Chemistry ETDs   Electronic Theses and Dissertations

2-1-2012

Function assignment within the haloacid dehalogenase superfamily
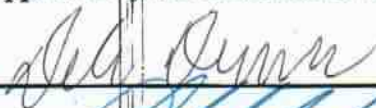
hua huang

Recommended Citation

Hua Huang
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:
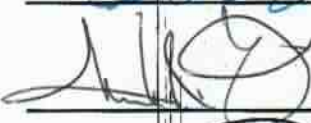
Debra Dunaway-Mariano, Chairperson
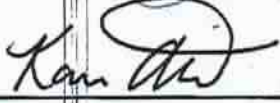
Patrick S Mariano

Martin L Kirk

Karen N Allen

FUNCTION ASSIGNMENT WITHIN THE HALOACID DEHALOGENASE SUPERFAMILY

BY

HUA HUANG

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

ACKNOWLEDGEMENTS

Foremost, I would like to express my greatest and sincerest gratitude to my research advisor- Professor Debra Dunaway-Mariano. It is she who opened the door for my exploration of biochemistry world, and during this journey, she always expressed her insight in knowledge and patience in teaching. With the help of her sustained encouragement and guidance, I successfully overcame the difficulties one by one. I believe the benefits I gained will live for my entire life.

The second person I appreciate much is Professor Patrick S. Mariano. Dr. Mariano is a nice and erudite professor. He granted me a bunch of useful advices in the inhibitor and substrate synthesis. I am also very grateful for having Professor Martin L. Kirk, Professor Karen N Allen in my doctoral committee and wish to thank them for their precious time in reading and evaluating my dissertation.

I also would like to extend my thanks for the collaborators-Dr. Bill Lu from Boston University and Dr. Rafael Toro from Albert Einstein Lab who helped us to solve the structures of the HAD enzymes.

The last but not the least is the current and ex-members in Professor Debra Dunaway-Mariano and Patrick Marino's research group. They are my colleagues and my friends. My appreciation covered Dr. Liangbing Wang, Dr. Zhimin Li, Dr. Jian Cao, Dr. Chunchun zhang, Dr. Jiwen zou, Dr. Hong Zhao, Dr. Danqi Chen, Mr. Min Wang, Ms. Li zheng, Mr. Jonh lathan, Ms. Sarah Toews …….. .Among them, I want to point out my

FUNCTION ASSIGNMENT WITHIN THE HALOACID DEHALOGENASE SUPERFAMILY

BY

HUA HUANG

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

FUNCTION ASSIGNMENT WITHIN THE HALOACID

DEHALOGENASE SUPERFAMILY

HUA HUANG

B.S., Chemistry, Lanzhou University, 2002

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

ABSTRACT

The HaloacidDehalogenase Enzyme Superfamily (HADSF) is a ubiquitous family of enzymes.Presently, more than 45, 000 deposited gene sequences encode proteins of the HADSF, and only a fraction of these have defined structure and/or function. The work described in this thesis focuses on function determination in several members of the HADSF. An integrated bioinformatic-protein structure-enzyme mechanism approach was used to differentiate and track D-glycero-D-manno-heptose-1,7-bisphosphate phosphatase (GHMB) and histidinol phosphatephosphatase (HisB)orthologues;2-keto-3-deoxynononic acid 9-phosphatephosphatase (KDN9PP) and 2-keto-3-deoxy-D-manno-octulosonic acid 8-phosphatephosphatase (KDO8PP)orthologues; inorganic pyrophosphatase and β-phosphoglucomutase (β-PGM) orthologues. In addition, a structure-function/bioinformatic analysis was carried-out on thebifunctional 1,3-diposphoglycerate acyltransferase/Cys-S-glyceryl-3-phosphate phosphatase (FKBH).

Each study began with the examination of the genome context of the encoding gene of the target HADSF member. Based on this analysis possible catalytic functions were posited. *In vitro* activity assays were then applied to test possible substrates. Having identified a potential physiological substrate the X-ray structure of the enzyme-substrate (or substrate analog) complex was determined. From this structure the substrate recognition residues were identified. These residues were replaced by site directed mutagenesis and the impact on substrate binding and catalysis was determined by measuring the steady-state kinetic constants k_{cat} and k_{cat}/K_m for each of the mutant enzymes. Residues shown to be important were used as sequence markers to identify among the sequence homologues identified in BLAST searches the most confidently defined orthologues. The final step used in the function annotation procedure was to examine the genome context of the encoding gene of each putative orthologue. These data were then used to formulate the proposal for *in vivo* function.

TABLE OF CONTENTS

x

LIST OF FIGURES

panel), HI-KDO8PP surface is shown with KDO for positioning reference (middle panel), and BT-KDN9PP surface shown with KDN (right panel).174

LIST OF TABLES

LIST OF ABBREVIATIONS

A	Adenosine/Alanine
AA	Amino Acid
ADP	Adenosine 5'-diphosphate
AMP	Adenosine 5'-monophosphate
ATP	Adenosine 5'-triphosphate
Arg/R	Arginine
Asn/N	Asparigine
Asp/D	Aspartic acid
ATCC	American Type Culture Collection
Bp	Base-Pair
Blast	Basic Local Alignment Search Tool
BSA	Bovine Serum Albumin
C	Cysteine
CoA	Co-enzyme A
cDNA	Complementary Deoxynucleic acid
CDP	Cytidine diphosphate
Da	Dalton
dATP	Deoxy adenosine 5'-triphosphate
DDT	Dichlorophenyltrichloroethene
DNA	Deoxyribonucleic acid

DI H2O	Distilled and deionized water
dGTP	Deoxy guanosine 5'-triphosphate
dTTP	Deoxy thymidine 5'-triphosphate
3-D	Three dimensional
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
E	Enzyme
EC	Enzyme Commission
E.coli	Escherichia Coli
EDTA	Disodium ethylenediamine tetraacetate
E-P	Phosphoenzyme
ESI-MS	Electrospray ionization Mass spectrometry
EXPASY	Expert protein analysis system
FAS	Fatty Acid synthase
F-6-P	Fructose-6-phosphate
FPLC	Fast Protein Liquid Chromatography
g	Gram
Gln/Q	Glutamine
Glu/E	Glutamte
Gly/G	Glycine
βG1P	β-D-Glucose-1-Phosphate
βG1,6bisP	β-D-Glucose 1,6-(bis)phosphate

GDH	a-Glycerophosphate dehydrogenase
GmhB	D-glycero-D-manno-heptose 1,7-bisphosphate phosphatase
GTP	Guanosine Triphosphate
GmhA	Phosphoheptose isomerase
h	Hour
HADSF	Haloalkanoic acid dehalogenase superfamily
H-bond	Hydrogen bond
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
HddA	D-glycero-D-manno-heptose 7-phosphate a-kinase
His/H	Histidine
HPLC	High Performance Liquid Chromatography
Ile/I	Isoleucine
IPTG	Isopropylthio-β-galactoside
K	Rate constant
Kb	Kilo-base pair
kcat	Enzyme turnover rate
Kd	Dissociation constant
Km	Michaelis-Menten constant
Ki	Inhibition Constant
kob	Observed rate constant
kDa	Kilo-Dalton
KDN	2-Keto-3-deoxy-D-glycero-D-galacto-nonomic acid

KDN9P	2-Keto-3-deoxy-D-glycero-D-galacto-nononate 9-phosphate
KDO	2-Keto-3-deoxy-D-manno-octulosonic acid
KDO8P	2-Keto-3-deoxy-D-manno-octulosonate 8-phosphate
L	Liter or Leucine
LB	Luria-Bertani
Lys/K	Lysine
LDH	Lactate dehydrogenase
LPS	Lipopolysaccharide
mg	Milligram
min	Minute
ml	Milliliter
mmole	Milimole
MS	Mass spectrum
MW	Molecular Weight
mRNA	Message Ribonucleic acid
Met/M	Methionine
ng	Nanogram
NADH	β-Nicotinamide adenine dinucleotide, reduced form
NCBI	National Center for Biotechnology Information
Neu5Ac	N-acetylneuraminate
Neu5Ac-9-P	N-acetylneuraminate 9-phosphate
OD	Optical density

OAH	Oxaloacetate Acetyhdrolase
ORF	Open reading frame
pfu	Plaque forming units
PDB	Protein Data Bank
PCR	Polymerase Chain Reaction
PAGE	Polyacrylamide gel electrophoresis
PEG	Phosphoenolpyruvate
Phe/F	Phenylalanine
βPGM	β-phosphoglucomutase
Pi	inorganic phosphate
PPi	Pyrophosphate
pI	Isoelectric point
PLP	Pyridoxalphosphate
PK	Pyruvate Kinase
PMM	Phosphomutase
PNPP	P-nitrophenyl phosphate
PPTase	Phosphopantetheinyl transferase
PPDK	Pyruvate phosphate Kinase
Pro/P	Proline
PS	Polysaccharide
RNA	Ribonucleic acid
RT	Room Temperature

rpm	Rotation per minute
s	Second
S	Substrate/Serine
SDS	Sodium dodecyl sulfate
S_N	Nucleophilic substitution
TCA	Tricarboxylic acid
Thr/T	Threonine
Tris	Tris[hydroxymethyl]aminomethane
Trp/W	Tryptophan
Tyr/Y	Tyrosine
U	Uridine
μM	Micromolar
UDP	Uridine diphosphate
UV	Ultraviolet
Val/V	Valine
Vmax	Maximum Velocity
V	Volume
WT	Wild type

HALOACID DEHALOGENASE ENZYME SUPERFAMILY FUNCTION

DISCOVERY

1.1 Introduction to the Haloacid Dehalogenase Enzyme Superfamily

The Haloacid Dehalogenase Enzyme Superfamily (HADSF) is a ubiquitous family of enzymes [1]. Presently, >45, 000 deposited gene sequences encode proteins of the HADSF, and only a fraction of these have defined structure and/or function. Numerous proteins from the HADSF are found in each organism (29 in *E. coli* and 185 in humans, for example.)

The HADSF was named after the first family member characterized: 2-haloalkanoate dehalogenase. Despite the name, the dehalogenases, which catalyze carbon group transfer, represent a minute fraction of the family [2-4]. The known reactions catalyzed by the HADSF are depicted in **Figure 1.1**. Most of these are phosphoryl transfer reactions of which the phosphohydrolase activities of the ATPases and the phosphatases are the most prevalent. Diversification of the HADSF catalytic scaffold to include phosphonate (P-C bond) hydrolysis (phosphonatases) and the transfer of phosphoryl groups between hexose hydroxyl substituents (phosphomutases) has occurred less frequently. phosphotransferases, having evolved to perform a multitude of different biochemical functions essential to cell growth and adaptation. Nature uses the phosphoryl group to enhance metabolite solubility, link monomers, facilitate protein-substrate

binding, couple catabolism to biosynthesis and movement, and as the on-off switch for signal transduction. Phosphotransferases therefore rank high among Nature's protein catalysts and a significant fraction of these phosphotransferases are found in the HADSF.
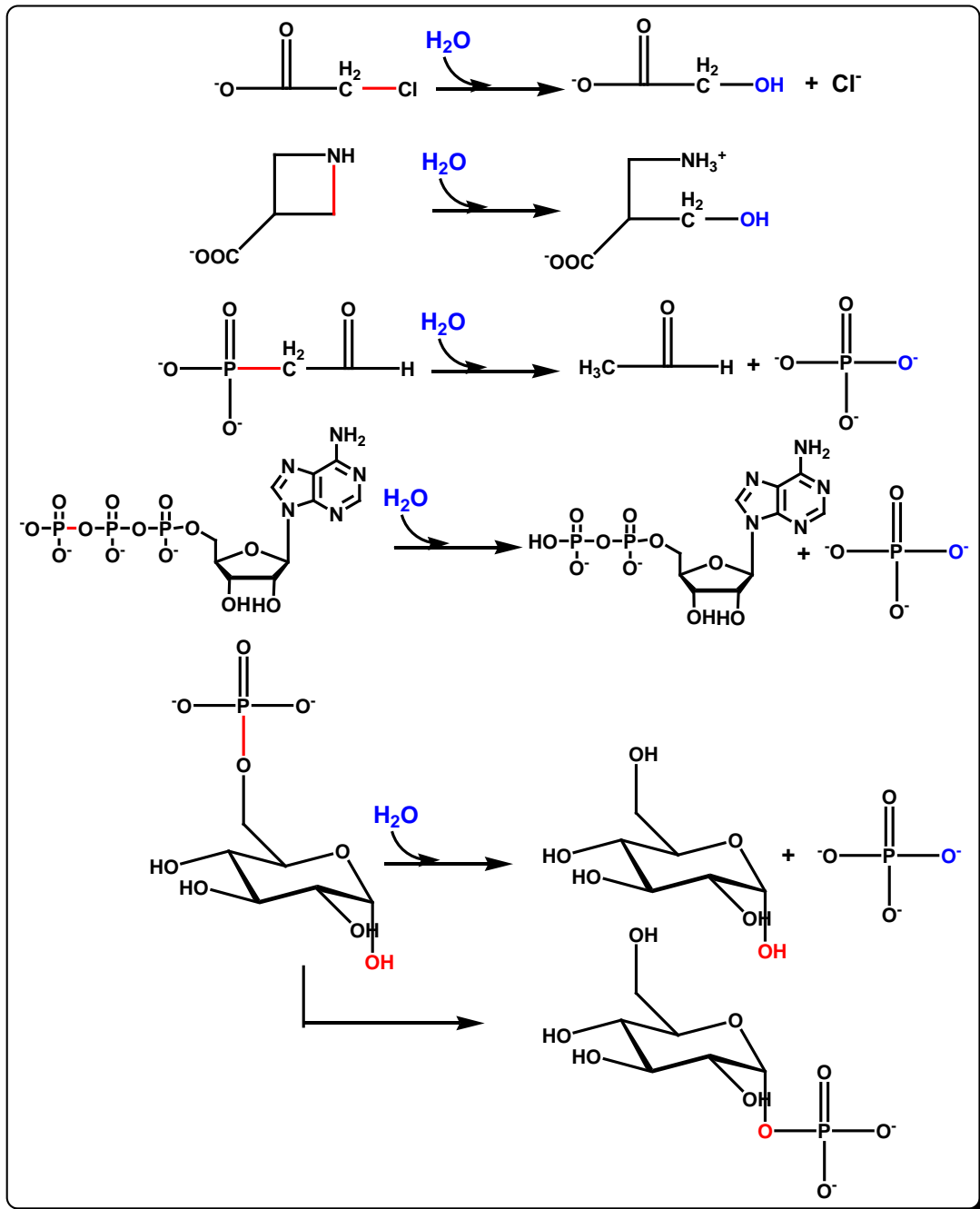
Figure 1.1: Known reactions catalyzed by the members of the HADSF.

All members of the HADSF possess an Alpha, Beta-core catalytic domain consisting of a central parallel Beta-sheet flanked by Alpha-helices on both sides [5,6]. The core catalytic domain of the HAD superfamily contains a three layered Alpha/Beta sandwich comprised of repeating Beta-Alpha units which adopt the topology typical of the Rossmannoid class of Alpha/Beta-folds. The central sheet is parallel and typically comprised of at least five strands (S1–S5) in a 54123 strand order (**Figure 1.2**). The HAD fold is distinguished from other Rossmannoid folds by two key structural motifs: First, immediately downstream of strand S1 is a unique, six residue structural motif that assumes a nearly complete single helical turn, termed the "squiggle". Second, downstream of the squiggle there is a Beta-hairpin turn formed by two strands projecting from the core of the domain. This structural motif constitutes the "flap". The squiggle, being close to a helical conformation, can be alternatively tightly or loosely wound, inducing a movement in the flap immediately juxtaposed to the active site and alternatively results in the closed and open states. Since the simplest structures (type C1) add the cap domain to the flap motif itself (**Figure 1.3A**), conformational changes in the squiggle allow the active site to be completely sealed in the closed state. Although the C2 caps have a different insertion site and thus likely lack mobility comparable to the C1 caps, together, the movement of the "squiggle-flap" and C2 cap allow dynamic ligand induced fit.

3
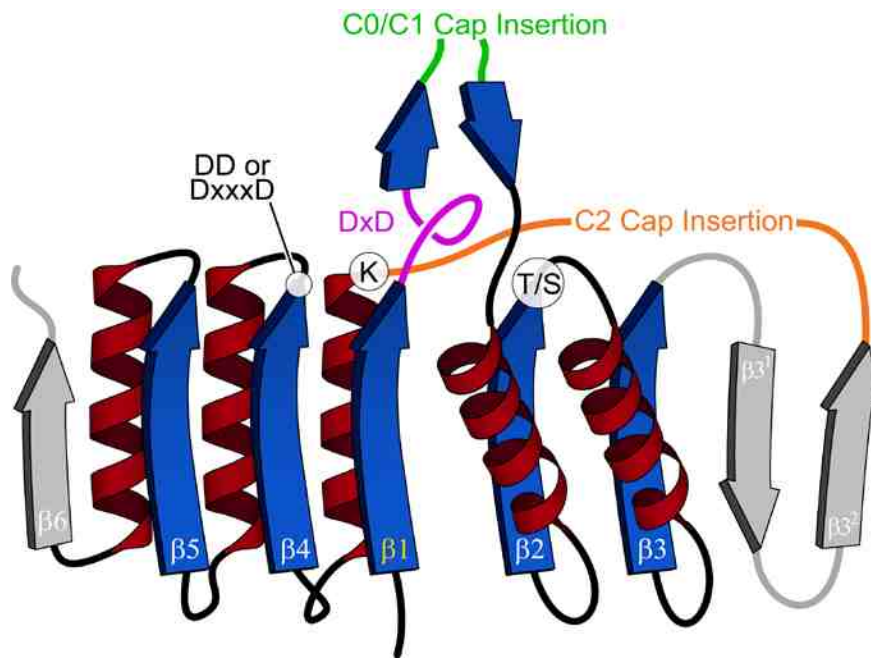
Figure 1.2: Topology diagram of the classic HADSF Rossmann core domain. Core strands conserved across all members are in blue; non-conserved elements that may have been absent from the ancestral state are in gray. The HAD C0/C1 cap insertion point is a bright green and the C2 cap insertion point is an orange line. The magenta loop represents

the conserved squiggle. Residues conserved across all members are labeled on their corresponding loops.

The concave surface of one face of the catalytic domain is the location of the active site (**Figure 1.3B**) where the substrate binds with its phosphoryl group pointed at the bottom and its leaving group projected upwards towards solvent. The core domain

4

catalyzes the transfer of the phosphoryl group from a specific phosphate ester or anhydride to an acceptor, which is most commonly a water molecule. The catalytic scaffold, or "active-site template" is formed by 4-loops that station the conserved catalytic residues" (**Figure 1.3B**). These include the two Asp residues of loop 1, the loop 2 Ser/Thr, the loop 3 Lys/Arg and the loop 4 Asp/Glu.
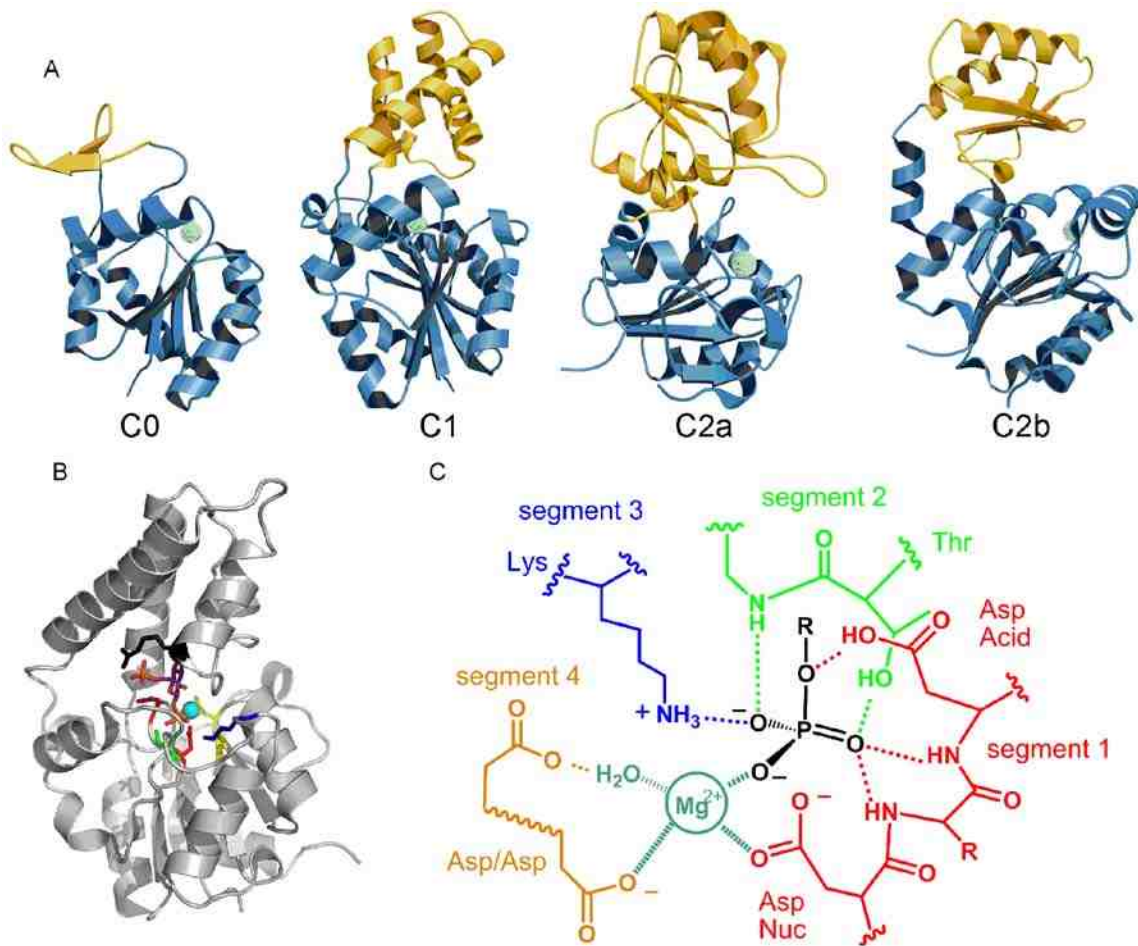
Figure 1.3: A) The HADSF subfamilies-C1 and C2 are distinguished by the insertion points of the cap (gold) in the Rossmann fold (blue) (C1 after b-strand 1, C2 after b-strand 1, C0 has a minimal insert in either site. B) liganded Beta-PGM with catalytic segments colored as in C) with cap residue in black.

The HADSF phosphotransferase reaction occurs in two steps (**Figure 1.4**). The first step involves phosphoryl transfer to one of the two loop 1 Asp residues (the Asp nucleophile) aided by proton transfer from the other loop 1 Asp residue (the Asp acid/base residue). The second step involves phosphoryl transfer from the aspartylphosphate to the acceptor (a water molecule in the case of the phosphohydrolases or hexose in the case of the phosphomutases) activated by the Asp acid/base residue.
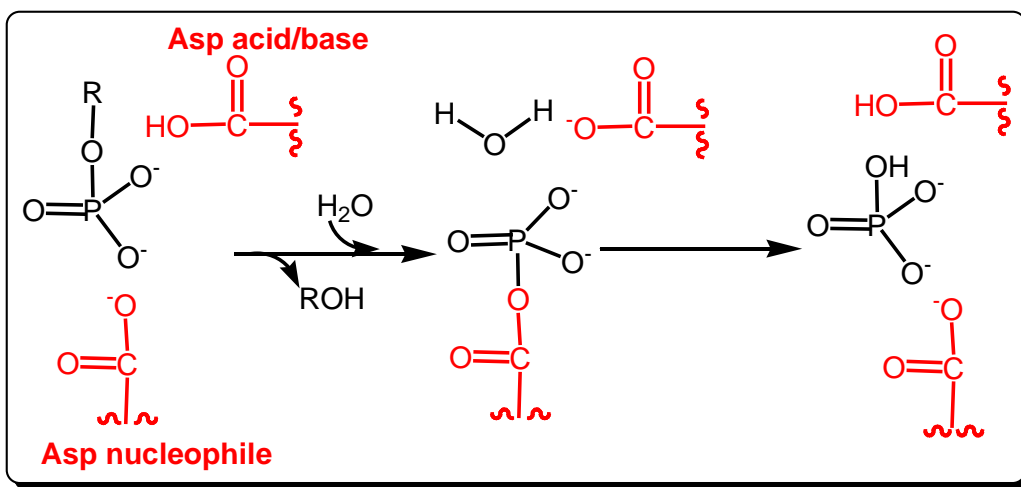
Figure 1.4: The two-step reaction pathways catalyzed.

The Mg^{2+} cofactor is coordinated by the carboxylate group of the loop 1 Asp nucleophile, the backbone amide C=O of the loop 1 Asp acid/base and by the carboxylate of an Asp (or Glu) located on loop 4 (**Figure 1.5**).

6
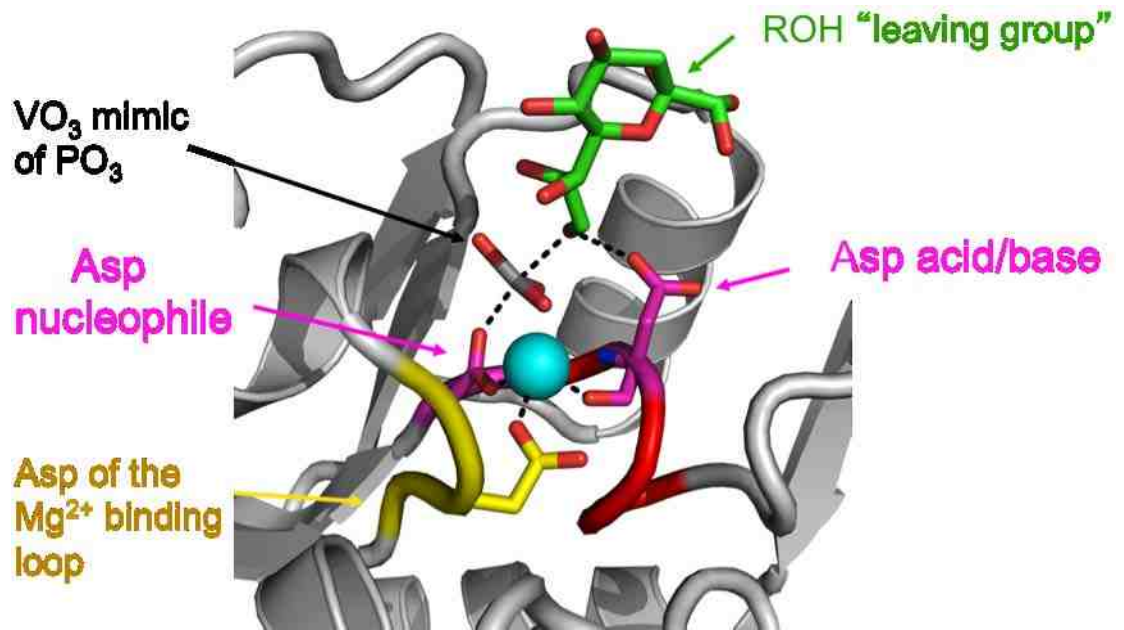
Figure 1.5: The Mg^{2+} (cyan sphere) binding site in a HADF phosphatase.

The Mg^{2+} functions to prevent charge repulsion between the Asp nucleophile and the transferring phosphoryl group and at the same time orient the Asp nucleophile, transferring phosphoryl group and the Asp acid/base for the reaction (**Figure 1.6**).
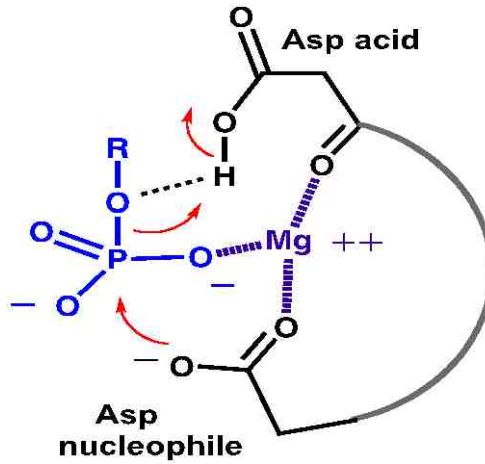
Figure 1.6: A depiction of the charge shielding and orientation functions of the HADF Mg^{2+} cofactor.

The transferring phosphoryl group is also activated for nucleophilic attack through hydrogen bond formation with the loop 2 Thr/Ser side chain, the loop 3 Lys and several backbone amide NHs of the catalytic scaffold (**Figure 1.7**).
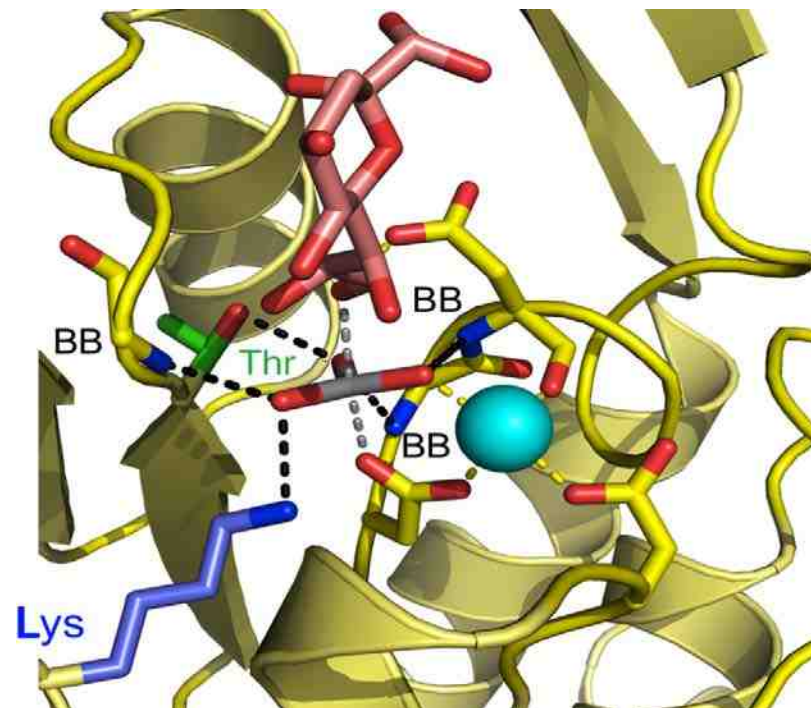
Figure 1.7: Activation of the transferring phosphoryl group (represented by the grey and red vanadate) by hydrogen bonds to the side chains Thr (green and red) and Lys (blue) and backbone amide NHs (BB). Hydrogen bonds are shown as black dashed lines.

The catalytic scaffold common to all HADSF phosphotransferases acts binds the tetrahedral phosphoryl group of the substrate weakly and through increased noncovalent bonding interactions stabilizes the trigonal bipyramidal phosphorane at the transition states. Crystallographic snap-shots of various liganded HADSF phosphotransferases indicate that the catalytic scaffold acts as an electrostatic mold for transition state stabilization (**Figure 1.8**).

9
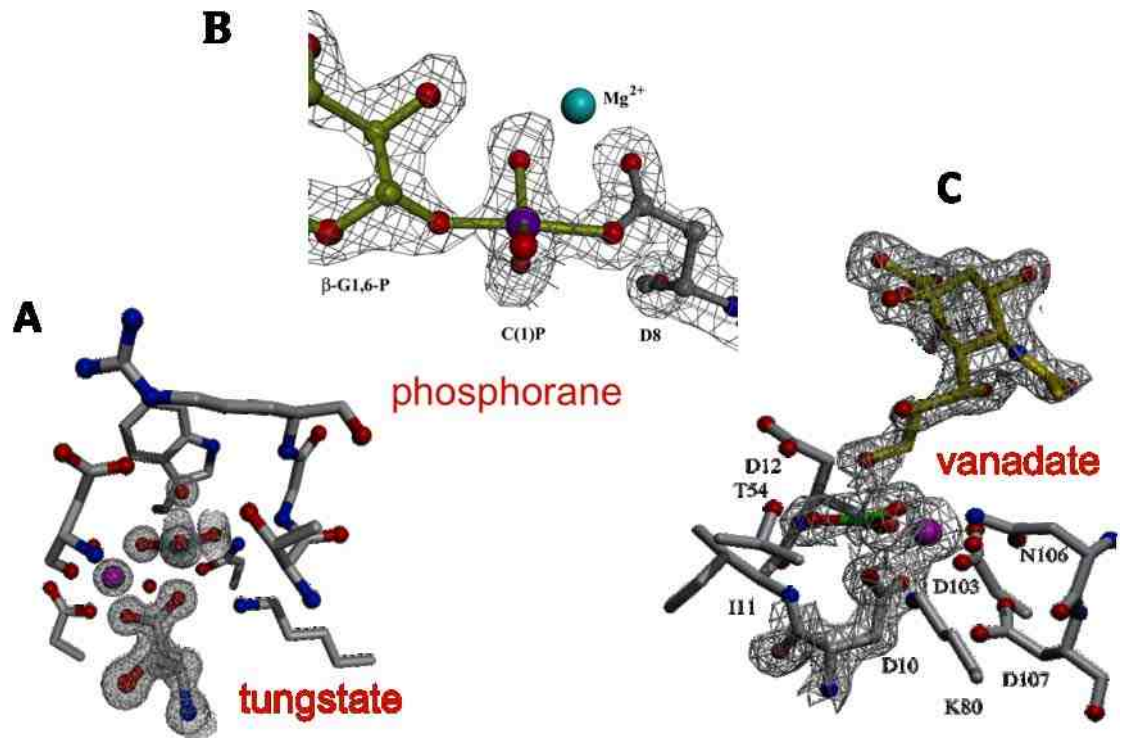
Figure 1.8: Crystallographic snapshots of (A) the hexose phosphate phosphatase BT4131 bound with Mg^{2+} and tungstate, (B) Beta-phosphoglucomutase bound with Mg^{2+} and Beta-glucose-1,6-bisphosphate and (C) KDN-9-phosphate phosphatase bound with Mg^{2+}, vanadate and neuramic acid.

1.2 The Structural Determinants of HADSF Phosphatases Substrate Recognition

The expansion of the HADSF can be attributed to phosphatases whose biochemical functions can be divided into four main categories: 1) metabolism, 2) regulation of metabolic pools, 3) cell housekeeping, and 4) nutrient uptake. Earlier work on the HADSF phosphatases showed that the physiological substrates run the full gamut in size and shape from phosphoproteins, nucleic acids, and phospholipids to phosphorylated disaccharides, sialic acids and terpenes to the smallest of the organophosphate metabolites, phosphoglycolate [7]. It is through the acquisition of structural accessories to the catalytic domain that the HADSF has succeeded in covering this vast range of substrate structure.

Whereas the catalytic scaffold binds the substrate phosphoryl group weakly at the ground state and tightly at the transition state it does not bind the leaving group (**Figure 1.9**). Therefore, the substrate must derive additional binding energy through favorable binding interactions with residues outside of the catalytic scaffold. This added binding energy augments the binding energy derived from binding the tetrahedral phosphoryl group such that under physiological conditions the HADSF phosphotransferase is able to target its physiological substrate.

11
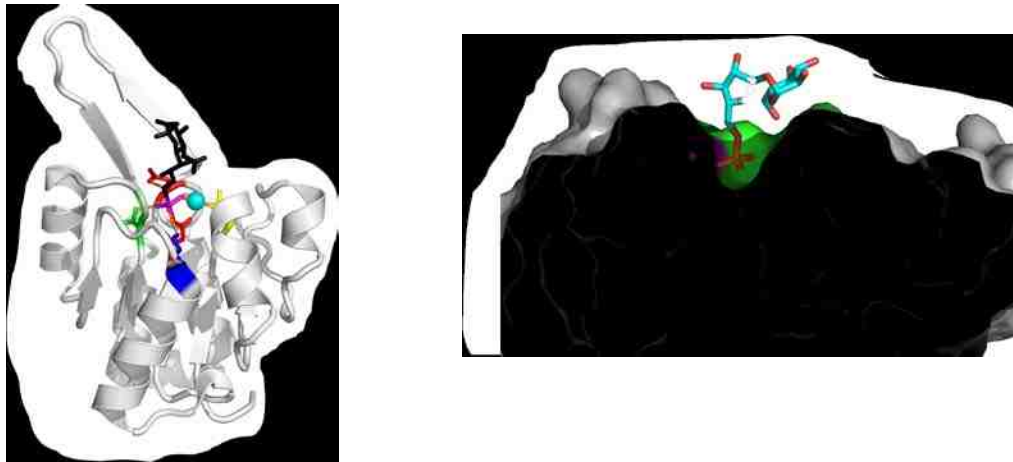
Figure 1.9: Snap-shots of C0 HADSF phosphatases illustrating the placement of the leaving group into the solvent.

The C0 phosphatases either form tetramers wherein the active site is formed at the subunit-subunit interface: one subunit contributes the catalytic residues and the other subunit contributes the leaving group binding residues (**Figure 1.10**).
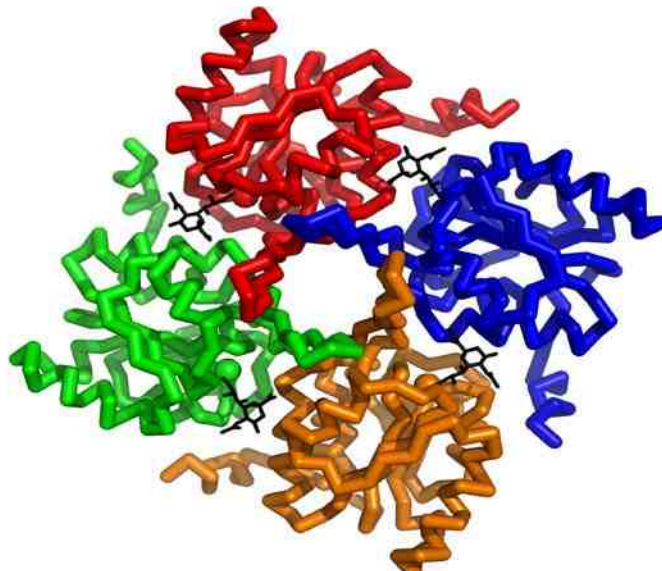
Figure 1.10: KDN-9-phosphate phosphatase bound with neuramic acid (black stick). Each subunit is shown in a different color.

Other C0 HADSF phosphotransferases do not rely on an interfaced subunit to bind the substrate leaving group. Those that act on large substrates such as a phosphorylated protein or a nucleic acid use the surface surrounding the catalytic site to dock with the substrate. Those that act on metabolites possess extended loops that bind the substrate-leaving group (**Figure 1.11**).
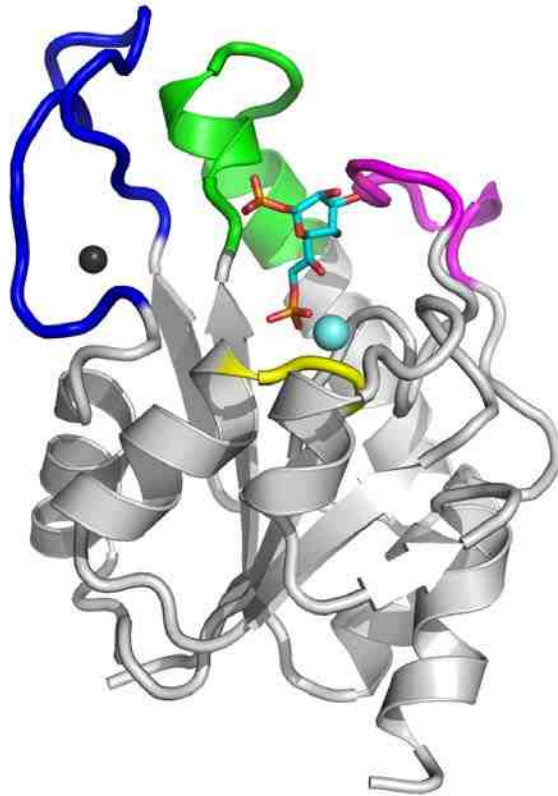
Figure 1.11: The C0 HADSF phosphatase GMHB bound with Mg^{2+} (cyan sphere) and

substrate (glycerol-manno-heptose 1,7-bisphosphate) (stick). The residues that bind the substrate leaving group are located on the extended loops colored blue, green or magenta.

