The cells were removed by centrifugation at 20,000 rpm for 45 min at 4 °C, and the supernatant containing PA5187 was passed through a GE Histrap FF crude 1 mL column which had been pre-equilibrate with 5 CV binding buffer (20 mM K^+HEPES, 500 mM

NaCl, 40 mM imidazole, pH=7.4) at 4 ^{o}C. The column was washed with 2 CV (10 mL) binding buffer and then eluted with 5 CV (25 mL) elution buffer (20 mM HEPES, 500 mM NaCl, 500 mM imidazole, pH 7.4). The column fractions were monitored at 280 nm and the desired fraction was concentrated by using a 10 kDa Macrosep centricon (Pall Filtron) device before dialyzing against 6 L 50 mM HEPES and 300 mM NaCl (pH 7.5) at 4 ^{o}C. The protein concentration was determined by the Bradford method and by measuring the absorption at 280 nm (ε = 56500 M^{-1} cm^{-1}, and 41320 M^{-1} cm^{-1} for PA5187 and PA5188 respectively) The protein purity was verified by SDS-PAGE analysis.

4.2.5 PA3302 (*R*-specific Enoyl-CoA Hydratase) Gene Cloning and Expression in *E. coli* and Recombinant Protein Purification (22)

The gene encoding the PA3302 (NCBI accession NP251992, Uni Prot accession: Q9LBK2) was amplified by PCR from *Pseudomonas aeruginosa PAO1* genomic DNA and ligated into the pET-14b vector, which adds a His_6 tag to the protein N-terminus. Oligonucleotide primers of PA5187 containing restriction the endo-nuclease cleavage sites *NdeI* and *HindIII* were used in the PCR reactions. Plasmid DNA was purified using a Qiaprep Spin Miniprep Kit.

Gene expression was carried out in *E. coli* BL21(DE3) competent cells. Following growth of the transformed cells for 6 h at 37 °C in 3 mL of LB media containing 50 μg/mL Ampicillin, induction was initiated with the addition with 0.2 mM isopropyl-1-thio-β-D-galactopyranoside. Following incubation for 12 h at 20 °C 170 rpm, the cells were harvested by centrifugation (6500 rpm for 15 min at 4 °C), resuspended in

50 mL of binding buffer (20 mM HEPES, 500 mM NaCl, 40 mM imidazole, pH 7.4) and lysed by a single pass through a French Press cell (12,000 psi). The cells were removed by centrifugation at 20,000 rpm for 45 min at 4 °C, and the supernatant containing PA5187 was passed through a GE Histrap FF (1 mL) column pre-equilibrated with 5 CV (25 mL) binding buffer (20 mM K^+HEPES, 500 mM NaCl, and 40 mM imidazole, pH 7.4) at 4 ^oC. The column was first washed with 2 CV (10 mL) binding buffer and then the PA5187 was eluted with 5 CV (25 mL) elution buffer (20 mM HEPES, 500 mM NaCl, 500 mM imidazole, pH 7.4). The column fractions were monitored at 280 nm and the desired fraction was concentrated by using a 10 kDa Macrosep centricon (Pall Filtron) device before dialyzing against 6 L 50 mM HEPES and 300 mM NaCl (pH 7.5) at 4 ^oC. The protein concentration was determined by the Bradford method and by measuring the absorption at 280 nm. The protein purity was verified by SDS-PAGE analysis.

4.2.6 Purification of Recombinant *S*-specific Enoyl-CoA Hydratase

The clone of S-specific enoyl-CoA hydratase (from rat mitochondrial, NCBI accession NP_511178, Uni Prot accession: P14604). The gene was ligated into pET-15b vector, and expressed in *E. coli* (13). The clone was a gift from Professor Peter Tonge of the State University of New York at Stony Brook. The transformed *E. coli* BL21(DE3) cells were cultured with mixing (220 rpm) at 37 °C for 6 h grown in 3 L of LB media containing 50 μg/mL Ampicillin before inducing with 0.2 mM isopropyl-1-thio-β-D-galactopyranoside. After incubation for 12 h at 20 °C (mixing at 170 rpm), the cells were harvested by centrifugation (6500 rpm for 15 min at 4 °C), resuspended in 50 mL of binding buffer (20 mM HEPES, 500 mM NaCl, 40 mM imidazole, pH 7.4) and lysed by a

single pass through a French Press cell (12,000 psi). The cell debri were removed by centrifugation at 20,000 rpm for 45 min at 4 °C, and the supernatant containing N-terminal His_6-tagged PA5187 was passed through a 1 mL GE Histrap FF pre-equilibrated with 5 CV (25 mL) binding buffer (20 mM HEPES, 500 mM NaCl, 40 mM imidazole, pH 7.4). The column was first washed with ? mL CV binding buffer (define) before eluting with 2 CV (10 mL) elution buffer (20 mM HEPES, 500 mM NaCl, 500 mM imidazole, pH 7.4). The column fractions were monitored at 280 nm and the desired fraction was concentrated by using a 10 kDa Macrosep centricon (Pall Filtron) device before dialyzing against 6 L 50 mM HEPES and 300 mM NaCl (pH 7.5) at 4 ^{o}C. The protein concentration was determined by the Bradford method and by measuring the absorption at 280 nm. The protein purity was verified by SDS-PAGE analysis.