The majority of the HADSF phosphatases, do however possess a cap domain. In the caped phosphatases the catalytic scaffold of the Rossmann-like domain alone binds the transferring phosphoryl group, whereas the cap domain alone binds the leaving group (**Figure 1.12**). Thus, HADSF phosphatase catalytic residues positioned by the catalytic scaffold are physically separated from the substrate recognition residues positioned on the cap domain.
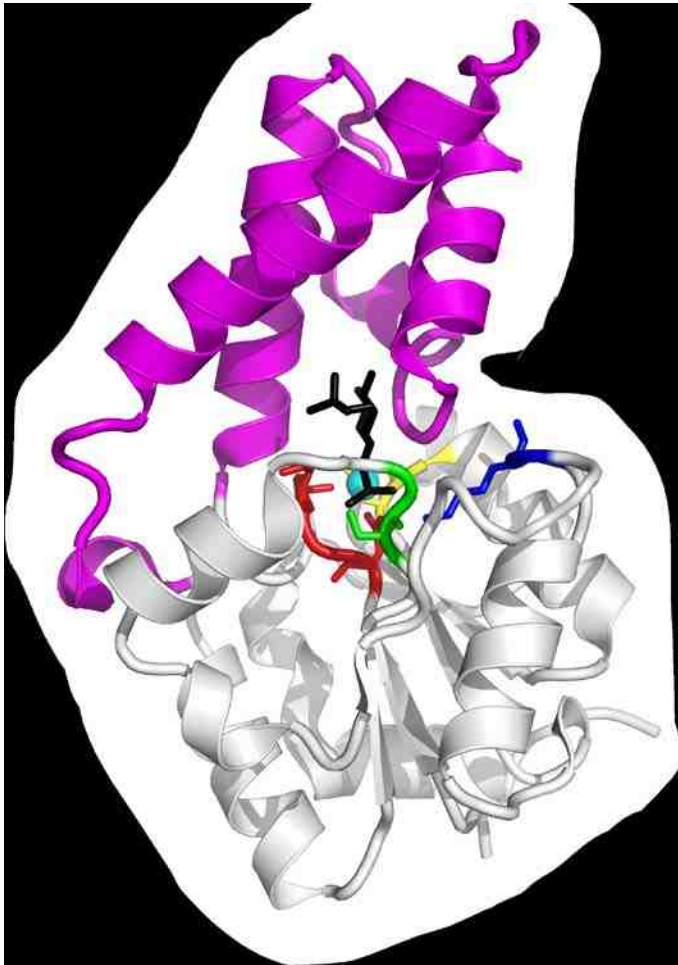
Figure 1.12: The C1 HADSF phosphotransferase b-phosphoglucomutatase bound with Mg^{2+} (cyan sphere) and Beta-glucose-1,6-bisphosphate (black stick). The cap domain is shown in magenta and the catalytic domain in gray.

The cap domain provides a large surface, which can accommodate numerous substrate recognition residues. Typically, however, only one or two cap residues participate in binding the physiological substrate. The cap domain serves as a "pegboard" (see **Figure 1.13**) for the placement/replacement of substrate recognition residues.
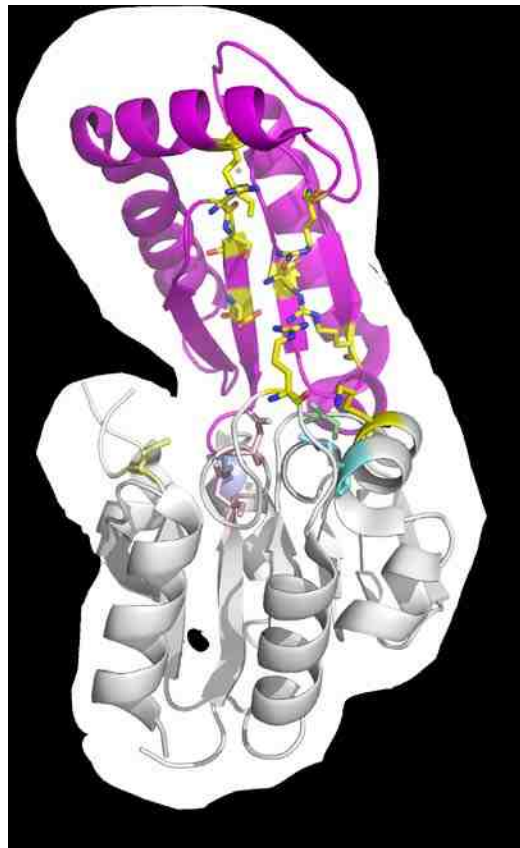
Figure 1.13: Structure of a-phosphomannose mutase showing the cap residues (yellow) that are potentially available to bind a substrate-leaving group.

Notably, the cap domain carries the potential to recognize more than one substrate, and this ability fosters promiscuity among the HADSF phosphatases. Promiscuity might directly support a general housekeeping role in the cell or alternatively, facilitate specialization through replacements that enhance the binding of the physiological substrate and suppress the binding of others. In addition, the cap residues can be tailored to specifically target two structurally distinct physiological substrates. Such tailoring would provide the HADSF phosphatase with two biochemical functions. For example, the housekeeper NagD [8] protects the cell from toxic glucoseamine 6-phosphate build-up during cell wall recycling, and siphons excess UMP from imbalanced ribonucleotide pools.

1.3 HADSF Phosphatase Function Determination

A challenge of this decade to the life sciences community is the accurate function annotation to the rapidly growing numbers of enzyme sequences and structures that are accumulating in the public databases. The classification of enzymes into enzyme families constitutes the first step in function assignment because members of a given family share a common chemical trait the knowledge of which facilitates function determination. An enzyme family grows by evolving new members of unique function via gene duplication and random mutation against a selection pressure. The inherited family chemical trait (exemplified in the phosphotransferase branch of the HADSF as the ability to stabilize trigonal bipyramidal phosphorane-like transition states) is the result of an inherited

16

backbone fold, which in turn supports the catalytic scaffold. Located on the catalytic scaffold are the residues that are inherited in the duplicated family member and which can be conserved and replaced as needed to adapt the inherited chemical trait to a novel biochemical context. It is through developing an understanding of the structural determinants of enzyme function that genomic enzymologists hope to expedite the process of function annotation.

Function assignment to a HADSF phosphatase entails identification of its physiological substrate, and in order for this function assignment to facilitate future assignments, the structural basis for substrate recognition must be determined. Moreover, knowledge of the key substrate recognition residues also allows the tracking orthologs found in the gene data banks thus expanding function well beyond that of the single enzyme that was subjected to the structure-function analysis. This is exemplified by my work on the HADSF phosphatase BT2127 reported in Chapter 2.

Organophosphates comprise a big piece (*viz.* ~35%) of the metabolome. Because of their predominance the HADSF phosphatases play a major role in regulating and recycling the organophosphate metabolite pool as well functioning in metabolic pathways. By discovering new functions among the HADSF phosphatases novel biochemical pathways can be revealed. This is exemplified by my work on the HADSF phosphatase FKBH reported in Chapter 3.

In addition to my investigation of BT2127 and FKBH, I have participated in the structure-function analysis of HADSF phosphatases GMHB, KDO-8-phosphate phosphatase and KDN-9-phosphate phosphatase. This work is reported in the Appendix

of this thesis.

1.4 Reference:

1)   Collet, J.F., Van Schaftingen, E., and Stroobant, V. (1998) A new family of phophotransferases related to P-type ATPase. *Trends Biochem Sci 23,284*

2)   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, 125-32*.

3)   Lahiri, S. D., Zhang, G., Dai, J., Dunaway-Mariano, D., and Allen, K. N. (2004) Analysis of the substrate specificity loop of the HAD superfamily cap domain. *Biochemistry 43*, 2812-20.

4)   by A S Baker, M J Ciocci, W W Metcalf, J Kim, P C Babbitt, B L Wanner, B M Martin, D Dunaway-Mariano (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*

5)   Ridder, I. S., Rozeboom, H. J., Kalk, K. H., Janssen, D. B., and Dijkstra, B. W. (1997) Three-dimensional structure of L-2-haloacid dehalogenase from Xanthobacter autotrophicus GJ10 complexed with the substrate-analogue formate. *J. Biol Chem 272*, 33015-22.

6)   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

18

cleavage and catalytic diversification within the HAD enzyme superfamily. *Biochemistry 39*, 10385-96

7) Jeremy D. Selengut(2001) MDP-1 Is a New and Distinct Member of the Haloacid Dehalogenase Family of Aspartate-Dependent Phosphohydrolases biochemistry (40),12704-12711

8) Lee W. Tremblay, Debra Dunaway-Mariano, and Karen N. Allen (2006) Structure and Activity Analyses of *Escherichia coli* K-12 NagD Provide Insight into the Evolution of Biochemical Function in the Haloalkanoic Acid Dehalogenase Superfamily *Biochemistry 45, 1183-1193*

CHAPTER TWO

DIVERGENCE OF STRUCTURE AND FUNCTION IN THE HALOACID

DEHALOGENASE (HAD) ENZYME SUPERFAMILY: *BACTERIODES*

THETAIOTAMICRON BT2127 IS AN INORGANIC PYROPHOSPHATASE

2.1 Introduction

Function annotation to the rapidly growing numbers of enzyme sequences and structures that are accumulating in the public databases is a challenge that should be met quickly so that the scientific community can effectively mine these databases. Ultimately, enzyme function annotation must be done by computer because experimental-based function determinations are far too time consuming. The Enzyme Function Initiative (http://enzymefunction.org/) was instituted in 2010 for the ultimate purpose of evolving current experimental strategies into computational-based function prediction. The work reported in this chapter is an early product of this initiative as it outlines enzyme function determination in a member of the Haloalkonate Dehalogenase enzyme superfamily (HADSF) that is based on an integrated bioinformatic-protein structure-enzyme mechanism approach which provides reliable tracking of orthologs.

The HADSF was selected to explore methods of function assignment because of its large size, its presence in all organisms represented in the NCBI genome bank and because of the diversity in the types of chemical reactions that its members catalyze. The initial focus is on the HADSF phosphatase subfamily. HADSF phosphatases can number

up to 30 in bacterial cells and up 200 - 300 in eukaryotic cells. A central question to be addressed is "how many unique *in vivo* functions are there to be discovered within this subfamily?". Bacterial (as opposed to eukaryotic) HADSF phosphatases were chosen for study, so to avoid splice variants and the complicating issues of protein partners and subcellular localizations as modulators of *in vivo* function. Function assignment to the bacterial HADSF phosphatase thus entails identification of its physiological substrate among the pool of organophosphate metabolites that exist in the cell, followed by defining the biological context of the catalyzed reaction.

Earlier work that involved the screening of 23 cytoplasmic *E. coli* HADSF phosphatases with an 80-compound library of known metabolites made clear the challenge of identifying the physiological substrate [1]. Specifically, most of the phosphatases displayed activities towards multiple substrates, which in numerous instances were structurally dissimilar. In addition, there also existed some degree of overlap in the substrate ranges between paired phosphatases. Nevertheless, the physiological substrates of several of the 23 *E. coli* HADSF phosphatases have been convincingly identified and their *in vivo* functions as phosphoserine phosphatase (SerB) [1,2],histidinol phosphate phosphatase (HisB) [3], trehalose-6-phosphate phosphatase (OtsB) [4], 3-deoxy-D-manno-octulosonate-8-phosphate phosphatase (YrbI) [5] , glycerol-manno-heptose 1,7-bisphosphate phosphatase (YaeD) [6] assigned. Notably, each of these phosphatases function in biosynthetic pathways. The *in vivo* function assignment was facilitated by gene context as well as by a narrow substrate range coupled and a high k_{cat}/K_m value for the physiological substrate. In addition, gene

knockout experiments have been used to show that the promiscuous phosphatases YniC [1] and YjjG [7] can function *in vivo* to remove toxic 2-deoxy-glucose-6-phosphate and toxic nonconical nucleoside-5'-monophosphates. Genetic experiments have also shown that YjjG performs in the thymidine salvage pathway as a dUMP nucleotidase [8].

With the objective in mind of gaining insight into how many of the HADSF phosphatases contained in two different species of bacteria share common function and how many do not, a side-by-side comparison of the HADSF phosphatases from *E. coli* and *Bacteroides* *thetaiotaomicron* was initiated. Whereas these two species derive from different phyla they share a common habitat: the human gut. *B.* *thetaiotaomicron* possesses 19 HADSF phosphatases. Two of these have been experimentally characterized. One is 3-deoxy-D-manno-octulosonate-8-phosphate phosphatase (BT1677) [9], ortholog to the *E. coli* YrbI of the 3-deoxy-D-manno-octulosonate pathway and the other one is 2-keto-3-deoxy-D-glycero-D-galactonate-9-phosphate phosphatase (BT1713) of the 2-keto-3-deoxy-D-glycero-D-galactonate pathway [10], which notably is not found in *E. coli*.

In collaboration with structural biologists in the Steve Almo lab at Einstein Medical School and the Karen Allen's lab at Boston University, I have carried out the structure-function analysis of the unknown HADSF member BT2127 (ExPASy accession # Q8A5V9) from *B.* *thetaiotaomicron*. In this chapter I report the BT2127 function, X-ray structure and biological range and contrast these with those of its closest homologue Beta-phosphoglucomutase (Beta-PGM) and with those of a HADSF Archeal inorganic pyrophosphatase.

2.2 Materials and Methods

2.2.1 General Method

All chemicals and buffers were purchased from Sigma-Aldrich. The sources of the gene cloning materials are as follows: primers, T4 DNA ligase, restriction enzymes (Invitrogen); *E. coli* BL21 (DE3) competent cells and *Pfu*, Turbo polymerases (Strategene); pET14b, pET23a and pET28a vector kits (Novagen); Qiaprep Spin Miniprep Kit (Qiagen). DEAE Sepharose was from Amersham Biosciences. Butyl and Phenyl-Sepharose resins were purchased from Sigma-Aldrich, whereas the Ni-NTA resin was from Qiangen. Snakeskin pleated dialysis tubing was purchased from Thermo Scientific. SDS-PAGE analysis was performed with a 12% acrylamide running gel and a 4% stacking gel (37.5:1 acrylmide:biacryamide ratio) (BioRad, Hercules, CA). Protein solutions were concentrated using a 10K Amicon Ultra Centrifugal filter (Millipore). The nucleotide sequence of each cloned gene or mutant gene was determined by the Center for Genetics in Medicine, University of New Mexico. Electro-spray mass-spectrometry (ES-MS) determinations were carried out by the University of the New Mexico Mass Spectrometry Facility. Protein concentrations were determined using the Bradford assay kit from Sigma-Aldrich.

2.2.2 Preparation of Recombinant Wild-type and Mutant BT2127

The DNA encoding the gene NP_811040 from *B. thetaiotaomicron* was amplified by PCR using the genomic DNA *B. thetaiotaomicron* (ATCC 29148D), *Pfu* Turbo DNA

polymerase and oligonucleotide primers (5'-GATTCCATCTAACCCA<u>CATATG</u> AGAA

AGAAAC) and (5'-CTTTTGCATAGTA<u>GGATCC</u>GTATTT ATAGGT) containing

restriction endonuclease cleavage sites *Nde*I and *Bam*HI. The pET-28A vector, cut by

restriction enzymes *Nde*I and *Bam*HI, was ligated to the PCR product that had been

purified and then digested with the same restriction enzymes. The ligation product was

used to transform *E. coli* BL21 (DE3) competent cells which were then grown on a

Kanamycin-containing agar plate. The selected colony was checked for BT2127

production and the isolated plasmid was sequenced to verify the correct gene sequence.

For BT2127 preparation, the transformed cells (9 L) were grown at 25 °C with agitation

at 200 rpm in Luria broth containing 40 μg/mL Kanamycin to an OD_{600} of 0.6-0.7, and

then induced for 12 h at 20 °C with 0.4 mM isopropyl Alpha-D-thiogalactopyranoside.

The cells were harvested by centrifugation (6500 rpm for 15 min at 4 °C) to yield 2.2 g/L

of culture medium. The cell pellet was suspended (1 g wet cells/10 mL) in ice-cold buffer

A (50 mM Tris (pH 7.6), 5 mM $MgCl_2$ and 1 mM DTT). The cell suspension was passed

through a French press at 1,200 PSIG before centrifugation at 20,000 rpm and 4°C for 45

min. The supernatant was loaded onto a 40 × 5 cm DEAE-Sepharose 50-120 column,

which was eluted with a 2 L linear gradient of NaCl (from 0 to 0.5 M) in buffer A. The

column fractions were analyzed by SDS-PAGE. The desired fractions were combined

and loaded onto 10 mL Ni-NTA Agarose column at 4 °C. After washing the column with

100 mL of buffer B (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole (pH 8.0)), the

enzyme was eluted with 200 mL elution buffer C (50 mM NaH_2PO_4, 300 mM NaCl, 250

mM imidazole (pH 8.0)). The column fractions were analyzed by SDS-PAGE, and the

desired fractions were combined and concentrated with an Amicon Ultrafiltration apparatus (PM10) before dialysis at 4 ^{o}C against buffer A. Yield: 11 mg BT2127/g wet cells.

The site directed mutagenesis was carried out using a PCR-based strategy, the BT2127-pET-28A clone as template and commercial primers. The mutant proteins were purified in the same manner as described for the wild-type BT2127.

2.2.3 BT2127 Molecular Weight Determination

The theoretical subunit molecular mass of recombinant BT2127 was calculated by using the amino acid composition, derived from the gene sequence, and the ExPASy Molecular Biology Server program Compute pI/MW. The subunit size of recombinant BT2127 was determined by SDS-PAGE analysis, which included the molecular weight standards from New England Biolabs Inc. The subunit mass was determined by MS-ES mass spectrometry. The molecular weight of native BT2127 was estimated by FPLC gel filtration column chromatography against protein standards (13.7-220 kDa from GE Healthcare). The 1.6 cm x 60 cm Sephacryl S-200HR column (GE Healthcare) was eluted at 4 ^{o}C with buffer D (50 mM HEPES, 100 mM NaCl (pH 7.5)) at a flow rate of 1 mL/min. The BT2127 molecular weight was derived from the measured elution volume by extrapolation of the plot of the elution volume of the molecular weight standard *versus* log molecular weight. BT2127 native mass was also analyzed at the HHMI Biopolymer/Keck Foundation Biotechnology Resource Laboratory at Yale University by size exclusion chromatography coupled with on-line laser scattering, refractive index,

and ultraviolet detection.

2.2.4 Kinetic Assay for Beta-Phosphoglucomutase Activity

Reaction solutions initially contained 7 M wild-type BT2127, 200 µM Beta-glucose-1-phosphate, 10 M Beta-glucose-1,6-bisphosphate, 2 mM MgCl$_2$, 0.4 mM NADP and 5.4 units/ml glucose-6-phosphate dehydrogenase in 50 mM K$^+$HEPES (pH 7.5). The formation of NADPH from reduction of glucose-6-phosphate was monitored at 340 nm ($\varepsilon = 6.2$ mM^{-1}cm^{-1}).

2.2.5 Steady-State Kinetic Constant Determinations

Initial velocities for BT2127 catalyzed hydrolysis of phosphate esters and anhydrides were measured at 25 ^{o}C using assay solutions that contained 1 mM MgCl$_2$, 1.0 unit/ml purine nucleoside phosphorylase, and 0.2 mM MESG in 50 mM Tris (pH 7.5). The reactions were monitored at 360 nm ($\varepsilon = 9.8$ mM^{-1} cm^{-1}). The steady-state kinetic parameters (K_m and k_{cat}) were determined by fitting the initial velocity data measured at varying substrate concentrations (ranging from $0.5K_m$ to $5K_m$) to **equation 1** using KinetAsyst I.

$$V_0 = V_{max} [S]/ ([S] + K_m) \qquad (\textbf{Equation.1})$$

V_0 is the initial velocity, V_{max} the maximum velocity, [S] the substrate concentration, and K_m the Michaelis constant for the substrate. The k_{cat} value was calculated from V_{max} and [E] according to the equation $k_{cat}=V_{max}/[E]$, where [E] is the enzyme concentration.

The steady-state competitive inhibition constant K_i was determined for

26

imidodiphosphate by fitting the initial velocity data, measured as a function of pyrophosphate (3, 5, 8, 10, 15, 20µM) and imidodiphosphate (0, 10, 35, 50µM) concentration to **Equation 2** using KinetAsyst I.

$V_0 = V_{max} [S] / (K_m(1+(1/K_i))+[S])$ (**Equation.2**)

Where [I] is the inhibitor concentration and K_i is the inhibition constant.

2.2.6 pH Rate Profile Determination

The steady-state kinetic constants k_{cat} and K_m for BT2127-catalyzed hydrolysis of pyrophosphate were measured at 25 ^{o}C as a function of reaction solution pH. Reaction solutions initially contained pyrophosphate ($0.5K_m$ to $10K_m$), 1 mM $MgCl_2$, 1.0 unit/ml purine nucleoside phosphorylase, 0.2 mM MESG in 50 mM buffer (MES, pH 5.0-6.0; HEPES, pH 6.5-7.5; Tris, pH 8.0-8.5). The k_{cat}/K_m data were fitted with **Equation 2** and the k_{cat} data were fitted with **Equation 3**.

$LogY = Log[C/(1 + [H]/K_b)]$ (2)

$LogY = Log[C/(1 + [H]/K_a + K_b/[H])]$ (3)

where Y is the k_{cat} or k_{cat}/K_m, [H] is the hydrogen ion concentration of the reaction solution, K_a and K_b are the apparent ionization constants, and C is the constant value of Y.

2.2.7 BT2127 Crystallization and Structure Determination

This work was carried out by X-ray crystallographers Patskovsky Yury and Rafael Toro. I prepared the wild-type and mutant proteins. The proteins were crystallized

by the sitting drop vapor diffusion method. In brief, the protein solution (usually 0.3 or 1 mL) was mixed with an equal volume of a precipitant solution and equilibrated at room temperature (294 oK) against the same precipitant solution in the clear tape-sealed 96-well *INTELLI-plates* (*Art Robbins Instruments, Sunnyvale, CA, USA).* Crystallization was performed using either a TECAN crystallization robot (TECAN US, Research Triangle Park, NC, USA) or a PHOENIX crystallization robot (Art Robbins Instruments) and three types of commercial crystallization screens: the WIZARD screen (Emerald BioSystems, Bainbridge Island, WA, USA), the INDEX and the CRYSTAL SCREEN I and II (both from Hampton Research, Aliso Viejo, CA, USA). A number of crystallization conditions produced diffraction-quality protein crystals starting within 24-72 h of incubation. The crystals were collected using Hampton Research cryogenic loops, quickly transferred in liquid nitrogen, and stored frozen in liquid nitrogen until X-ray analysis and/or data collection. Only in one case (PDB code 3QU2) was the cryoprotectant glycerol employed. Where necessary, the crystallization conditions were optimized manually using 24-well Cryschem sitting drop plates (Hampton Research). The crystallization conditions for each crystal structure are listed in the Table **SI1** (Supporting Information)

The X-ray diffraction data from majority of frozen crystals were collected at 100 oK on the Beamline *X29A* (NSLS, Brookhaven National Laboratory) using the wavelength of the radiation of either 0.93 or 1.08 Angstroms. The X-ray diffraction data from the crystal of the D13A mutant were collected at 100 oK using an RU-200 rotating-anode X-ray generator (λ=1.5418Å) coupled to a Rigaku R-AXIS IV area detector. All

diffraction data were processed and scaled using the HKL2000 software package (Otwinowski and Minor, 1997). The first out of 13 crystal structures (PDB 3QU2) reported here was solved by molecular replacement using coordinates for the PDB structure 3DV9 as a search model and the program MOLREP (the CCP4 program package suit, *www.ccp4.ac.uk*). The other 12 structures were solved using coordinates for the structure 3QU2 as a search model. Each structure was refined using REFMAC 5.03 (CCP4 suit, Murshudov *et al*, 1997) and the resulting model was fixed manually using COOT visualization and refinement software (Emsley *et al*, 2010). The data collection and refinement statistics for these structures are shown in the Tables SI2-SI5 of Supporting Information.

2.2.8 Bioinformatic Analysis

Putative orthologues to BT2127 and the pyrophosphatase from *Themococcus onnurineus* TON0002 were identified by carrying out BLAST searches of the NCBI microbial genome bank (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi) and selecting those sequences in which the key marker residues are conserved. The multiple sequence alignments were made in Cobalt (http://www.ncbi.nlm.nih.gov/tools/cobalt/) and displayed in ESPript (http://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi).

In order to compare biological ranges of BT2127 and Beta-PGM a more automated approach was employed. BLAST searches were performed for BT2127, Beta-PGM, maltose phosphorylase, and trehalose phosphorylase using a locally installed copy of BLAST against the NCBI non-redundant protein database (ftp://ftp.ncbi.nih.

gov/blast/db/) with default parameters. The resulting hits were filtered for query coverage >90% and >30% sequence identity. A reciprocal BLAST search, using an e-value threshold of 1e-10, was performed for each filtered hit. A hit was considered successful if the reciprocal BLAST search identified the initial query sequence. For each successful hit, taxonomic information was downloaded from NCBI taxonomy website (http://www.ncbi. Nlm.nih.gov/Taxonomy/taxonomyhome.html/). For the selection of β-PGM orthologues, only species with a maltose or trehalose phosphorylase gene were considered. A phylogenetic tree for the orthologues was generated using iTOL - interactive Tree Of Life (http://itol.embl.de/) and visualized using FigTree (http://tree.bio.ed.ac.uk/software/fig tree/).

2.3 Results and Discussion

2.3.1 BT2127 Substrate Specificity Profile

Because the closest characterized structural homolog of BT2127 is Beta-phosphoglucomutase (Beta-PGM), the ability of purified recombinant BT2127 (see SDS-PAGE gel in **Figure 2.1**) to catalyze the conversion of Beta-glucose-1-phosphate to glucose-6-phosphate in the presence of the cofactor Beta-glucose-1, 6-bisphosphate was tested. No activity was observed above the detection limit of one catalytic turnover per hour.
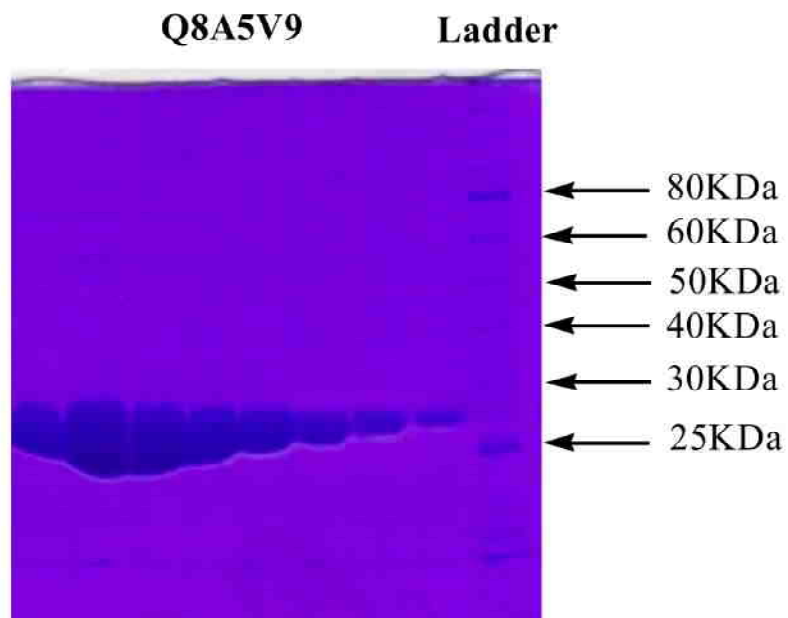
Figure 2.14: Commassie blue stained SDS-PAGE gel of purified BT2127 (Q8A5V9).

Next, the BT2127 phosphatase activity was tested using a structurally diverse chemical library of 21 organophosphate metabolites as the substrate screen (**Table 2.1**). The most active substrates identified by this screen were further analyzed to determine k_{cat} and k_{cat}/K_m values (**Table 2.2**) at the pH optimum of 7.5 (pH rate profile is shown in **Figure 2.2**).

Table 2.1: The apparent first order rate constants for BT2127 catalyzed hydrolysis of phosphate esters and anhydrides at pH 7.5 and 25 $^{\circ}$C. Reaction solutions initially contained 0.3 mM substrate, 8.4 μM BT2127, 1 mM MgCl$_2$, 1.0 unit/mL purine nucleoside phosphorylase and 0.2 mM MESG in 50 mM Tris (pH 7.5). The k_{obs} value was calculated by dividing the initial velocity of the reaction by the enzyme concentration.

Substrate	k_{obs} (min^{-1})	Substrate	k_{obs} (min^{-1})
Pyrophosphate	19	□□-fucose-1-P	0.10
Imidodiphosphate	0.06	GMP	0.38
Glyceraldehyde 3-P	0.17	UDP	0.17
Glycerate-3-P	0.31	ATP	0.20
D-erythrose-4-P	0.20	dATP	0.18
Dihydroxyacetone phosphate	0.30	Farnesylpyrophosphate	0.059
Glycerol-2-P	0.49	PLP	NA[a]
Glucose-1-P	0.12	CTP	NA[a]
Glucose-6-P	0.19	Serine-3-P	NA[a]
Fructose-6-P	0.32	Threonine P	NA[a]
□ß-fucose-6-P	0.21		

[a] NA represents no detectable activity.

Table 2.2: Steady-state kinetic constants for BT2127-catalyzed hydrolysis of pyrophosphate and selected phosphate monoesters at 25 $^{\circ}$C and pH 7.5. See Materials and Methods for Details.

Substrate	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (M^{-1}s^{-1})
pyrophosphate	0.32 ± 0.01	0.0036 ± 0.0004	9×10^4
glycerol-1-phosphate	0.35 ± 0.02	4.8 ± 0.3	7×10^1
D-ribose-5-phosphate	0.20 ± 0.02	3.8 ± 0.3	6×10^1
fructose-6-phosphate	0.066 ± 0.002	6.4 ± 0.2	1×10^1
uridine-5-phosphate	0.15 ± 0.01	2.6 ± 0.4	6×10^1
p-nitrophenylphosphate	0.0019 ± 0.0004	0.05 ± 0.01	4×10^1
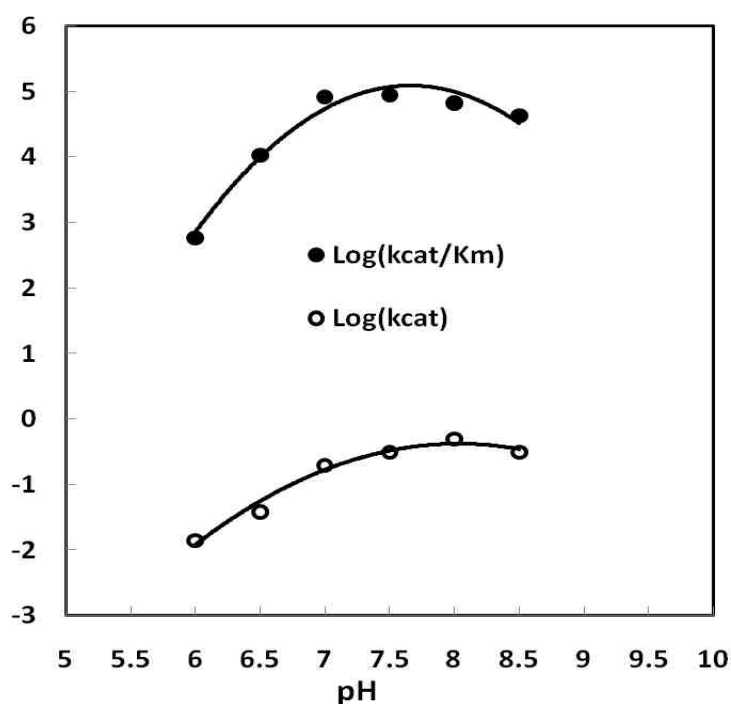
Figure 2.15: The pH rate profiles measured for BT2127 catalyzed hydrolysis of pyrophosphate. (See the Materials and Methods section for details).

The k_{cat} value is governed by the rate-limiting step of the two-step reaction: phosphoryl transfer from the bound substrate to the Asp nucleophile followed by hydrolysis of the aspartyl-phosphate intermediate (**Figure 2.3**). Variation in the k_{cat} value with substrate structure indicates that the first step is rate-limiting (or at least partially

rate-limiting) and that some substrates bind more productively than others. The largest k_{cat} value was measured for the substrate inorganic pyrophosphate (0.3 s^{-1}). Inorganic pyrophosphate exists in the cell as the magnesium complex, and it undergoes spontaneous hydrolysis to two molecules of orthophosphate at the reported rate of 2.8 x 10^{-10} s^{-1} (at 25 $^{\circ}$C, pH 8.5) [11]. BT2127 thus increases the hydrolysis rate ~1 x 10^{9}-fold.
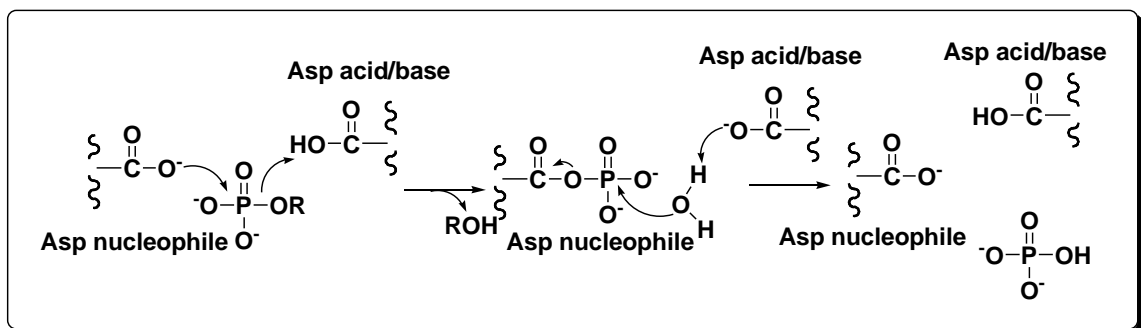
Figure 2.16: The HADSF phosphatase chemical pathway.

The k_{cat}/K_m value (also known as the substrate specificity constant) is determined by the substrate binding affinity as well as by the efficiency that the bound substrate is converted to product. Because the concentration of the substrate in the cell is likely to be sub-saturating, the k_{cat}/K_m value is most useful for identifying substrates that have physiologically relevant activities. HADSF phosphatases that target a single

34

physiological substrate typically display a k_{cat}/K_m value in the range of 1 x 10^5 to 1 x 10^7 M^{-1} s^{-1}, whereas those that function as regulators of the level and composition of organophosphate metabolite pools typically have K_m values in the mini-molar range and k_{cat}/K_m values in the 1 x 10^3 to 1 x 10^4 M^{-1} s^{-1} range. The k_{cat}/K_m value of ~1 x 10^5 M^{-1} s^{-1} measured for inorganic pyrophosphate qualifies it as a candidate for physiological substrate. Furthermore, the fact that the screen did not identify other substrates with physiologically relevant activities suggests that BT2127 might indeed function *in vivo* as an inorganic pyrophosphatase.

2.3.2 BT2127 Structure Determination

The structure of BT2127 was examined in order that we might identify the structural determinants of substrate recognition. A total of 13 X-ray structure determinations were carried out with wild-type enzyme and the D11N, D13A, D13N and E47A, E47D, and E47N mutants. The crystallization conditions (**Table 2.3**) and the crystallographic and refinement statistics are reported in Supporting Information (**Tables 2.4-2.7**). The different crystallization conditions led to different crystal packing and to different ligands in the active site. All crystallizations were carried-out in the presence of 5 mM MgCl₂, however only six of the structures (PDB codes 3QU2, 3QU4, 3QU9, 3QUQ, 3QUT and 3QX7) contained Mg^{2+} within the active site. In two of the structures (PDB codes 3QU7 and 3QYP) Ca^{2+} derived from the crystallization solution, replaced the Mg^{2+}. Co-crystallization of each enzyme was attempted with pyrophosphate as well as with its analog imidodiphosphate (a competitive inhibitor with a $K_i = 13 \pm 1$ µM

(**Figure 2.4**), but without success. Crystal soaking with these two ligands also failed to place the intended ligand in the active site. Three structures derived from the pyrophosphate-soaked crystals (PDB codes 3QYP, QX7 and 3QU7) contained the hydrolysis product phosphate.
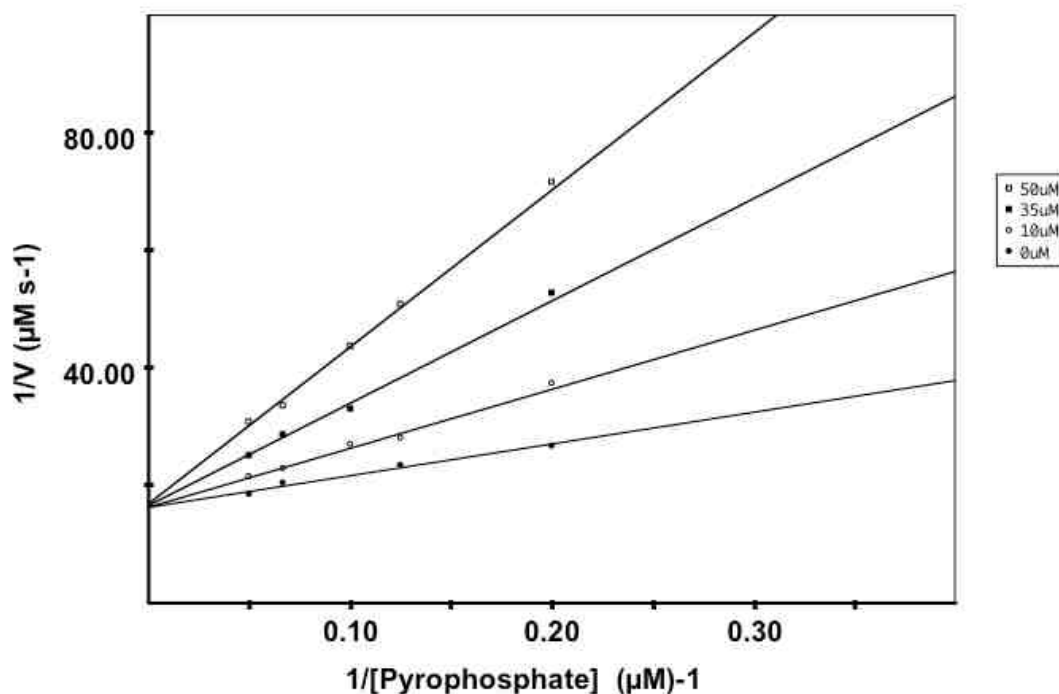
Figure 2.17: Inhibition plot for BT2127 with Imidodiphosphate under different fixed concentrations of pyrophosphate.

The structures of the BT2127 mutants showed that the wild-type native conformation, including the active site residue side chains, was retained. Therefore, the reduction in catalytic activity in the mutants (*vide infra*) is attributed to a loss of side chain function and not to an alteration of the native conformation.

36

Table 2.3: A list of the crystallization conditions used for wild-type and mutant BT2127.

BT2127	PDB code	Crystallization conditions
WT	3QU2	0.1 M sodium citrate, pH 5.6, 30% PEG4000, 0.2 M ammonium acetate, 5 mM, $MgCl_2$, 294 ^{o}K
WT	3QXG	5 mM $MgCl_2$, 294 ^{o}K
WT	3QUQ	0.1 M Bis-Tris, pH 6.5, 25% PEG3350, 0.2 M $MgCl_2$, 294 ^{o}K
WT	3QX7	0.056 M sodium phosphate monobasic, pH 8.2, 1.34 M potassium phosphate dibasic, 5 mM $MgCl_2$, 294 ^{o}K
D11N	3QU5	0.1 M Bis-Tris, pH 6.5, 25% PEG3350, 0.2 M $MgCl_2$, 294 ^{o}K
D13A	3QU4	0.1M MES, pH 6.0, 20% PEG8000, 0.2 M calcium acetate,5 mM $MgCl_2$, 291 ^{o}K
D13N	3QU9	0.1 M potassium sodium tartrate, pH 8.0, 0.1 M imidazole, 0.2 M NaCl, 5 mM $MgCl_2$, 294 ^{o}K
D13N	3QU7	0.1 M sodium acetate, pH 4.5, 30% PEG400, 0.2 M calcium acetate,5 mM $MgCl_2$, 294 ^{o}K
D13N	3QUT	0.15 M Malic acid, pH 7, 20% PEG3350, 5 mM $MgCl_2$, 294 ^{o}K
E47D	3R9K	0.1 M Tris-HCl, pH 8.5, 30% PEG4000, 0.2 M $LiSO_4$, 5 mM $MgCl_2$, 294 ^{o}K
E47N	3QYP	0.2M MES, pH 6.0, 20% PEG8000, 0.2 M calcium acetate,5 mM $MgCl_2$, 294 ^{o}K
E47N	3QUC	0.1 M Tris-HCl, pH 8.5, 30% PEG4000, 0.2 M $LiSO_4$, 5 mM $MgCl_2$, 294 ^{o}K
E47A	3QUB	0.1 M Tris-HCl, pH 8.5, 30% PEG4000, 0.2 M $LiSO_4$, 5 mM $MgCl_2$, 294 ^{o}K

Table 2.4: Data collection and refinement statistics for the wild-type BT2127 crystals.

PDB ID	3QUQ	3QU2	3QX7	3QXG
Cap orientation	Opened	Closed	Closed	Closed
Space group	P212121	C2	P3121	C2
Unit cell dimensions (Å)	-	-	-	-
-a	45.20	137.60	84.58	114.07
-b	65.34	71.76	84.58	70.59
-c	75.62	114.53	79.27	76.32
Cell angles (degrees)	-	-	-	-
-alpha	90.00	90.00	90.00	90.00
-beta	90.00	105.80	90.00	119.77
-gamma	90.00	90.00	120.00	90.00
Molecules per asymmetric unit	1	4	1	2
Solvent content	38.42	51.25	61.05	50.18
Matthew's Coefficient	2.00	2.52	3.16	2.47
X-ray source	NSLS X29A	NSLS X29A	NSLS X29A	NSLS X29A
Wavelength	1.0750	1.0809	0.9791	0.9791
Resolution	49.45-1.65	40.00-1.94	20.00-2.00	50.00-1.24
Reflections	28,023	79,445	22,599	147,352
Completeness (%)	99.2 (92.1)	99.3 (93.1)	95.00	99.0 (86.0)
I/sigma (I)	6.10 (3.00)	3.40 (1.20)	7.40 (1.00)	7.40 (0.90)
Rsym (Rmerge)	0.092 (0.530)	0.124 (0.800)	0.078 (0.900)	0.056 (?)
Rwork (Rfree)	16.4 (17.8)	23.4 (27.5)	19.1 (24.3)	15.2 (18.2)
Rfree reflections (%)	683 (3.2%)	2349 (3%)	683 (3.2%)	4418 (3.0%)
Average B factor (overall)	23.69	20.31	58.11	22.23
-amino acid residues	22.54	19.78	58.15	20.26
	33.12	25.21	56.95	35.98
Root-mean square deviation	-	-	-	-
-bond lengths	0.011	0.010	0.011	0.012
-bond angles	1.293	1.273	1.293	1.411
Ramachandran	99.1 (0.9)	99.1 (0.9)	98.7 (1.3)	98.8 (1.2)
Number of solvent molecules	98	649	98	525

Table 2.5: Data collection and refinement statistics for the BT2127 D11N crystals.

PDB ID	3QU5
Mutation	D11N
Cap orientation	Closed
Space group	P21
Unit cell dimensions (Å)	-
-a	42.80
-b	77.21
-c	70.76
Cell angles (degrees)	-
-alpha	90.00
-beta	90.41
-gamma	90.00
Molecules per asymmetric unit	2
Solvent content	45.44
Matthew's Coefficient	2.25
Wavelength	0.9786
Resolution	40.00-1.24
Reflections	129,987
Completeness (%)	96.3 (68.4)
I/sigma (I)	8.00 (1.40)
Rsym (Rmerge)	0.043 (0.550)
Rwork (Rfree)	13.0 (16.0)
Rfree reflections (%)	3775 (3%)
Average B factor (overall)	22.35
-amino acid residues	19.62
-waters	35.36
Root-mean square deviation	-
-bond lengths	0.011
-bond angles	1.407
Ramachandran favored/allowed (%)	99.4 (0.6)
Number of solvent molecules	764

Table 2.6: Data collection and refinement statistics for the BT2127 D13 mutant crystals.

PDB ID	3QU4	3QUT	3QU7	3QU9
Mutation	D13A	D13N	D13N	D13N
Cap orientation	Closed	Opened	Closed	Closed
Space group	P1	P212121	C2221	P3121
Unit cell dimensions (Å)	-	-	-	-
-a	69.94	45.47	72.55	84.17
-b	76.22	65.40	136.87	84.17
-c	95.59	76.17	114.83	79.36
Cell angles (degrees)	-	-	-	-
-alpha	89.97	90.00	90.00	90.00
-beta	89.97	90.00	90.00	90.00
-gamma	90.00	90.00	90.00	120.00
Molecules per asymmetric	8	1	2	1
Solvent content	47.95	38.42	53.48	59.60
Matthew's Coefficient	2.36	2.00	2.64	3.04
X-ray source	NSLS	NSLS	NSLS	NSLS
Wavelength	1.5418	0.9791	0.9792	1.0750
Resolution	50.00-2.00	50.00-1.50	50.00-1.90	40.00-1.87
Reflections	133,986	37,187	45,325	27,292
Completeness (%)	93.2 (83.9)	100.0 (100.0)	99.8 (99.5)	99.1 (92.8)
I/sigma (I)	4.50 (0.60)	7.40 (3.00)	7.70 (2.00)	8.10 (1.00)
Rsym (Rmerge)	0.101 (0.730)	0.067 (0.600)	0.053 (0.680)	0.069 (0.820)
Rwork (Rfree)	21.8 (28.6)	14.5 (17.2)	19.0 (22.6)	20.9 (25.2)
Rfree reflections (%)	3211 (3)	1152 (3.1)	1408 (3.1)	826 (3.2)
Average B factor (overall)	36.17	25.13	47.74	51.00
-amino acid residues	36.24	23.78	47.63	50.92
-waters	35.32	35.55	49.74	52.02
Root-mean square deviation	-	-	-	-
-bond lengths	0.009	0.011	0.009	0.011
-bond angles	1.177	1.387	1.202	1.352
Ramachandran Favored/allowed(%)	98.3 (1.7)	98.7 (1.3)	98.7 (1.3)	97.0 (3.0)
	982	243	275	110

Table 2.7: Data collection and refinement statistics for BT2127 E47 mutant crystals.

PDB ID	3QUB	3R9K	3QUC	3QYP
Mutation	E47A	E47D	E47N	E47N
Cap orientation	Closed	Closed	Closed	Closed
Space group	P3121	P3121	P3121	P21
Unit cell dimensions (Å)	-	-	-	-
-a	85.38	84.60	84.72	42.94
-b	85.38	84.60	84.72	77.18
-c	78.92	79.24	79.21	70.80
Cell angles (degrees)	-	-	-	-
-alpha	90.00	90.00	90.00	90.00
-beta	90.00	90.00	90.00	90.21
-gamma	120.00	120.00	120.00	90.00
Molecules per asymmetric unit	1	1	1	2
Solvent content	59.60	59.60	59.60	43.50
Matthew's Coefficient	3.04	3.04	3.04	2.17
X-ray source	NSLS X29A	NSLS X29A	NSLS X29A	NSLS X29A
Wavelength	1.0750	1.0750	1.0750	0.9792
Resolution	40.00-1.90	50.00-1.80	50.00-1.87	50.00-1.60
Reflections	27,369	30,588	27,575	61,148
Completeness (%)	99.1 (99.6)	99.7	97.2 (95.3)	95.3 (69.0)
I/sigma (I)	4.60 (1.00)	8.50 (1.20)	4.30 (1.20)	9.50 (1.00)
Rwork (Rfree)	19.8 (24.3)	19.0 (22.7)	21.6 (28.3)	18.5 (21.7)
Rfree reflections (%)	846 (3.2%)	966 (3.2%)	703	1776
Average B factor (overall)	43.06	44.14	57.09	29.12
-amino acid residues	42.58	43.57	57.04	28.16
-waters	47.80	49.74	58.06	36.31
Root-mean square deviation	-		-	-
-bond lengths	0.010	0.011	0.010	0.012
-bond angles	1.253	1.225	1.256	1.343
Ramachandran favored/allowed	99.1 (0.9)	98.3 (1.7)	98.7 (1.3)	99.6 (0.4)
	191	192	127	467

Quaternary Structure. The theoretical mass of BT2127 minus its N-terminal Met is 26,983 Da compared to 26,970 Da determined by mass spectrometry. The SDS-PAGE analysis (**Figure 2.1**) gave an estimated subunit mass of 27 kDa. The native mass was determined by using molecular size gel filtration chromatography to be 27-28 kDa. Thus, BT2127 is a monomer. Seven of the structure determinations revealed only a single protein molecule in the asymmetric cell (PDB codes 3QU9, 3QUB, 3QUC, 3QUQ, 3QUT, 3QX7 and 3R9K).

Tertiary Structure. BT2127 possess the conserved HADSF Alpha/Beta Rossmann-like catalytic domain comprised of a six-stranded parallel Beta sheet (Beta1, and Beta4-Beta8) surrounded by six helices (Alpha1-Alpha3, and Alpha5-Alpha7) and the type C1A tetra-Alpha-helical cap domain inserted between Beta1 and Alpha1. Well-characterized members of the C1A class include *Lactobacillus lactis* Beta-PGM (Z-score 10.2) [12] and *Bacillus cereus* phosphonatase (Z-score 7.8) [13]. The superposition of the structure of BT2127 with that of Beta-PGM is shown in **Figure 2.5**. The C1A type cap domains of BT2127 and Beta-PGM superimpose with a RMSD of 2.4 Å (13.4% identity), whereas the two catalytic domains superimpose with a RMSD of 1.9 Å (26.7% identity). The closest homolog (Z-score 14.7) is the HADSF member PDB code 3DV9 from *Bacteroides vulgatis*, which is annotated "putative Beta-PGM". As detailed below I posit that this protein is not Beta-PGM but rather a pyrophosphatase, orthologue to BT2127.
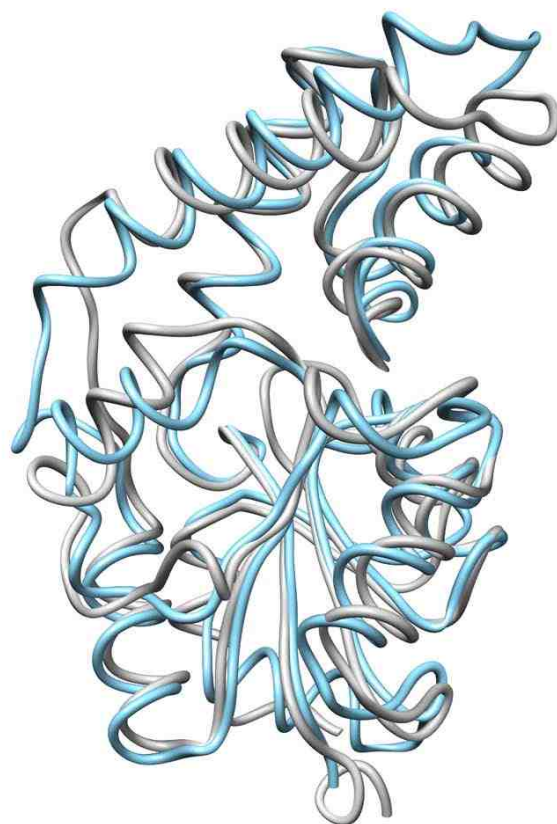
Figure 2.18: Superposition of BT2127 (PDB code 3QX7) (gray) and *L. lactis* Beta-PGM (PDB code 1O08) (cyan).

Cap Domain-Catalytic Domain Association. CA1 cap motion has been most thoroughly studied with *L. lactis* Beta-PGM for which crystallographic snapshots of the enzyme in the cap-open and cap-closed conformations have been obtained (see superposition in **Figure 2.6**) [14,15], The snapshots of BT2127 reveal a cap-closed conformation that is similar to that of Beta-PGM (**Figure 2.5**) and a cap-open conformation that is not as open as that observed for Beta-PGM (**Figure 2.6**).

43
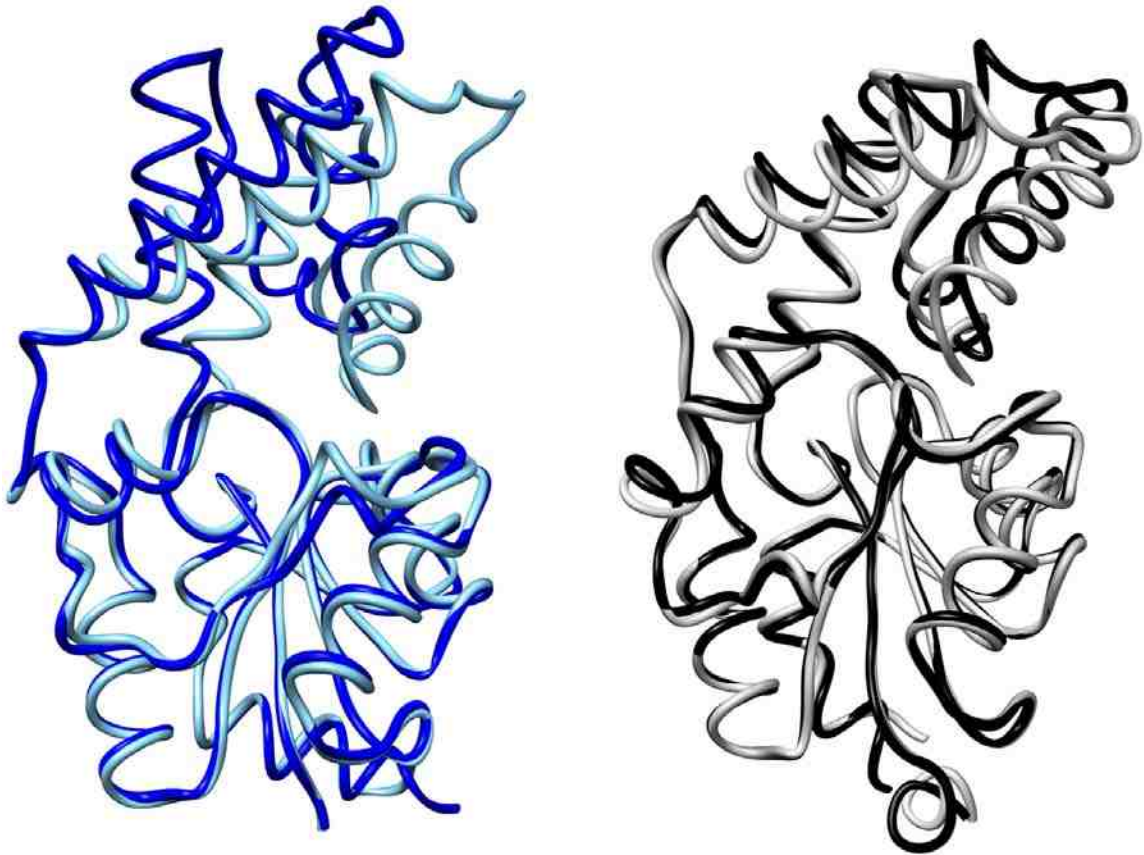
Figure 2.19: (Left) Superposition of the cap-open (royal blue) and cap-closed (cyan) conformations of *L. lactis* Beta-PGM (PDB codes 1ZOL and 1O08). (Right) Superposition of the cap-open (black) and cap-closed (gray) conformations of BT2127 (PDB codes 3QUQ and 3QX7).

In the case of Beta-PGM the cap and catalytic domains associate for catalysis and dissociate to allow ligand exchange. An earlier crystallographic based analysis of the "clam-like" domain movement in Beta-PGM using DynDom (Hayward 1997) indicated That the cap and catalytic domain move as rigid spheres and that the movement is primarily the result of changes in backbone conformations (psi and phi angles) at

domain-domain linker 1 hinge residues Thr114-Asp15-Thr16-Ala17 [16]. The more modest domain movement observed in BT2127 is attributed to changes in phi and psi angles at linker 1 hinge residues Ser19-Met20-Pro21 and linker 2 hinge residues Pro87-Glu88-Ala89-Glu90-Arg91. The rotation angle between the cap and catalytic domains in BT2127 is 16.3 degrees, with a translation of -0.6 Å and a closure percentage of 99.6%. As seen with Beta-PGM, the BT2127 cap and core domains must dissociate in order for pyrophosphate to bind to the active site and phosphate to depart from it because only in the cap-open conformation is there solvent access to the active site.

Crystallographer Jeremiah Farelli used the program Voidoo (http://xray.bmc.uu .se/usf/voidoo.html) was to calculate the volume of the active sites of BT2127 (189.8 Å^3) and Beta-PGM (285.1 Å^3) in their cap-closed conformations. Next, the volumes of pyrophosphate (99.4 Å^3) and Beta-glucose-1,6-bisphosphate (242.5 Å^3) were calculated using the molinspiration server (http://www.molinspiration.com/). Whereas pyrophosphate is small enough to fit in the active site of BT2127, Beta-glucose-1,6-bisphosphate is not. Models of the BT2127 bound with pyrophosphate and Beta-glucose-1,6-bisphosphate were generated to show the fit of the ligand in the 3-dimensional space available within the active site cavity (**Figure 2.7**). Whereas the pyrophosphate can be easily docked in a productive binding mode without steric clash, the larger Beta-glucose-1,6-bisphosphate ligand cannot fit without the cap dissociating to some extent from the core domain.
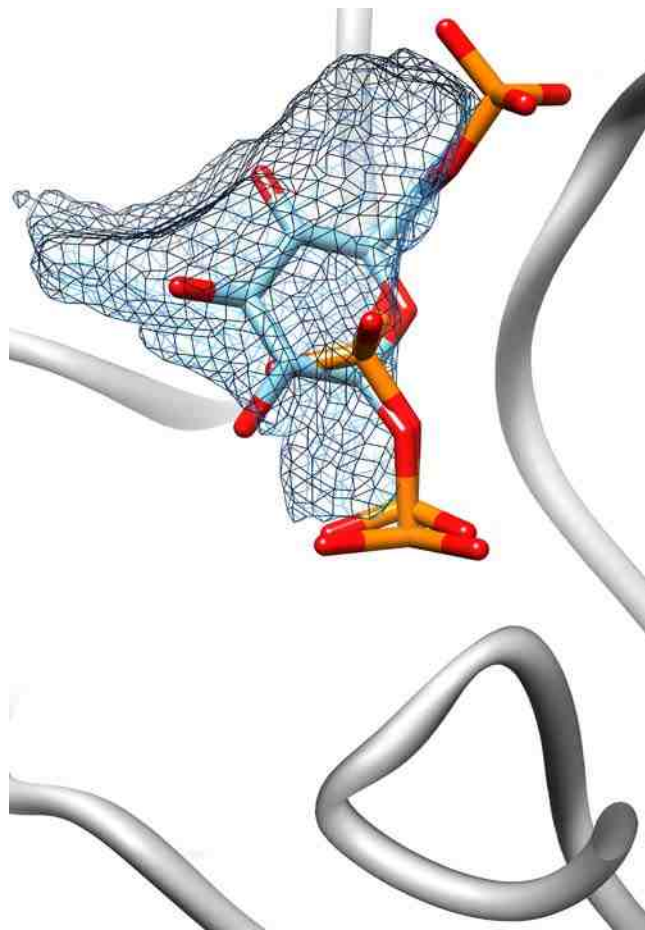
Figure 2.20: The active site of BT2127 in the cap-closed conformation (PDB code 3QX7) showing the unfilled area (mesh) calculated in Voidoo and the pyrophosphate and Beta-glucose-1,6-bisphosphate ligands modeled in Coot. Oxygen atoms are colored red, carbon atoms cyan and phosphorus atoms orange.

I posit that the cap-open and cap-closed conformations observed in the BT2127 crystal structures are not artifacts of crystal packing because the different crystallization conditions led to different crystal packing and different ligands in the active site. Substrate binding and product release would occur with the cap-open conformer whereas

46

catalysis would occur with the cap-closed conformer. The low level activity that is observed with the larger substrates (**Table 2.1** and **2.2**) might be rationalized by catalytic turnover in a conformer in which the cap domain is partially dissociated to enlarge the active site cavity.

Active Site. The active site scaffold of the HADSF phosphotransferase catalytic domain consists of a 4-loop platform on which the conserved catalytic residues are located: the Asp nucleophile and Asp acid/base on loop 1; the Thr/Ser and Lys hydrogen bond donors to the transferring phosphoryl group on loop 2 and 3, respectively; the Asp/Glu Mg^{2+} ligand on loop 4 [17]. **Figure 2.8** shows the residues surrounding the Mg^{2+} cofactor in the BT2127-Mg complex and the interacting residues in the BT2127 E47N-Ca-phosphate complex. As observed with other HADSF phosphotransferases, the Mg^{2+} is coordinated to the loop 1 Asp nucleophile (Asp11) carboxylate group and Asp acid/base (Asp13) backbone amide C=O (2.2 Å). However, the BT2127 loop 4 Asp171 is not positioned to coordinate the Mg^{2+} but instead is positioned to form a hydrogen bond (2.7 Å) to one of the two Mg^{2+} water ligands. It is the loop 4 Asn172 that forms a coordination bond with the Mg^{2+} (2.3 Å). A second deviation from the classical HADSF Mg^{2+} binding site [18, 19] is the identity of the sixth ligand, Glu47 (2.0 Å), which is located on the specificity loop of the cap domain (Thr50-Arg49-Gly48-Glu47).

From the structure of BT2127 E47N bound with phosphate and Ca^{2+} at the catalytic site we observe that an oxygen atom of the phosphate ligand replaces a water ligand (**Figure 2.8**). The phosphate ligand also is engaged in numerous hydrogen bond interactions with the active site residues: loop 2 Thr113 and loop 3 Lys147 side chains,

and loop 1 Asp13, loop 2 Gly114 & Ser115 backbone amide NHs.
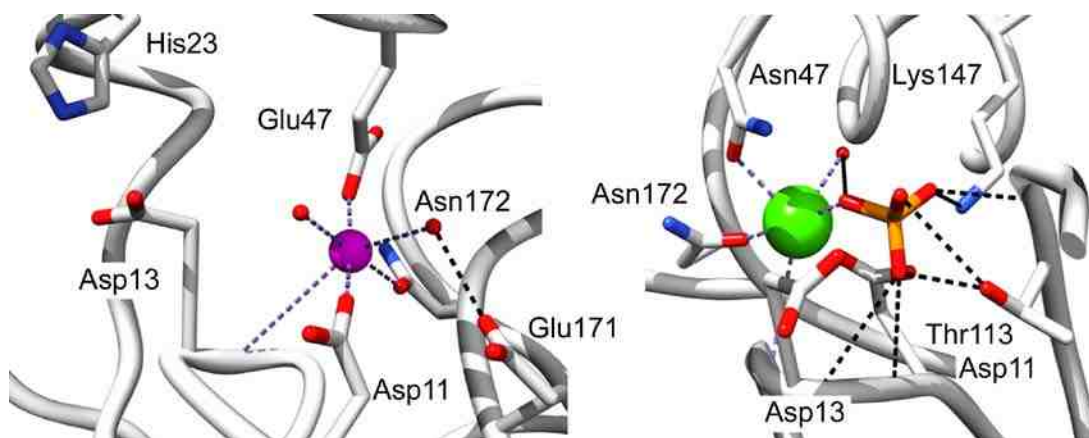
Figure 2.21: (Left) The Mg^{2+} binding site observed structure of wild-type BT2127 bound with Mg^{2+} (PDB code 3QUQ). (Right) The phosphate binding site observed in the structure of the BT2127 E47N mutant bound with phosphate and Ca^{2+} (PDB code 3QYP). Oxygen atoms are colored red, nitrogen atoms blue and phosphorus atoms orange. The Mg^{2+} is shown as a magenta sphere, the Ca^{2+} as a green sphere and the water molecules are represented as red spheres. Coordination bonds are shown as dashed purple lines and hydrogen bonds as dashed black lines.

In order to best illustrate the key interactions in the context of the pyrophosphate substrate, a model of the BT2127 active site (cap-closed conformer) with pyrophosphate bound in an orientation that superimposes the transferring phosphoryl group on the phosphate ligand of the BT2127 E74N-Ca-phosphate structure was made (by Jeremiah

Farelli) (**Figure 2.9**). Only one P-O-P rotomer conformation placed the phosphate-leaving group within the vacant space defined by the Voidoo analysis (*vide supra*). Here it can engage in hydrogen bond formation with a water molecule that is bound to cap domain Trp27 and with the backbone amide NH of cap domain residue Gly48 and catalytic domain Ser115. Moreover, the Asp13 acid/base is positioned to donate a hydrogen bond to a non-bridging oxygen atom, and possibly to the bridging oxygen atom, of the phosphate-leaving group.
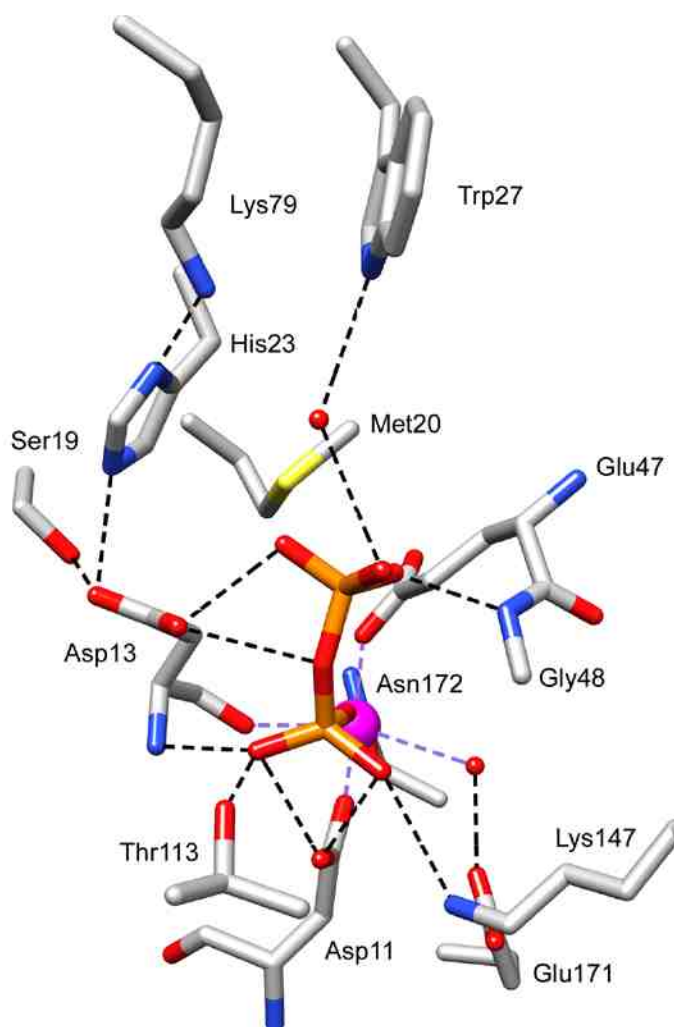
Figure 2.22: The BT2127 active site with pyrophosphate (manually docked). The Mg^{2+} is shown as a magenta sphere and the water molecules are represented as red spheres. Oxygen atoms are colored red, nitrogen atoms blue and phosphorus atoms orange. Coordination bonds are shown as dashed purple lines and hydrogen bonds as dashed black lines.

It is noteworthy that at pH 7.5 the pyrophosphate is monoprotonated ($pK_a = 9.3$) and if it were to bind with the protonated phosphate group in the leaving-group position, then there would be no need of acid catalysis. On the other hand, it has been suggested that the Asp acid/base of the HADSF phosphotransferases is protonated by the substrate phosphate ester, which binds to the active site in the monoprotonated form and upon coordination with the Mg^{2+} loses its proton [16]. Thus, in order for the Asp13 to be protonated the pyrophosphate would have to be fully ionized.

Interestingly, by replacing the pyrophosphate bridging oxygen atom with an "NH" as in imidodiphosphate tight binding to BT2127 is retained ($K_i = 13 \pm 1\mu M$) but catalytic turnover is greatly diminished (**Table 2.1**). We infer from this observation that the Asp13 is not able to protonate the bridging nitrogen atom most likely owing to steric restrictions. In contrast, several of the 23 *E. coli* HADSF phosphatases examined by Kuznetsova et al [1] showed significant hydrolase activity with imidodiphosphate but curiously, not with pyrophosphate.

The transferring phosphate group of the modeled pyrophosphate is positioned to coordinate with the Mg^{2+} and to form hydrogen bonds with the side chains of loop 2

Thr113 and loop 3 Lys147 and with the backbone amide NHs of loop 1 Asp13 and loop 2

Gly114 (**Figure 2.10**).
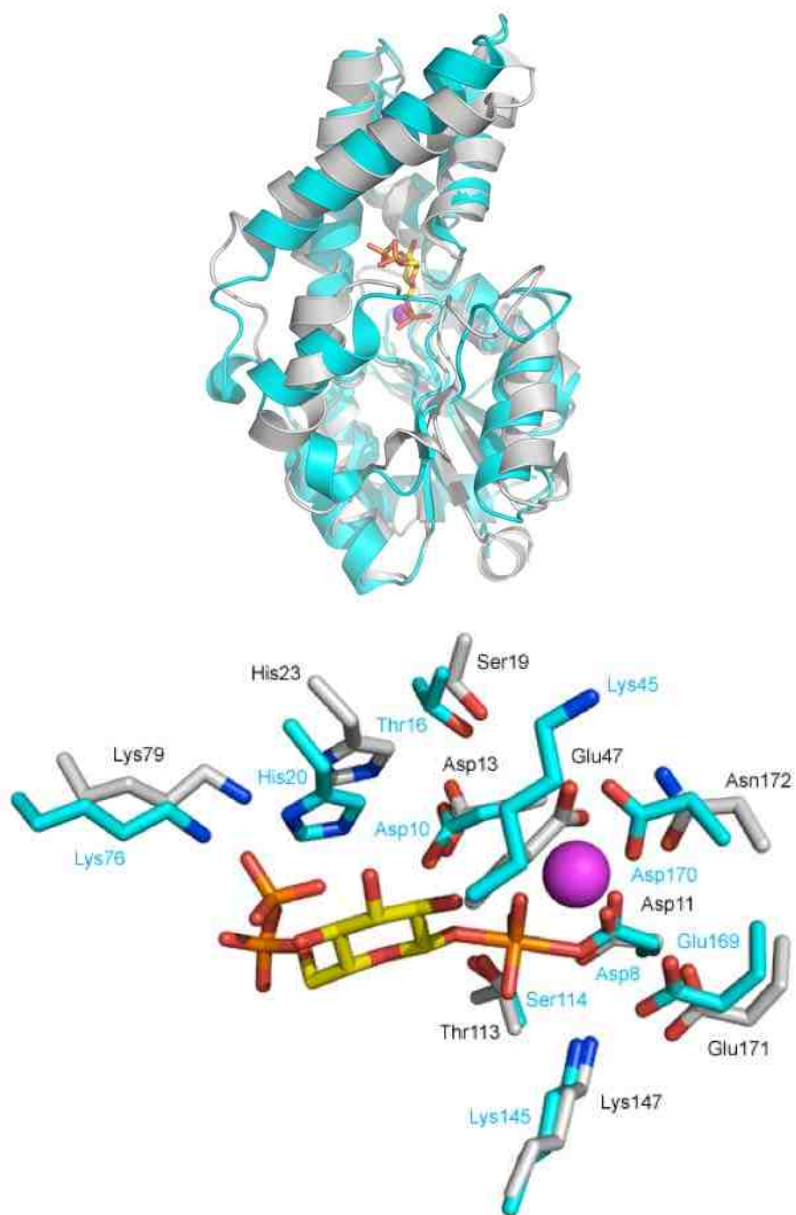
Figure 2.23: (Top) Superposition of BT2127 (gray) bound with phosphate and Mg^{2+} (PDB code 3QX7) and β-PGM (cyan) bound with Mg^{2+} and β-glucose-1,6-bis-phosphate (PDB code 1O08). (Bottom) Zoom in on the active site.

An interesting side note is the position that phosphate ligand assumes in the structure of wild-type BT2127-Mg-phosphate complex (PDB code 3QX7). Superposition of this structure with the structure of *L. lactis* Beta-PGM bound with Beta-glucose-1,6-bisphosphate places the phosphate ligand at the same position that the non-transferring phosphoryl group is found (i.e., C6-phosphate) (**Figure 2.10**). It is tempting to speculate that once formed the phosphate product in the BT2127 active site it might diffuse into this vacant space to make room for the water nucleophile, which will undergo in-line attack at the phosphorus atom of the aspartylphosphate intermediate in the second step of the reaction (**Figure 2.3**). Alternatively, the movement of the phosphate into the cap domain might trigger cap dissociation.

2.3.3 Site Directed Mutagenesis of BT2127 Active Site and Domain-Domain Interface Residues

The individual contributions that the active site residues make to catalysis were evaluated by site-directed mutagenesis coupled with steady-state kinetic analysis of the mutant enzymes in catalysis of pyrophosphate hydrolysis. The mutants that did not display activity above background are listed in **Table 2.8** as not active ("NA"). The mutants that displayed detectable yet very low activity were evaluated by measuring the initial velocity of the reaction using a concentration of pyrophosphate (300μM) that we

assume to be saturating. The ratio of the initial velocity and the enzyme concentration gives an approximate k_{cat} value (*i.e*, turnover rate). For the mutants that displayed higher activity, k_{cat} and K_m values were measured.

Table 2.8: Steady-state kinetic constants for wild-type and mutant BT2127-catalyzed hydrolysis of pyrophosphate at 25 ^{o}C and pH 7.5. See Materials and Methods for Details.

BT2127	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m (M^{-1}s^{-1})
Wild-type	1.9 ($\pm$ 0.06) x 10^1	3.6 $\pm$ 0.4	8.9 x 10^4
D11N	NA[a]	------	------
D13A	NA[a]	-------	------
D13N	NA[a]	-------	------
K147A	~2 x10^{-3} [b]	-------	------
N172A	7.4 ($\pm$ 1.8) x 10^{-1}	4.5 $\pm$ 0.3	3 x 10^3
T113A	4.0 ($\pm$ 0.03) x 10^{-1}	5.9 $\pm$ 0.2	1 x 10^3
S115A	8.4 ($\pm$ 0.12) x10^{-2}	3.6 $\pm$ 0.4	4 x 10^2
T50A	4.1 ($\pm$ 0.18) x 10^{-1}	2.7 $\pm$ 0.7	3 x 10^3
R49A	~3 x 10^{-2} [b]	-------	-------
K79A	~2 x 10^{-2} [b]	-------	-------
H23A	3.3 ($\pm$ 0.06) x 10^{-1}	1.2 ($\pm$ 0.08) x 10^1	5 x 10^2
M20A	~5.1 x 10^{-2} [b]	-------	-------
M20K	2.4 ($\pm$ 0.04) x 10^{-1}	2.5 ($\pm$0.3) x 10^1	2 x10^2
M20L	5.8 ($\pm$ 0.04) x 10^{-2}	3.5 $\pm$ 0.1	3 x 10^2
E47A	~3 x 10^{-3}	-------	-------
E47Q	4.8 ($\pm$ 0.4) x 10^{-1}	6 $\pm$ 1	1 x 10^3
E47N	~3 x 10^{-3} [b]	-------	-------
E47D	~3 x10^{-3} [b]	-------	-------
S80A	1.7 ($\pm$ 0.04) x 10^{-1}	7.1 $\pm$ 0.9	4.1 x 10^2
Q117A	~3 x 10^{-2} [b]	-------	-------
Y76F	~2 x 10^{-2} [b]	-------	-------
S19A	~2 x1 0^{-2} [b]	-------	-------

[a] NA represents no activity detected above the detection limit k_{cat} < 0.001 min^{-1}.

Table 2.9: The apparent first order rate $(k_{obs}$ (min^{-1})) constants for BT2127 catalyzed hydrolysis of phosphate esters and anhydrides at pH 7.5 and 25 ^{o}C. Reaction solutions initially contained 300 μM substrate, 8.4 μM mutant BT2127, 1 mM MgCl₂, 1.0 unit/mL purine nucleoside phosphorylase and 0.2 mM MESG in 50 mM Tris (pH 7.5). The k_{obs} value was calculated by dividing the initial velocity of the reaction by the enzyme concentration.

Substrate	wild-type	E47A	E47N	M20A	M20L	M20K
Ppi	1.9×10^{1}	4.7×10^{-4}	5.5×10^{-4}	8.4×10^{-4}	3.1×10^{-3}	6.5×10^{-3}
D-ribose5P	2.0×10^{-1}	2.5×10^{-4}	7.3×10^{-4}	1.3×10^{-3}	2.1×10^{-3}	2.3×10^{-4}
glycerol1p	3.5×10^{-1}	4.4×10^{-4}	4.6×10^{-4}	4.8×10^{-3}	4.2×10^{-3}	4.0×10^{-3}
PNPP	5.8×10^{-3}	4.2×10^{-4}	3.8×10^{-4}	2.1×10^{-5}	2.6×10^{-3}	4.6×10^{-3}
fructose6p	5.3×10^{-3}	2.5×10^{-4}	9.5×10^{-5}	4.4×10^{-4}	1.6×10^{-3}	3.6×10^{-4}
UMP	1.5×10^{-1}	$2.16 \times 10^{-}$	2.6×10^{-4}	3.16×10^{4}	1.6×10^{-3}	3.7×10^{-3}
imidodip	$1.03 \times 10^{-}$	6.2×10^{-5}	2.2×10^{-4}	6.0×10^{-3}	3.1×10^{-3}	4.4×10^{-3}
glucose1p	2.0×10^{-3}	1.1×10^{-4}	9.5×10^{-5}	8.4×10^{-4}	1.3×10^{-3}	2.9×10^{-3}
glucose6p	3.2×10^{-3}	1.9×10^{-4}	2.9×10^{-4}	2.8×10^{-4}	6.2×10^{-4}	3.7×10^{-3}
□ßG1,6p	4.1×10^{-3}	3.0×10^{-5}	2.8×10^{-5}	6.3×10^{-4}	6.7×10^{-4}	1.5×10^{-4}
ß-fucose1p	3.5×10^{-3}	2.3×10^{-4}	9.2×10^{-4}	2.5×10^{-4}	2.1×10^{-3}	1.6×10^{-3}
a-fucose1p	1.7×10^{-3}	5.5×10^{-4}	3.2×10^{-4}	2.3×10^{-4}	1.9×10^{-3}	9.1×10^{-4}
glycerol2p	$8.16 \times 10^{-}$	2.1×10^{-4}	2.7×10^{-4}	1.6×10^{-4}	1.6×10^{-3}	3.7×10^{-3}
serine3p	3.3×10^{-6}	1.4×10^{-4}	4.1×10^{-4}	3.5×10^{-4}	1.0×10^{-3}	3.5×10^{-3}

Firstly, we tested the impact of Ala and/or Asn replacement of the key catalytic residues Asp11 and Asp13, both of which are stringently conserved (see multiple sequence alignment of putative BT2127 orthologues in **Figure 2.11**). Not surprisingly,

given that the Asp nucleophile and Asp acid/base are known to be essential to HADSF phosphatase catalysis, Ala and/or Asn replacement of BT2127 Asp11 or Asp13 removed all detectable activity (**Table 2.8**). However, the observation of a phosphate ligand in the active sites of pyrophosphate-soaked crystals of BT2127 D13N indicates that a single turnover reaction had occurred during the 15 min soaking time or within the frozencrystal during data collection.