4.2.7 Synthesis of *trans*-2,3-Alkenoyl-CoA Thioesters (C6, C8, C10)

The trans-2,3-hexenoyl-CoA, trans-2,3 octenoyl-CoA and trans-2,3-decenoyl-CoA were synthesized using previous reported method (8). In a standard preparation, *trans*-2,3-hexenoic acid, *trans*-2,3-octenoic acid or *trans*-2,3 decenoic acid (1 mmol) was dissolved in tetrahydrofuran (2 mL), which have been freed of water and ether peroxides by distillation over $LiAlH_4$. The acids were neutralized with triethylamine (1.1 mmol) and allowed to react with ethyl chloroformate (1.1 mmol). After 10 min at 25^{o} the solution was rapidly filtered through a Pasteur pipette containing a glasswool plug and then added dropwise over a period of 10 min to a solution of CoA (Li^+ salt; 40 μmol) dissolved in a mixture (5 mL) of water and tetrahydrofuran (3:2 v/v) that had been adjusted to pH 8 by the addition of $NaHCO_3$ (40 mg) and that had been purged with N_2. In order to optimize the yield of product, phase separation in the reaction mixture was prevented by the addition of a few drops of water. After a total reaction time of 20 min, the pH of

the reaction mixture was adjusted to approximately 3 and the tetrahydrofuran was removed by evaporation under vacuum. The trans-2,3-hexenoyl-CoA and trans-2,3-octenoyl-CoA were used without further purification after the excess alkanoic acid was removed by extraction with pentane. The trans-2,3-decenoyl-CoA was purified by first precipitating it by saturating the crude aqueous concentrate (pH 1) with NaCl and then collecting and washing the precipitate with pentane.

The products were verified by mass spectral analysis. The observed m/z of trans-2,3-hexenoyl-CoA, trans-2,3-octenoyl-CoA and trans-2,3-decenoyl-CoA are 860.3 (calct 859.1) 888.2 (calct 887.1), and 916.5 (calct 915.5) respectively. which matches the theoretical M_w of hexenoyl-CoA (859.1), octenoyl-CoA (887.1).

4.2.8 Enzyme Kinetics.

Activity Assays. Thioesterase activity was measured using the DTNP coupling assay. The reaction solutions (1 mL) contained specified concentrations of substrate and 1 mM DTNB in 0.15 M KCl and 50 mM K^+HEPES (pH 7.5). The progress of the reaction was monitored at 412 nm. The acyl-CoA dehydrogenase activity was measured at 600 nm by coupling dihydroflavinadenine dinucleotide $FADH_2$ formation to the reduction of 2,6-dichlorphenol-indophenol (DCPIP) (19). The reaction solutions initially contained 50 mM K^+HEPES (pH 7.5), a specified concentration of enzyme, 150 μM DCPIP, 100 μM phenazine methosulphate (PMS) and 10 μM flavinadenine dinucleotide (FAD). The 3-hydroxyacyl-CoA dehydrogenase (PA5186) activity was measured by generating the *S*- or *R*-3-hydroxyacyl-CoA substrate *in situ* by incubation of a specified concentation of the trans-2,3-alkenoyl-CoA (*vide infra*) with 5 μM (*S*-specific) enoyl-CoA hydratase or 5 μM of the

(*R*-specific) enoyl-CoA hydratase PA3302 in 0.5 mL of 50 mM K^+HEPES for 5 min before adding NAD to a final concentration of 0.5 mM. The reaction was initiated by the addition of a small aliquot (20 μL) of the 3-hydroxyacyl-CoA dehydrogenase (PA5188) stock solution (0.5μM). The reaction progress was monitored at 340 nm.

Steady-state kinetic constant determinations. The initial velocity of the reaction was measured as a function of the concentration of the substrate, varied in a range of 0.5 to 10-fold K_M. The initial velocity data were fitted with the enzyme kinetic module in Sigmaplot to define V_{max} and K_M. The k_{cat} was calculated from the ratio of V_{max} and enzyme concentration.

$$V = V_{max}[S]/([S] + K_M)$$

where V is the initial velocity, V_{max} the maximum velocity, [S] the substrate concentration, K_m the Michaelis constant.

4.2.9 Size exclusion chromatography

Size exclusion column chromatography was performed at 4 ^oC on ÄKTA P-920 FPLC system and Hiprep 16/60 Sephacryl S-200 HR column (GE Healthcare) pre-equilibrated with 50 mM HEPES (pH 7.5), 500 mM NaCl, and 1 mM DTT. The column was calibrated using the gel-filtration low molecular weight and high molecular weight gel filtration calibration kit from Amersham Biosciences. Purified protein (2 mL, ca. 200 mM) was applied to the column.

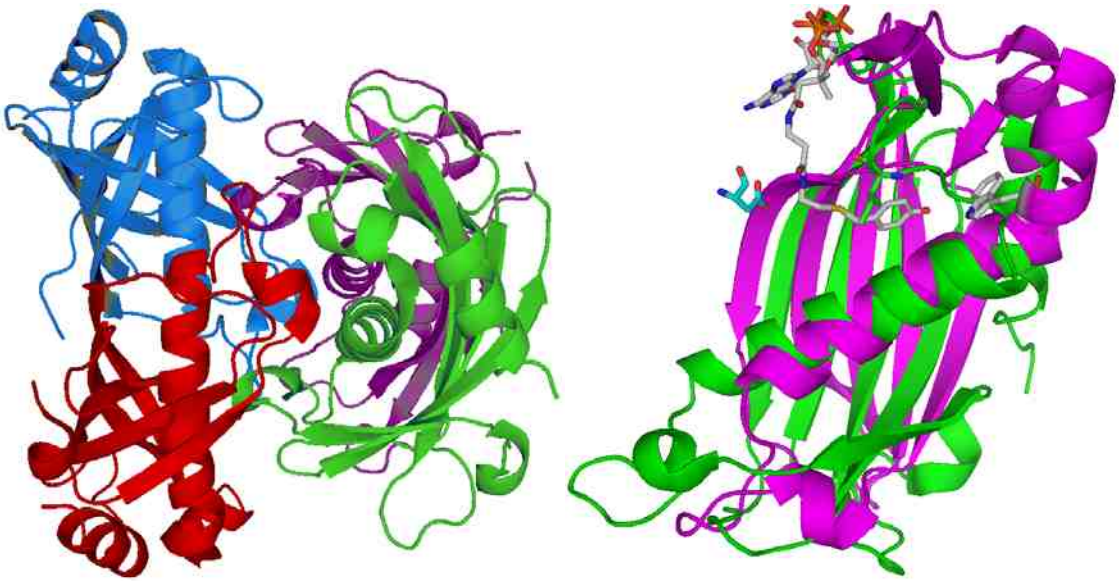
4.3 Results and Discussion

4.3.1 Comparison of the PA5185 X-ray Structure and Amino Acid Sequence to Clade AA Hotdog-fold Thioesterases

Figure 4.1. Sequence alignment of PA5185 and five other clade AA homologues. The X-ray structure of PA5185 was superimposed with the X-ray structures of these hotdog-fold family members (PDB ID: 1LO8, 2OAF, 2ALI, 2FUJ, and 2CYE) using the Dali structural alignment server. The structures 1LO8, 2OAF, 2ALI, 2FUJ, and 2CYE share 14%, 19%, 31%, 28% and 28% sequence identity with BT3352 and have an RMSD of 1.9, 1.7, 1.8, 1.7 Å and 2.1 Å, respectively. The figure was prepared using CLC sequence viewer.