Ala replacement of the residues that contribute to the positioning of the Asp13 acid-base: Ser19, His23, Ser115 and Met20 (each of which is stringently conserved) reduced the turnover rate 950, 60, 230 and 370-fold, respectively. The Met20 was also replaced by Leu and by Lys. The M20L and M20K mutants were found to be 330 and 80-fold lower in turnover rate compared to that of the wild-type BT2127. Because the Met20 side chain restricts both the volume and the polarity of the region available to accommodate the substrate leaving-group we tested the activities of M20A, M20L and M20K towards the same panel of phosphate esters that were used to screen the wild-type enzyme. The objective was to determine if creating additional space (Ala for Met with Leu as the control) at the binding site might facilitate catalyzed hydrolysis of a larger substrate.

lcl|12539

```
                                                          1
lcl|12539        ................................MRK...K..........................
gi|29347537      ................................MRK...K..........................
gi|253570881     ................................MRK...K..........................
gi|298383766     ................................MRK...K..........................
gi|153808810     ................................MKK...K..........................
gi|260174129     ................................MKK...K..........................
gi|237720029     ................................MKK...K..........................
gi|160885231     ................................MKK...K..........................
gi|294644124     ................................MKK...K..........................
gi|53715098      ................................MFQ...EAIAQYLKQNHYESI..........
gi|255011243     ................................MFQ...EAITQYLQQNHYQSI..........
gi|60683056      ...............................................................
gi|167764833     .........MCNIKELPLDEQLSVLYSDISNPLFSIYQVFK...EAISRYLREHCYDNI....
gi|329957690     ................................MPK...ESITRYLKKHCYADI..........
gi|189463942     ................................MFQ...EAIARYLQSSGHSQI..........
gi|224536081     ................................MFQ...EAISRYLQSSCYSRI..........
gi|160889281     MRIEKWLHNKCHBNRSPILNSLFI.YPAFLNSQSPIYQVPE...NSIARYLEKHCHPYI...
gi|317475556     ................................MFE...VAISRYLKHHGHSDI..........
gi|329960268     ................................MFE...ESITRYLEKHCHSKI..........
gi|218128650     MTIFLFF........SKFEYRFCVILCPIQQQI.....MFE...VAISRYLKHHGHSDI...
gi|319903121     ................................MFE...ESIARYLEKHCHSNI..........
gi|212691210     ................................MFK...EAINNYLHAHGYRSI..........
gi|150006590     ................................MFK...EAINNYLHTHGYESI..........
gi|198275658     ................................MFQ...QEINQYLTTHHYTRI..........
gi|333031526     ................................MYK...QQIINYLTQHKQSNL..........
gi|189459664     ................................MFQ...QAIENYLTTHHPDKL..........
gi|154493629     ................................MIQ...EAIAQYLQKQRQVLF..........
gi|218262603     ................................MIQ...EAIARYLQKQQQTFL..........
gi|150006815     ................................MIQ...EAIIRYLEKHRQESL..........
gi|313205499     ................................MFT...AEIEKFIRSKNYDTF..........
gi|260885482     ..............................MMNDYIQAIQNYLKDHGFEQL...........
gi|298372101     ................................MNG.....ISKSFEKFYYYIMLNNNLLSFADG
gi|294674461     ...............................................................
```

              β1   α1   α2   α3
lcl|12539
              10   20   30   40   50   60   70

```
lcl|12539      ..LKAVLFDMDCVFMSLPYESEACHQVMETHCLDLSRHEAAMBECRTGASTHNIVFQRELGKEATQREI
gi|29347537    ..LKAVLFDMDCVFMSLPYESEACHQVMETHCLDLSRHEAAMBECRTGASTHNIVFQRELGKEATQREI
gi|253570881   ..LKAVLFDMDCVFMSLPYESEACHQVMETHCLDLSRHEAAMBECRTGASTHNIVFQRELGKEATQREI
gi|298383766   ..LKAVLFDMDCVFMSLPYESEACHQVMETHCLDLSRHEAAMBECRTGASTHNIVFQRELGKEATQREI
gi|153808810   ..LKAVLFDMDCVFMSLPYESEACHKVMKSHCLTLSRHEAAMBECRTGASTHNIVFQRELGREATQREI
gi|260174129   ..LKAVLFDMDCVFMSLPYESEACHKVMKSHCLTLSRHEAAMBECRTGASTHNIVFQRELGREATQREI
gi|237720029   ..LKAVLFDMDCVFISLPYESEACHKVMKSHCLTLSCEHAAMBECRTGASTHNIVFQRELGKEATQREI
gi|160885231   ..LKAVLFDMDCVFISTPYESEACHTVMKSHCLTLSRHEAAMBECRTGASTHNIVFQRELGREATQREI
gi|294644124   ..LKAVLFDMDCVFISTPYESEACHTVMKSHCLTLSRHEAAMBECRTGASTHNIVFQRELGREATQREI
gi|53715098    .QLKAVLFDMDCVPFDSLPYESEACHKTMEAAHCLNLSRHEAAMBECRTGAGTHNIVCQRQLGRDATQREI
gi|255011243   .QLKSALFDMDCVFMSLPYESAEACHKTMEAAHCLHLSKHEAALHECRTGAGTHNIVCQRQLGRDATPREI
gi|60683056    ....MFDMDCVPFISTPYESAEACHKTMEAAHCLNLSRHEAAMBECRTGAGTHNIVCQRQLGRDATQREI
gi|167764833   .RLKAVLFDMDCVFMSTPYESADACHTVMERHCLHLSRHEAAMBECRTGASTHNIVYQRQYCKDATPRMI
gi|329957690   .RLKAVLFDMDCVFMSTPYESADACHKVMERHCLHLSRHEAAMBECRTGAATHNLVYQRQYCKDATPDMI
gi|189463942   .QLKSVLFDMDCVFMSTPYESADACHKVMSRHDLHLSRHEAALHECRTGAATHNIVYQRQYCKDATPRMI
gi|224536081   .QLKSVLFDMDCVFMSTPYESADACHKVMSRHCLHLSRHEAALHECRTGAATHNIVYQRQYCKDATPRMI
gi|160889281   .QLKAVLFDMDCVFMSTPYESADACHTVMERHCLHLSRHEAAMBECRTGASTHNIVYQRQYCKDATPRMI
gi|317475556   .RLKAVLFDMDCVFMSTPYESADACHKVMRRGCLHLSRHEAAMBECRTGAATHNIVYQRQYCQDATPELI
gi|329960268   .QLKAILFDMDCVFMSTPYESADACHTVMERHCLHLSRHEAALHECRTGAATHNIVYQRQYCKDATPRMI
gi|218128650   .RLKAVLFDMGCVFMSTPYESADACHKVMERHCLHLSRHEAAMBECRTGAATHNIVYQRQYCQDATPELI
gi|319903121   .RLKAVLFDMDCVFMSTPYESADACHKVMEHHCLHLSRHEAATHECRTGAATHNIVYQRQYCKDATPRMI
gi|212691210   .DLKAVLFDMDCVFFISTPNEAEACHKIMERPGCGLSRHEAAMBECRTGASTHNIVSRRHERGHDATEREI
gi|150006590   .DLKAVLFDMDCVFFISTPMEAEACHKIMKRPGCGLSRHEAAMBECRTGASTHNIVSRRHERGHDATEREI
gi|198275658   .NLKAVLFDMDCVFFISLKMEATACHEAMEQYGNSLEKHEAALHECRTGAATHNIVEQRHERGYEADEKEI
gi|333031526   .NLKAVLFDMDCVFFISLPMEAEACHKVMERHNLDMNKHVVMMNECRTGRGTHNLVCQRHERGHDATDRREV
gi|189459664   .NLKAVLFDMDCVFFISLKMEAAACHEAMEQYGMHLSPEEAMHECRTGAATHNIVSTRQRHHEATEREI
gi|154493629   .SPKAVLFDMDCVFTISLRPEARACYETATHHQLISTPELPLLYKERTGESTHNELYQKTFLRDATDEEK
gi|218262603   .APKAVLFDMDCILYISLRPEARACYETATHHQLISTPELFLLYEERTGRETHNELYQKTFQRDATDEEK
gi|150006815   .SPKAVLFDMDCVFTISLRPEARACHEVATLRQLTERDEDFLMPECRTGHSTHNELYQRTFQRDATAREK
gi|313205499   .KPKAVLFDMDCVFFISLKPEATACYEALHEKNLPPTEYQLSMNECRTGASTHDEIFVKEHCKKATEQRE
gi|260885482   .APEAVLFDMDCVFFTSLRIANELCQATAAAYNLEMLPDREVTLECCTSTLMENTLMRRHQKQATPREL
gi|298372101   KNIKAVLFDMDCVFFISTPYEAAQACQRALSERGLPPSLYDELHECRTADSTDEAFIRAYCHTADERTK
gi|294674461   .MIRAVLFDMDCVFYIGLKMEAVALVQBMKEFGINMTAQDDALATZCAKGVDTHRYMVKQQQGREITEARS
```

56

αII

QQ QQQQQ QQQQQ
220

lcl	12539	QTLCDSWDTIML..............
gi 29347537	QTLCDSWDTIML..............
gi 253570881	QALCDSWDTIML..............
gi 298383766	QALCDSWDTIML..............
gi 153808810	QALCHAWDGL...DL...........
gi 260174129	QALNDTWDMMTENDI...........
gi 237720029	QALNDTWDMMTENDI...........
gi 160885231	QALSDHWDTLFEKNT...........
gi 294644124	QALSDHWDTLFEKNT...........
gi 53715098	QALCESWEKLVRLLH...........
gi 255011243	QALCESWEALARFLHQDPLETTIADN
gi 60683056	QALCESWEKLVRLLH...........
gi 167764833	QAFCDNWEAVRDALA..SHE......
gi 329957690	QAFCDDWEIVRKHLA..SQE......
gi 189463942	QAFCDNWEKLHEAFG..QINRTL...
gi 224536081	QDPCDHWENLRKAFH...........
gi 160889281	QALCDDWERLQTALD..ASNSLRHKD
gi 317475556	QALNENWEKLQQALI..D........
gi 329960268	QALNANWESLHTALN..LTHLP....
gi 218128650	QALNENWEKLQQALI..D........
gi 319903121	QDPCDNWEKLHGKLN..RLHTPLPY.
gi 212691210	PDFNKNWETLQSALKQD.........
gi 150006590	PDFNKNWBTLQSALKQD.........
gi 198275658	SAPNENWETCRQSLQEVSL.......
gi 333031526	QALADNWDLLQNDILYCSKNSGKKKK
gi 189459664	SDPSKNWKDFYQAIK...........
gi 154493629	TALAKDWDNLMNAIK..IS.......
gi 218262603	TALAHDWNNLMNIIK..VS.......
gi 150006815	HNLAKDWKQIIELAK..STRLNHYSK
gi 313205499	KELYDNWEAIAS..............
gi 260885482	QALATHWEHIRAAIAACAPTK.....
gi 298372101	QHLEQIMQIILT..............
gi 294674461	QALSDSWKDLPDVKGQVDVI......

Figure 2.24: Multiple alignment of the amino acid sequences of BT2127 orthologues.

Likewise, we wished to determine if the placement of positively charged side chain (Lys for Met with Leu as the control) that might favorably interact with the phosphate-leaving group might enhance activity towards pyrophosphate. The results (reported in **Table 2. 9**) indicate that the substrate specificity of the mutants was not significantly different than that of wild-type BT2127.

The coordination of the Mg^{2+} by the loop 1 Asp nucleophile and by the backbone

amide C=O of the Asp acid-base is observed in all structurally characterized HADSF phosphatases and an Asp or Glu residue located on loop 4 adds a third ligand. In some cases an Asp or Glu also located on loop 4 forms a hydrogen bond with a water ligand bound to the Mg^{2+} cofactor. To my knowledge the collaboration between a loop 4 amide (Asn172) and cap domain specificity loop carboxylate (Glu47) to augment the Mg^{2+} binding by the two loop 1 Asp residues has not previously been reported. The Ala replacement of the stringently conserved Asn172 resulted in only a 26-fold reduction in turnover rate. In contrast, Ala replacement of the stringently conserved Glu47 resulted in ~6000-fold decrease in turnover rate and the same reduction is observed with the E47D and E47N mutants. The E47 might be required for Mg^{2+} binding and/or for productive binding of the cap and catalytic domains (E47 remains coordinated to the Mg^{2+} in the cap-open conformation). Notably, the structures of the three E47 mutants (PDB codes 3QUB, 3R9K, 3QUC, 3QYP) crystallized in the presence of Mg^{2+} conserve the positions of the Asp11 and Asn172 side chain and the Asp13 backbone amide C=O yet the Mg^{2+} is absent, and the enzyme is in the cap-closed conformation.   In contrast, the E47N mutant crystallized in the presence of Ca^{2+} (PDB code 3QYP) contains Ca^{2+} and it is coordinated to the loop 1 and 4 ligands as well as to the Asn47. The larger coordination sphere of Ca^{2+} apparently compensates for the shortened reach of the Asn side chain (see **Figure 2.8**). The replacement of Glu47 with Gln would allow Mg^{2+} coordination and this is reflected in the comparatively higher activity of this mutant which has a turnover rate that is decreased only 40-fold from that of the wild-type enzyme (**Table 2.8**). These findings suggest that E47 rather than the loop 4 Asn172 plays the key role in Mg^{2+} binding. The

59

E47-Mg^{2+} coordination is the likely reason why the cap domain does not dissociate from the catalytic domain to the extent observed with ß-PGM.

The Thr113 and Lys147 activate the transferring phosphoryl group via hydrogen bond formation (**Figures 2.8** and **2.9**). Both are stringently conserved, and we found that the turnover rate was reduced 42-fold in the case of the T113A mutant and 9,500-fold in the case of the K147A mutant.

The stringently conserved cap domain residue Lys49 is not close enough to the catalytic site for direct interaction with pyrophosphate. It does however, form a hydrogen bond with stringently conserved cap domain His23, which in turn forms a hydrogen bond with the catalytic domain Asp13 acid/base (**Figure 2.9**). Whereas the turnover rate of the K49A mutant was found to be 950-fold slower than that of the wild-type BT2127, that of the H23A mutant was reduced only 60-fold. Thus, Lys49 appears to contributing to catalysis in some way beyond simply serving as a hydrogen bond partner to His23 (*vide infra*).

Additional mutants were examined to evaluate the contributions made by residues located at the domain-domain interface, which appear from the structures to stabilize the cap-closed conformation through hydrogen bond formation with a partner located on the opposing domain (**Figure 2.12**).
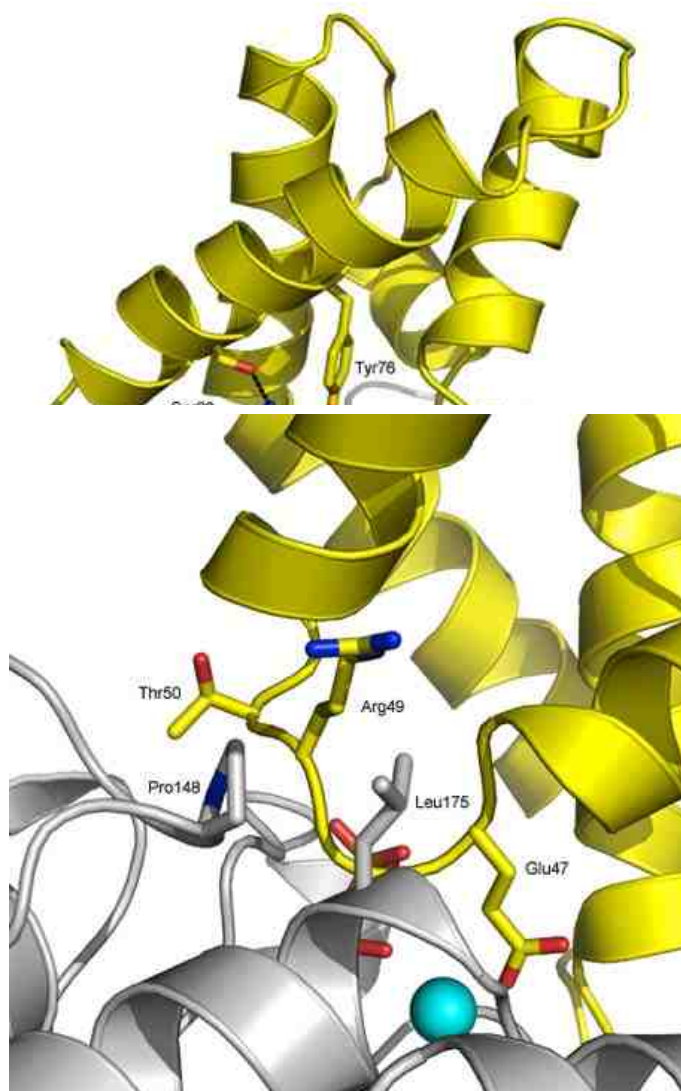
Figure 2.25: Interactions at the BT2127 cap domain (yellow)-catalytic domain (gray) interface.

Firstly, the stringently conserved Gln117 of the catalytic domain forms a hydrogen bond with cap domain residue Ser80 (replaced with Thr, Ala or Arg in 8 out of 33 sequences). Ala replacement of Gln117 reduced the turnover rate 630-fold whereas Ala replacement of Ser80 reduced the turnover rate 110-fold. Secondly, the stringently

conserved Tyr76 located on the inter-domain linker 2 is almost close enough (3.7 Å) for hydrogen bond formation with the backbone amide NH of catalytic domain Gln117 and Gly116 (highly conserved; replaced with Ala in 1 out of 33 sequences). The conservative replacement of Tyr76 with Phe reduced the turnover rate 950-fold. Thirdly, the highly conserved (31 out of 33 sequences) Arg49 and Thr50 (32 out of 33 sequences) are located on the cap domain specificity loop (Thr50-Arg49-Gly48-Glu47). The Arg49 side chain projects into solvent where the nonpolar hydrocarbon region of the side chain packs against the side chains of the catalytic domain Pro148 (stringently conserved) and Leu175 (conservatively replaced by Met in 5 out of 33 sequences), an interaction that might facilitate the proper positioning of Glu47 within the active and/or influence the association of the cap and core domains. Ala replacement of Arg49 reduced the turnover rate 630-fold. Thr50 is seen in the crystal structures to be a bit too far away (4.3 Å) from the stringently conserved catalytic domain Gly114 amide C=O for hydrogen bond formation. Nevertheless, Ala replacement of Thr50 reduced catalytic turnover 50-fold, suggesting that the solution conformation might allow hydrogen bond formation or alternatively, that the Thr50 contributes to the conformation of the cap domain substrate specificity loop.

2.3.4 BT2127 Biological Range and *In Vivo* Function

Based on the structure-function analysis described in the previous sections we selected the catalytic residues Asp11, Asp13, Thr113 and Lys147 as well the metal binding residues Asp171, Asn172 and Glu47 to curate the list of BT2127 homologues identified by BLAST searches. An alignment of 33 putative homologues was generated (see **Figure 2.11**) that conserved 46 residues (including the 7 marker residues) out of 244 residues total (19% conservation). The lowest pair-wise sequence identity between BT2127 and a putative sequence homolog is 47%. Many of these sequences have N-terminal extensions (~20 amino acids) when compared to the BT2127 sequence. One of these is the orthologue from *Bacteroides vulgatis* (ExPASy accession code A6L7P8). The supposition of its X-ray structure (PDB code 3DV9) with that of BT2127 (not shown) reveals that N-terminal extension forms an a-helix at the surface of the catalytic domain located at the edge of the ß-sheet, distant from the active site.

Although the marker residues used in the curation of the BT2127 homologues are by definition stringently conserved, 14 sequences could be identified that replaced a marker residue, Asn172 with Asp (multiple sequence alignment shown in **Figure 2.13**). These sequences are more divergent (28-30% pair-wise sequence identity with BT2127) and all but one of them also possesses the conservative replacement of the Met12 of the Asp11-Met12-Asp13 loop 1 catalytic motif with Leu.
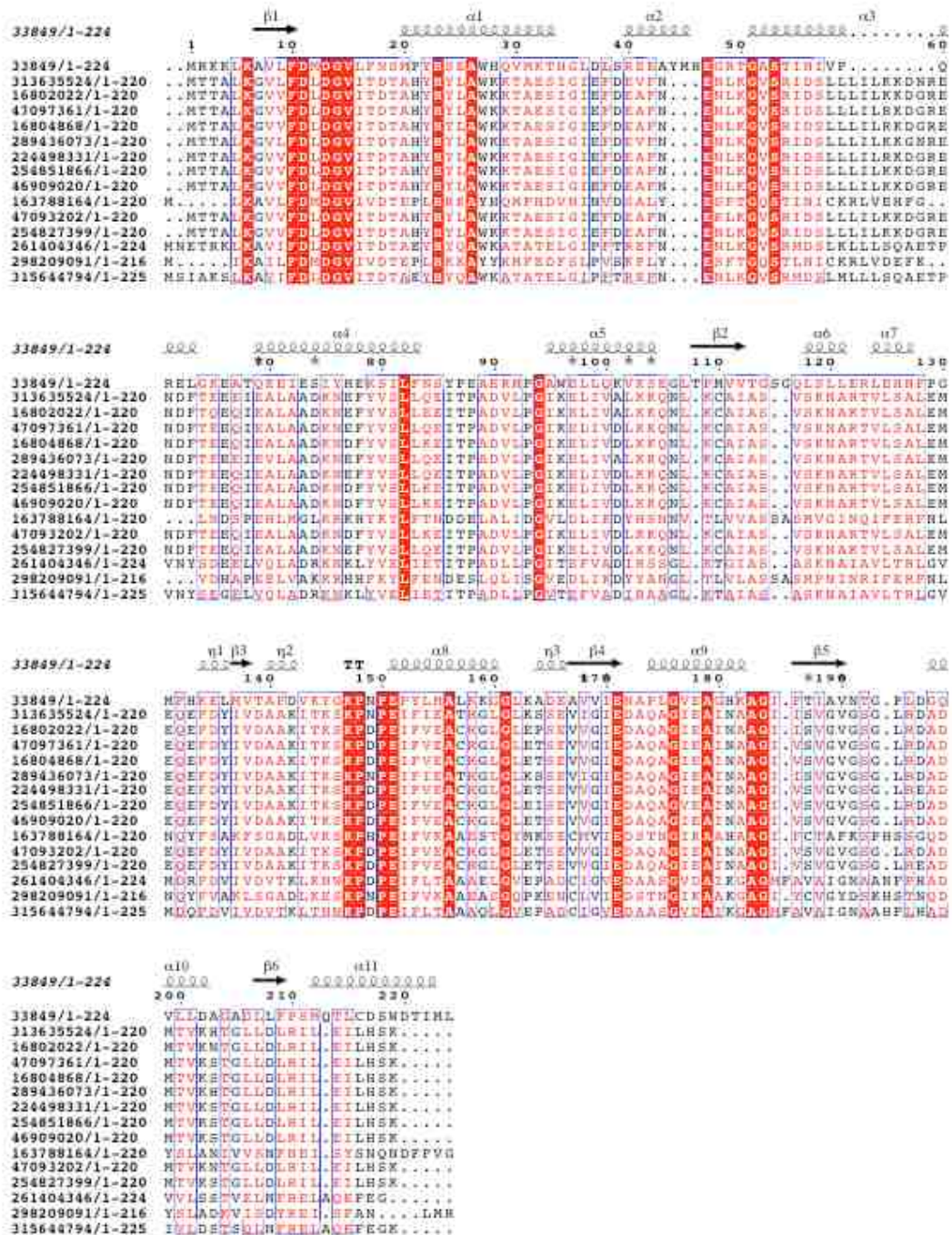
Figure 2.26: Alignment of the BT2127 sequence with the sequences of more distantly related homologues, which do not conserve the marker residue Asn172.

Although the residue that separates the Asp nucleophile and Asp acid/base on loop 1 is not known to have a direct catalytic role, it has been our observation that it is typically conserved within function families that have clear sequence boundaries. Moreover, highly or stringently conserved supporting residues identified in the previous section such as Tyr76, Thr50, Gln117, Ser19, Met20 and K49 are not conserved. Nevertheless, the 14 sequences might in fact be pyrophosphatases, but such annotation should be first supported by an *in vitro* activity assay of a representative member. For now we will refer to them as "BT2127-like" homologues.

The biological range of the putative BT2127 orthologues is restricted to the phylum Bacteroidetes/Chlorobi (**Figure 2.14**). BT2127 orthologues are found in all Bacteroides species/strains represented in the gene databanks (with the exception of *B. pectinophilus* that instead possesses at BT2127-like homologue) as well in the *Parabacteroid*es species/strains and *Prevotella* species/strains. *Bacteroides*, *Parabacteroides* and *Prevotella* colonize the human cavity from mouth to colon, yet occupy different regions along this environmentally stratified track. The BT2127-like homologues are also contained in the Bacteroidetes/Chlorobi phylum where they are found in select species that occupy a diverse range of habitats: *Capnocytophaga ochracea* (human mouth); *Chlorobium luteolum* and *Pelodictyon phaeoclathratiforme* (sulfide rich equatic environments), *Croceibacter atlanticus* and *Flavobacteria bacterium* BAL38, BBFL7, ALC-1 (sea water); *Listeria innocua*, *Listeria marthii*, *Listeria monocytogenes* and *Listeria seeligeri* (ubiquitous in the environment), *Paenibacillus sp.* Y412MC10 and

65

Paenibacillus vortex (heterogeneous and complex environments such as soil and rhizosphere).
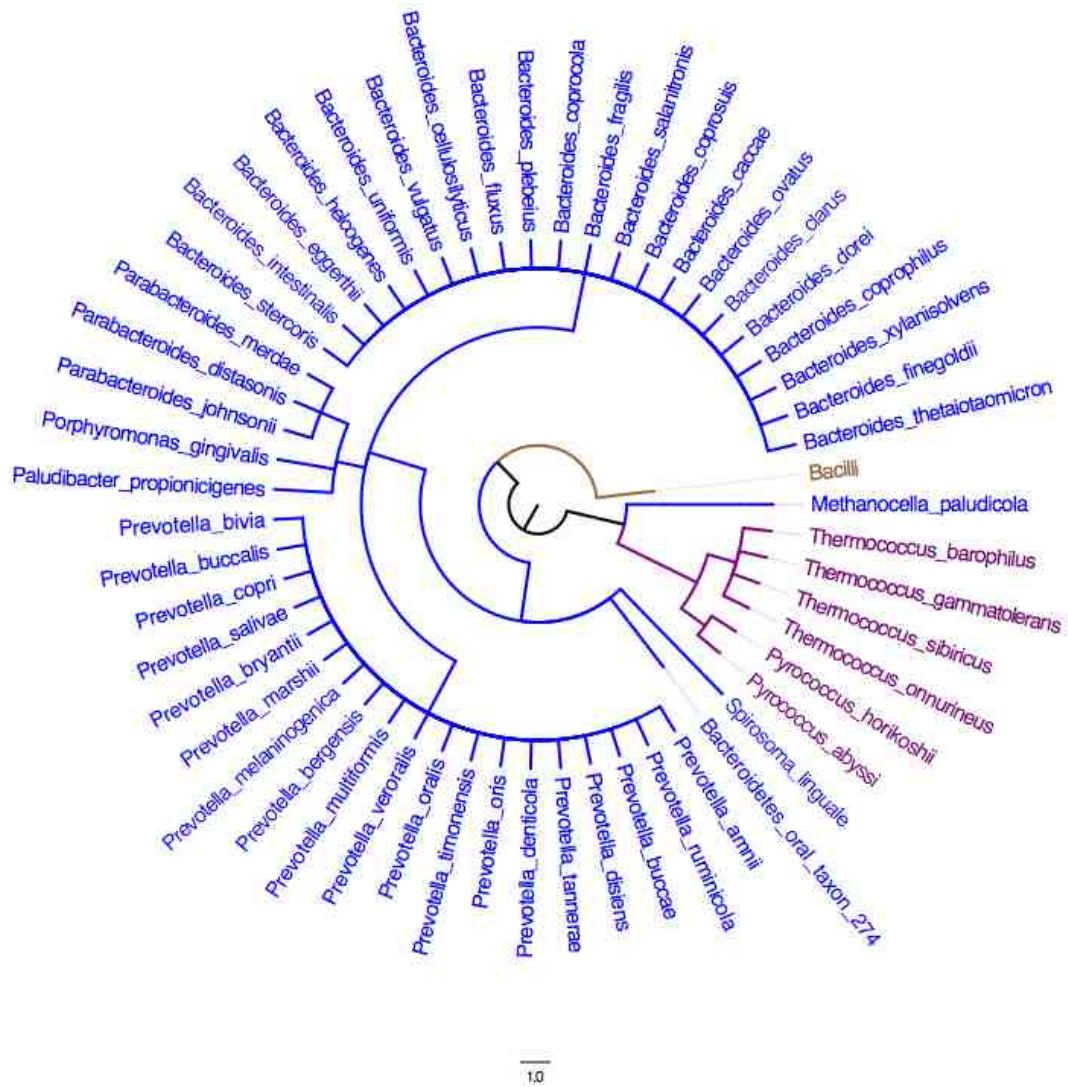
Figure 2.27: A phylogenic representation of the biological range of the putative orthologues of BT2127 (blue), Beta-PGM (brown) and the archeal pyrophosphatase TON0002 (purple). This tree was created by Chetanya Pandya.

The substrate specificity profile and the active site structure strongly indicate that inorganic pyrophosphate is the physiological substrate for BT2127. Independent

66

supportive evidence for the *in vivo* function of BT2127 as inorganic pyrophosphatase derives from gene context analysis. Pyrophosphate is derived from nucleoside 5'-triphosphates as they undergo nucleotidyl transfer reactions catalyzed by ligases, synthases, nucleotidyl transferases and DNA and RNA polymerases. The hydrolysis of the pyrophosphate product drives these reactions forward and at the same time replenishes the orthophosphate pool.

The gene neighborhoods of BT2127 and the 33 putative orthologues were examined for the bacterial genomes that have been mapped. The BT2127 gene overlaps the fkp gene that encodes the bifunctional L-fucose kinase/L-fucose-1-phosphate guanyltransferase. The products of the kinase-catalyzed reaction are Beta-L-fucose-1-phosphate (a poor substrate for BT2127, k_{cat} ~0.1 min^{-1}) and ADP whereas the products of the transferase-catalyzed reaction are pyrophosphate and Beta –L-fucose-GDP. The Beta –L-fucose-GDP is in turn used as a precursor for the decoration of surface capsular polysaccharides and glycoproteins with L-fucose for colonizing the mammalian intestine [19]. Thus, we posit that the biochemical function of BT2127 is inorganic pyrophosphate hydrolysis, which supports the biological function of host colonization. Intriguingly, several genes upstream and on the same DNA strand as BT2127 is the lysyl-tRNA synthase gene lysS (BT2122). Thus, BT2127 might also assist protein synthesis by removal of the synthase pyrophosphate product.

In *Bacteriodes fragilis* the putative pyrophosphatase orthologue encoding gene (BF3814) is next to the gene (BF3815) that encodes a putative □Asparagines synthase. Located a few genes upstream and on the same DNA strand is the lysyl-tRNA synthase

gene (BF3610). This same gene context is observed for *Bacteroides helcogenes* (*viz.* Bache 3336 = lysyl-tRNA synthase gene, Bache 3341 = pyrophosphatase gene, Bache 3343 =Asparagines synthase). In *Parabacteroides distasonis* (BDI 0146 = lysyl-tRNA synthase gene, BDI 0150 = pyrophosphatase gene) and *Prevotella denticola* (1697 = pyrophosphatase gene, 1699 = lysyl-tRNA synthase gene). Finally, we note that the *Paludibacter propionicigenes* genome encodes two BT2127 homologs, one having 56% sequence identity (Paplr 3043) and the other 28% sequence identity (Paplr 2273) with BT2127. The Palpr 3043 gene is located on a different DNA strand, opposite of the gene encoding DNA polymerase (Palpr 3030). Curiously, it is the more distant homolog (Paplr 2273) that is co-located on the same strand with cysteinyl-tRNA synthase.

2.3.5 Structural Determinants in the Divergence of Function in BT2127 and ß-PGM

Beta–PGM and maltose phosphorylase or trehalose phosphorylase collaborate in the utilization of maltose or trehalose as a carbon and energy source. Beta -PGM and BT2127 share 30% sequence identity and a common cap domain fold (tetra-a-helical) in addition to the highly conserved catalytic domain fold. Most significantly, they share the cap His20-Lys76 hydrogen bond diad motif that in Beta-PGM is the responsible for the discrimination between Beta-glucose-1-phosphate and solvent as reactants with the aspartyl phosphate intermediate (**Figure 2.15**) [16].
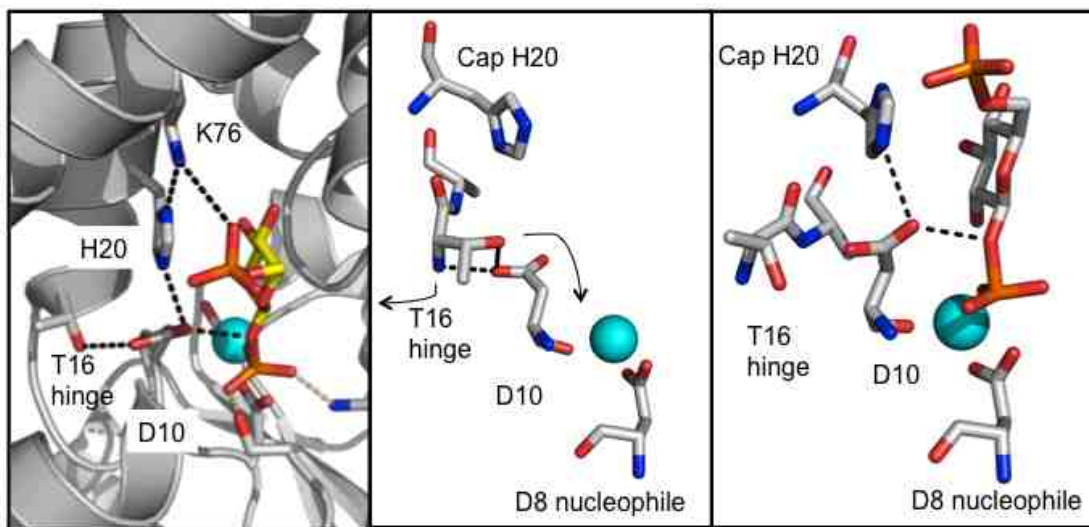
Figure 2.28: Left: The hydrogen bond network (black dashed line) of *L. lactis* Beta-PGM. Mg^{2+} cyan sphere and Beta-glucose-1,6-bisphosphate stick (oxygen red, carbon yellow, phosphorus orange, nitrogen blue). Center: The placement of Asp10 in the cap-open conformation (arrows indicate the direct direction of movement of the Thr16 hinge backbone and Asp10 side chain in going to the cap-closed conformation. Right: The placement of Asp10 in the cap-closed conformation.

Briefly, in the cap domain open conformation the Beta–PGM Asp acid/base side chain assumes a rotamer conformation, stabilized by hydrogen bond formation with the backbone amide NH and side chain of the linker hinge residue Thr117, which places it outside of the active site [14]. Upon substrate induced cap closure the His20-Lys76 diad and the Beta-glucose 1,6-bisphosphate P-O-C oxygen atom engage the Asp acid/base in hydrogen bond formation, placing it into position for catalysis (see **Figure 2.15**) [15].

As described above, BT2127 conserves the His23-Lys79 diad, and in both the cap domain-open and closed conformations the Asp13 side is in the same conformation, engaged in hydrogen bond formation with linker residue Ser15. In the cap-closed conformation the His23-Lys79 diad is close enough to the Asp13 carboxylate for hydrogen bond formation with the His23, however in the cap-open conformation it is too far away. The cap domain His-Lys diad appears to be a vestige of the common ancestry between BT2127 and Beta-PGM.

Having determined the biological range of the BT2127 orthologues we next examined the biological range of Beta-PGM using the same approach applied to BT2127. The specifics will be published separately, but here we present the final result that the putative Beta-PGM orthologues are primarily found in the phylum Firmicutes and especially in the genera Bacilli (as represented in **Figure 2.14**), however they can also be found in certain species of the phyla Proteobacteria, Actinobacteria and Bactoidetes/Chlorobi (not represented in **Figure 2.14**). Notably, a bacterial species that contain orthologues of both BT2127 and Beta-PGM were not found.

2.3.6 Evolution of Pyrophosphatases Within the HADSF and Other Superfamilies

Inorganic pyrophosphatases are essential to cellular function. The large (660-770 amino acid) membrane-bound pyrophosphatases are limited in range to plants and certain bacteria, wherein they function as proton pumps. The soluble pyrophosphatases are ubiquitous and they are found into two distinct fold families (type I and type II). The type I pyrophosphatases are widely distributed in all three kingdoms of life. This is an ancient family with high sequence divergence. The more recently characterized family is the type II family. Inorganic pyrophosphatases of the type II family are restricted in range to select lineages of bacteria [20].

The type I pyrophosphatases are single domain, Beta-barrel proteins whereas the type II pyrophosphatases are two-domain Alpha/Beta-proteins in which the active site is located at the domain-domain interface. Type I and II pyrophosphatases employ three or four Mg^{2+} cofactors to activate the transferring phosphoryl group and water nucleophile and to stabilize the phosphate-leaving group [20,21];

Most recently, a novel pyrophosphatase (TON0002) from *Thermococcus onnurineus* was discovered [22]. This Archeal pyrophosphatase is a member of the HADSF, which prompted us to take a closer look at its structure and biological range. Like BT2127, TON0002 possesses a C1 tetra-alpha-helical cap domain, yet this enzyme has evolved some unique structural features as inferred from the structure of its close sequence homolog (60% identity) from *Pyrococcus horikoshii* (PDB code 2OM6). The superposition of the structures of BT2127 and the *P. horikoshii* enzyme is shown in **Figure 2.16**.
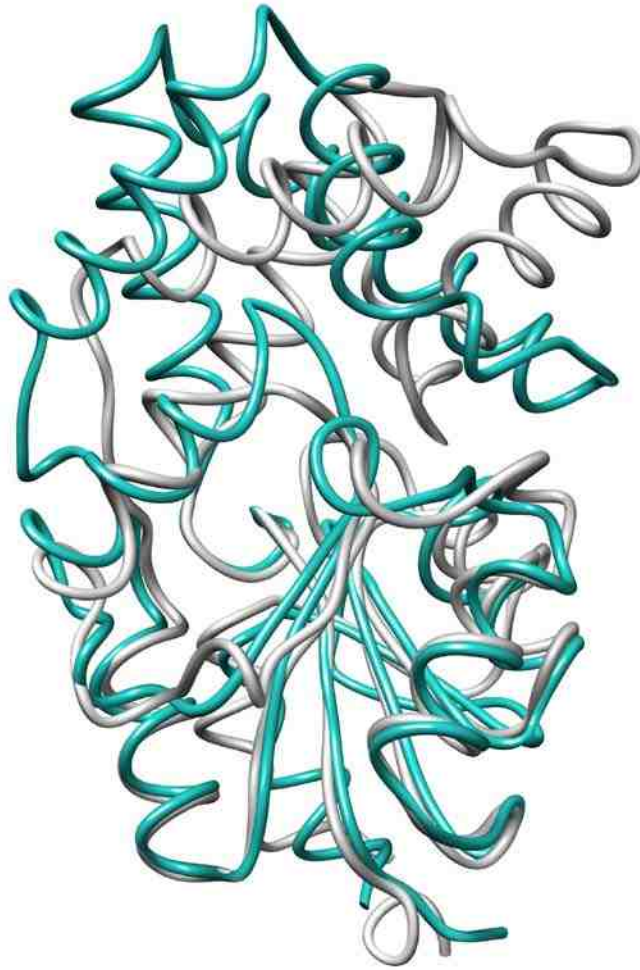
Figure 2.29: Superposition of the structures of BT2127 (gray) (PDB code 3QX7) and the putative pyrophosphatase from *P. horikoshii* (teal) (PDB code 2OM6).

The two sets of alpha-helices that form the TON0002 cap domain are splayed apart thereby expanding the area above the catalytic site. The active site would be open to solvent on two sides of the protein if it were not for: (1) the insertion of a hydrophobic loop (*viz.* loop 2) from the catalytic domain between the paired a-helices which seals one of the two openings and fills part of the void and (2) the long side chains of Arg48, Lys52

and Arg55 from the cap Alpha-helix 2 that partially occlude the opening on the opposite side.

A model was made (by Jeremiah Farelli) of the active site with Mg^{2+} and pyrophosphate bound by using the positions of the Mg^{2+} and the phosphate (to define the cofactor and transferring phosphoryl group) in BT2127 structure and the position of the sulfate ligand in the *P. horikoshii* enzyme structure to define the position of the phosphate leaving group (**Figure 2.17**). Remarkably, the loop 1 Asp acid/ base and loop 2 Thr/Ser are not conserved and instead Trp12 and Gly122, respectively are observed at these positions. The loop 1 Asp nucleophile is Asp10 and the loop 3 Lys is Lys158. The counterparts to the BT2127 loop 4 Glu171 and Asn172 are Gly182 and Asp183, respectively. In the model, the phosphoryl leaving group is positioned for favorable interaction with Arg48, Lys52 and Arg55 of the cap domain a-helix 2 whereas the transferring phosphoryl group is positioned to coordinate the Mg^{2+} and to favorably interact with catalytic domain residues Asn123 (loop 2) and Lys158 (loop 3).
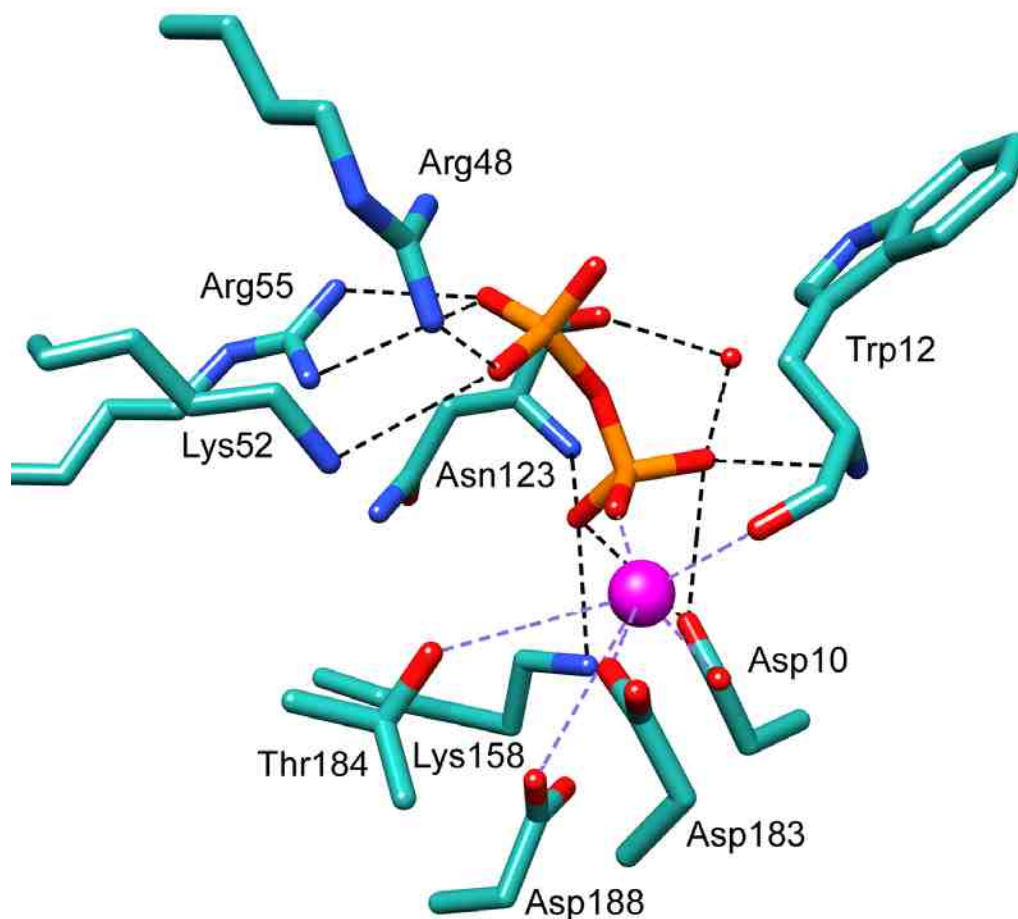
Figure 2.30: The active site of the putative pyrophosphatase from *T. onnurineus* , BT2127 modeled with pyrophosphate (manually docked) and Mg^{2+} (derived from the superposition of BT2127 PDB code 3QUQ). The Mg^{2+} is shown as a magenta sphere and the water molecules are represented as red spheres. Oxygen atoms are colored red, nitrogen atoms blue and phosphorus atoms orange. Coordination bonds are shown as dashed purple lines and hydrogen bonds as dashed black lines.

Because the sequence identity is high (60%) and because the key active site residues identified in the *P. horikoshii* enzyme structure are conserved in the *T.*

onnurineus pyrophosphatase we assume that the *P. horikoshii* enzyme is also a pyrophosphatase. The active site model suggests that the backbone amide NH of Asn123 substitutes for the conserved loop 2 Thr/Ser in activating the transferring leaving group and that cap domain residues Arg48, Lys52 and Arg55 stabilize the phosphate leaving group. In the second partial reaction (**Figure 2.3**), the loop 1 Asp acid/base activates the water nucleophile for attack on the aspartylphosphate. If it were to assume a slightly altered side chain rotamer conformation from that observed in the structure, the Asn123 could orient (but not deprotonate) the water nucleophile.

The biological range of the Archeal pyrophosphatase orthologues was tracked so that we might compare it to the BT2127 orthologues. A BLAST search carried out using the *T. onnurineus* pyrophosphatase identified three orders of Archea that produce homologues: Thermococcales (53-80% sequence identity), Desulfurococcales (31-41% sequence identity) and Thermoproteales (29-33% sequence identity). A sequence alignment was made (all sequences conserved Trp12) and those sequences that did not conserve the marker residues Arg48, Lys52, Arg55 and Asn123 were omitted. The alignment of suggested orthologues (orders Thermococcales and Thermoproteales) is shown in **Figure 2.18** whereas the biological range is contrasted with those of the BT2127 orthologues in **Figure 2.14**. Inspection of the gene contexts associated with these putative orthologues revealed that in some of the thermophiles a gene encoding DNA polymerase was found either on the same DNA strand next to pyrophosphates gene, or across from it on the opposite strand. This suggests a possible biochemical

context for the pyrophosphatase, i.e. aiding the polymerase by removing its product, pyrophosphate. On the other hand, the TON0002 $k_{cat}/K_m \sim 4 \times 10^2 \ M^{-1} \ s^{-1}$ is so low.

Figure 2.31: Multiple alignment of the amino acid sequences of TON0002 orthologues.

One might question its physiologically relevance especially given that the *T. onnurineus*

genome encodes a type A inorganic pyrophosphatase.

2.3.7 Conclusion

A BLAST search of the *B. thetaiotaomicron* genome using the *E. coli* type I

inorganic pyrophosphatase and the *Bacillus subtilus* type II inorganic pyrophosphatase

sequences as query failed to identify a homologue. Thus, the C1-type HADSF member

BT2127 appears to be the only known soluble inorganic pyrophosphatase in *B.

thetaiotaomicron*. The C1-type class appears to be the most prevalent and functionally

versatile of the three HADSF classes (C0, C1, C2) [23]. Aside from the many

phosphatases that have evolved within this class, dehalogenase, phosphonatase, Beta-

PGM and the newly identified pyrophosphatase activites have evolved as well. By

comparing these structures we gain insight into the key structural changes that might be

responsible for the change in function. In the case of the inorganic pyrophosphatase

BT2127 and its closest relative Beta-PGM the key structural changes are seemingly

directed at the how large the leaving group can be (phosphate vs glucose-phosphate) and

the position of the Asp acid/base in the cap-open vs cap-closed conformation. The

BT2127 pyrophosphatase active site volume precludes binding the Beta-glucose-1, 6-

bisphosphate. The two key residues in space restriction are the cap domain residues

Tyr76 and Glu47. The Beta-PGM activity requires the movement of the Asp acid/base

out of the active site when the cap opens to allow the association of the Beta-Glucose-1-

Phosphate for reaction with the aspartylphosphate. In both BT2127 and Beta-PGM the Asp acid/base forms a hydrogen bond with the linker hinge residue. In Beta-PGM this hydrogen bond only occurs in the cap-open conformation and it holds the Asp acid/base outside of the active site. In the BT2127 pyrophosphatase the hydrogen bond is present in both the cap-open and cap-closed conformations and the Asp acid/base remains in the active site.

2.4 References

1) Kuznetsova E, Proudfoot M, Gonzalez CF, et al (2006) Genome-wide analysis of substrate specificities of the Escherichia coli haloacid dehalogenase-like phosphatase family. *J Biol Chem. 281:36149–36161*.

2) 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*.

3) Erumbi S. Rangarajan, Ariane Proteau, John Wagner, Ming-Ni Hung, Allan Matte, and Miroslaw Cygler (2006) E. coli histidinol phosphate phosphatase: Structural snapshots along the reaction pathway *J.Biol.Chem 281(49) 37930-37941*

4)   Toms C Joseph, Lawrance Anbu Rajan, Nirmala Thampuran and Roswin James(2010)Functional characterization of trehalose biosynthesis genes from E. coli: an osmolyte involved in stress tolerance. *Mol Biotechnol 46(1):20-5*

5)   Wu J, Woodard RW (2003) Escherichia coli YrbI is 3-deoxy-D-manno-octulosonate 8-phosphate phosphatase *J Biol Chem.278(20):18117-23*

6)   Wang L, Huang H, Nguyen HH, Allen KN, Mariano PS Dunaway-Mariano D(2010) Divergence of biochemical function in the HAD superfamily: D-glycero-D-manno-heptose-1,7-bisphosphate phosphatase (GmhB), *biochemistry (49) 1072-1081*

7)   Björn Titz, Roman Häuser, Anne Engelbrecher, Peter Uetz(2007) The Escherichia coli protein YjjG is a house-cleaning nucleotidase in vivo *FEMS Microbiology Letters 270(1) 49-57*

8)   Bernard Weiss(2007) YjjG, a dUMP Phosphatase, Is Critical for Thymine Utilization by *Escherichia coli* K-12 *JOURNAL OF BACTERIOLOGY, 189(5) 2186-2189*

9)   Zhibing Lu, Liangbing Wang, Debra Dunaway-Mariano, and Karen N. Allen (2009)Structure-Function Analysis of 2-Keto-3-deoxy-D-glycero-D-galactonononate-9-phosphate Phosphatase Defines Specificity Elements in Type C0 Haloalkanoate Dehalogenase Family Members *THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 284(2), 1224–1233*

10) Wang L, Lu Z, Allen KN, Mariano PS, Dunaway-Mariano D (2008) Human symbiont Bacteroides thetaiotaomicron synthesizes 2-keto-3-deoxy-D-glycero-D-galacto-nononic acid (KDN) *Chem Biol.15(9):893-7.*

11) Stockbridge RB, Wolfenden R(2011) Enhancement of the rate of pyrophosphate hydrolysis by nonenzymatic catalysts and by inorganic pyrophosphatase *J Biol Chem 286(21):18538-46*

12) Sushmita D. Lahiri, Guofeng Zhang,Debra Dunaway-Mariano, and Karen N. Allen(2002) Caught in the Act: The Structure of Phosphorylated *ß*-Phosphoglucomutase from *Lactococcus lactis Biochemistry 41,8351-8359*

13) Marc C. Morais,Wenhai Zhang,Angela S. Baker,Guofeng Zhang,Debra Dunaway-Mariano, and Karen N. Allen(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 39,10382-10396*

14) Guofeng Zhang, Jianying Dai, Liangbing Wang, and Debra Dunaway-Mariano(2005) Catalytic Cycling in Beta-Phosphoglucomutase: A Kinetic and Structural Analysis *Biochemistry 44, 9404-9416*

15) Lahiri SD, Zhang G，Dunaway-Mariano D Allen KN. (2003) The pentacovalent phosphorus intermediate of a phosphoryl transfer reaction. *Science. 299(5615):2067-71.*

16) Dai J, Finci L, Zhang C, Lahiri S, Zhang G, Peisach E, Allen KN, Dunaway-Mariano D(2009) Analysis of the structural determinants underlying

discrimination between substrate and solvent in beta-phosphoglucomutase catalysis.*Biochemistry 45(25) 7818-7824*

17) Allen KN, Dunaway-Mariano D(2004) Phosphoryl group transfer: evolution of a catalytic scaffold. *Trends Biochem Sci.29(9):495-503*

18) Zhang G, Morais MC, Dai J, Zhang W, Dunaway-Mariano D, Allen KN (2004) 50 investigation of metal ion binding in phosphonoacetaldehyde hydrolase identifies sequence markers for metal-activated enzymes of the HAD enzyme superfamily. *Biochemistry.43(17):4990-7.*

19) Michael J. Coyne, Barbara Reinap, Martin M.Lee and Laurie E. Comstock(2005) Human Symbionts Use a Host-Like Pathway for Surface Fucosylation *Science 307, 1778-1781*

20) Ahn S, Milner AJ, Fütterer K, Konopka M, Ilias M, Young TW, White SA. (2001) The "open" and "closed" structures of the type-C inorganic pyrophosphatases from Bacillus subtilis and Streptococcus gordonii. *J Mol Biol. 313(4):797-811*

21) Fabrichniy IP, Lehtiö L, Tammenkoski M, Zyryanov AB, Oksanen E, Baykov AA, Lahti R, Goldman A. (2007) A trimetal site and substrate distortion in a family II inorganic pyrophosphatase. *J Biol Chem. 282(2):1422-31.*

22) Lee H-S, Cho Y., Kim Y-J, Lho T-O, Cha S-S, Lee J-H, Kang S-G, (2009) A novel inorganic pyrophosphatase in *Thermococcus onnurineus* NA1. *FEMS Microbiology Letters, 300 (1): 68-74*

23) Burroughs AM, Allen KN, Dunaway-Mariano D, Aravind L. (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(5):1003-34.*

CHAPTER THREE

STRUCTURE, CATALYSIS AND FUNCTION OF FKBH

3.1 Introduction

FKBH is comprised of three sequential structural domains. The center domain is a capless HADSF domain whose closest characterized homologue is the mammalian phosphatase MDP (ExPASY accession # Q86V88, 176 amino acids). The physiological substrate of MDP is not known although it has been speculated to be a protein phosphorylated at a tyrosine residue [1]. FKBH catalyzes the three-step reaction shown in **Figure 3.1**, the product of which is a glyceryl-holo acyl carrier protein/domain or glyceryl-CoA. The C-terminal domain catalyzes the autoacylation of its Cys residue using the metabolite 1,3-diphosphoglycerate as substrate [2] (see **Figure 3.1**). The C(3) phosphoryl group is then removed by the HAD domain. The glyceryl unit is transferred to the pantetheine thiol of a charged acyl carrier protein. The biological function of FKBH varies according to the biosynthetic context of the synthase that acts on the glyceryl-holo acyl carrier protein or domain. The glycerated ACP can either modified to serve as a starter unit for some polyketides biosynthesis (**Figure 3.1A**)[3-5], or it can be oxidized into a glycolate unit which can act as an extender unit in some other polyketide products (**Figure 3.1B**)[6-11], or the glycerate unit can be transferred to pantetheine arm of CoA to produce a glycerate CoA, which can be applied to construct other large molecules (**Figure 3.1C**)[12].
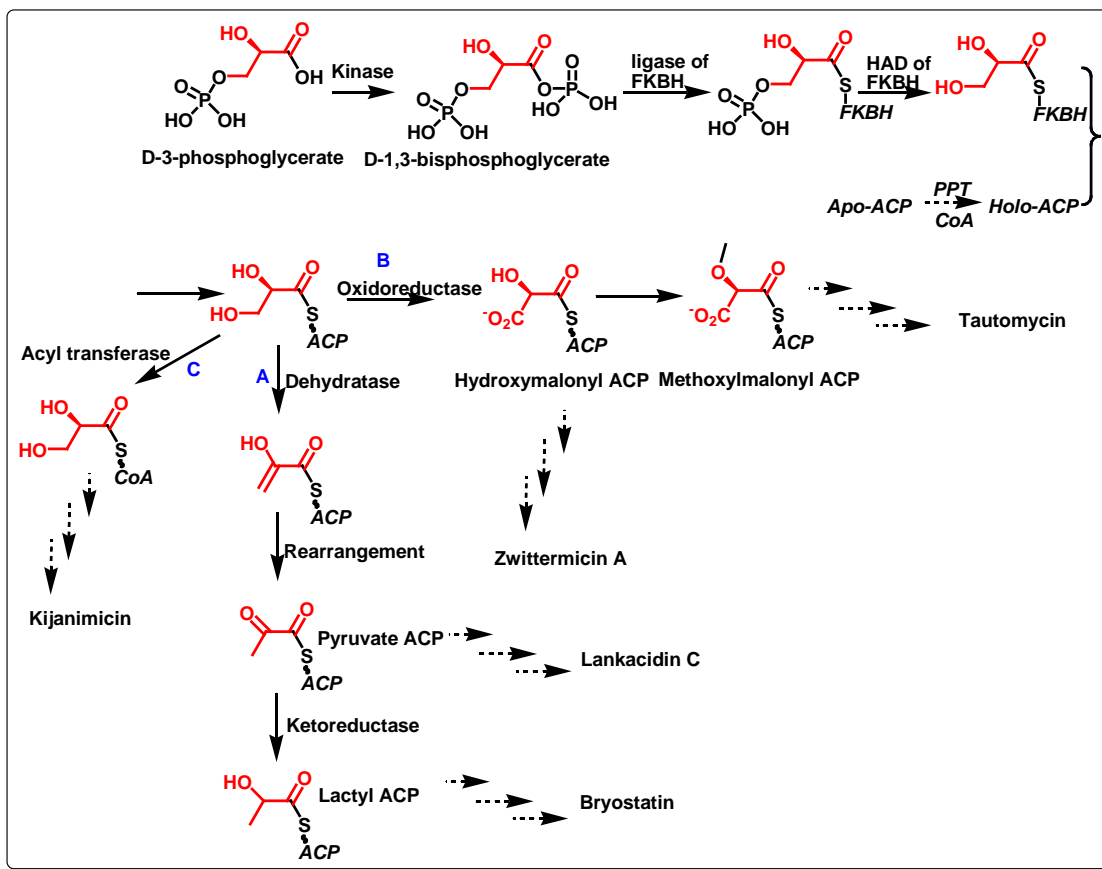
Figure 3.32 Pathway for FKBH generation of the glyceryl unit and pathways for its incorporation into natural products.

The *Bacteriodes fragilis* strain YCH46 genome encodes two putative FKBH homologues: BF0824 (529 amino acids) and BF1531 (575 amino acids). Curiously, the encoding genes are not located within polyketide biosynthetic gene clusters. Instead they are found within two respective capsular polysaccaharide loci (BF0803-BF0823 and BF1529-BF1554). *B. fragilis* lives closely associated with mucosal surface of the human

colon where it us under constant scrutiny by the host's immune system. Variation in the polysaccharide makeup of the *B. fragilis* outer capsule is important for evasion as well as for opportunistic pathogenesis. Three different strains of *B. fragilis* (638R, NCTC 9343, YCH46) have been sequenced to show that the genomes of all three contain ten distinct capsular polysaccaharide loci. Remarkably, when the ten polysaccharide loci of strain NCTC were compared with those of 638R, it was discovered that none are shared. Strain YCH46 shares one locus with strain NCTC 9343 and a second, different locus with strain 638R. Thus, in total there are 28 unique polysaccharide loci represented among these three strains.

The fact that the BF0824 and BF1531 genes are located within capsular polysaccaharide loci (**Figure 3.2** and **3.3**) suggests that the FKBH protein products are not functioning in the context of a polyketide synthase or a nonribosomal peptide synthase. Indeed, a BLAST search of the deposited *B. fragilis* genomes using polyketide synthase I and II sequences and a nonribosomal peptide synthase sequence as query failed to retrieve homologues. The glyceryl unit produced by BF0824 and BF1531 might be used in the biosynthesis of a novel capsular polysaccharide sugar moiety.
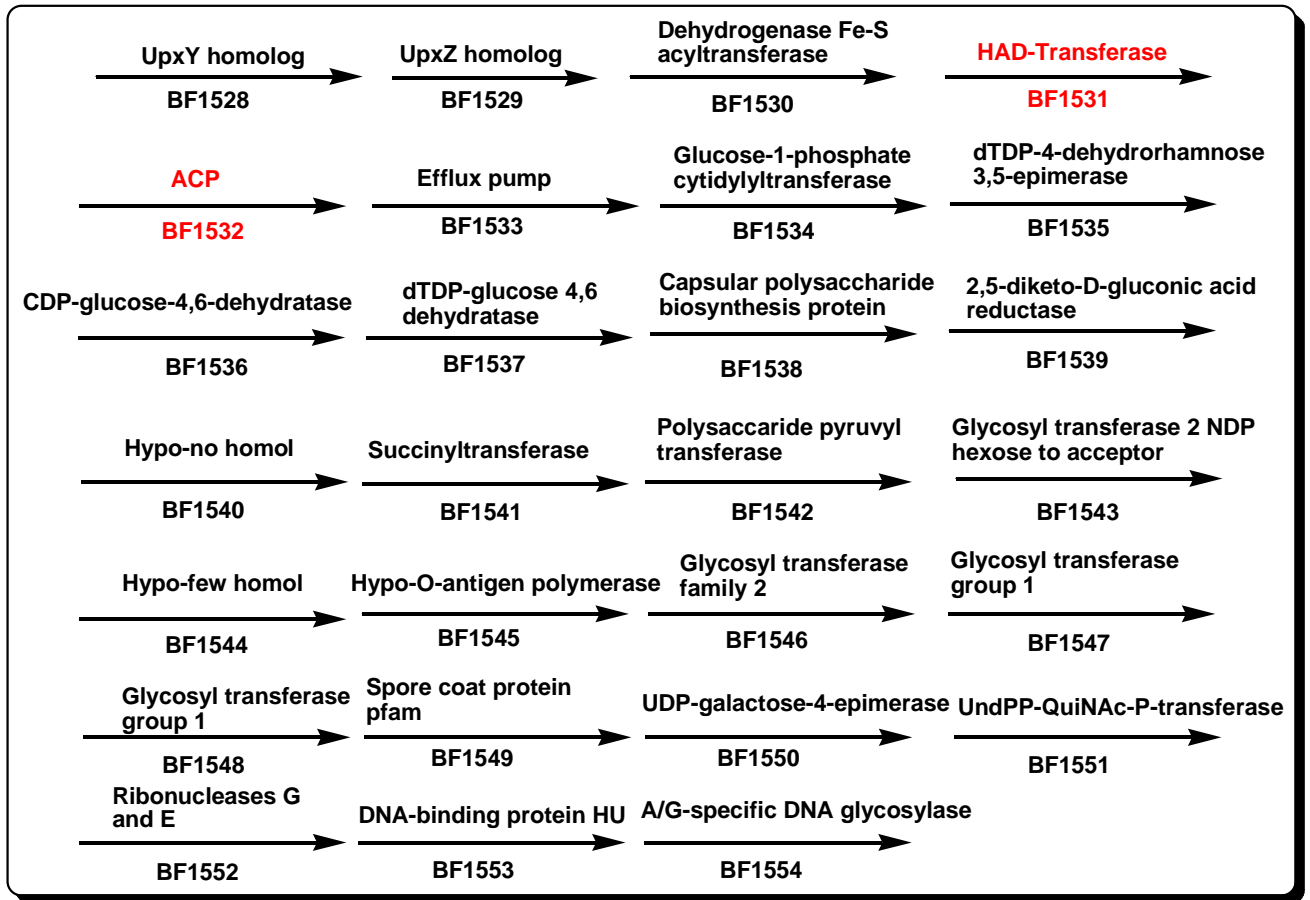
Figure 3.33 The capsular polysaccharide loci that contains the FKBH homolog BF1531 gene. The gene annotations are based on the identity of the closest sequence homolog of known Function
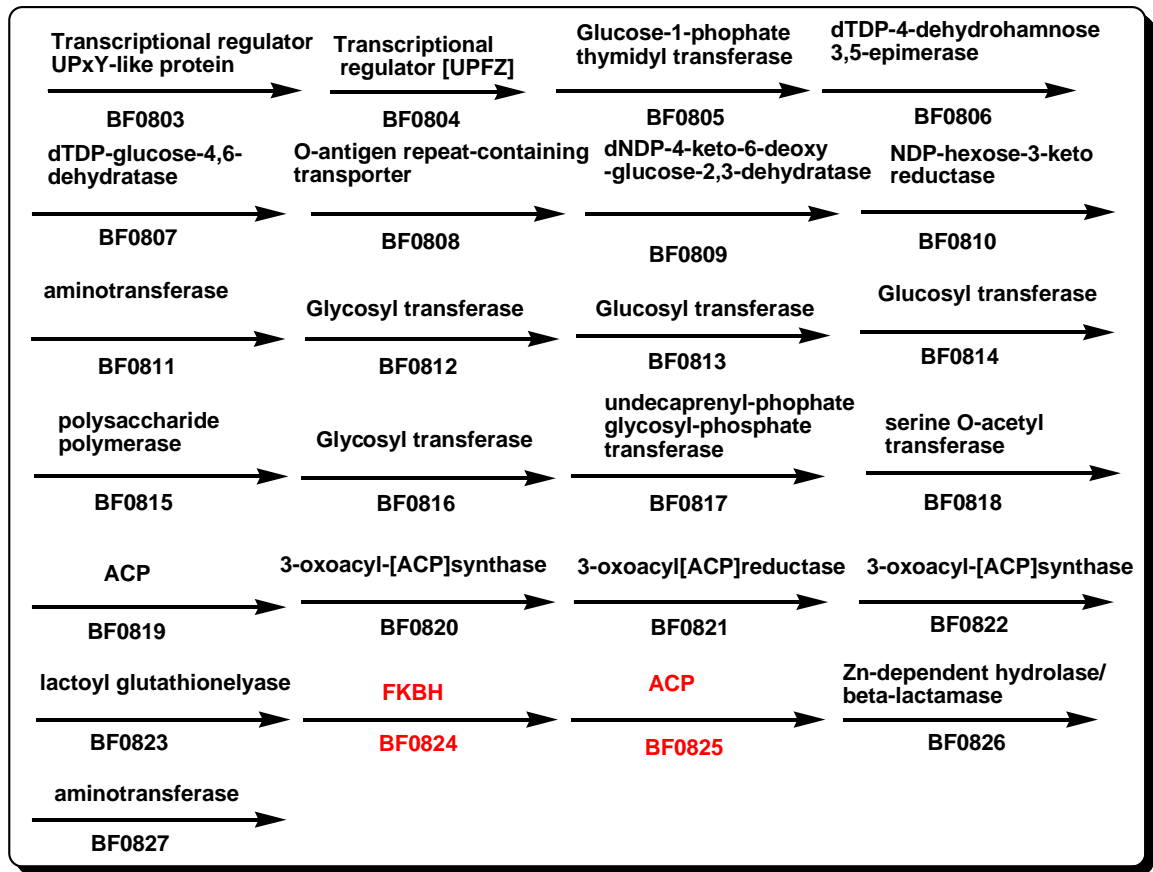
Figure 3.34 The capsular polysaccharide loci that contains the FKBH homolog BF1531 gene. The gene annotations are based on the identity of the closest sequence homolog of known function.

The work reported in this Chapter focuses on the structures and functions of BF0824 and BF1531. The X-ray structure determination of BF1531 was carried out in Professor Steve Almo's laboratory at Albert Einstein School of Medicine. The (attempted) X-ray structure determination of BF0824 was carried out in Professor Karen Allen's laboratory at Boston University. The SAXS (small angle X-ray diffraction) determinations were carried out in collaboration with Professors Karen Allen and Hiro Tsuruta.

3.2 Materials and Methods

3.2.1 General Methods

All chemicals and buffers were purchased from Sigma-Aldrich. The sources of the gene cloning materials are as follows: primers, T4 DNA ligase, restriction enzymes (Invitrogen); *E. coli* BL21 (DE3) competent cells and *Pfu*, Turbo polymerases (Strategene); pET14b, pET23a and pET28a vector kits (Novagen); Qiaprep Spin Miniprep Kit (Qiagen). Genomic DNA (strain YCH46) from *B. fragilis* strain YCH46 was gift from Dr. Masahira Hattori of the University of Tokushima. DEAE Sepharose was from Amersham Biosciences. Butyl and Phenyl-Sepharose resins were purchased from Sigma-Aldrich, whereas the Ni-NTA resin was from Qiangen. Snakeskin pleated dialysis tubing was purchased from Thermo Scientific. SDS-PAGE analysis was performed with a 12% acrylamide running gel and a 4% stacking gel (37.5:1 acrylmide:biacryamide ratio) (BioRad, Hercules, CA). Protein solutions were concentrated using a 10K Amicon Ultra Centrifugal filter (Millipore). The nucleotide sequence of each cloned gene or mutant gene was determined by the Center for Genetics

in Medicine, University of New Mexico. Electro-spray mass-spectrometry (ES-MS) determinations were carried out by the University of the New Mexico Mass Spectrometry Facility. Protein concentrations weredetermined using the Bradford assay kit from Sigma-Aldrich.

3.2.2 Preparation of Wild-Type Recombinant BF0824 (ExPasY #Q64Y51)

To prepare the N-terminal His$_6$-tagged protein the BF0824 gene was amplified by PCR using the genomic DNA (strain YCH46 from *B. fragilis*) and *Pfu* DNA polymerase. Oligonucleotide primers, which contain restriction endonuclease cleavage sites- *BamhI* (GGCGCAACTTGGATCCACATTCTCCAACTC) and *NheI* (GAGTCTTAATACT GCT AGC ATGAAGTATTTTAT) were used in the PCR reactions. The purified PCR product was digested with restriction enzymes *BamhI* and *NheI* and the digested product was ligated into vector pET28a that had been digested with the same enzymes. The ligation product wasused to transform *E. coli* BL21(DE3) competent cells. The confirmed colony of *E. coli* (BL21) cells was used to inoculate 10 mL of LB medium containing 30 µg/mL karamycin at 37 $^{\circ}$C and 250 RPM. The 10 mL culture was then used to inoculate 10 L of fresh LB medium containing 30 µg/mL karamycin and the culture was grown at 25 °C with shaking at 200 RPM. After 6 h (O.D ~0.6 at 600nm)IPTG was added to a final concentration of 1 mM. The mixture was incubated at 20 °C and 200 RPM for 14 h before harvesting the cells by centrifugation at 6500 x g for 15 min. The 10 g of wet cells were suspended in 100 mL of lysis buffer (50 mM NaH$_2$PO$_4$, 300 mM NaCl, 10 mM imidazole; pH 8.0). Thecell suspension was passed through a French press at 1,200 PSIG and then centrifuged at 20000 x g at 4°C for 45 min. The supernatant was loaded onto a

Ni-NTA Agarose column (25 mL, 4°C) that had been pre-equilibrated with the lysis

buffer. The column was eluted with 300 mL of wash buffer (pH 8.0;50 mM

NaH_2PO_4,300 mM NaCl, 20 mM imidazole) and then eluted with 400 mL elution buffer

(pH=8.0;50 mM NaH_2PO_4, 300 mM NaCl, 250 mMimidazole). The protein was dialyzed

for 3 h against three changes of 2.5 L buffer A (pH 7.6; 50 mM Tris, 5 mM Mg^{2+}, 1 mM

DTT). The protein purity was verified by SDS-PAGE analysis, the desired fractions were

combined and concentrated by using a Centricon (10kDa, Pall Filtron) at 4 ^{o}C, The

concentration of protein was measured by using the Bradford method (12). Yield: 25mg

protein/gm wet cell.

A similar procedure was used to prepare C-terminal His_6-tagged BF0824.The

oligonucleotide primers used are: GAGTCTTAATACTGCT AGC ATGAAGTATTT

TAT(NheI site) and ATTTTCTCTTCCCTCGAGGTTCATATTTATTTTAT(XhoI site).

Following digestion, the gene was ligated into the pET 23a vector. Ampicilin was used in

the growth media in place of the karamycin. All other steps were the same as described

above. Yield: 15 mg protein/gm wet cell.

Two different cloning strategies were tried in the cloning of the native (i.e., no

His_6 tag) BT0824. Firstly, the oligonucleotide primers that contain NheI and BamH I

restriction sites (*vide supra*) were used in the PCR reaction and the digested product was

ligated to the linear pET 23a vector, which was used to transform the *E. coli* (BL21) cells

as previously described. Following cell lysis, the supernatant was loaded onto a 40 cm ×

5 cm DEAE Sepharose column at 4 ^{o}C, which was eluted with a 2 L linear gradient of

KCl (0 to 0.5 M) in buffer A. The column fractions were analyzed by SDS-PAGE. The

90

desired protein fractions were combined, adjusted to 20% $(NH_4)_2SO_4$ (W/V), and loaded onto a 18 cm × 3 cm Butyl-Sepharose column that had been pre-equilibrated with buffer A containing 20% $(NH_4)_2SO_4$. The column was eluted at 4 ^{o}C first with a 0.5 L linear gradient of $(NH_4)_2SO_4$(20% to 15%) in buffer A, and then with 0.5 L buffer A containing 15% $(NH_4)_2SO_4$. The column fractions were analyzed by SDS-PAGE. The desired protein fractions were combined and concentrated at 4 °C using a 10K Amicon Ultra Centrifugal filter (Millipore). The concentrated protein fractions were loaded onto a Sephadex G-75 column and then the column was eluted with buffer A. The column fractions were analyzed by SDS-PAGE. The desired protein fractions were concentrated and stored at -80°C. Yield: 12mg protein/gm wet cell.

For the second cloning method the IMPACT (Intein Mediated Purification with an affinity Chintin-binding Tag) system was employed. This system (vector pTXB1 (NEB#N6707) with NdeI and SapI restriction sites) utilizes the inducible self-cleavage activity of protein splicing elements to cleave the affinity tag from recombinant protein. The base pairs for the two Nde I restriction sites were substituted using the following primers: GTTATGATGACATTT<u>CGTATG</u>TCCCTTTGG,CCAAAGGGA<u>CATACG</u>AAATGTCATCATAAC;GAAGATGGGGCAAAAG<u>CGTATG</u>TACAATG,CATTGTA<u>CATACG</u>CTTTTGCC CCATCTTC.The original stop codon was replaced using the Primers ATAAATATGAAC<u>TGC</u>TATGGAAGAGAAAAT,ATTTTCTCTTCCATA<u>GCA</u>GTTCATATTTAT. Finally, the oligonucleotide primers containing the NdeI (GTCTTAATACTGCT<u>CATATG</u>AAGTA TTTTATT) and SapI (CCAAT ATTTTC T<u>GC</u>TC<u>TTC</u> AGCAGTTCATATTT) restriction sites were used to clone the modified

gene into pTXB vector digested with the corresponding restriction enzymes. The recombinant vector was used to transform*E.coli* ER2566 (NEB# C2566) cells for culturing as previously described. The cell pellet was suspended in200 mL of column buffer (20 mM Tris-HCl pH 8.5, 500 mM NaCl, 1 mM EDTA).The cell lysate-derived supernatant was loaded onto a chitin bead (NEB#S6651) column(10 mL) at 4°C. The column was washed with 20 bed volumes of the column buffer before applying the on-column cleavage buffer (50 mM DTT, 20 mM Tris-HCl pH 8.5, 500 mM NaCl, 1 mM EDTA). Following 24 h at 4°C the native protein was eluted from the column with 50 mL of column buffer. The eluant was concentrated and dialyzed at 4 °C against 3 changes of buffer A (pH 7.6; 50 mM Tris-HCl, 5 mM MgCl$_2$, 1 mM DTT). Yield: 20mg protein/gm wet cell.

3.2.3 Preparation of Wild-Type Recombinant BF1531 (ExPasY # Q64W45)

All attempts to clone the *B. fragilis* (strain YCH46) BF1531 gene for over-expression in *E. coli* using a genomic DNA/PCR-based strategy failed. Therefore, a codon-optimized gene was synthesized. The protein production group at Eli Lilley (under Steve Sauder and Steve Burley) synthesized 5 constructs: aa 2-377, 2-575, 17-377, 17-575 and 216-377.These gene constructs were expressed in an *E. coli* Met auxotroph to produce the Seleno-Met protein products for crystal structure determination, which was carried out in Steve Almo's laboratory. The gene constructs were also sent to our lab for kinetic analysis. The full length construct in the pET30 vector was used to transform*E. coli*BL21(DE3) codon+RIL cells. The confirmed colony was used to inoculate 10 mL of

LB medium containing 30 µg/mL karamycin and 20 µg/mL chloramphenicol at 37 °C. The 10 mL culture was then used to inoculate 6 L of fresh LB medium containing 30 µg/ml karamycin and 20 µg/ml chloramphenicol. The culture was grown at 25 °C with shaking at 200 RPM. After ~6 h (O.D ~0.6 at 600nm), IPTG was added to a final concentration of 1 mM. The mixture was incubated at 20 °C with shaking at 200 RPM for 14 h, before harvesting the cells by centrifugation at 6500 x g for 10 min. The 20 g of wet cells were suspended in 200 mL of lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole; pH 8.0). The cell suspension was passed through a French press at 1,200 PSIG and then centrifuged at 20000 x g at 4 °C for 45 min. The supernatant was loaded onto a Ni-NTA Agarose column (10 mL), which had been pre-equilibrated with the lysis buffer. The column was eluted with 150 mL of wash buffer (pH 8.0; 50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole) and then eluted with 400 mL elution buffer (pH 8.0; 50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole). The eluant wasanalyzed by SDS-PAGE and the desired fractions were combined and concentrated by using a Centricon (10kDa,Pall Filtron) at 4°C. Finally, the protein was dialyzed at 4 °C against three changes of 2.0L buffer A (pH 7.6; 50mM Tris, 5mM $MgCl_2$, 1 mM DTT). The final concentration of protein was measured by using the Bradford method. Yield: 10mg protein/gm wet cell.

3.2.4 Preparation of BF0824 Site Directed Mutants and Truncation Mutants

Several gene constructs that encode BF0824 minus one or two of the three domains were prepared for the purpose of structure determination. Truc1 (1-330) was

prepared using two primers with NheI(GAGTCTTAATACT<u>GCTAGC</u>ATGAAG TATTTTAT) and BamHI (GTTCCGTCTTTTTC<u>GGATCC</u>TCTTACGTGATGG) restriction sites; Truc 2 (331-529) was prepared using two primers with NheI (GGTGATTAAT<u>GCTAGC</u>TTGGCAGGGGC) and BamH I (GGCGCAACTT<u>GGATC C</u>ACATTCTCCAACT C) restriction sites and Truc 3 (161-529) was prepared using two primers with NheI (GGCTTTTCGCAAA<u>GCTAGC</u>A CTTGTAAAACAAG) and BamH I (GGCGCAACTT<u>GGATCC</u>ACATTCTCCAACTC) restriction sites. The PCR products were cloned into a pET 28a vector (adds an N-terminal His_6tag). The protein preparation procedures used were the same as those used for the His_6 tagged full length BF0824 protein.

Because the full length BF0824 was found to undergo cleavage at the central-C-terminal linker during crystallization amino acid replacements were made in hopes of stabilizing the linker. The K350K351was replaced with E351V352 and K334K335 was replaced with 334A335L. Two pairs of oligonucleotide primers were used to make the quadruple–mutation, (CGGAAGA GGAT<u>GCACTA</u>AAGACGGAAC AATATAAAGC, GCTTTATATTGTTCCGTCTT<u>TAGTGC</u>AATCCTCTTCCG;GCAGGACACAGGAA<u>G AAGTG</u>AAATTTACAGATCTCGGTGC,GCACCGAGATCTGTAAATTT<u>CACTTC</u>TT CCTGTGTCCTGC). The protein preparation procedures used were the same as those used for the His_6 tagged full length BF0824 protein.

Three key catalytic residues (Asp 178, Asp180 and Cys488) were replaced with Ala. The primers used for generation of the D178A, Asp180 and Cys488 mutanta are GTTTGGTTCTT<u>GCT</u>TTGGATAATACACTGTGGAG and CTCCACAGTGTATTA

TCCAA<u>AGC</u>AAGAACCAAAC;GTTTGGTTCTTGATTTG<u>GCT</u>AATACACTGTGGA

G and CTCCACAGTGTATT<u>AGC</u>CAAATCAACAACCAAAC;GATTCATTGTTGT

TAAGC<u>TCT</u>AGGATCTTAGG and CCTAAGATCCT<u>AGA</u>GCTTAACAACAATG

AATC, respectively. The protein preparation procedures used were the same as those

used for the His$_6$ tagged full length wild-type BF0824 protein.

3.2.5 Preparation of the Apo and Holo Acyl Carrier Proteins BF0825 (ExPasY #Q64Y50) and BF0819 (ExPasY #Q64Y56)

The BF0825 (Q64Y50) gene was cloned into the pET23a vectorusing the primers

TAAATATGAACT<u>CATATG</u>GAAGAGAAAATATTGGAG (Nde I site) and CAAA

ACGAACAACCTC<u>CTCGAG</u>TTTCCCCTGAAGTATC (Xho I site) and the

methodology used for preparation of the BF0824-pET23a construct (for preparation of

the C-terminal His6-tagged protein product). The same cell culture (except that

amphicillin was used instead of kanamycin) and protein purification steps were followed.

Yield: 8 mg protein/gm wet cell. ES-MS analysis identified the protein product as the apo

BF0285. Theoretical mass: 9388 Da   vs observed mass 9415 Da.

The BF0819 (Q64Y56) gene was cloned into the pET 14b vector using the

primers TTTAAAATAATAAT<u>CATATG</u>GAACTTAAAAC (Nde I site) and CAGG

GGTTT<u>GGATCC</u>AAATATCAGACATC (Bamh I site). The N-terminal His6-tagged

protein was prepared using the same methodology used in the preparation of the His$_6$-

tagged BF0824. Yield: 5mg protein/gm wet cell. ES-MS analysis identified the protein

product as the apo BF0285. Theoretical mass: 9842 Da   vs observed mass 9870 Da.

The holo-BF0825 was prepared as follows. To a 1 mL plastic tube 10 μL of apo-BF0825 (8.53 mM in 50 mM Tris-HCl pH 7.6), 2.5 μL BF1558 (0.65 mM in50 mM Tris-HCl pH 7.6) and 70 μl CoA (20 mM) were added. The resulting solution was incubated at 25 °C for 10 min after which 50 μL of 10 % formic acid was added. ES-MS analysis identified the protein product as the apo BF0285. Theoretical mass: 9727 Da   vs observed mass 9755 Da.

The holo-BF0819 was prepared as follows. To a 1 mL plastic tube 30 μL of apo-BF0819 (1.52 mM in 50 mM Tris-HCl pH 7.6), 1.5 μL BF1558 (0.65 mMin 50 mM Tris-HCl pH 7.6) and 7 μl CoA (20 mM) were added. The resulting solution was incubated at 25 °C for 10 min after which 50 μL of 10 % formic acid was added. ES-MS analysis identified the protein product as the apo BF0285. Theoretical mass: 10181 Da   vs observed mass 10210 Da.

3.2.6 Preparation of the Phosphopantethienyl Transferase BF1558 (ExPasY # Q64W18)

The BF1558 gene was cloned into the vector pET28a for expression in *E. coli* BL21 (DE3) cells using the primers CCTGATATCAGCTTTT<u>CATATG</u>GCTTTGTTACGG) (NdeI site) and TAATAGGATACGCT<u>CTCGAG</u>TTGTATTTTCTTCAT (XhoI site) in the PCR reactions. The purified PCR product was digested with restriction enzymes *NdeI* and *XhoI* and then the digested product was ligated to vector pET28a which had been digested with same enzymes. The ligation product was used to transform *E. coli* BL21 (DE3) competent cells (Stratagene). The confirmed colony of *E.coli* (BL21)

cells was used to inoculate 10mL of LB medium containing 30 µg/ml karamycin at 37°C. The 10mL culture was then used to inoculate 4L of fresh LB medium containing 30 µg/ml karamycin and the culture was grown at 25 °C with shaking at 200 RPM. After 6 h (O.D ~0.6 at 600nm) IPTG was added to a final concentration of 1mM. The mixture was incubated at 20 °C and 200 RPM for 14 h, before harvesting the cells by centrifugation at 6500 x g for 10min. The 15 g of wet cells were suspended in 150 mL of lysis buffer (50mM NaH$_2$PO$_4$, 300 mM NaCl, 10mM imidazole; pH 8.0). The cell suspension was passed through a French press at 1,200 PSIG and then centrifuged at 20000 x g at 4°C for 45min. The supernatant was loaded onto a Ni-NTA Agarose column (10 ml), which had been pre-equilibrated with the lysis buffer. The column was eluted with 300 mL of wash buffer (pH 8.0; 50mM NaH$_2$PO$_4$, 300mM NaCl, 20mM imidazole) and then eluted with 400mL of elution buffer (pH 8.0; 50mM NaH$_2$PO$_4$, 300mM NaCl, 250mM imidazole). The desired fractions were combined and concentrated at 4°C by using a Centricon (10 kDa, Pall Filtron). The protein was dialyzed at 4°C against three changes of 2.5L buffer A (pH 7.6; 50mM Tris-HCl, 5mM MgCl$_2$, 1mM DTT). The protein purity was verified by SDS-PAGE analysis. The concentration of protein was measured by using the Bradford method. Yield: 12mg protein/g wet cell.