PA5185 belongs to the hotdog-fold superfamily of thioesterase. Despite the low pair-wise sequence identity (see sequence alignment of structural homologues in Figure 4.3), the tertiary and quaternary structure of PA5185 (PDB 1BVQ) is quite similar to that of the 4-hydroxybenzoyl-CoA (4-HB-CoA) thioesterase from *Pseudomonas* sp. strain

CBS-3 (PDB 1LO8) (7), which is the prototype of the type AA clade of the hotdog-fold thioesterases (7, 20, 21). The trait of the type AA clade is the utilization of the catalytic Asp/Glu (nucleophile or general base) (loop) and Gly (at helix N-terminus; backbone amide NH polarizes the thioester C=O) on the loop-helix catalytic motif of the same subunit. The trait of the clade AB, the prototype of which is the *Arthrobacter* sp. strain SU 4-HB-CoA thioesterase, is the utilization of the catalytic Asp/Glu from one subunit (positioned on the helix of the loop-helix catalytic motif) and the catalytic Gly of the helix N-terminus of the opposing subunit.



(a) (b)

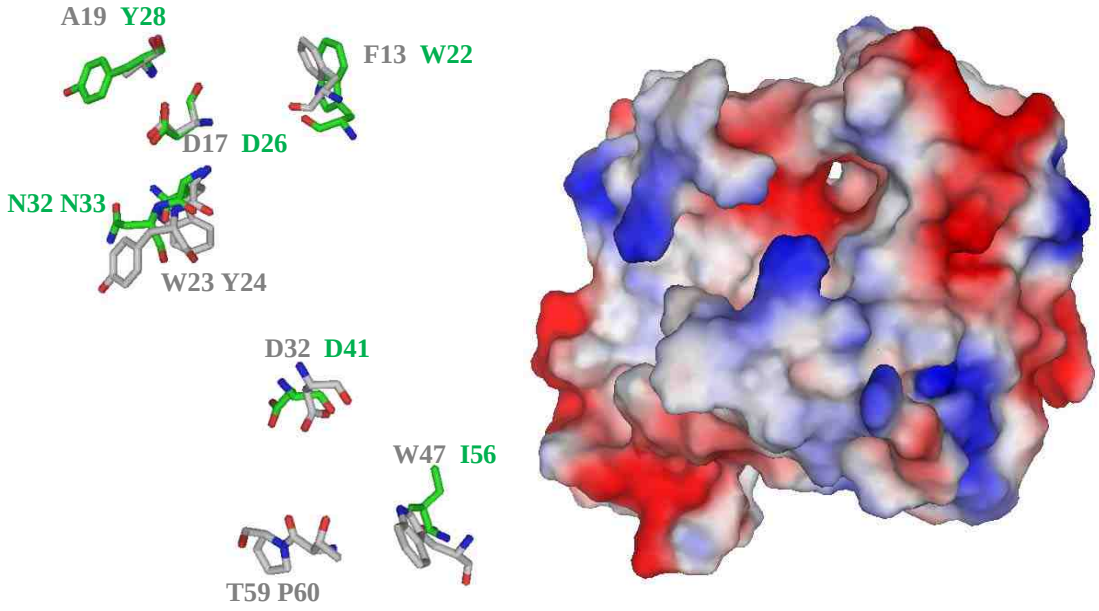


Figure 4.2 (a) The Pymol-generated stereoview of the face-to-face tetramer of PA5185. The subunits are colored individually to show the dimer units (blue and red; green and magenta). (b) The Pymol generated stereoview of the superimposed active site residues of PA5185 (green) and *Pseudomonas* 4-HB-CoA thioesterase (7) (gray). (c) The Pymol-generated stereoview of the superimposed active site residues of PA5185 (green), *Pseudomonas* 4-HB-CoA thioesterase (grey). (d) The Pymol-generated representations of the electrostatic topological features of PA5185.

The superposition of the structures of PA5185 and the *Pseudomonas* 4-HB-CoA thioesterase bound with the inhibitor 4-hydroxybenzyl-CoA show that the CoA binding site residues Ile61, Phe68, Arg89, Lys90 and Ser91 in *Pseudomonas* 4-HB-CoA thioesterase correspond to Val66, Tyr73, Asn93, Ser94 and Ser95 in PA5185 and thus, are not conserved. Nevertheless, the surface topology (finger-like electropositive projection encircling the entrance to catalytic site; Figure 4.4d) is consistent with the 4-HB-CoA thioesterase binding site for CoA

(as opposed to ACP) thiol unit of the PA5185 acetyl/propionyl-CoA substrate (see following section). The active site residues (with the exception of the catalytic Asp and Gly) of the two thioesterases are also dissimilar. Moreover, the binding pocket for the 4-hydroxybenzoyl unit of the 4-HB-CoA substrate in the 4-HB-CoA thioesterase is larger than the PA5185 binding pocket for the acetyl/propionyl unit. This is consistent with PA5185 substrate specificity profile (see following section).

4.3.2 Native molecular weight of PA5185

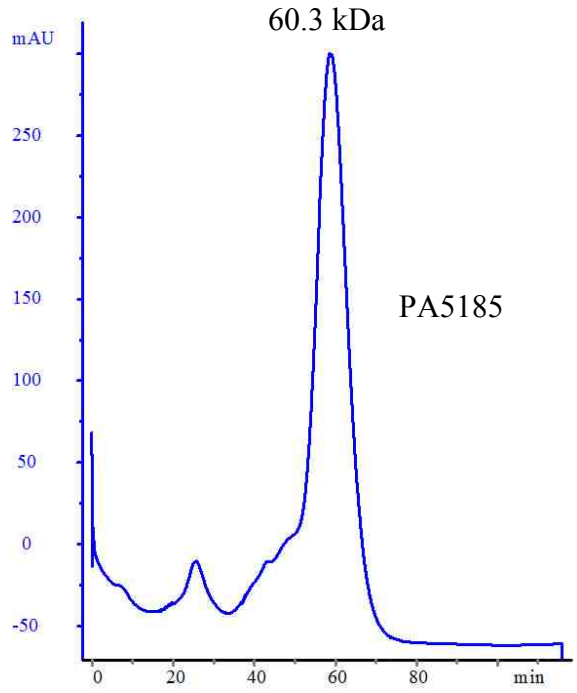


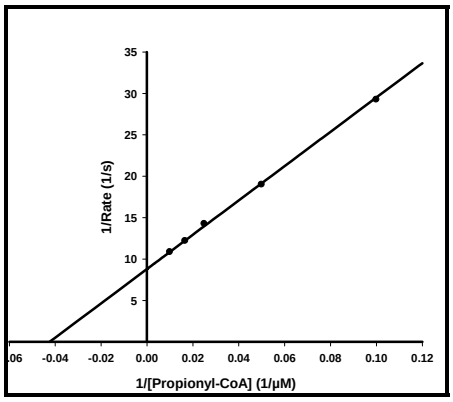
Figure 4.3. The FPLC elution profile for the chromatography of recombinant PA5185 on a size exclusion column, calibrated with protein molecular weight standards .

The oligomeric state of the PA5185 in solution was determined by size exclusion chromatography (SEC), which defined a Native mass of 60.3 kDa (Figure 4.5 and Table 4.4). The mass spectral analysis of the PA5185 defined a subunit mass of 16744 Da, which is consistent with the theoretical mass of 16745 Da. The ratio of the Native and subunit masses of 3.6 suggests that PA5185 is a tetramer, which is consistent with its X-ray structure.

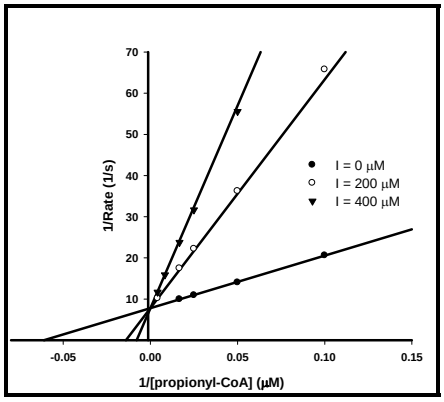
4.3.3 PA5185 Substrate Specificity Profile

Purified PA5185 looses activity quickly (even when stored at -88 ^{o}C) and was therefore subjected to substrate activity analysis soon after isolation. PA5185 exhibited narrow substrate specificity, with only C2-C4 acyl-CoAs hydrolyzed. Propionyl-CoA is the best substrate with K_m= 24 μM, and k_{cat}= 0.57 s^{-1} (k_{cat}/K_m = 2.4 x 10^4) (Figure 4.4a and Table 4.1), followed by acetyl-CoA (k_{cat}/K_m = 1.6 x 10^3). Acetyl-CoA is the expected end product of β-oxidation of fatty acids of even carbon number (common form) whereas propionyl-CoA would be generated by β-oxidation of fatty acids of odd carbon number (rare form). The 10-fold smaller K_m value measured for the propionyl-CoA vs acetyl-CoA might be due to tighter binding of the propionyl group driven by the larger area for desolvation. The acetyl-CoA is most likely the targeted substrate and the relatively large K_m value (160 μM) is consistent with its role as a housekeeper in releasing CoA from acetyl-CoA, a common metabolite, only in the event of its accumulation. PA5185 is very sensitive to the carbon chain length of the substrate: the K_m and k_{cat} for malony-CoA, butyryl-CoA and 3-hydroxybutyryl-CoA are 880 μM & 0.018 s^{-1}, 320 μM & 0.082 s^{-1} and 180 μM & 0.036 s^{-1}, respectively. Longer acyl-CoAs are not substrates and this

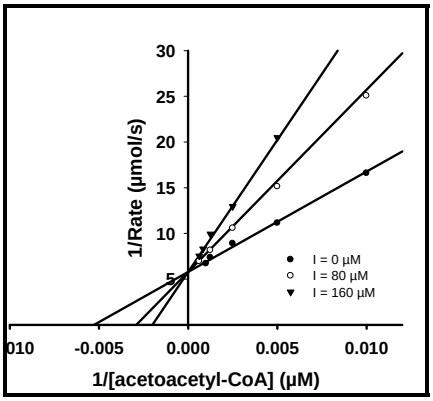
critical for PA5185's function in supporting the fatty acid β-oxidation pathway proposed below. The activity of propionyl-CoA thioesterase is modestly inhibited by the propionyl-CoA products CoASH (a competitive inhibitor K_i = 98 ± 8 μM) and propionate (a noncompetitive inhibitor K_i = 69 ± 7 μM) (Figure 4.6b and c). The substrate analog 2-keto-3-propanyl-CoA, was also found to be a competitive inhibitor vs propionyl-CoA: Ki = 60 ± 0.9 μM. (Figure 4.9d)



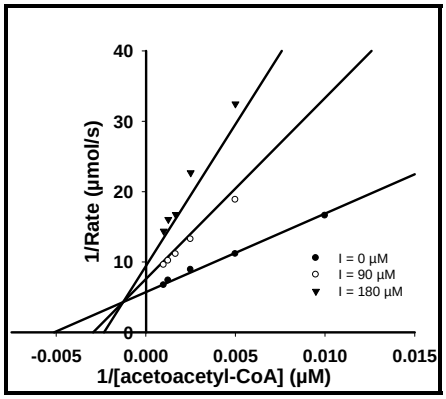
(a)



(b)



(c)



(d)

Figure 4.4. Lineweaver-Burk plot of initial velocity data from (a). PA5185 catalyzed propionyl-CoA hydrolysis reaction in 50 mM HEPES (pH 7.5, 25 ^{o}C), (b) CoA inhibition of PA5185 catalyzed propionyl-CoA hydrolysis, (c) propionate inhibition of PA5185 catalyzed propionyl-CoA hydrolysis (d) 2-keto-3-propanyl-CoA inhibition of PA5185 catalyzed propionyl-CoA hydrolysis.