3.2.7 Preparation ofthe Putative 3-Oxyoacyl-ACP SynthaseBF0820 (ExPasY #Q64Y55)and Attempted Preparation of the Putative 3-Oxyoacyl-ACP Synthase BF0822 (ExPasY # Q64Y53).

The BF0820 (Q64Y55) gene was cloned into the PET-28a vector using the primers GTTGTAAGAACAAAGGA<u>CATATG</u>GCTTTTATAG (Nde I site) and ATCAA

AGAAAAA<u>GGATCC</u>TATATTTCCATAAAG (Bamh I site). The N-terminal His$_6$-tagged protein was prepared using the same methodology used in the preparation of the His$_6$-tagged BF0824. Yield: 15 mg protein/g wet cell.

The BF0822 (Q64Y53) gene was successfully cloned however it failed to express in *E.coli* cell lines BL21(DE3), BL21 codon-RIL and NEB. Several attempts were tried to express this gene. The two primers with Nhe I (CTACAGGG<u>GC TAGC</u>TTGGTAATAGATGGAG) and HindIII (GAAATTTAAAATCA<u>GGATCC</u>ATA ACTCAATATTC) restriction site were designed to amplify the gene by PCR. The gene product was inserted into different vectors to test expression. The pET28a vector was used for production of the N-terminal His$_6$-tagged protein. The pET23a vector was used for production of the C-terminal His$_6$-tagged and taggless protein. The chaperone plasmid set (Takara, Code 3340, pGro7) was used to aid the folding of the target protein. The BF0822 gene was inserted into a pCold II DNA vector and co-expressed with the chaperones-GroEL and GroES. The pETDuetvector (from Novagen) was utilized to express the BF0820 and BF0822 genes on the same vector. Finally the codon-optimized gene was synthesized at Eli Lilleyand inserted into an expression vector. None of the approaches resulted in BF0822 gene expression.

3.2.7 Preparation of the Putative Serine O-Acetyl Transferase BF0818 (ExPasY # Q64Y57)

The BF0818 (Q64Y57) gene was cloned into the pET23a vector using the primers ATGGAAATAATTAA<u>CATATG</u>AAGAATATAGC (Nde I site) and TTTTCAATA

AATG TCTCGAGTTCCATTTTATTATT (Xho I site) and the methodology used for preparation of the BF0824-pET23a construct (for preparation of the C-terminal His6-tagged protein product). The same cell culture and protein purification steps used in the preparation of the His$_6$-tagged BF0824 were followed. Yield: 10 mg protein/g wet cell.

3.2.8 Preparation of the Putative Malonyl-CoA: Acyl Carrier Protein Transacylase BF2258 (ExPasY # Q64U24).

BF2258 (Q64U24) gene was cloned into the pET23a vector using the primers TTAAAAAACAAACATATGAAAGCATTTGTATTC (Nde I site) and GAGTGGTCT ATCA ATACTCGAGTGCTATTCC (Xho I site) and the methodology used for preparation of the BF0824-pET23a construct (for preparation of the C-terminal His$_6$-tagged protein product). The same cell culture and protein purification steps used in the preparation of the His$_6$-tagged BF0824 were followed. Yield: 12 mg protein/g wet cell.

3.2.9 BF0824 (ExPasY #Q64Y51) Molecular Mass Determination

The theoretical subunit molecular mass of N-terminal His$_6$-tagged recombinant BF0824 was calculated by using the amino acid composition, derived from the gene sequence, and the ExPASy Molecular Biology Server program Compute pI/MW. The subunit size of recombinant BT2127 was determined by SDS-PAGE analysis, which included the molecular weight standards from New England Biolabs Inc. The subunit mass was determined by MS-ES mass spectrometry. The native molecular was estimated by FPLC gel filtration column chromatography against protein standards (13.7-220 kDa

from GE Healthcare). The 1.6 cm x 60 cm Sephacryl S-200HR column (GE Healthcare) was eluted at 4 ^{o}C with bufferD (50 mM HEPES, 100 mM NaCl (pH 7.5)) at a flow rate of 1 mL/min. The BT2127 molecular weight was derived from the measured elution volume by extrapolation of the plot of the elution volume of the molecular weight standard *versus* log molecular weight. BT2127 native mass was also analyzed at the HHMI Biopolymer/Keck Foundation Biotechnology Resource Laboratory at Yale University by size exclusion chromatography coupled with on-line laser scattering, refractive index, and ultraviolet detection.

3.2.10 BF1558 (ExPasY # Q64W18) Activity Assay

<u>BF0825.</u> A reaction solution initially containing 1 mM apo-BF0825, 21 μM BF1558, 16 mM CoA, 1 mM DTT, 5 mM MgCl$_2$ and 50 mM Tris-HCl (pH 7.6) was incubated at 25 ^{o}C for 10 min and then 50 μL of 10% formic acid were added to terminate the reaction. The holo-BF0825 was detected by ES-MS analysis (m/z at 9, 755 Da).

<u>BF0819.</u> A reaction solution initially containing 1.2 mM apo-BF0819, 24.5 μM BF1558, 3.5 mMCoA ,1 mM DTT, 5 mM MgCl$_2$ and 50 mM Tris-HCl (pH 7.6) was incubated at 25 ^{o}C for 10 min and then 50 μL of 10% formic acid were added to terminate the reaction. The holo-BF0819 was detected by ES-MS analysis (m/z at 10, 210 Da).

3.2.11 Reaction of BF1531 (ExPasY # Q64W45) with 1,3-Diphosphoglycerate

A reaction solution initially containing 300 μM C-terminal His$_6$-tagged BF1531, 1.8 mM ATP, 1.8 mM 3-phosphoglycerate, 20 units of phosphoglyceratekinase, 1 mM DTT, 50 mM MgCl$_2$ and 50 mM Tris-HCl (pH 7.6) was incubated at 25 ^{o}C for 10 min

and then 50 μL of 10% formic acid were added to terminate the reaction. The glyceryl-BF1531 was confirmed by ES-MS analysis (m/z at 67, 359 Da).

3.2.12 Preparation of Glyceryl-BF0824 (ExPasY # Q64Y51) and Reaction with Holo-ACP BF0825 and Holo-ACP BF0819.

To prepare glyceryl-BF0824 400 μL of N-terminal His$_6$-tagged BF0824 (390 μM in 50 mM Tris-HCl pH 7.6), 33 μL 3-phosphoglycerate (28 mM), 66 μL ATP (14 mM), 30 μL MgCl$_2$ (800 mM) and 5 μL 3-diphosphoglycerate-kinase (20 units) were combined and the resulting solution was incubated at 25°C for 30 min. The solution was clarified by centrifugation at 14000 rpm for two min. The supernatant was stored at -20°Cwhile the product, glyceryl-BT0824 was confirmed by LC-MS. The frozen was reaction solution was thawed and a 1 mL aliquot containing ~0.29mM glyceryl-BF0824 was mixed with 200 μL Holo-0825 (1.48mM). The resulting solution was incubated at 37 ^{o}C for 5 min. The reaction product was validated by ES-MS analysis.

The transfer of the glyceryl moiety from glyceryl-BF0824 to the holo-BF0825 or holo-BF0819 was tested by incubating the reaction mixture comprised 22 μM holo -BF0825 or holo-BF0819 and 74 μM glyceryl-BF0824 in 50 mM Tris-HCl pH 7.6 at 25°C or 37 ^{o}C for 20 min. The reaction solutions were directly subjected to ES-MS analysis.

3.2.13 Reaction of Glyceryl-BF0825 with BF0820

Glyceryl-BF0825 (100 μL 0.24 mM in 50 mM Tris-HCl, pH 7.6) was incubated with the putative 3-oxoacyl-[acyl carrier protein] synthase BF0820 (39441Da, 7 μL, 1.0

mM in 50 mM Tris-HCl, pH 7.6) at 37 oC for 30 min. The solution was clarified by centrifugation and then directly subjected to ES-MS analysis. The Glyceryl-BF0820 was observed (39530.6Da) with a slow rate.

3.2.14 Reaction of Glyceryl-BF0825 with Coenzyme Aand BF0818

Glyceryl-BF0825 (50 µL 0.24 mM in Tris pH 7.6 buffer) was incubated with the putative Serine-O-Acyl Transferase BF0818 (24268Da, 2 µL, 0.3 mM in Tris pH 7.6) and Coenzyme A (2µL, 4mM in Tris pH 7.6 buffer)at 25 ^{o}C for 30 min. The solution was frozen and then directly subjected to ES-MS analysis. No detectable glyceryl-CoA was found (m/z at 766 Da for CoA).

3.2.15 Reaction of Holo-BF0825 with Various Acyl-CoAs Catalyzed by BF2258.

The relative rates of BF2258 catalyzed acyl-transfer to the ACP holo-BF0825 were measured by carrying out an HPLC-based fixed time assay. The initial reaction solutions contained 500 µM holo-BF0825, 500µM acyl-CoA, 50µM BF2258, 5 mM MgCl$_2$, 1 mM DTT and 50 mM Tris-HCl (pH 7.6). The reaction solutions were incubated at 25 ^{o}C for 15 min before chromatographing them on a reversed-phased C18 column (100 x 4.6 mm; ThermoScientific ODS HYPERSIL) using a linear gradient of solvent A (100 mM Na$_3$PO$_4$ and 75 mM sodium acetate; pH 4.6) and solvent B (70% A in methanol) and a flow-rate at 1.5 ml/min. The acyl-BF0825 products were verified by ES-MS analysis.

102

3.2.16 Structure Determination of Crystallization the C-terminal Domain BF1531 Truncation Mutant (Residues 3-381).

This work was carried out in Steve Almo's laboratory at Albert Einstein School of Medicine. PurifiedC-terminal His_6-tagged Se-modified BF1531 (residues 3-381) was crystallized from 35% pentaerythritol propoxylate, 0.05 M HEPES, 0.2 M potassium chloride, pH 7.5, via the hanging drop vapor diffusion method, at 293.0 K. Diffraction data were collected at 100 K on NSLS beamline X4A using an ADSC QUANTUM 4 CCD area detector, located in Brookheaven National Laboratory and processed using the DENZO [13] software suite. The His_6-tagged Se-modified BF1531 crystallized in space group P21212 with unit cell dimensions a =75.74 Å, b = 126 Å and c = 41.62 Å. Data were collected to 1.71 Å resolution. Data collection, refinement, and final model statistics are reported in **Table 3.1**. Phases were determined by the SAD method using PHENIX [14] yielding a solution with one enzyme protomer in the asymmetric unit. Alternating rounds of manual rebuilding were performed using the molecular graphics program COOT[15]followed by minimization and simulated annealing in PHENIX.

The final model of the BF1531 partial structure, included residues 3-376 in the protomer, 275 water molecules, one tungstate ion, and one Mg^{2+} ion, was refined to 1.71 Å resolution with a R_{work} of 18.3% and R_{free} of 20.9%. Analysis of the Ramachandran plot showed that 97.5% of residues fall in the most favored regions with 2.5% in the additionally allowed and generously allowed regions. Figures were produced using Molscript[16], and were rendered in POVRAY.

3.2.17 BF0824 Small Angle X-ray Diffraction (SAXS) Data Collection and Processing

This work was carried out by Dan Saltzberg (Karen Allen laboratory) in collaboration with the Hiro Tsuruta laboratory. SAXS data were collected at beam line 4-2 at the Stanford Synchrotron Radiation Lightsource (Palo Alto, CA). Concentrations of 5, 10 and 20 mg/mL BF0824 in 50 mM Tris-HCl were each centrifuged at 14,000 g for 5 min before placing on the sample holder block maintained at 10 °C. Samples were exposed to radiation for 15 exposures, each 1 s in duration while passing through a capillary flow cell at a rate of 1μL/s and the scattering data collected on a MarCCD165 at a distance of 1.5mm from the sample flow cell. Two buffer blanks were collected using the same procedure. Each exposure was reduced by radially averaging each image to produce the scattering intensity as a function of the magnitude of the momentum transfer vector ($q = 4\pi*\sin\theta/\lambda$). Outlier exposures, possibly a result of bubbles in the beam path, were indicated computationally and then manually inspected. Those exposures clearly different from the majority were discarded.

1D scattering profiles for each concentration of protein were constructed by subtracting (1 – volume fraction protein) * 100% of the buffer scattering intensity from the averaged protein sample intensity. A Guinier plot was constructed to look for signs of aggregation. At extremely low scattering angles, evidence of interparticle interactions were observed at higher concentrations. Low-angle data that did not conform to the Guinier approximation was discarded. The 1D scattering profiles at 20, 10, and 5 mg/mL FkbH were extrapolated to zero concentration using the PRIMUS data analysis package

of Svergun [17]. The radius of gyration was estimated using the Guinier approximation for points 9 through 19 ($0.017 < q < 0.026$) and found to be 51.6 +/- 0.06 angstroms. The pairwise distance probability distribution (p(r))curve was constructed with GNOM [17]and found to reach zero at 185 Å. The R_g as calculated from the p(r) curve is 51.3 Å.

Ab-initio molecular envelopes for the full length BF0824 were calculated from the zero concentration extrapolation described above at a resolution range from $0.017 < q < 0.311$ (1/370 to 1/20 Angstroms). The program DAMMIN [18]was used using default values for simulated annealing to create bead models fit to the extrapolated dataset. Two models were created, one enforcing no symmetry (P1) and one enforcing a two-fold rotation axis (P2) because of the dimer assumed from the gel filtration solution data and the packing in the crystal structure of residues 3-376 of the ortholog BF1531. Forty-six independent DAMMIN simulations were performed with each symmetry and the resulting bead models superimposed, averaged and filtered using the programs SUPCOMB, DAMAVER and DAMFILT [19-20]. The final bead models for both P1 and P2 symmetry were converted to surface maps at 20Å resolution using pdb2vol from the SITUS software suite.

"Threading" homology models of the two N-terminal domains of BF0824 FkbH were based on the crystal structure of homologue BF1531 (PDB ID: 3NVB, XX% similarity) and constructed using MODELLER [21]. A set of homology models for the C-terminal transferase domain of FkbH was produced by the phyre2 protein fold recognition algorithm[22] using the final 169 residues of BF0824 (residues 360-529). The top 20 templates used in the modeling were all acetyl-transferases of the GNAT

superfamily. The homology model based on the structure with the highest sequence identity, 22%, was selected (PDB 2OH1) allowing 90% of the residues to be modeled at >90% confidence, with the low confidence areas at the N- and C- termini of the modeled domain.

The homology model was fitted into the SAXS envelopes using COLORES from the SITUS software suite. The top six fits were recorded.

3.3 Results and Discussion

3.3.1 BF0824 (ExPasY #Q64Y51) and BF1531 (ExPasY # Q64W45) X-ray Structure Analysis

BF0824 (529 amino acids) and BF1531 (575 amino acids) share 45 % sequence identity. An InterProScan (http://www.ebi.ac.uk/Tools/services/web_iprscan/) analysis of the two sequences predicts three domains: the N-terminal domain (residues 1-216 in BF1531) belongs to the GDSL hydrolase superfamily, the central domain (residues 220-372 in BF1531) belongs to the HADSF type C0 and a C-terminal domain (residues 375-575 I BF1531) belongs to the acyl-CoA N-acyl transferase family. The X-ray structure determinations of full length and truncated BF0824 and BF1531 constructs were pursued in the Karen Allen (BF0824) and Steve Almo (BF1531) laboratories. I carried out the BF0824 genetic engineering and protein preparations for the BF0824 X-ray structure determinations. I also provided full length BF0824 for the SAXS experiments.

BF0824. His6-tagged BF0824 was found having a mass of 63,685 Da compared to which had the theoretical mass of 63675 Da. The SDS-PAGE analysis gave an estimated subunit mass of 63 kDa. The native mass measured by using molecular size gel filtration chromatography is ~116 kDa consistent with the data from the Keck Foundation Biotechnology Resource Laboratory ~125 kDa. These results indicate that protein BF0824 exist as a homodimer. The mass for the three truncated BF0824: Truc 1(1-330), Truc 2 (331-529) and Truc 3 (161-529AA) were determined to confirm their structures: theoretical mass of Truc 1 is 40971 Da observed mass is 40971 Da; the theoretical mass of Truc 2 is 25052Da and the observed mass is 25184Da; the theoretical mass of Truc 3 is 44442 Da and the observed mass is 44482 Da.

Full-length BF0824 was crystallized and an X-ray data set was collected. SDS-PAGE analysis of the crystalline protein revealed that the protein had undergone cleavage and that only the truncated protein had crystallized. The C-terminal domain had been lost. Whereas the central domain could be resolved by molecular replacement using a HADSF homologue structure the N-terminal domain could not be resolved by molecular replacement. The truncation constructs Truc1 (1-330) Truc 2 (331-529) Truc 3 (161-529) failed to crystallize. The linker mutant (334A335L/E351V352) designed to reduce the rate of C-terminal domain loss proved to undergo linker cleavage albeit at a reduced rate.

BF01531. Five BF01531 constructs Truc1 (2-377), Truc2 (2-575), Truc3 (17-377), Truc4 (17-575) and Truc5 (216-377). Only Truc1 and Truc3 crystallized. The structure determination was carried out on Truc2 which contains the N-terminal and central

domain. The crystallization and refinement statistics are provided in **Table 3.1**. The structure was refined to 1.7 Å resolution (pdb code: 3NVB). The structure includes one Mg^{2+} and one tungstate ion, which occupy the cofactor and phosphate binding sites, respectively, found in all HAD phosphotransferases.

The structure is composed of two separate domains connected by one linker region (**Figure 3.4**). The N-terminal domain belongs to the SGNH hydrolase, or GDSL hydrolase superfamily based on DALI [23] search results, which has an unknown function in BF1531. The most similar structure to the N-terminal domain identified

Table 3.10. Crystallographic Data Collection and Refinement Statistics

Data Collection Statistics	
Resolution (Å)	37.87-1.71(1.75-1.71)
X-ray Source	SYNCHROTRON
Wavelength (Å)	0.97915
Space group	P21212
Cell dimension (Å)	a=75.74 b=126 c= 41.62
Reflections Observed (unique)	42477 (xxx)
Completeness (%)	xxx (xxx)
R_{merge}[a] (%)	xxx (xxx)
I/σ (I)	xxx (xxx)
Redundancy	xxx (xxx)
Refinement Statistics	
No. of protein residues/water atoms per asu	374/270
No. of other ligands	2
Number of reflections (work/free)	42477/2144
R_{work}/R_{free} (%)	18.4/20.9
Resolution (Å)	37.87-1.71
Average B-factor (Å^2)	22.1
Protein	21.6
Manganese /tungstate	22.1
Water	27.7
Root mean square deviation	
Bond length (Å)	0.008
Bond angle (°)	1.140

a. $R_{merge} = \sum_{hkl}\sum_i | I_{hkl, i} - <I_{hkl}> | / \sum_{hkl}\sum_i | I_{hkl, i} |$, where $<I_{hkl}>$ is the mean intensity of the multiple $I_{hkl, i}$ observations for symmetry-related reflections. Data for the highest resolution shell is in parenthesis by DALI is a lipase/acylhydrolase from*Enterococcus faecalis* (PDB 1YZF) with rmsd 2.5 Å and 13 % sequence identity The C-terminal domain (central domain in full-length protein) is identified as C0 class of HAD members (**Figure 1A**) [24] which confers a phosphohydrolase activity. The HAD domain shares similar active-site residues with other HAD members.
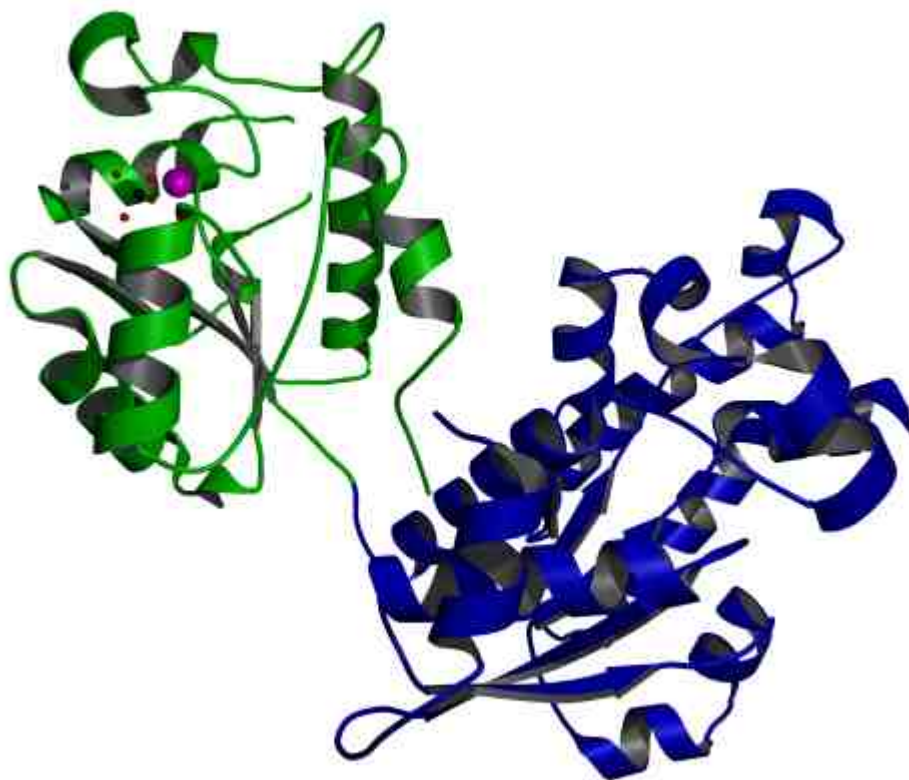
Figure 3.35. Ribbon diagram of the BF1531 partial structure consisting of a HAD domain (green) and GDSL hydrolase domain (blue). The Mg^{2+} cofactor is shown as a magenta sphere.

The BF1531 HAD domain has the typical active-site residues seen in family members: two Asp residues (D226 and D228, the nucleophile and general acid/base catalyst, respectively) from motif 1, a Ser residue (S277) from motif 2, a Lys residue

(K311) from motif 3 (which serve to position the transferring phosphoryl group and nucleophile) and one Asp residue (D334) from motif 4, which bind the Mg^{2+} cofactor.
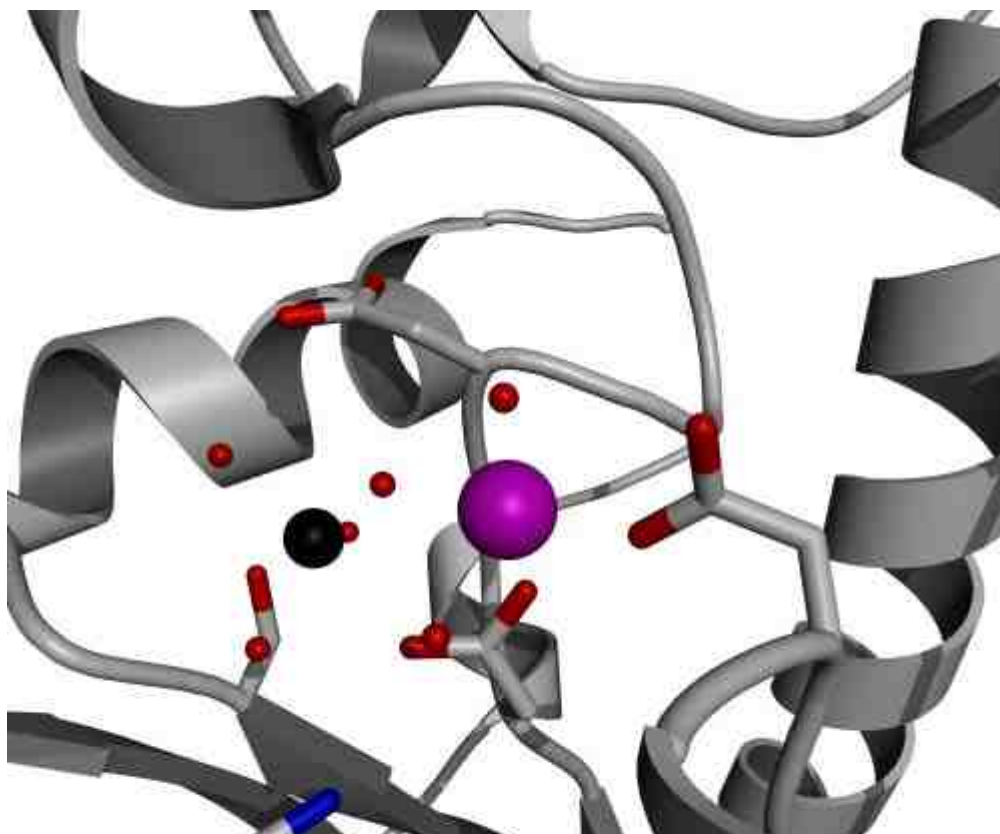
Figure 3.36. Zoom-in on the active site of the HAD domain. Red spheres are water molecules, the magenta sphere is Mg^{2+} and the black sphere is tungstate.

In the HAD domain active site, there is one magnesium ion and one tugstate ion. The Mg^{2+} is coordinated by the side-chain of D226, backbone carbonyl group of D228, the side-chain of D334, two water molecules (one of which is positioned by the side chain of motif 1 D226 and the other by the side chain of motif 4 residue D335) and one

oxygen atom from the tungstate ion. The tungstate ion is 2.5 Å from the Asp nucleophile carboxylate.
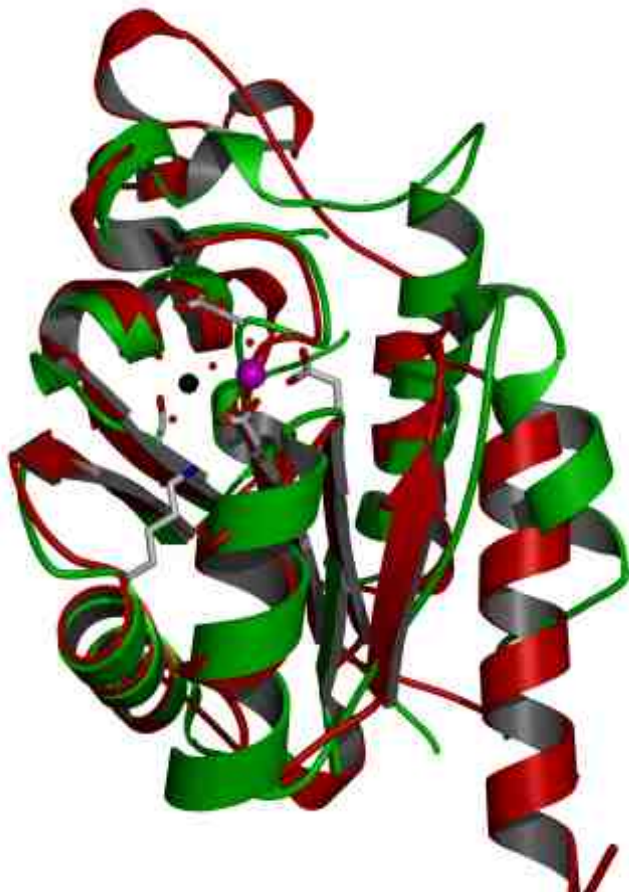
Figure 3.37. Structural overlay between MDP-1 (red) and BF1531 HAD domain (green) shown as a ribbon diagram. The active site residues, waters, and tungstate ligand from the BF1531 HAD domain are shown as ball and stick.

Based on Dali search results, MDP-1 is the most similar in structure to the BF1531 HAD domain. The two structures share RMSD 1.87 Å and an identical active site. The two structures are superimposed in **Figure 3.6** and their active sites are superimposed in **Figure 3.7**.
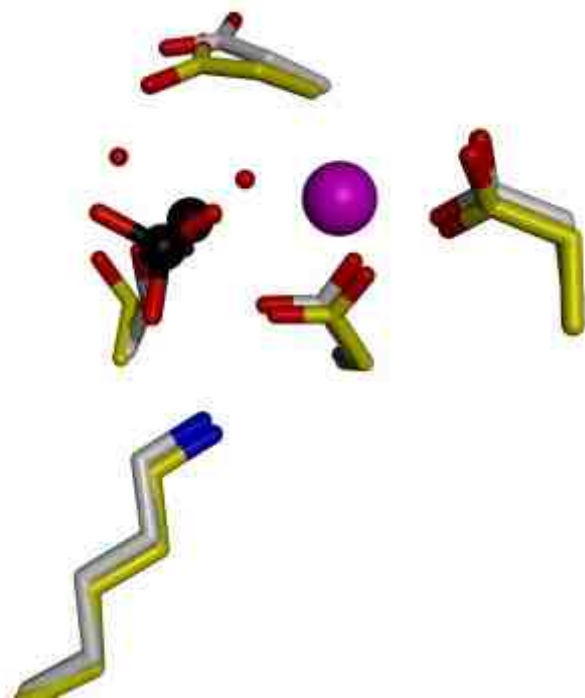
Figure 3.38. Active site overlay between MDP-1 (yellow) and BF1531 HAD domain (grey). The active site residues, Mg^{2+} cofactor (magenta sphere) and tungstate ligand from the BF1531 HAD domain and MDP1 domain are shown as ball and stick.

3.3.2 BF0824/BF1531 SAXS Analysis

Because the structure determination of the full length BF0824 and BF1531 was unsuccessful SAXS was used to determine a structural envelope that we could use to generate a model of the native protein. The PHYRE webserver was used to construct a

threading model of the C-terminal acyl transferase domain. The transferase domain was placed at a complementary interface with the HAD domain and positioned such that the catalytic cysteine was in close proximity (16Å) to the nucleophile Asp of the HAD domain (**Figure 3.8**).

| **N-term** | | **HAD** | | **Transferase** |
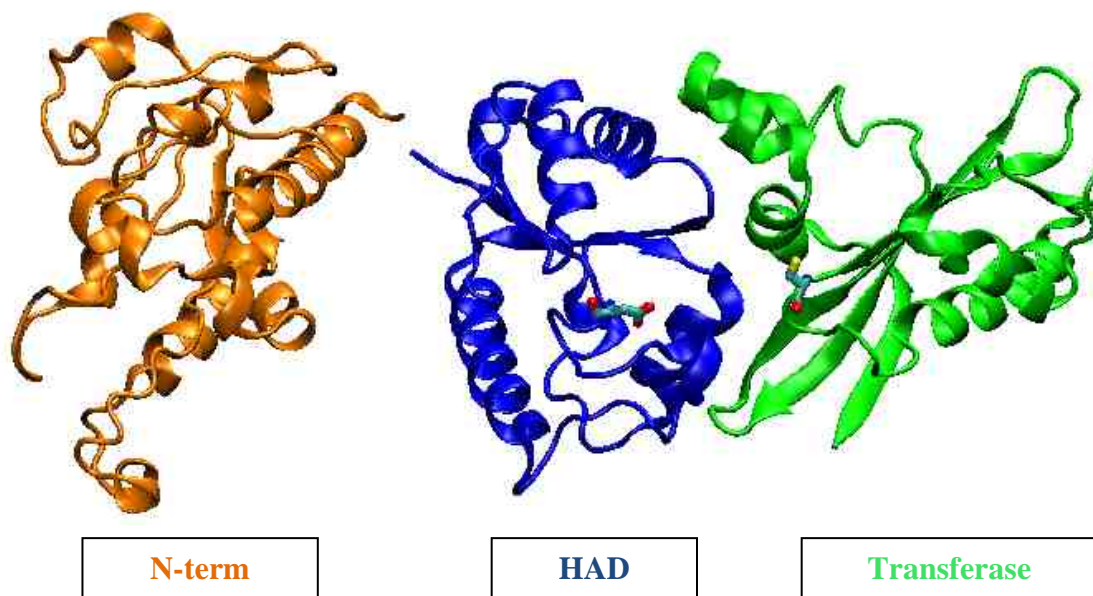
Figure 3.39. All-atom model of BF0824 monomer

A model of the BF0824 dimer was constructed by applying the P2 symmetry axis from the 3NVB PDB file (see **Figure 3.9**)
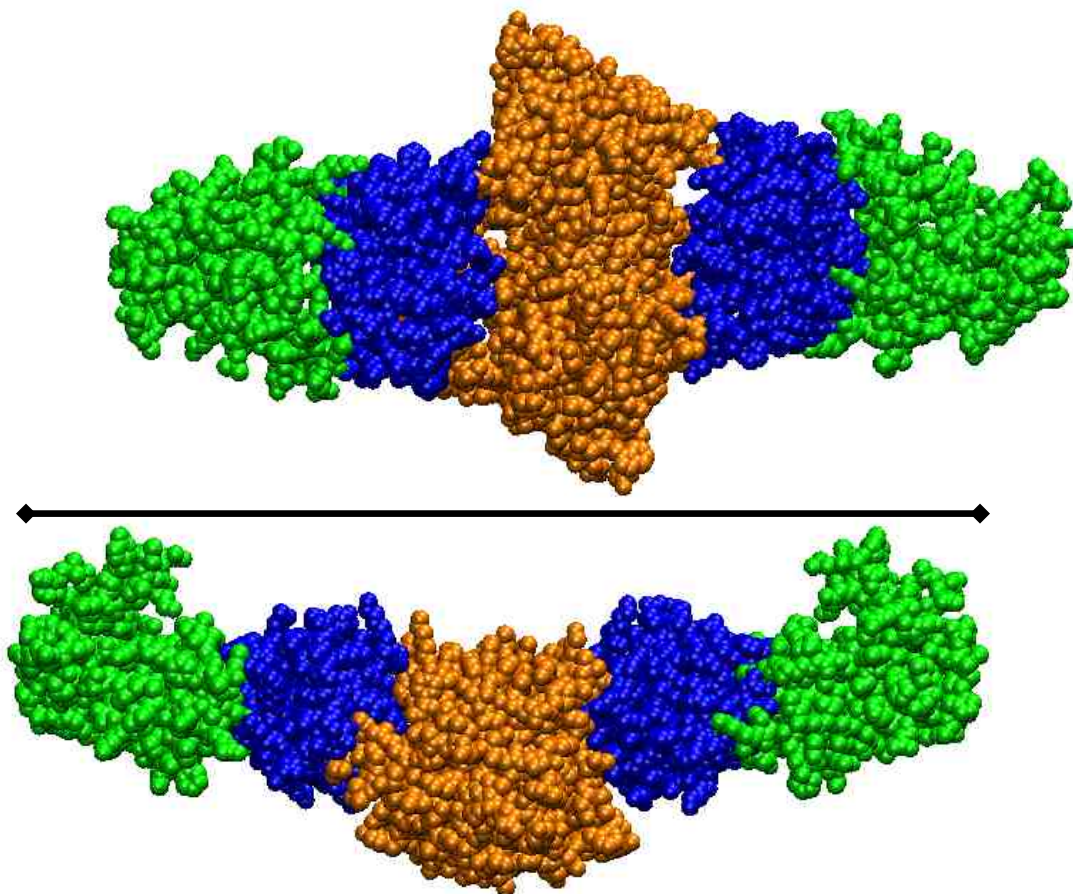
Figure 3.40. Rough all-atom model of BF0824 dimer from homology model of C-terminal (orange), HAD (blue) and acyl transferase (green) domains. The dimer is 160 Å long.

The overall shape of the rough all-atom dimer model has a consistent shape with the SAXS data, indicating that the oligomeric state of the protein is likely correct and that the transferase domains extend away from the central structure at either end of the dimer (**Figure 3.10**).
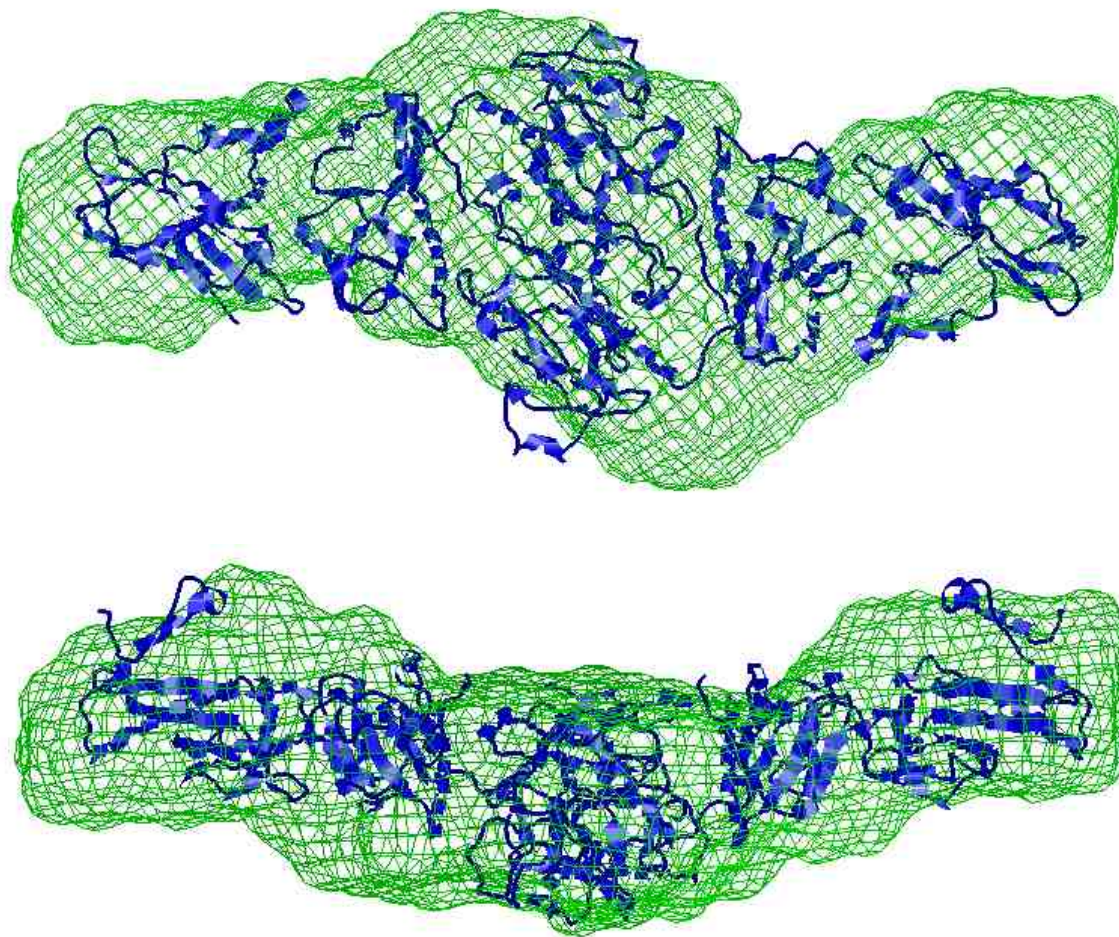
Figure 3.41. Ribbon models of BF0824 all-atom model with P2 SAXS envelope.

.   The docked model indicates the N-terminal does not play a role in catalysis but rather it serves as a dimerization domain. As described below, the biological range of BF0824/BT1531 homologues was determined. A multiple sequence alignment of these homologues shows that many of these homologues are either missing the N-terminal domain or possess an N-terminal domain from a different protein fold family.