Table 4.1. Steady-state kinetic constants measured for PA5185-catalyzed acyl-CoA metabolite thioester hydrolysis.

Substrate	k_{cat} (s^{-1})	K_m (uM)	k_{cat}/K_m ($M^{-1}s^{-1}$)
Short Chain Fatty Acyl-CoAs			
Acetyl-CoA	$(2.6 \pm 0.1) \times 10^{-1}$	$(1.6 \pm 0.1) \times 10^{2}$	1.6×10^{3}
n-Propionyl-CoA	$(5.7 \pm 0.2) \times 10^{-1}$	$(2.4 \pm 0.1) \times 10$	2.4×10^{4}
Malonyl-CoA	$(1.8 \pm 0.1) \times 10^{-2}$	$(8.8 \pm 0.9) \times 10^{2}$	2.0×10
n-Butyryl-CoA	$(8.2 \pm 0.2) \times 10^{-2}$	$(3.2 \pm 0.2) \times 10^{2}$	9.3×10
β-Hydroxybutyryl-CoA	$(3.6 \pm 0.1) \times 10^{-2}$	$(1.8 \pm 0.2) \times 10^{2}$	1.9×10^{2}
2-Butenoyl-CoA	$< 1 \times 10^{-4}$	--------	--------
Crotonyl-CoA	$< 1 \times 10^{-4}$	--------	--------
β-Methylcrotonyl-CoA	$< 1 \times 10^{-4}$	--------	--------
Tiglyl-CoA	$< 1 \times 10^{-4}$	--------	--------
n-Hexanoyl-CoA	$< 1 \times 10^{-4}$	--------	--------
n-Octanoyl-CoA	$< 1 \times 10^{-4}$	--------	--------
Medium-Long Chain Fatty Acyl-CoAs			
Lauroyl-CoA	$< 1 \times 10^{-4}$	--------	--------
Myristoyl-CoA	$< 1 \times 10^{-4}$	--------	--------
n-Palmitoyl-CoA	$< 1 \times 10^{-4}$	--------	--------
Stearoyl-CoA	$< 1 \times 10^{-4}$	--------	--------

Table 4.2. Steady-state kinetic constants measured for wild-type and mtuant PA5185 catalyzed hydrolysis of n-propionyl-CoA in 50 mM HEPES (pH 7.5, 25 ^{o}C).

	k_{cat} (s^{-1})	K_m (uM)	k_{cat}/K_m ($M^{-1} s^{-1}$)
WT	$(5.7 \pm 0.2) \times 10^{-1}$	$(2.4 \pm 0.1) \times 10$	2.4×10^4
D26A	$< 1 \times 10^{-4}$	--------	--------
D26N	$< 1 \times 10^{-4}$	--------	--------
D41A	$(1.5 \pm 0.1) \times 10^{-1}$	$(3.7 \pm 0.1) \times 10^2$	4.1×10^2

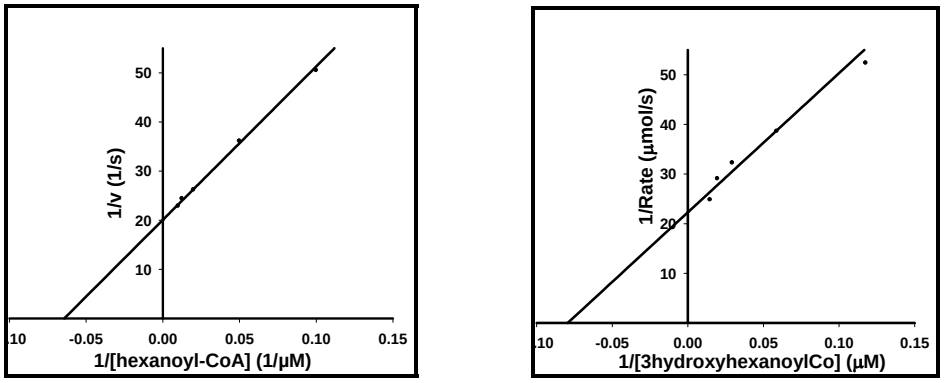
Site-directed mutagenesis was used to prepare catalytically inactive PA5185 for future structure determination of the substrate-bound enzyme. Steady-state study showed that both D26A and D26N were inactive in catalysis of propionyl-CoA (confirming the role of Asp26 as the catalytic nucleophile or base, whereas the D41A mutant significant retained activity (k_{cat}/K_m value of 4.1×10^2) (Table 4.2). D26A and D26N were both sent to Dr Karen's lab to carry out cocrystallization with propionyl-CoA.

4.3.4 Characterization of the Acyl-CoA Dehydrogenase PA5187.

To place the *in vitro* function of PA5185 in a biochemical context, its "neighborhood" gene PA5187 was cloned, expressed and purified. Its genebank annotation is "acyl-CoA dehydrogenase". The acyl-CoA dehydrogenase activity of PA5187 was tested against a series of acy-CoA metabolites (Table 4.3; Figure 4.7). PA5187 showed wide activity over C4 to C14 chain acyl-CoA substrates with decant catalytic efficiency. The k_{cat}/K_m value varied from 10^3 M^{-1} s^{-1} to 10^4 M^{-1} s^{-1}. The best substrate for PA5187 is n-hexanoyl-CoA, with K_m=15 μM, k_{cat}= 0.5 s^{-1}, and $k_{cat}/K_m = 3.3 \times 10^4$ M^{-1} s^{-1}. The Km value did not vary much although lauroyl-CoA did have the lowest K_m=9.2 μM. The k_{cat} value is more sensitive to the elongation of substrate's carbon chain. The maximum k_{cat} value was

observed for C6 to C8 acyl-CoAs. Overall, the largest substrate specificity constants (k_{cat}/K_m) were obsrved for medium chain length acyl-CoAs, thus suggesting that PA5187 partners with PA5185 in the β-oxidation of medium chain length fatty acids.