3.3.3 BF0824 (ExPasY #Q64Y51) and BF1531 (ExPasY #Q64W45) *In vitro* Function Determination

The catalytic function of FKBH is to autoacylate the catalytic Cys residue of the C-terminal acyl transferase domain using 1,3-diphosphoglycerate and then remove the C(3)phosphate from the Cys-S-glyceryl-3-phosphate using the HAD domain catalytic residues. The final step is the transfer of the glyceryl moiety to an acceptor. First we showed using ES-MS to monitor the reaction that in the presence of 1,3-diphosphoglycerate (generated in situ) BF0824 and BF1531 form the Cys-S-glyceryl adduct. The mass spectrum of N-terminal His_6-tagged BF0824 before and after incubation (at a concentration of with 140 μM) with 3-phosphoglycerate (767 μM), ATP (924 μM), $MgCl_2$(11 mM) and glycerate kinase (20 units) in Tris-HCl (50 mM, pH 7.6, 25 ^{o}C) for 30 min, is shown in Figure 3.11. The mass spectrum of C-terminal His_6-tagged BF1531 before and after incubation (at a concentration of with 300 μM) with 3-phosphoglycerate (1.8 mM), ATP (1.8 mM), $MgCl_2$(5 mM) and glycerate kinase (20 units) in Tris-HCl (50 mM, pH 7.6, 25 ^{o}C) for 10 min, is shown in Figure 3.12.These results show that BF0824 and BF1531 are in fact functional FKBH enzymes.
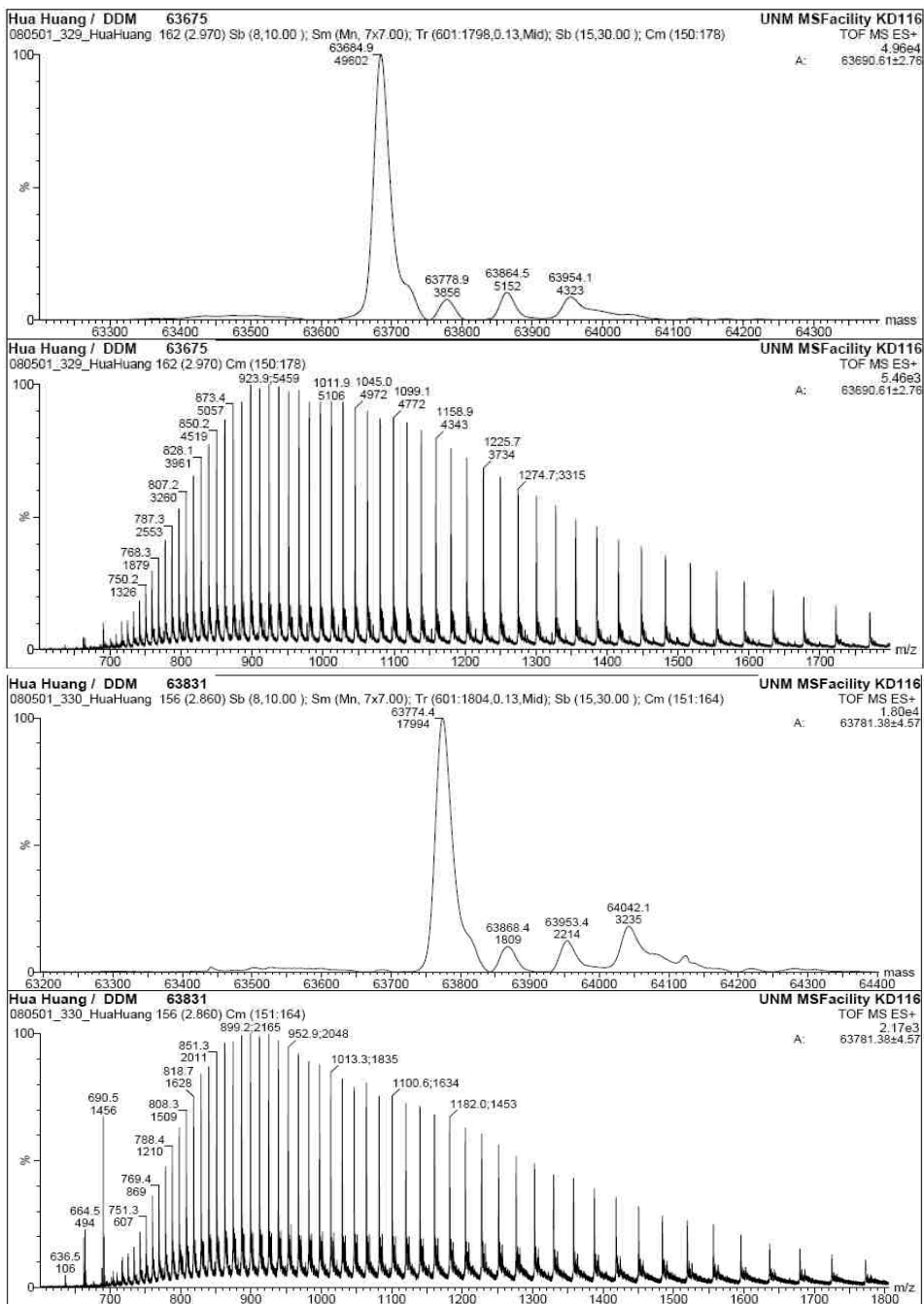
Figure 3. 42. The mass spectrum of N-terminal His₆-tagged BF0824 before (above) and

after (below) incubation (at a concentration of with 140 μM) with 3-phosphoglycerate

(767 μM), ATP (924 μM), MgCl$_2$ (11 mM) and glycerate kinase (20 units) in Tris-HCl

(50 mM, pH 7.6, 25 ^{o}C) for 30 min.
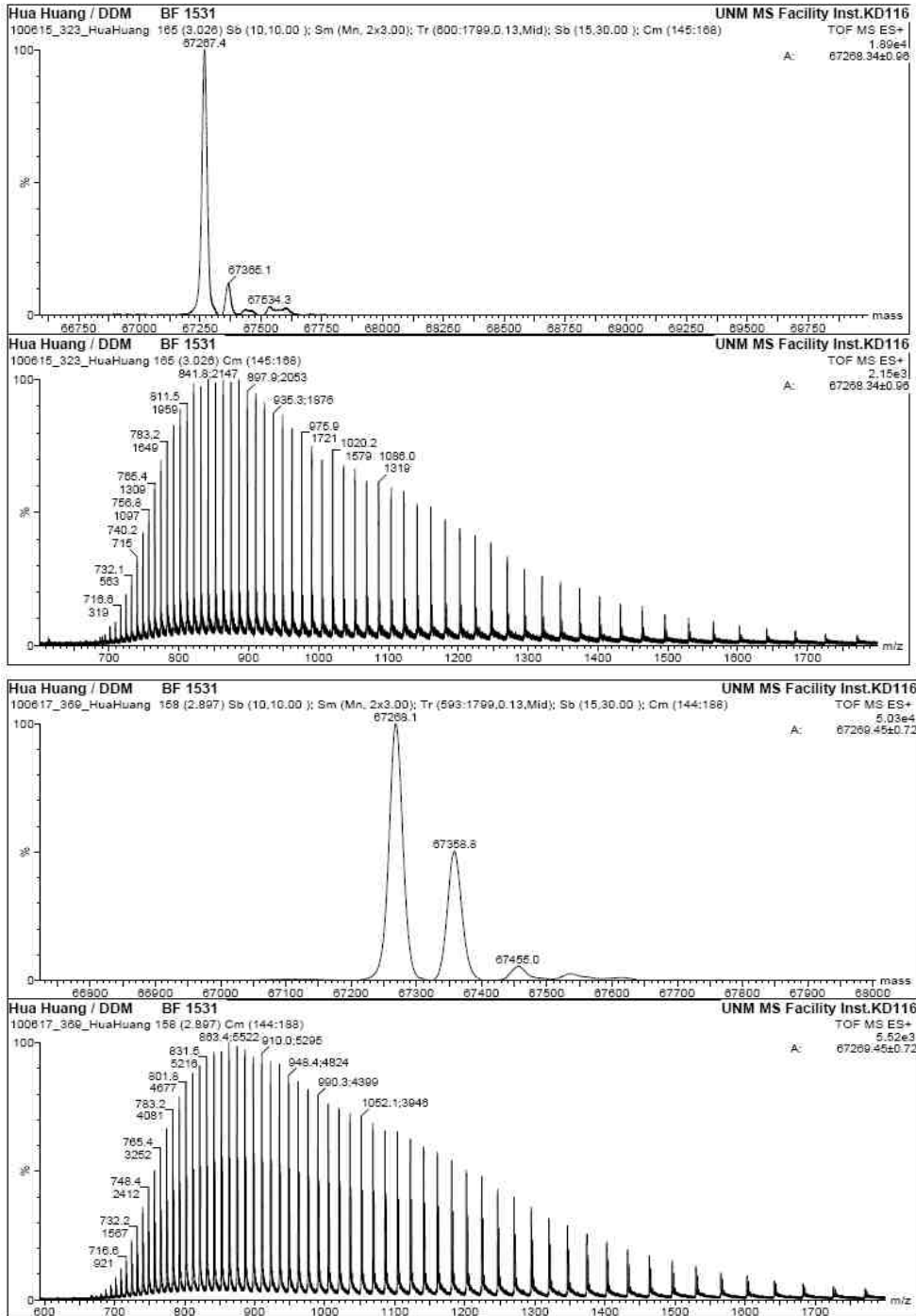
Figure 3.43. The mass spectrum of C-terminal His6-tagged BF1531 before(above) and after (below) incubation (at a concentration of with 300 μM) with 3-phosphoglycerate (1.8 mM), ATP (1.8 mM), MgCl2 (5 mM) and glycerate kinase (20 units) in Tris-HCl (50 mM, pH 7.6, 25 °C) for 10 min

The next question to be addressed is what is the identity of the glyceryl unit acceptor. The likely candidate is the ACP, which is encoded by a gene located next to the FKBH gene (BF0825 in the case of BF0824 and BF1532 in the case of BF1532). For now only the transfer of the glyceryl unit from glyceryl-BF0824 to holo-BF0825 was tested. Recombinant BF0825 was isolated as the apo-ACP and the converted to the holo-ACP by reaction with CoA catalyzed by recombinant phosphopantetheinyl transferase BF1558.

The ES-MS spectrum of holo-BF0825 before and after incubation (at a concentration of 22 μM) with glyceryl-BF0824 (71 μM)in 50 mM Tris-HCl (pH 7.6) at 25 °C for 20 min shows that BF0825 is the acceptor. Notably, the recombinant ACP BF0819 (encoding gene is just upstream of that of the BF0824 gene) is a substrate for the recombinant phosphopantetheinyl transferase BF1558, however the holo-BF0819 is not a substrate for glyceryl-BF0824. The ES-MS spectrum of holo-BF0825 before and after incubation (at a concentration of 22μM) with glyceryl-BF0824 (37μM) in 50 mM Tris-HCl (pH 7.6) at 25 °C and also at 37 °C for 20 min shows that holo-BF0825 is unchanged (**Figure 3.13**). Thus, the glyceryl-BF0824 discriminates between its ACP acceptor and an ACP that is encoded by the same capsular polysaccharide loci.

The conclusion drawn from these findings is that BF0824 catalyzes the glycerylation of the holo-ACP BF0825 using 1,3-diphosphoglycerate as the source of the glyceryl unit. The next step was to determine the identity of the acyl transferase that receives the glyceryl unit from glyceryl-BF0825. The putative acyl-transfereases encoded by the polysaccharide loci are: putative 3-oxyoacyl-ACP synthase BF0820 (a type III fatty acid synthase homolog), putative 3-oxyoacyl-ACP synthase BF0822 (a type III fatty acid synthase homolog), putative serine O-acetyl transferase BF0818 and putative malony-CoA: ACP transacylase BF2258. Attempts to express the clone of BF0822 failed, however the other three acyl transferases were successfully prepared for activity testing. The reactions were carried out under the conditions detailed in Materials and Methods and the reaction solutions were subjected to ES-MS analysis. Disappointingly, no product was observed and presently the recipient of the BF0825 glyceryl unit remains unknown.
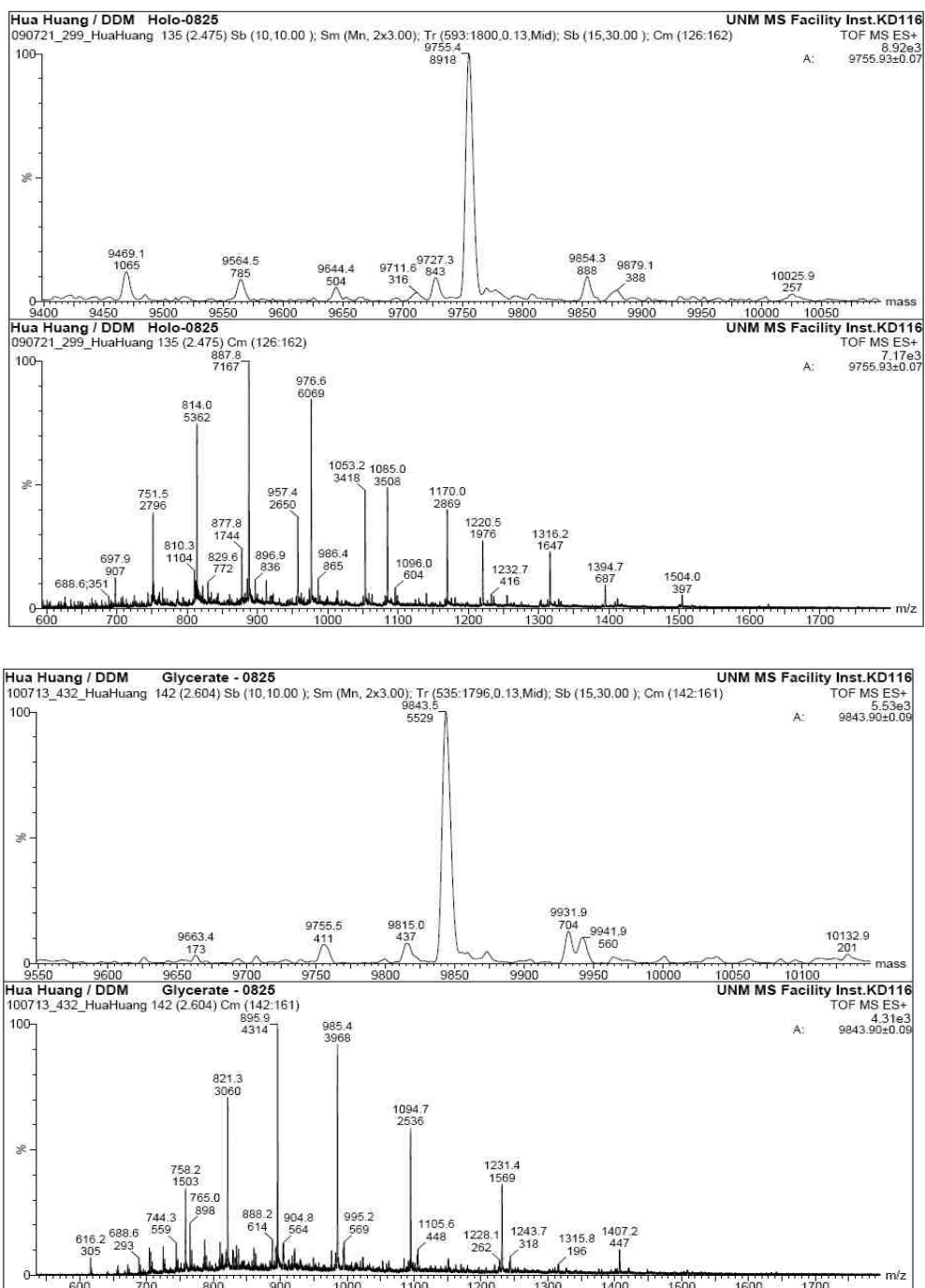
Figure 3.44. The ES-MS spectrum of holo-BF0825 before and after incubation (at a concentration of 22 μM) with glyceryl-BF0824 (71 μM) in 50 mM Tris-HCl (pH 7.6) at 25 °C for 20mins.

123

3.3.4 BF0824 (ExPasY #Q64Y51) and BF1531 (ExPasY #Q64W45) Biological Range.

To date reports of FKBH homologues have focused on FKBH domains that function within a polyketide synthase. BF0824 and BF1513 are not associated with a polyketide synthase and instead appear to contribute a glyceryl unit to a capsular polysaccharide. By tracking homologues via BLAST searches of the NCBI genome database the biological range of FKBH could be examined. Moreover, inspection of the gene context of FKBH genes in annotated genomes provides insight into possible FKBH functions.

The FKBH biological range search was limited to prokaryotes. None were found in Archea. Within the Bacteria FKBH is found primarily in Acintobacter and Bacillales. within Bacteroidetes/Chlorobi FKBH is found in *Bacteroides fragilis* (not all strains), *Prevotella marshii, Parabacteroides distasonis, Bacteroides fluxus, Chitinophaga pinensis, Paludibacter propionicigenes, Bacteroides eggerthii* and *Prevotella ruminicola*. Within Proteobacteria FKBH is present in select species of alpha, gamma, delta and epsilon division species.

Inspection of FKBH gene contexts within the genomes of these bacteria provides some insight into the applications of FKBH function. FKBH is most commonly represented as a free-standing protein that is not linked to a polyketide synthase. The glyceryl moiety is apparently used in a variety of biosynthetic contexts that are yet to be discovered.

3.4 Reference:

1)   Ezra Peisach, Jeremy D Selengut, Debra Dunaway-Mariano and Karen N Allen (2004) X-ray crystal structure of the hypothetical phosphotyrosine phosphatase MDP-1 of the haloacid dehalogenase superfamily *Biochemistry 43(40), 12770-12779*

2)   Dorrestein, P. C.; Van Lanen, S. G.; Li, W.; Zhao, C.; Deng, Z.; Shen, B.; Kelleher(2006) , N. L, The Bifunctional Glyceryl Transferase/Phosphatase OzmB Belonging to the HAD Superfamily That Diverts 1,3-Bisphosphoglycerate into Polyketide Biosynthesis, *J. Am. Chem. Soc. 128(32) ,10386-10387*;

3)   M.Hildebrand, L E. Waggoner and Margo Haygood(2004). *bryA*: An Unusual Modular Polyketide Synthase Gene from the Uncultivated Bacterial Symbiont of the Marine Bryozoan *Bugula neritina. Chemistry & Biology 11:1543-1552*

4)   L J. Walton, C.Corre, Gregory L. Challis(2006), Mechanisms for incorporation of glycerol-derived precursors into polyketide metabolites, *J Ind Microbiol Biotechnol 33:105-120*

5)   C Zhao, J M.Loughlin, Zixin Deng(2010), Oxazolomycin Biosynthesis in *Streptomyces albus* JA3453 Featuring an "Acyltransferase-less" Type I Polyketide Synthase That Incorporates Two Distinct Extender Units,*The Journal of Biological Chemistry 285(26), 20097-20108*

6)   K Wu, L Chung, W P. Revill, L Katz and Christopher D. Reeves(2000), The

FK520 gene cluster of *Streptomyces hygroscopicus* var. *ascomyceticus* (ATCC 14891) contains genes for biosynthesis of unusual polyketide extender units, *Gene 251:81-90*

7)   T Yu, L Bai and Heinz G. Floss(2002), The biosynthetic gene cluster of the maytansinoid antitumor agent ansamitocin from Actinosynnema pretiosum, *PNAS 99(12), 7968-7973*

8)   S F. Haydock, A N. Appleyard, Peter F. Leadlay(2005), Organization of the biosynthetic gene cluster for the macrolide concanamycin A in Streptomyces neyagawaensis ATCC 27449, *Microbiology 151:3161-3169*

9)   J.Ligon, S. Hill, J Beck and Thomas Schupp(2002), Characterization of the biosynthetic gene cluster for the antifungal polyketide soraphen A from Sorangium cellulosum So ce26, *Gene 285:257-267*

10) B M. Kevany, D A.Rasko and Michael G. Thomas(2009), Characterization of the Complete Zwittermicin A Biosynthesis GeneCluster from *Bacillus cereus*, *Applied and Environmental Microbiology   75(4): 1144-1155*

11) Y A. Chan, M T. Boyne and Michael G. Thomas(2006), Hydroxymalonyl-acyl carrier protein (ACP) and aminomalonyl-ACP are two additional type I polyketide synthase extender units, *PNAS 103(39):14349-14354*

12) H. Zhang, Hung-wen Liu(2007), Elucidation of the Kijanimicin Gene Cluster: Insights into the biosynthesis of Spirotetronate Antibiotics and Nitrosugars, *JACS 129,14670-14683*

13)   Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, *Method Enzymol276*, 307-326.

14)   Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr D Biol Crystallogr66*, 213-221.

15)   Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot, *Acta Crystallogr D Biol Crystallogr66*, 486-501.

16)   Kraulis, P. J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures, *J. Appl. Cryst.24*, 946-950.

17)   Svergun, D. I. (1992) Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria, *J Appl Crystallogr25*, 495-503.

18)   Svergun, D. I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing (vol 76, pg 2879, 1999), *Biophys J77*, 2896-2896.

19)   Kozin, M. B., and Svergun, D. I. (2001) Automated matching of high- and low-resolution structural models, *J Appl Crystallogr34*, 33-41.

20)   Volkov, V. V., and Svergun, D. I. (2003) Uniqueness of ab initio shape determination in small-angle scattering, *J Appl Crystallogr36*, 860-864.

21) Kelley, L. A., and Sternberg, M. J. E. (2009) Protein structure prediction on the Web: a case study using the Phyre server, *Nat. Protocols4*, 363-371.

22) Holm, L., Kaariainen, S., Rosenstrom, P., and Schenkel, A. (2008) Searching protein structure databases with DaliLite v.3, *Bioinformatics24*, 2780-2781.

23) Burroughs, A. M., Allen, K. N., Dunaway-Mariano, D., and Aravind, L. (2006) Evolutionary genomics of the HAD superfamily: understanding the structural adaptations and catalytic diversity in a superfamily of phosphoesterases and allied enzymes, *J Mol Biol361*, 1003-1034.

APPENDIX

A.1 Published Collaborative Work

A1.1 Divergence of biochemical function in the HAD superfamily: D-glycero-D-manno-heptose-1,7-bisphosphate phosphatase (GmhB).

Wang L, Huang H, Nguyen HH, Allen KN, Mariano PS, Dunaway-Mariano D.

Department of Chemistry and Chemical Biology, University of New Mexico, Albuquerque, New Mexico 87131, USA.

Abstract

D-Glycero-D-manno-heptose-1,7-bisphosphate phosphatase (GmhB) is a member of the histidinol-phosphate phosphatase (HisB) subfamily of the haloalkanoic acid dehalogenase (HAD) enzyme superfamily. GmhB supports two divergent biochemical pathways in bacteria: the D-glycero-D-manno-heptose-1-alpha-GDP pathway (in S-layer glycoprotein biosynthesis) and the l-glycero-D-manno-heptose-1-beta-ADP pathway (in lipid A biosynthesis). Herein, we report the comparative analysis of substrate recognition in selected GmhB orthologs. The substrate specificity of the l-glycero-D-manno-heptose-1-beta-ADP pathway GmhB from Escherichia coli K-12 was evaluated using hexose and heptose bisphosphates, histidinol phosphate, and common organophosphate metabolites. Only D-glycero-D-manno-heptose-1-beta,7-bisphosphate ($k(cat)/K(m) = 7 \times 10^6$ M^{-1} s^{-1}) and D-glycero-D-manno-heptose-1-alpha,7-bisphosphate ($k(cat)/K(m) = 7 \times 10^4 M^{-1}$ s^{-1}) displayed physiologically significant substrate activity. ^{31}P NMR analysis demonstrated

that E. coli GmhB selectively removes the C(7) phosphate. Steady-state kinetic inhibition studies showed that D-glycero-D-manno-heptose-1-beta-phosphate (K_{is} = 60µM, and K_{ii} = 150µM) and histidinol phosphate (K_{is} = 1 mM, and K_{ii} = 6 mM), while not hydrolyzed, do in fact bind to E. coli GmhB, which leads to the conclusion that nonproductive binding contributes to substrate discrimination. High catalytic efficiency and a narrow substrate range are characteristic of a well-evolved metabolic enzyme, and as such, E. coli GmhB is set apart from most HAD phosphatases (which are typically inefficient and promiscuous). The specialization of the biochemical function of GmhB was examined by measuring the kinetic constants for hydrolysis of the alpha- and beta-anomers of D-glycero-d-manno-heptose-1-beta,7-bisphosphate catalyzed by the GmhB orthologs of the l-glycero-D-manno-heptose-1-beta-ADP pathways operative in Bordetella bronchiseptica and Mesorhizobium loti and by the GmhB of the D-glycero-D-manno-heptose-1-alpha-GDP pathway operative in Bacteroides thetaiotaomicron. The results show that although each of these representatives possesses physiologically significant catalytic activity toward both anomers, each displays substantial anomeric specificity. Like E. coli GmhB, B. bronchiseptica GmhB and M. loti GmhB prefer the beta-anomer, whereas B. thetaiotaomicron GmhB is selective for the alpha-anomer. By determining the anomeric configuration of the physiological substrate (D-glycero-D-manno-heptose-1,7-bisphosphate) for each of the four GmhB orthologs, we discovered that the anomeric specificity of GmhB correlates with that of the pathway kinase. The conclusion drawn from this finding is that the evolution of the ancestor to GmhB in the HisB subfamily provided for specialization toward two distinct biochemical functions.

A.1.2 Structural determinants of substrate recognition in the HAD superfamily member D-glycero-D-manno-heptose-1,7-bisphosphate phosphatase (GmhB) .

Nguyen HH, Wang L, Huang H, Peisach E, Dunaway-Mariano D, Allen KN.

Department of Physiology and Biophysics, Boston University School of Medicine, Boston, Massachusetts 02118-2394, USA.

Abstract

The haloalkanoic acid dehalogenase (HAD) enzyme superfamily is the largest family of phosphohydrolases. In HAD members, the structural elements that provide the binding interactions that support substrate specificity are separated from those that orchestrate catalysis. For most HAD phosphatases, a cap domain functions in substrate recognition. However, for the HAD phosphatases that lack a cap domain, an alternate strategy for substrate selection must be operative. One such HAD phosphatase, GmhB of the HisB subfamily, was selected for structure-function analysis. Herein, the X-ray crystallographic structures of Escherichia coli GmhB in the apo form (1.6 Å resolution), in a complex with Mg^{2+} and orthophosphate (1.8 Å resolution), and in a complex with Mg^{2+} and D-glycero-D-manno-heptose-1-beta-7-bisphosphate (2.2 Å resolution) were determined, in addition to the structure of Bordetella bronchiseptica GmhB bound to Mg^{2+} and orthophosphate (1.7 Å resolution). The structures show that in place of a cap domain, the GmhB catalytic site is elaborated by three peptide inserts or loops that pack to form a concave, semicircular surface around the substrate leaving group. Structure-

guided kinetic analysis of site-directed mutants was conducted in parallel with a bioinformatics study of sequence diversification within the HisB subfamily to identify loop residues that serve as substrate recognition elements and that distinguish GmhB from its subfamily counterpart, the histidinol-phosphate phosphatase domain of HisB. We show that GmhB and the histidinol-phosphate phosphatase domain use the same design of three substrate recognition loops inserted into the cap domain yet, through selective residue usage on the loops, have achieved unique substrate specificity and thus novel biochemical function.

A.2 Manuscript of Collaborative Work Submitted for Publication

A.2.1 Evolution of Substrate Specificity in Sialic Acid Phosphatases within Bacterial Species[†], [††]

Kelly D. Daughtry, Liangbing Wang#, Hua Huang#, XXX, Steve C. Almo††, Debra Dunaway-Mariano#*, and Karen N. Allen§*

Department of Physiology and Biophysics, Boston University School of Medicine, Boston, Massachusetts 02118-2394

#Department of Chemistry and Chemical Biology, University of New Mexico, Albuquerque, NM 87131-0001

††Department of Medicine, Albert Einstein College of Medicine, Bronx, NY, USA

§Department of Chemistry, Boston University, Boston, MA 02215- 2521

Running Title: Evolution of Phosphatase Specificity

Abbreviations

HADSF, Haloalkanoate acid dehalogenase superfamily; KDO, 2-keto-3-deoxy-D-manno-octulosonic acid; KDO8P, 2-keto-3-deoxy-D-manno-octulosonic acid 8-phosphate, KDO8PP, 2-keto-3-deoxy-D-manno-octulosonic acid 8-phosphate phosphohydrolase; KDN, 2-keto-3-deoxynononic acid ; KDN9P, 2-keto-3-deoxynononic acid 9-phosphate; BT, Bacteroides thetaiotaomicron;

Abstract

Defining substrate specificity in enzyme systems is crucial to understanding evolution and selection processes over time. Two paralogous enzymes (with 28% overall sequence identity) that work on substrates with similar structures, 2-keto-3-deoxy-D-manno-octulosonate 8-phosphate (KDO8P) phosphatase and 2-keto-3-deoxy-9-O-phosphonononic acid phosphatase (KDN9P) phosphatase from *Bacteriodes thetaiotaomicron*, are both members of the C0 subfamily of the Haloalkanoate Dehalogenase Superfamily (HADSF). The two phosphotransferases have similar three-dimensional structures, sugar phosphate substrates, and active site composition, with 28% overall sequence identity). Previously, kinetic analyses show that *B. thetaiotaomicron* KDN9P phosphatase has a 60 fold greater specificity (k_{cat}/K_m) for KDN9P over KDO8P, while KDO8P phosphatase is 20 fold more efficient for KDO8P over KDN9P. In order to uncover the structural origins of substrate specificity the crystal structures of KDN9PP from *B. thetaiotaomicron* and KDO8PP from *Haemophilus influenza* were solved in the presence of the corresponding intermediate/transition state analog complex, KDN/vanadate or KDO/vanadate at 2.1 Å and 1.8Å resolution, respectively. Sequence analysis, examination of the ligand interactions, solvent accessibility, and comparison with the structure of the unliganded KDO8PP from *B. thetaiotaomicron* (1.8 Å resolution) shows that the previously identified specificity residues Glu 56 and Lys 67 in KDN9PP are replaced by Gly and Ala respectively in *B. thetaiotaomicron* KDO8P phosphatase. The role of these residues in delineating specificity was further demonstrated by mutational analysis and kinetics, with the Glu56

134

to Ala mutation in KDN9P phosphatase showing a switch in specificity from KDN9P to KDO8P. Moreover, the bioinformatic analysis of bacterial genomes lead to the identification of a subset of KDO8PP sequences with altered specificity determinants (Arg and Asp in place of Gly and Ala). Structural and kinetic analysis of hypothetical KDO8PP and KDN9PP enzymes confirm the identity of four KDO8PPs that can be recognized by these markers throughout gram-negative bacteria. Bioinformatic analysis of bacterial genomes indicates that the Bacteroides phylum contains KDO8PP enzymes with Gly/Ala specificity determinants, while the remainder of gram-negative bacteria (across all phyla) contains the Arg specificity determinant (similar to *H. influenzae*). Overall, structural and kinetic analysis in the presence of ligand was utilized to uncover specificity determinants which were not apparent from sequence analysis alone.

Although the HADSF was initially named for 2-haloacid dehalogenase, the first well characterized family member [1], most extant members are phosphohydrolases. The conserved HADSF three-dimensional structsandwich comprised of repeating units which adopt the typical Rossmannoid fold [2]. The four conserved motifs of the HADSF are juxtaposed in tertiary but not primary structure and contain catalytic residues and, in the phosphohydrolase members, residues that coordinate the Mg^{2+} (**Figure A.1**). Motif 1 (DxD) comprises the nucleophilic Asp residue, and the Asp+2 residue contributes the backbone carbonyl for cofactor binding, and acts as general acid/base (**Figure A.1**). Motif 2 contains a Ser or Thr residue which coordinates the transferring phosphoryl group. Motif 3 contains a Lys/Arg residue which forms a salt bridge to both nucleophile and transferring phosphoryl group, positioning them with respect to one another. In Motif

4 Asp/Glu (and very rarely Asn/Gln) coordinate the magnesium cofactor. The HADSF can be classified by the location and topology of the inserted caps [2], C0 caps have only a small inserted region between motifs 1 and 2, and C1 caps have a larger insert/domain in this region, whereas C2 caps have an insert/domain between motifs 2 and 3. The cap domain introduces a specificity loop into the active site in C1 and C2 HADSF members [3-4], while little is understood about the mechanism of specificity within C0 sub-family members. Thus C0 members are interesting to study to understand HADSF specificity elements.
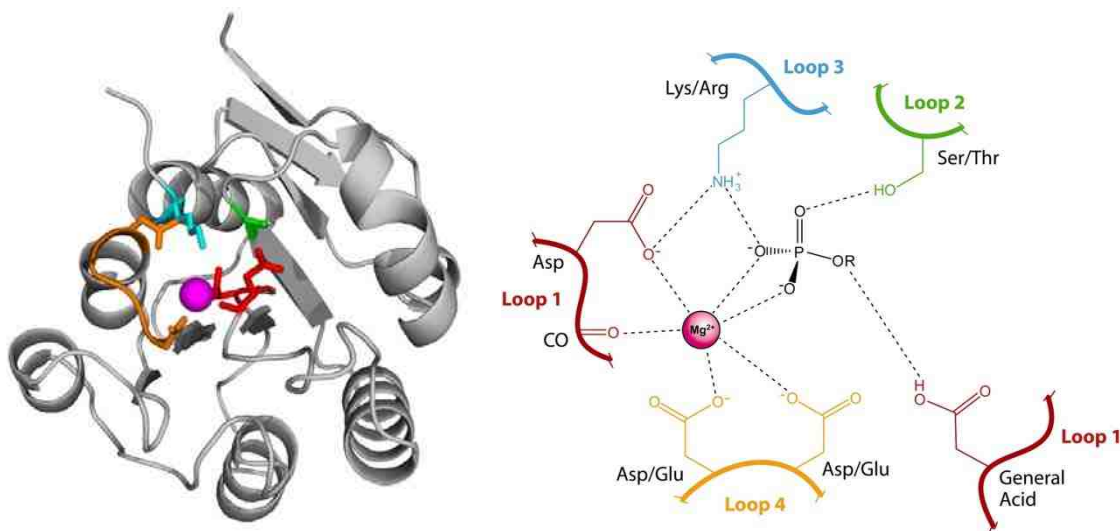
Figure A.45: HADSF Structure, Figure generated by Pymol [53]

B. thetaiotaomicron is a gram negative symbiotic human gut bacterium, that has evolved to acquire and hydrolyze otherwise indigestible dietary polysaccharides [5]. BT1677 is a 2-keto-3-deoxy-D-manno-octulosonate 8-phosphate phosphatase (KDO8PP) which produces KDO, a saccharide integral to lipid A and thus crucial for gram negative cell survival [6-8]. BT1713 was recently identified as a 2-keto-3-deoxy-9-*O*-

phosphononic acid phosphatase (KDN9PP) [9]. KDN is a sialic acid derivative commonly used in cell wall polysaccharides in higher level organisms [10-11].

Eukaryotic synthesis of KDN is a two step process, entailing the condensation of mannose 6-phosphate and phosphoenolpyruvate to produce KDN9P, which is then dephosphorylated by KDN9PP to produce KDN [12]. In general, bacterial synthesis of KDN differs in that KDN is directly produced from the condensation of mannose and phosphoenolpyruvate [13]. Notably, in *B. thetaiotaomocron* KDN is synthesized via the eukaryotic two-step process (**Figure A.2**) [14]. The 8-carbon sugar KDO8P is formed from the condensation of D-arabinose 5-phosphate and phosphoenolpyruvate [15]. KDN9P is formed by the condensation of mannose 6-phosphate and phosphoenolpyruvate. Hydrolysis of the phosphate group is catalyzed in the next step by the corresponding phosphatase. The synthases preceding the phosphatases in the KDN and KDO pathways derive from different fold families (**Figure A.2**). Thus, biosynthesis of KDO and KDN are similar structurally yet are biosynthesized by distinct pathways.
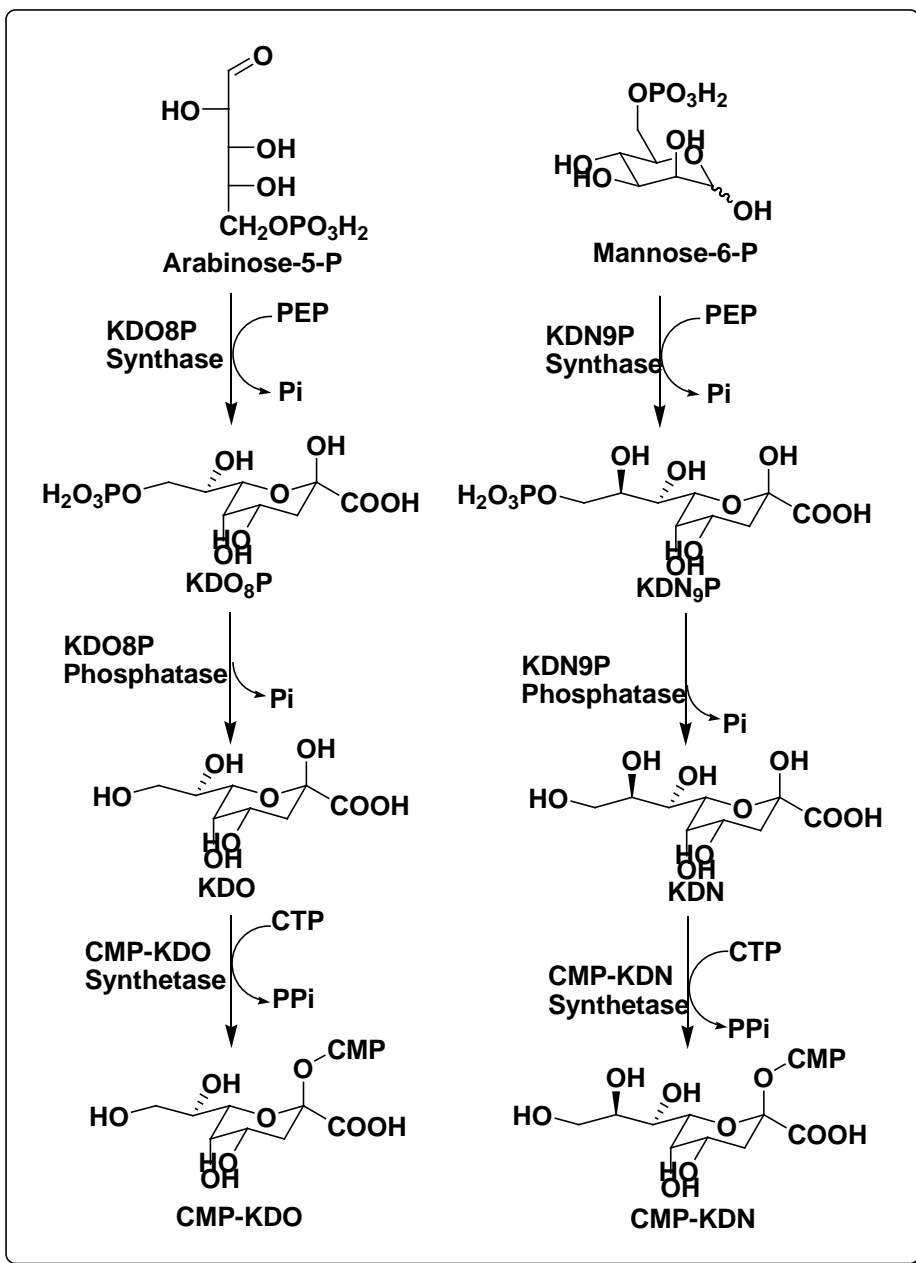
Figure A.46: KDO and KDN biosynthesis. Schematic of KDO (left panel) and KDN (right panel) biosynthesis. Figure generated with ChemDraw

The means of conferring specificity in the two phosphatases is not clear from primary sequence alignments since there is not a distinct specificity loop provided by the C0 cap, as there is in C1 or C2 cap members. An N-acetylneuraminic acid liganded structure of KDN9PP was recently published [9], but structural overlays of unliganded KDO8PP and liganded KDN9PP do not lead to the clear identification of specificity determinants. An *E. coli* KDO8PP c-terminal truncation mutant structure with ligand has also been determined to modest resolution [16], but a wild-type liganded structure is not available for comparison. While this work is informative, it does not assist in elucidating the basis of specificity.

The object of this work was to elucidate the structural means of KDO8PP and KDN9PP specificity in *B. thetaiotaomicron*, and determine the evolutionary importance in all gram-negative bacteria using bioinformatic techniques.

Methods

Unless otherwise stated all chemicals were obtained from Sigma-Aldrich and all primers, competent cells, ligases, polymerases and restriction enzymes from Invitrogen.

Cloning and site-directed mutagenesis

The gene encoding BT1713 was previously cloned [9].The gene encoding BT1677 was amplified from *Bacteriodes thetaiotaomicron* VPI-5482 genomic DNA with primers 5'-CCG GCG GGG AGA ATT GGA GGC TAA TAT AA 3' (forward) and 5'-GCT GGC AGA CAG TCA GGC ATT GCA GGA AAT 3' (reverse). After

amplification, the clone was modified to include restriction sites for Nde1 and BamH1 using the primers 5' GTAAAAG**CATATG**AGCACCATCAATTATGATTATCCCGC 3' (Nde1) and 5'TTATTAAAG**GGATCC**ACAACTCACCAGCCGAAAGCATCTTCC GCC 3' (BamH1) (bold letters indicate restriction sites). After amplification, the clone was digested with restriction enzymes Nde1 and BamH1. The fragment was cloned into a pET15b vector (Invitrogen) containing a thrombin protease site.

BT1713 and BT1666 genes bearing site-directed mutations were prepared by a PCR-based strategy with commercial primers and the wild-type BT1713 or BT1677 plasmid serving as a template. The gene sequence was confirmed by DNA sequencing.