The native PA5187 in solution was determined by size exclusion chromatography to be ~152 kDa compared to the subunit mass of 66601 determined by mass spectral analysis (Figure 4.3a and Table 4.4), thus defining the protein as dimer in solution.



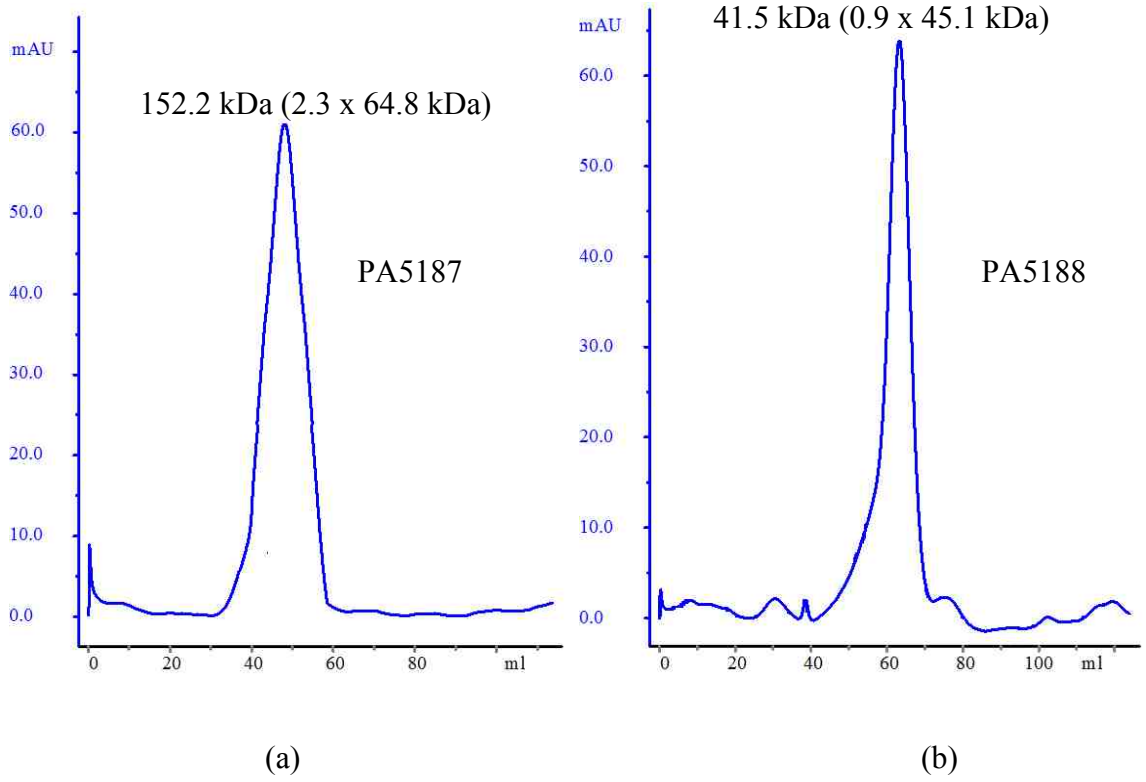
(a) (b)

Figure 4.5. Lineweaver-Burk plot of the initial velocity data from (a). PA5187 catalyzed dehydrogenation of hexanoyl-CoA in 50 mM HEPES (pH 7.5, 25 ^{o}C), and (b). PA5188 catalyzed dehydrogenation of S-specific 3-hydroxylhexanoyl-CoA in 50 mM HEPES (pH 7.5, 25 ^{o}C)

Table 4.3. Steady-state kinetic constants determined for PA5187 acyl-CoA dehydrogenase. All data are obtained at 25 ^{o}C, 50 mM HEPES buffer pH 7.5. DCPIP (2,6

dichlorophenol indophenols), PMS (Phenazine methosulphate) and FAD were used coupling system.

Substrate	k_{cat} (s^{-1})	K_m (uM)	k_{cat}/K_m ($M^{-1}s^{-1}$)
n-butyryl-CoA	$(8.8 \pm 0.6) \times 10^{-2}$	$(3.0 \pm 0.2) \times 10$	3.0×10^3
n-hexanoyl-CoA	$(5.0 \pm 0.1) \times 10^{-1}$	$(1.5 \pm 0.1) \times 10$	3.3×10^4
n-octanoyl-CoA	$(5.1 \pm 0.1) \times 10^{-1}$	$(2.1 \pm 0.2) \times 10$	2.3×10^4
n-decanoyl-CoA	$(2.8 \pm 0.2) \times 10^{-1}$	$(6.8 \pm 0.7) \times 10$	4.1×10^3
Lauroyl-CoA	$(1.2 \pm 0.1) \times 10^{-1}$	9.2 ± 0.8	1.4×10^4
Myristoyl-CoA	$(8.0 \pm 0.3) \times 10^{-2}$	$(1.1 \pm 0.1) \times 10$	8.0×10^3
n-Palmitoyl-CoA	$< 1 \times 10^{-4}$	--------	--------
Palmitoleoyl-CoA	$(6.8 \pm 0.2) \times 10^{-2}$	$(8.9 \pm 0.9) \times 10$	7.8×10^2
Stearoyl-CoA	$< 1 \times 10^{-4}$	--------	--------
Oleoyl-CoA	$< 1 \times 10^{-4}$	--------	--------
Linoleoyl-CoA	$< 1 \times 10^{-4}$	--------	--------
Arachidonyl-CoA	$< 1 \times 10^{-4}$	--------	--------



(a) (b)

Figure 4.6. SEC of recombinant expressed and purified proteins (a) PA5187; (b) PA5188.

Table 4.4. Size exclusion chromatography evaluation table

Enzymes	Monomer		Sec	Oligomer
	$M_{w,\ calc}$ (kDa)	$M_{w,\ obs}$ (kDa)	x-mer_{obs}	
PA5185	16.7	60.3	3.6	Tetramer
PA5187	64.8	152.2	2.3	Dimer
PA5188	45.1	41.5	0.9	Monomer

4.3.5 Charaterization of the 3-hydroxyacyl-CoA dehydrogenase PA5188.

Another PA5185 "neighborhood" protein PA5188 is annotated as 3-hydroxyacyl-CoA dehydrogenase. It is the third enzymes in beta oxidation cycle following the enoyl – CoA hydratase. The mass spectral analysis of the PA5188 shows that the Mw of the protein is 45089.4 Da. It matches perfectly with the theoretical Mw of 45088.1 Da, computed by ExPASy. PA5188 is not very stable, and it will automatically precipitate in room temperature, so thrombin was used to remove the histag at 4 °C overnight. The oligomeric state of the PA5188 in solution was determined by size exclusion chromatography too. In the solution used (50 mM Hepes, 500 mM NaCl, 1 mM DTT), PA5188 stay as monomer. SEC results showed a calculated molecular weight at about 41.5 kDa (figure 4.3b and table 4.4), which is 0.9 times less than the monomer molecular weight showed by mass spectral analysis(45.1 kDa).

Table 4.5. PA5188 putative beta hydroxyacyl-CoA dehydrogenase substrate screen table.

All data are obtained at 25 ^{o}C, 50 mM Hepes buffer pH=7.5. Echs was used as coenzyme to product s-specific 3-hydrxoyacyl-CoA, and the increase of absorbance at 340 nm was monitored by UV-Vis spectrometer.

Substrate	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1} s^{-1}$)
(S) 3-hydroxybutyryl-CoA	$(1.6 \pm 0.1) \times 10^{-1}$	$(3.5 \pm 0.1) \times 10^{2}$	4.5×10^{2}
(R) 3-hydroxybutyryl-CoA	$< 1 \times 10^{-4}$	--------	--------
(S) 3-hydroxyhexanoyl-CoA	4.5 ± 0.3	$(1.2 \pm 0.1) \times 10$	3.8×10^{5}
(R) 3-hydroxyhexanoyl-CoA	$< 1 \times 10^{-4}$	--------	--------
(S) 3-hydroxyoctanoyl-CoA	$(7.5 \pm 0.3) \times 10^{-1}$	$(9.8 \pm 1.0) \times 10$	7.6×10^{3}
(R) 3-hydroxyoctanoyl-CoA	$< 1 \times 10^{-4}$	--------	--------
(S) 3-hydroxydecanoyl-CoA	$(7.4 \pm 0.2) \times 10^{-2}$	$(2.0 \pm 0.2) \times 10^{2}$	3.8×10^{2}
(R) 3-hydroxydecanoyl-CoA	$< 1 \times 10^{-4}$	--------	--------

The kinetic data of PA5188 was evaluated by enoyl-CoA hydratase assisted coupling assay. Enoyl-CoA hydratase is the second enzyme in the beta-oxidation cycle, it convert the enoyl-CoA, which is the protduct of acyl-CoA dehydrogenase, to 3-hydroxyacyl-CoA. The S-specific coupling enzyme echs1 is a gift from Dr Tonge' lab at SUNY, while the R-specific enoyl-CoA hydratase, $PhaJ1_{PA}$ (22), was cloned from *Pseudomonas aeruginosa PAO1*, and purified with histag. The trans hexenoyl-CoA, octenoyl-CoA and decenoyl-CoA were organic synthesized from trans hexenoic acid, octenoic acid and decenoic acid respectively, and the substrates were confirmed by mass spectral. The kinetic substrate screen data was shown on table 4.5. PA5188 showed very high activity over S-specific 3-hydroxyhexanoyl-CoA, k_{cat}=4.5 s^{-1} and K_m= 12 μM and k_{cat}/K_m= 3.8 x 10^5 M^{-1} s^{-1}. The smaller substrate, 3-hydroxybutyryl-CoA, and longer substrate, 3-hydroxydecanoyl-CoA, both bind much less tightly to PA5188 than 3-

hydroxyhexanoyl-CoA with K_m value of 350 μM and 200 μM respectively. PA5188 showed decant activity over another medium long chain substrate 3-hydroxyoctanoyl-CoA, with 8 times greater K_m and 6 fold slower rate than 3-hydroxyhexanoyl-CoA. PA5188 seems to be very "S-specific" protein. It cannot recognize all the R-specific substrates, which testified that it is involved in the beta oxidation pathway.

4.5 Conclusion

Hot dog fold thioesterase PA5185 act as a "house keeper" protein to regulate the level of propionyl-CoA and release propionate and CoA. PA5185 is a very specific thioesterase, and it only good for very short chain fatty acyl-CoA, such as acetyl-CoA, propionyl-CoA and butyryl-CoA. Its tiny active site may make CoASH as a good inhibitor for the protein. This may explain the slow turn over number of propionyl-CoA. Together with the propionyl-CoA is ending metabolite of beta oxidation pathway, we hypothesize this thioesterase may act as a "house keeper", it can be expressed when CoASH is in low level, and be stopped when CoASH resumed or even act as a regulator for beta oxidation pathway and other propionyl-CoA related pathway. PA5187 catalyze the dehydrogation of short chain to median long chain acyl-CoA. It exhibit wild substrate specificities, which may works on both even number fatty acyl-CoAs and odd number fatty acyl-CoAs to degrade them to propionyl-CoA and acetyl-CoA as end product. The PA5188 converts exclusively S-specific 3-hydroxyacyl CoA instead of R-specific substrate to 3-ketoacyl CoA for further degradation. Based on these results, the three target protein, PA5185, PA5187 and PA5188, may belong to the beta oxidation pathway.

There are still a lot of works need to be down to investigate the function and mechanism of PA5185. Enzyme-ligand complex is needed to further analysis the substrate binding, the mechanism of reaction, and product release, this is essential to better understanding how PA5185 sense the level of CoASH, and how it regulate the other pathway. To verify these proteins biological functions, the putative iron containing alcohol dehydrogenase need to be investigated. Because, there is no room for PA5186 in beta oxidation pathway, it is very possible that this protein is related to the downstream pathway that start from propionyl-CoA.

4.6 Reference

1. Kunau, W. H., Dommes, V., Schulz, H. (1995) beta-oxidation of fatty acids in mitochondria, peroxisomes, and bacteria: a century of continued progress, *Prog. Lipid Res.,* 34, 267-342.

2. Halarnkar, P. P., Blomquist, J. G., (1989) Comparative aspects of propionate metabolism. *Comp. Biochem. Physiol.B.,* 92, 227-231.

3. Uchida, Y.; Izai, K.; Orii, T.; Hashimoto, T. (1992) Novel fatty acid beta-oxidation enzymes in rat liver mitochondria. II. Purification and properties of enoyl-coenzyme A (CoA) hydratase/3-hydroxyacyl-CoA dehydrogenase/3-ketoacyl-CoA thiolase trifunctional protein, *J. Biol. Chem.,* 267, 1034-1041.

4. Carpenter, K.; Pollitt, R. J.; Middleton, B. (1992) Human liver long-chain 3-hydroxyacyl-coenzyme A dehydrogenase is a multifunctional membrane-bound beta-oxidation enzyme of mitochondria, *Biochem. Biophys.Res. Commun.,* 183, 443-448.

5. Yang SY & He XY. (1999) Molecular mechanism of fatty acid b-oxidation enzyme catalysis, *Adv. Exp. Med. Biol*., 466, 133–143.

6. Schulz H. (2004) Fatty acid oxidation. In Encyclopedia of Biological Chemistry (Lennarz W & Lane MD, eds),pp. 90–94. Elsevier, San Diego, CA.

7. Thoden, J. B., Holden, H. M., Zhuang, Z., and Dunaway-Mariano, D. (2002) X-ray crystallographic analyses of inhibitor and substrate complexes of wild-type and mutant 4-hydroxybenzoyl-CoA thioesterase, *J Biol Chem 277*, 27468-27476.

8. Thoden, J. B., Zhang, Z., Dunaway-Mariano, D., Holden, H. M., (2003) The structure of 4-hydroxybenzoyl-CoA thioesterase from Arthrobacter sp. strain SU, J. Biol. Chem., 278, 43709-43716.

9. Mary C. Hunt, Stefan E.H. Alexson (2002) The role Acyl-CoA thioesterases play in mediating intracellular lipid metabolism, *Progress in Lipid Research*, 41, 99–130.

10. Hauge, J.G., Crane, F.L. and Beinert, H. (1956) On the mechanism of dehydrogenation of fatty acyl derivatives of coenzyme A. III. Palmityl CoA dehydrogenase, *J. Biol. Chem.,* 219 727-733.

11. Beinert, H. (1963) in The Enzymes, eds. Boyer, P. D., Lardy,H. & Myrback, K. (Academic, New York), Vol. 7, pp. 447-473.

12. Kim JJ, Wang M, Paschke R. (1993) Crystal structures of medium-chain acyl-CoA dehydrogenase from pig liver mitochondria with and without substrate, *Proc Natl Acad Sci U S A.*, 15, 90(16), 7523-7527.

13. Wu, W.-J.; Feng, Y.; He, X.; Hofstein, H. A.; Raleigh, D. P.; Tonge, P.J., (2000) Stereospecificity of the Reaction Catalyzed by Enoyl-CoA Hydratase, J. Am. Chem. Soc., 122, 3987-3994.

14. Reeta Davis, Arun Chandrashekar, Tumkur R. Shamala, (2008) Role of (R)-specific enoyl coenzyme A hydratases of Pseudomonas sp in the production of polyhydroxyalkanoates, *Antonie van Leeuwenhoek,* 93, 285–296

15. Steinman, H. M.; Hill, R. L., (1975) Bovine liver crotonase (enoyl coenzyme A hydratase). EC 4.2.1.17 L-3-hydroxyacyl-CoA hydrolyase. Methods Enzymol., 35, 136-151.

16. Yang SY, He XY, Schulz H. (2005) 3-Hydroxyacyl-CoA dehydrogenase and short chain 3-hydroxyacyl-CoA dehydrogenase in human health and disease, *FEBS J.,* 272(19), 4874-83.

17. Birktoft JJ, Holden HM, Hamlin R, Xuong NH, Banaszak LJ. (1987) Structure of L-

3-hydroxyacyl-coenzyme A dehydrogenase: preliminary chain tracing at 2.8-A resolution, *Proc Natl Acad Sci U S A.,* 84(23),8262-8266.

18. Chruszcz M, Zimmerman MD, Wang S, Koclega KD, Zheng H, Evdokimova E, Kudritska M, Cymborowski M, Savchenko A, Edwards A, Minor W. (2008) Function-Biased Choice of Additives for Optimization of Protein Crystallization: The Case of the Putative Thioesterase PA5185 from *Pseudomonas aeruginosa* PAO1, *Cryst Growth Des*., 8(11), 4054-4061.

19. Engel PC (1981) Butyryl-CoA dehydrogenase from Megasphera elsdenii. Methods Enzymol 71: 495–508.

20. Benning, M. M., Wesenberg, G., Liu, R., Taylor, K. L., Dunaway-Mariano, D., and Holden, H. M. (1998) The three-dimensional structure of 4-hydroxybenzoyl-CoA thioesterase from Pseudomonas sp. Strain CBS-3, J Biol Chem 273, 33, 572-579.

21. Zhuang, Z., Song, F., Zhang, W., Taylor, K., Archambault, A., Dunaway-Mariano, D., (2002) Kinetic, Raman, NMR, and site-directed mutagenesis studies of the Pseudomonas sp. strain BS3 4-hydroxybenzoyl-CoA thioesterase active site, Biochemistry, 41, 11152-11160.

22. Tsuge T, Taguchi K, Seiichi T, Doi Y.(2003) Molecular characterization and properties of (R)-specific enoyl-CoA hydratases from Pseudomonas aeruginosa: metabolic tools for synthesis of polyhydroxyalkanoates via fatty acid beta-oxidation, Int J Biol Macromol. 31(4-5), 195-205.