Expression and Purification

Recombinant BT1713 wild-type and mutant was expressed and purified as previously described [9]. Recombinant BT1677 was purified from *Escherichia coli* BL-21(DE3) cells transformed with the pET15b vector harboring the BT1677 sequence. Cells were grown in LB media at 37 °C until $A_{600} = 1.0$ when IPTG was added to a final concentration of 1 mM. Cells were harvested by centrifugation after 4 h. Cell pellets were resuspended in 20 mM Hepes pH 7.5 and 100 mM NaCl and were lysed by sonication. Cell debris was removed by centrifugation for 45 min. The soluble fraction was filtered using a 0.22 µm sterile filter and applied to a TALON Metal Affinity Resin (Clontech) gravity column (5 mL) equilibrated in 20 mM Hepes pH 7.5, 100 mM NaCl and 10mM imidazole. BT1677 was eluted using a step gradient of imidazole (30 mM, 100 mM, 250 mM, and 500 mM). Fractions containing BT1677 were analyzed by SDS-PAGE, pooled,

and dialyzed against 2L of 20 mM Hepes pH 7.5, 100 mM NaCl 10 mM imidazole, and 1 mM DTT. The dialyzed protein was cleaved using Tagzyme DAPase (Qiagen) for 2 hours at 37 °C with 225 rpm shaking. The cleavage mixture was applied to the TALON Metal Affinity Resin (Clontech) gravity column (5mL) pre-equilibrated in 20 mM Hepes pH 7.5, 100 mM NaCl and 10 mM imidazole. The flow-through was analyzed by SDS-PAGE, pooled and dialyzed against 2L of 20 mM Hepes buffer, pH 7.5, 50 mM NaCl, and 10 mM MgCl$_2$. Purified protein was concentrated to 30 mg/mL using Amicon Ultra Concentrators. Typical yields were 3.75 mg / g cell paste of enzyme that was > 90% pure as assessed by SDS/PAGE.

All BT1713 mutants were isolated and purified to homogeneity (as judged by SDS-PAGE analysis) from transformed and induced *E. coli* cells using the same protocol used for wild-type BT1713.

Crystallization, Data Collection and Structure Solution BT1677 (BT-KDO8PP)

Crystals of BT1677 (BT-KDO8PP) were obtained by the vapor diffusion method using hanging drop geometry. Drops were equilibrated against reservoir solution containing 30% polyethylene glycol (PEG) MME 550, 40 mM MgCl$_2$, 100 mM Hepes, pH 7.5 at 18 °C. The hanging drops contained equal volumes (1 µl each) of the reservoir solution and a 30 mg/mL protein solution in 20 mM Hepes buffer, pH 7.5, 50 mM NaCl, and 10 mM MgCl$_2$. Small crystals (0.1mm x 0.1mm x 0.05mm) appeared within 2 days. The crystals belonged to the space group P2$_1$2$_1$2$_1$ with cell dimensions a=37.517 Å b= 91.349 Å c=170.115 Å. There were four molecules in the asymmetric unit and the

solvent comprised 35% of the unit cell. Surface solvent was removed using Paratone, and the resulting crystals were flash-cooled in liquid nitrogen. Data was collected at 100 K using a single wavelength at the X12C beamline of the National Synchrotron Light Source (Brookhaven National Laboratory, Upton, NY). For data collection, beamline X12C was equipped with a Brandeis 2 x 2 CCD detector. Data were collected to 1.8 Å resolution and processed with the program DENZO and SCALEPACK (**Table A.1**).

Table A.11 Data collection and refinement statistics for BT-KDO8PP, BT-KDN9PP liganded with Mg^{2+}, VO_3^- and KDN and HI-KDO8PP liganded with KDO and VO_3^-. Values in parentheses are for the highest resolution shell, NA indicates not applicable

Data Collection	BT-KDO8PP	BT-KDN9PP/ KDN/VO$_3^-$	HIKDO8PP/ KDO/VO$_3^-$
Space group	P2$_1$2$_1$2$_1$	P2$_1$2$_1$2	I4
Cell dimension (Å)	a = 37.52 b = 91.35 c = 170.12	a=81.30 b=106.32 c=74.15	a = b = 79.85 c = 52.15
Molecules / ASU	4	4	1
Wavelength (Å)	1	1.54	1.54
Resolution (Å)	85.1 - 1.8 (1.86 – 1.8)	28.5 - 2.1 (2.2 – 2.1)	29.5 - 1.8 (1.89 - 1.8)
Observed reflections	822,448	220,160	120,635
Unique reflections	55,354	38,011	14,537

Completeness (%)	91.1 (79.7)	95.5 (99.0)	95.6 (91.4)
Rmerge (%)	6.1 (43.2)	7.6 (48.3)	5.73 (38.13)
$<I/>$	19.6 (2.79)	6.29 (2.5)	13.68 (2.40)
Redundancy	5.3 (4.8)	5.4 (2.73)	7.9 (5.7)
Twin Law			-k,-h,-l
Twin Fraction			0.33

Refinement			
R / Rfree	17 / 22.1	19.67 / 26.45	14.52 / 17.60
Average Temperature Factor $(Å^2)$			
Protein /Water	23.47 / 29.0	26.4 / 26.0	18.3 / 26.0
Mg / Ligand	16.9 / NA	30.4 / 47.1	16.0 / 21.6
Rms deviation from ideal			
Bonds	0.013	0.008	0.0286
Angles	1.376	1.115	2.352

The structure of unliganded wild-type BT-KDO8PP was determined to 1.8 Å resolution with phases determined by the molecular replacement method with the program Phaser in the CCP4 suite [17-18]. The search model was the protomer of BT-KDN9PP (BT_1713), leading to a clear solution for four protomers (forming the expected tetramer). The structure was refined using the program CNS [19] using data between 50 and 1.8 Å, for which $F \geq 2\sigma(F)$. An energy minimization molecular dynamics cycle was followed by a simulated annealing molecular dynamics cycle at 5000 K which was followed by alternating cycles of positional and individual temperature factor refinement. The resulting model was manually inspected and modified with the program COOT [20]. In the final stages of the refinement, water molecules were added

to the model using an electron-density acceptance criteria of $d \geq 3\sigma(d)$ in the *F*o-*F*c difference electron-density map

BT1713 liganded with KDN/VO$_3^-$

Crystals of BT1713 (BT-KDN9PP) were obtained by vapor diffusion in hanging drops. Drops were equilibrated against reservoir solution containing 19% PEG 3350 and 100 mM magnesium formate at room temperature. The hanging drops contained equal volumes (1µl each) of the reservoir solution and a 29 mg/mL protein solution in 10 mM Hepes buffer, pH 7.5 and 10 mM MgCl$_2$. Small crystals (0.15 mm x 0.15 mm x 0.01 mm) appeared in 1 day. Crystals were soaked in 50 □L of 50 mM KDN, 20 mM sodium vanadate, 22% PEG 3350 and 100 mM magnesium formate for one week before data collection. Surface solvent was removed by dragging the crystal through paratone then flash-cooling in liquid nitrogen. The crystals belonged to the space group P2$_1$2$_1$2 with cell dimensions a=81.296 Å b=106.324 Å c=74.147 Å. Data was collected at a single wavelength on the Bruker Microstar-H running at 2.7 kW with Helios multi-layer optics and a Platinum 135 CCD detector at Bruker AXS, Madison, Wisconsin, USA. Data were collected to 2.1 Å, reduced using SAINT and scaled using SADABS from the PROTEUM2 software suite (**Table A.1**).

The structure of liganded BT-KDN9PP was determined to 2.1 Å resolution by the molecular replacement method with the program MOLREP in the CCP4 suite. The search model was the protomer of KDN9PP, leading to a clear solution for four protomers (the expected tetramer). Successive rounds of model refinement were performed with the

144

program PHENIX [21] with data between 50 and 2.1 Å, for which $F \geq 2\sigma(F)$ with alternating cycles of positional and individual temperature factor refinement. The resulting model was inspected and modified with the program COOT. In the final stages of refinement, water molecules were added to the model with electron-density acceptance criteria of $d \geq 3\sigma(d)$ in the Fo–Fc difference electron-density map. Ligand molecules (KDN and VO_3^-) were added when the R_{free} was below 30%. A composite omit map with coefficients 2Fo-Fc was calculated using the program CNS.

BT1713 Glu56Ala and Glu56Ala Lys67Ala

Crystals of BT1713 Glu56Ala mutant were obtained by vapor diffusion using hanging drop geometry. Drops were equilibrated against reservoir solution containing 100 Mm Bis-Tris buffer, pH 6.2, 200 mM $MgCl_2$, and 23% PEG 3350. Drops contained equal volumes (1µl each) of seed stock solution suspended in mother liquor and 20 mg/mL protein in 50 mM Hepes buffer, pH 7.5 and 5 mM $MgCl_2$. Seed stock was prepared by adding one crystal to 500 µL of mother liquor and crushing using Seed Bead from Hampton Research. Crystals appeared within one day (0.170 x 0.075 x 0.03 mm) belonging to the space group $P222_1$. A single crystal was dragged through paratone to remove surface solvent for cryo-protection, and placed directly in the liquid nitrogen stream. Data was collected at a single wavelength (1.54 Å) on a Bruker MICROSTAR micro-focus rotating anode with Helios optics and Platinum 135 CCD detector, using a four circle kappa goniometer, at Boston University. Data was collected to 2.28 Å

resolution, reduced using SAINT and scaled using SADABS from the PROTEUM2 software suite [24] (**Table A.1**).

The structure of BT1713 mutants, E56A and E56A/K67A double mutant, were determined to 2.28 and 2.05 Å resolution respectfully by the molecular replacement method with the program MolRep in the CCP4 suite. The search model was the protomer of BT1713. KDN9PP-E56A/K67A is twinned (twin fraction = 0.035, twin law l, -k, h). Twin refinement was included in the rounds of refinement, resulting in improved maps than without twin refinement for KDN9PP-E56A/K67A. Otherwise, the refinements of KDN9PP-E56A were the same. Phenix refine was used to perform iterative cycles of atomic positional refinement, atomic displacement parameter refinement and real space atomic position refinement. Each round of Phenix was followed by manual inspection in COOT. Waters were added to the model with electron-density acceptance criteria of $d \geq 3\sigma(d)$ in the Fo–Fc difference electron-density map.

HI1679 (HI-KDO8PP)

Crystals of HI1679 (HI-KDO8PP) were obtained by vapor diffusion in hanging drops. Drops were equilibrated against reservoir solution containing 21% PEG 3550, 100 mM Tris, pH 8.5 at 17 °C. The hanging drops contained equal volumes (1 µl each) of the reservoir solution and an 11 mg/mL protein solution in 50 mM Hepes buffer, pH 7.6, 5 mM $MgCl_2$ and 1 mM DTT. Small crystals (0.15 mm x 0.1 mm x 0.05 mm) appeared in within 1 day. Crystals were soaked in 50 µL of 100 mM Tris pH 8.5, 30% PEG 3350, 20mM vanadate, and 20 mM KDO for 3 days at room temperature before data collection.

Surface solvent was removed by dragging the crystal through Paratone then flash-cooling in liquid nitrogen. The crystals belonged to the space group I4 with cell dimensions a= 79.846 Å b= 79.846 Å c= 52.151 Å. There was one molecule in the asymmetric unit and the solvent comprised 46% of the unit cell. Data was collected at a single wavelength (1.54 Å) on a Bruker MICROSTAR micro-focus rotating anode with Helios optics and Platinum 135 CCD detector, using a four circle kappa goniometer, at Boston University Department of Chemistry. Data was collected to 1.8 Å resolution, reduced using SAINT and scaled using SADABS from the PROTEUM2 software suite [24] (**Table A.1**).

The structure of ligand-bound HI-KDO8PP was determined at a 1.8 Å resolution by the molecular replacement method with the program MolRep in the CCP4 suite. The search model was the protomer of HI1679 (PDB ID: 1K1E with 7 C-terminal residues manually removed) [22], leading to a clear solution for the protomer. The presence of twinning was confirmed by examination of the cumulative intensity distribution, with a twinning fraction of 0.33 and twin law –k, -h, -l. The structure was refined using the program REFMAC [23] with the data between 29.46 and 1.8 Å, for which $F \geq 2\sigma(F)$. Restrained refinement with hemihedral twin refinement was performed. The resulting model was inspected and modified with the program COOT. In the final stages of refinement, water molecules were added to the model with electron-density acceptance criteria of $\delta \geq 3\sigma(\delta)$ in the Fo–Fc difference electron-density map. Vanadate and 2-keto-3-deoxy-octulosonic acid (KDO) were added to the model in the final round of refinement.

Determination of Steady State Kinetic Constants

BT1713 Activity

Reaction mixtures (1 mL) at 25 °C and initially containing BT1713 (0.04µM for KDN9P; 1.0µM for KDO8P), varying concentrations of KDN9P (50µM-560µM) or KDO8P (100µM-1500µM), 10 U of NeuNAc aldolase[9], 10 U of lactate dehydrogenase, 2 mM $MgCl_2$, and 0.2 mM NADH in 50 mM K^+HEPES (pH 7.0) were monitored at 340 m ($\Delta\varepsilon = 6200M^{-1}$ cm^{-1}). Initial velocity data were fitted to **equation 1** using the computer program KinetAsyst I in order to define the steady state kinetic constants.

$$V_0 = V_{max}[S]/([S] + K_m)\ (\textbf{Equation 1})$$

where V_0 is the initial velocity, V_{max} the maximum velocity, [S] the substrate concentration, and K_m the Michaelis-Menten constant for the substrate. The k_{cat} value was calculated from V_{max} and [E] according to the equation $k_{cat}=V_{max}/[E]$, where [E] is the protein subunit concentration in the assay reaction

Result and Analysis:

Bioinformatic Analysis

The HI-KDO8PP amino-acid sequence was used to search for additional KDN9PP or KDO8PP sequences using BLAST [24] within each phylum of bacteria. Sequences were separated using the specificity sequence identifier: KDO8PP T/SxGxxxxxxxRxxA, KDO8PP T/SxRxxxxxxxRxxD/E/T/S/Q/N, KDN9PP T/SxExxxx xxxRxxK, where T/S indicates motif 2 of the HADSF. With each sequence identified, the bacterial genome

was then solely searched for additional C0 HAD family members using the initially identified C0. Additional sequences were classified based on the above specificity sequence identifiers.

Variability Analysis

Protein variability was computed with multiple sequence alignments of KDN9PP, KDO8PP (with Gly, KDO8PP-Gly) and KDO8PP (with Arg, KDO8PP-Arg) using the Protein Variability Server [25]. Variability is calculated using the Shannon entropy equation [26], with a score of 0 indicating absolute conservation across all sequences and 4.322 indicating that all 20 natural amino acids are found at that position(a variability score ≤ 2 is considered conserved). A random selection of sequences from each group was chosen as input and results mapped onto the corresponding three-dimensional structures of BT-KDN9PP, BT-KDO8PP, and HI-KDO8PP. Electrostatic potential surfaces of BT-KDO8PP, BT-KDN9PP and HI-KDO8PP were calculated using GRASP [30]. KDO8P was manually docked into BT-KDO8PP structure using the phosphate binding site and overlays with HI-KDO8PP/KDO/VO_3^- as a guide. The structure was then minimized using an energy minimization molecular dynamics cycle (200 steps) in the program CNS [18].

Structural Comparison of BT-KDN9PP and BT-KDO8PP

The structure of unliganded B. thetaiotaomicron KDO8PP with Mg^{2+}-bound was solved and refined to a resolution of 1.8 Å (**Table A.1**) for comparison with BT-

KDN9PP, which has previously been solved (PDB ID 3E8M, [9]). BT-KDO8PP and BT-KDN9PP have 28% primary sequence identity and an overall RMSD of 1.6 Å for the protomer (**Figure A.3**). Both enzymes were crystallized as tetramers in the asymmetric unit, with the active site at the interface of two subunits, consistent with other KDO8PPs [15, 22]. Thus, one subunit acts as the HAD core or 'catalytic' domain, containing the HAD conserved motifs, and the other subunit acts as a 'cap' domain, and is expected to contribute residues to ligand the substrate leaving group. When comparing unliganded structures of KDN9PP with KDO8PP, there exists a very similar protomer structure and active-site makeup. Notable differences are seen in the C-terminus with the C-terminus of BT-KDN9PP being alpha helical and extended away from the core domain, while the C-terminus of BT-KDO8PP does not have clear electron density associated with it; possibly this temporal or spatial disorder indicates flexibility in this region. Two other KDO8PP structures have previously been solved from E. coli (PDB ID 2R8E [15]) and H. influenza (PDB ID 1K1E, [22]), both containing extended C-terminal tails which are juxtaposed to the active site of the adjacent monomer, indicating a possible role in functionality of the C-terminus of KDO8PPs in specificity, exclusion of bulk solvent or catalytic positioning of the substrate [15]. Indeed, a truncation mutant of the last seve167) yields a 15-fold reduction in kcat compared to wild-type BT-KDO8PP (**Table A.2**) with a 7 fold reduction in Km. Similar results are seen for E. coli KDO8PP [15]. This highlights the importance of the C-terminus to substrate turnover in KDO8PP enzymes.

Table A.12: BT-KDN9PP and BT-KDO8PP wild-type steady-state kinetics. Steady-state kinetic constants for BT-KDN9PP and BT-KDO8PP catalyzed hydrolysis of KDN9P or KDO8P in 50 mM K$^+$HEPES containing 2 mM MgCl$_2$ (pH 7.0, 25°C).

Enzyme	Substrate	k_{cat} (s^{-1})	K$_m$ (mM)	k_{cat}/K$_m$ (M^{-1}s^{-1})
BT-KDO8PP	KDN9P	0.061 ± 0.005	0.89 ± 0.19	69
	KDO8P	1.5 ± 0.3	0.10 ± 0.04	1.5 x 10^4
A167 truncation	KDO8P	0.10 ± 0.1	0.76 ± 0.01	1.3 x 10^2
BT-KDN9PP	KDN9P	1.2 ± 0.1	0.105 ± 0.09	1.1 x 10^4
	KDO8P	0.092 ± 0.003	0.43 ± 0.07	2 x 10^2
E56A	KDN9P	0.026 ± 0.002	0.41 ± 0.01	6 x 10^1
	KDO8P	0.109 ± 0.02	0.207 ± 0.09	5 x 10^2
E56A / K67A	KDN9P	0.004 ± 0.0008	0.99 ± 0.03	4.1
	KDO8P	0.062 ± 0.002	0.196 ± 0.009	3.1 x 10^2
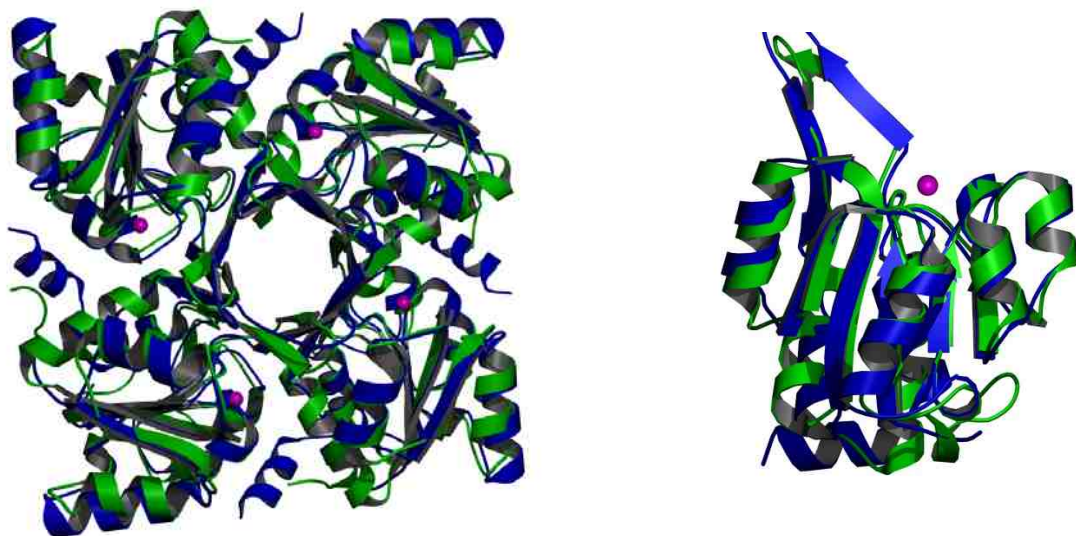
Figure A.47: Overlay of BT-KDN9PP and BT-KDO8PP. The RMSD of Cα of the monomers is 1.6 Å. BT-KDN9PP is shown in blue, BT-KDO8PP green (Mg^{2+} depicted as magenta sphere). figures generated using MOLSCRIPT [54]

Kinetic Comparison of BT-KDN9PP and BT-KDO8PP

With the similarities noted between these two enzymes, one could assume each enzyme would be able to readily dephosphorylate the sugar phosphate substrate of the other. Each enzyme shows high catalytic efficiency toward its physiological substrate with the $k_{cat}/K_m = 1.2 \times 10^4 \, M^{-1} \, s^{-1}$ for BT- KDN9PP against KDN9P, and $k_{cat}/K_m = 1.5 \times 10^4 \, M^{-1} \, s^{-1}$ for BT-KDO8PP against KDO8P (**Table A.2**, [9]). Steady-state kinetics of the two enzymes with the swapped substrate gives a large decrease in k_{cat}/K_M (KDO8PP has a 22 fold decrease in k_{cat}/K_m against KDN9P versus KDO8P, and KDN9PP has a 60 fold

152

decrease in k_{cat}/K_M against KDO8P versus KDN9P) (**Table A.2**). Clearly the enzymes are able to differentiate KDO8P and KDN9P via hydrogen-bonding, electrostatic, or steric interactions with the substrate.

Structural Analysis of BT-KDN9PP with KDN and VO$_3^-$

Analysis of the ligand bound structure clarifies specificity determinants by highlighting primary structure differences between BT-KDN9PP and BT-KDO8PP. The most obvious difference is found in two BT-KDN9PP ligand binding residues. Glu56 of the core subunit and Lys67* of the adjacent (cap) subunit in BT-KDN9PP occur as Gly63 and Ala74* respectively in BT-KDO8PP. Met20 in KDN9PP is Val in KDO8PP. These three differences open up a pocket in BT-KDO8PP for ligand binding. Also, Asn106 in KDN9PP is Pro113 in KDO8PP. These changes, which potentially could affect the ability of KDO8PP to discriminate against KDN9P were thus tested by site-directed mutagenesis.

Structure Determination and Kinetics of E56A and E56A/K67A BT-KDN9PP

The X-ray crystallographic structure of E56A and E56A/K67A BT-KDN9PP mutants were solved (**Table A.3**) to demonstrate that the mutations do not affect overall stability of the enzyme. Both structures are similar to wild-type BT-KDN9PP, with RMSD values of 0.23 Å, and 0.0 Å for E56A and E56A/K67A, respectively (**Figure A.4**). We can conclude that changes seen in activity (kcat and Km) are solely due to the replacement of a residue rather than global structural changes.

153
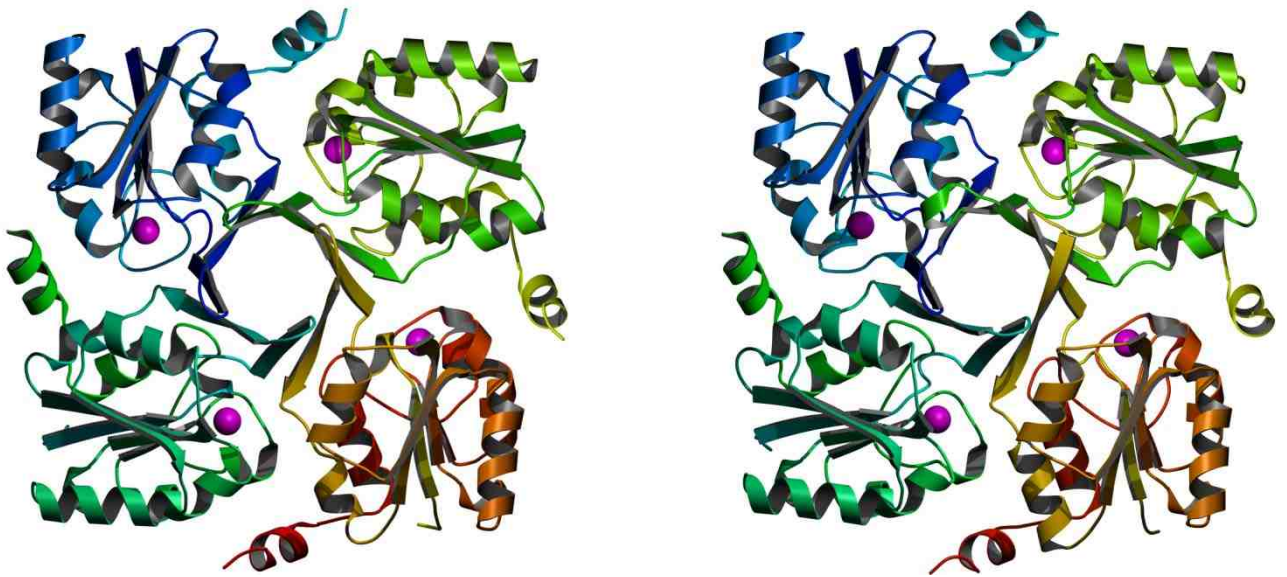
Figure A.48: BT-KDN9PPmutant structures. Rainbow colored ribbon diagram of BT-KEN9PP Glu56Ala (right panel) and BT-KDN9PP Glu56Ala / Lys67Ala (left panel). Magnesium is shown as magenta spheres.

Table A.13: BT-KDN9PP mutant data collection and refinement statistics. Data collection and refinement statistics for BT-KDN9PP Glu56Ala and Glu56Ala / Lys67Ala. Values in parentheses are for the highest resolution shell.

Data Collection	BT-KDN9PP-Glu56Ala	BT-KDN9PP-Glu56Ala-Lys67Ala
Space group	$P222_1$	$P2_1$
Cell dimension (Å)	a=88.94 b=94.12 c=161.50	a=72.37 b=116.77 c=75.18
Molecules / ASU	8	8
Wavelength (Å)	1.54	1.54
Resolution (Å)	2.28	2.05
Observed reflections	388,814	832,804
Unique reflections	62,511	76,062
Completeness (%)	99.9 (100)	97.4 (92.7)
R_{merge} (%)	8.78 (46.48)	13.08 (47.69)
$<I/\sigma>$	11.6 (2.2)	14.12 (3.73)
Redundancy	6.21	10.66
Twin Law	None	l, -k, h
Twin Fraction	0.0	0.067
Refinement		
R / R_{free}	19.54 / 24.93	21.27 / 25.29
Average B Factor ($Å^2$)	25.76	19.05
Rms deviation from ideal		
Bonds	0.007	0.006
Angles	1.015	0.941

Steady-state kinetics were previously reported for BT-KDN9PP-E56A mutant against the substrate KDN9P [9]. The BT-KDN9PP-K67A protein is unstable and could not be assayed. The BT-KDN9PP-E56A is a stable mutant, with $k_{cat}/K_m = 6 \times 10^1 \ M^{-1} \ s^{-1}$ against KDN9P as substrate **(Table A.2)** showing a 200 fold decrease compared to wild-type. When KDO8P is supplied as substrate, $k_{cat}/K_m = 5 \times 10^2 \ M^{-1}s^{-1}$ (2 fold increase compared to wild-type). With this single mutation, the specificity of BT-KDN9PP has

shifted to favor KDO8P over KDN9P (ratio of KDO8P/KDN9P k_{cat}/K_m is 8 fold, **Table A.2**). Double mutants of E56A and K67A were assayed to test for further enhancement of the specificity switch between the two sugars. The E56A/K67A mutant of KDN9PP is 80 fold more specific for KDO8P over KDN9P (k_{cat}/K_m of 3.1 x 10^2 *vs.* 4.1 $M^{-1}s^{-1}$, respectively, **Table A.2**). The change in specificity is due to a large loss of binding affinity of KDN9P as reflected by K_m, due to the absence of critical ligand binding residues (Glu56 and Lys67).

BT-KDO8PP and BT-KDN9PP Analysis

Together the X-ray crystallographic and mutational kinetic data highlight the importance of two critical residues within each active site. In KDN9PP, Glu56 and Lys67 provide key ligand stabilizing interactions to position the substrate KDN9P for catalysis (**Figure A.5**). When these residues are mutated to Ala, the selectivity of the KDN9PP changes to prefer KDO8P as substrate. This correlates with the presence of Gly and Ala at the homologous positions in KDO8PP, indicating their importance in substrate specificity. The production of KDN is uncommon in bacteria, and it is normally found in eukaryotic organisms [27], with the exception of a few bacteria such as *S. fredii* and *K. ozaenae* [28-29]. The pathway utilized in these bacterium has not been studied. It is thus interesting that *B. thetaiotaomicron* has acquired the ability to produce KDN, probably through gene duplication. It is tempting to speculate that environmental pressures played a role in this evolution. *B. thetaomicron* is a human gut symbiont, the most abundant found in human feces [5, 30]. Because Lipid A, of which KDO is an integral part, is a

potent stimulator of the human innate immune response [31], *B. thetaiotaomicron* must circumvent immunity in order to survive in the gut. It is known in other bacterial species (eg. *H. influenzae, Salmonella O48)* that sialyated bacterial cells do not up regulate the innate immune response [32-34]. The use of KDN to sialate the *B. thetaiotaomicron* cell surface and mimic human cells would prevent eliciting the innate immune response to Lipid A.
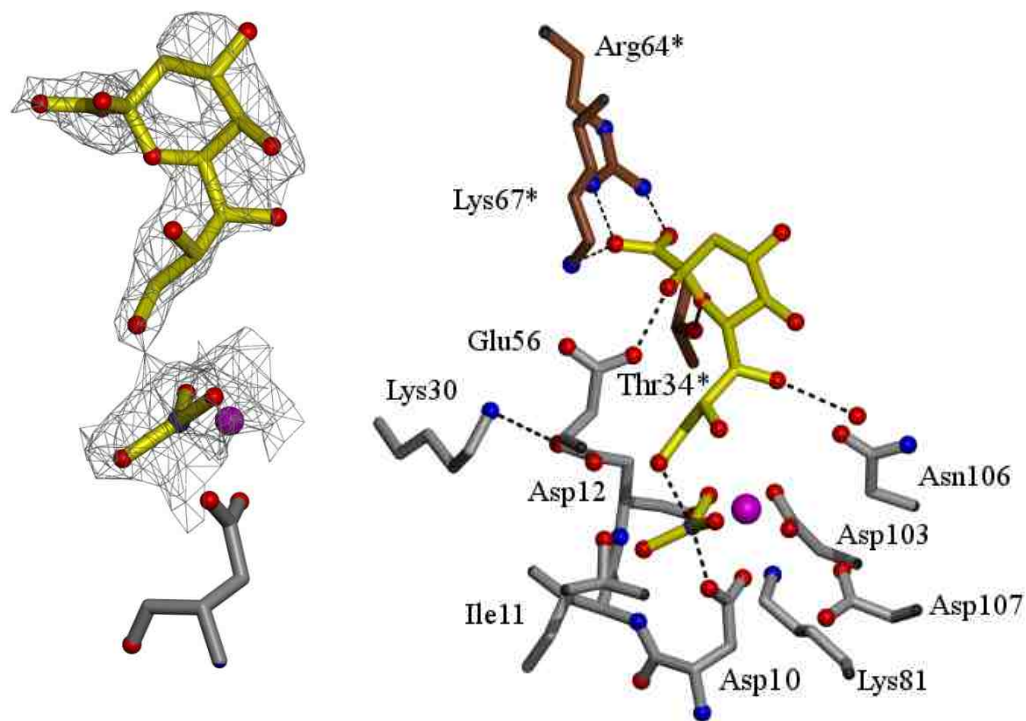
Figure A.49: Structure of KDN9PP in complex with KDN and VO_3^-. 2Fo-Fc composite omit electron density map contoured at 1.0 sigma (dark gray cages, left panel) shown for KDN and VO_3^-,(yellow), and Mg^{2+} is depicted as magenta sphere, residue Asp10 shown in gray. Hydrogen bonds to KDN (ball and stick, yellow) are shown (dashed line, right panel). The catalytic domain (gray residues) and 'cap domain' (brown residues with *) are shown, otherwise colored as in left panel.

Bioinformatic Analysis

To elucidate the evolutionary relationship between KDO8PP and KDN9PP enzymes, bioinformatic techniques were employed to analyze available bacterial genome

sequences in NCBI. 6 displays a representative sample of the sequences collected and analyzed, separated by specificity marker. Sequences were separated using the specificity markers for BT-KDO8PP, and BT-KDN9PP determined herein (*vide supra*). Out of a total of 4,303 bacterial genome sequences available, a random sample of 687 sequences was analyzed for KDO8PP and KDN9PP by phylum. It is important to note that over 2,000 genomes are found in the proteobacteria phylum alone and thus was sub-divided into classes. All bacteria labeled as non-classified were excluded from analysis.

All species in the Bacteroides phylum contain a KDO8PP gene with Gly and Ala at the two specificity residue positions. Only a subset of Bacteroides species (11 out of 77 analyzed) also contain a KDN9PP gene, confirmed by the presence or absence of KDN9P synthase, the preceding enzyme in the KDN pathway [12, 14], which is found adjacent to the KDN9PP gene. KDN9PP orthologues were identified in a subset of the Bacteroides phylum, and in the Chlorobi phylum (**Table A.4**). The Chlorobi phylum contains KDO8PP enzymes with Arg at the first specificity position and consists of green sulfur photoautotrophic bacteria which are considered most similar to Bacteroides [35], and the need for KDN production is not clear. All other gram-negative bacteria do not contain a gene for KDN9PP, as assessed by BLAST searches of each bacterial genome for which a KDO8PP gene was identified.

.

Table A.14: Bioinformatic Summary. All phyla of bacteria are listed with number of genome sequences analyzed, percent of sequences in phyla analyzed, number of C0

Phylum	Genomes	Percent	C0	Specificity	Enzyme
Actinobacteria	25	6.8	1	E	KDN9PP
Aquificae	9	75	1	R	KDO8PP
Bacteroides	77	54	1 or 2	G	KDO8PP
				E	KDN9PP
Chlorobi	12	86	2	R	KDO8PP
				E	KDN9PP
Chlamydiae	6	13	1	R	KDO8PP
Cyanobacteria	16	12	1	G / S	KDO8PP
Deferibacteres	2	100	1	R	KDO8PP
Elusomicrobia	2	40	1	R / G	KDO8PP
Fibrobacteres	3	75	2	R	KDO8PP
				E	KDN9PP
Firmicutes	17	1.5	1	R	KDO8PP
Fusobacteria	20	65	1	K	KDO8PP
Gemmatimonadetes	1	100	1	R	KDO8PP
Alphaproteobacteria	5	1.1	1	E	KDN9PP
Betaproteobacteria	98	32	1	R	KDO8PP
Gammaproteobactera	288	26	1	R	KDO8PP
Delta/Epsilon	88	56	1	R	KDO8PP
Zetaproteobacteria	1	100	1	R	KDO8PP
Nitrospirae	2	100	1	R	KDO8PP
Planctomycetes	5	83	1	R	KDO8PP
Spirochaetes	4	1.3	1	N / S	KDO8PP
Synergistetes	6	100	1	R	KDO8PP

gi|29347087|ref|NP_810590.1|
MRTVNIKDGYAIQLAVKKGLHIAII**TG**G**RTEAVRIRFAALGVK--DLYMG 83
gi|265765968|ref|ZP_06094009.1
MRTVNVKDGYAIQLAVKKGLRIAII**TG**G**RSDVVRKRFIGLGVS--DLYFG 83
gi|150005032|ref|YP_001299776.
MRTVNIKDGYALQLAVKCGLHVAII**TG**G**NTEAVRKRYEGLGIK--DVYLA 81
gi|237713692|ref|ZP_04544173.1
MRTVNIKDGYAIQLAVKKGLHIAII**TG**G**RTEAVRIRFEGLGVK--DLYMG 83
gi|224540513|ref|ZP_03681052.1
MRTVNIKDGYALHLACKQGLLLGII**TG**G**RSEAVRKRFMALGIPSEDIYMA 85
gi|262340978|ref|YP_003283833.
VRQMFAKDGYAMQLAKKKGYNLCII**TR**G**SDLMVFRRLRGLNIR--YIYQG 79
gi|167763293|ref|ZP_02435420.1
LRTVNIKDGYALHLAARHEIPLAII**TG**G**RTEAVRKRFRALGILPENIYMG 85
gi|288928245|ref|ZP_06422092.1
LRTVNIKDGYAIQFAQKVGLRICII**TG**G**DTKAVRKRFEGLGVE--DIYMK 81
gi|218130320|ref|ZP_03459124.1
LRTVNIKDGYALHLAAKHEIPLAII**TG**G**RTEAVRRRFLALGIPAENIYMG 85
gi|198276869|ref|ZP_03209400.1
MRSVNIKDGYALQLAVKCGLHVAII**TG**G**KTEAVRKRYEGLGIK--DVYLG 81

gi|29347123|ref|NP_810626.1|
WKKFNTSDSAGIFWAHNKGIPVGIL**TG**E**KTEIVRRRAEKLKVDY--LFQG 76
gi|294970963|gb|EFG46856.1|
SVRVHRGDGMGVSRLVKAEFPFLIL**SK**E**RNPVVTRRAEKLRVE---VAQG 84
gi|254403203|ref|ZP_05018149.1
FVSVHRGDGLGIAALRDSGLKMLIL**ST**E**QNPVVAARARKLKIP---VLHG 86
gi|72160415|ref|YP_288072.1|
TVVVNRGDGMGVSLLLKAGIRVCIL**ST**E**VNSVVRARARKLGVP---VMHG 85
gi|239979820|ref|ZP_04702344.1
RVSVHRGDGLGIAALRDAGIPLLIL**ST**E**QNPVVTARARKLRIP---VLHG 83
gi|29829917|ref|NP_824551.1|
FVSVHRGDGLGIAALRKSGLTMLIL**ST**E**QNPVVAARARKLKIP---VLHG 84
gi|271964804|ref|YP_003339000.
MVAVSRSDGMGIALLRRSGVKLMIM**ST**E**HNPVVAARARKLGVP---VLQG 83
gi|256778605|ref|ZP_05517068.1
LVAVHRGDGLGIAALRRAELALLIL**ST**E**KNPVVAARGRKLQVP---VLHG 84
gi|254377988|ref|ZP_04993443.1
LVSVHRGDGLGIAALRRSGLPLLIL**ST**E**QNPVVAARARKLRVP---VLHG 84
gi|78189040|ref|YP_379378.1|
LKRYSIRDGMGVERLRNAGIETCIM**TG**E**RSPNVQKRAEKLCMKW--LYLG 88

gi|16273566|ref|NP_439821.1|
IKSFHVRDGLGIKMLMDADIQVAVL**S**G**R**DSPILRRRIADLGIK--LFFLG 80
gi|149190591|ref|ZP_01868860.1
LKTFHTRDGYGIKALMSAGVEVAII**T**G**R**RSQIVENRMTALGIS--LIYQG 95
gi|260857325|ref|YP_003231216.
LKAFNVRDGYGIRCVLTSDIEVAII**T**G**R**KAKLVEDRCATLGIT--HLYQG 98
gi|170765508|ref|ZP_02900319.1
LKAFNVRDGYGIRCALTSDIEVAII**T**G**R**KAKLVEDRCATLGIT--HLYQG 98
gi|54310339|ref|YP_131359.1|
LKTFHTRDGYGIKSLMNAGIEIAII**T**G**R**QSAIVENRMTALGIK--HIYQG 95
gi|24114487|ref|NP_708997.1|
LKAFNVRDGYGIRCALTSDIEVAII**T**G**R**KAKLVEDRCATLGIT--HLYQG 98
gi|168231927|ref|ZP_02656985.1
LKAFNVRDGYGIRCALTSNIEVAII**T**G**R**KAKLVEDRCATLGIV--HLYQG 98
gi|15803738|ref|NP_289772.1|
LKAFNVRDGYGIRCALTSDIEVAII**T**G**R**KAKLVEDRCATLGIT--HLYQG 98
gi|218691488|ref|YP_002399700.
LKAFNVRDGYGIRCALTSDIEVAII**T**G**R**KAKLVEDRCATLGIT--HLYQG 98
gi|262273848|ref|ZP_06051661.1
LKAFHTRDGYGVKSLMNAGVEIAII**T**G**R**KSAIVERRMAALGIQ--HIYQG 95

Figure A.50: Sample alignment of KDO8PP and KDN9PP sequences from bacteria. HAD motif two is colored green, specificity residues are highlighted in blue (KDO8PP-Gly), pink (KDN9PP, Glu), and purple (KDO8PP-Arg).

Within the Gram-positive genomes, bacteria in the Actinobacteria phylum were found to possess KDN9PP genes, all with Glu at the first specificity position, and Lys at the second specificity position. The absolute conservation of the two specificity residues in KDN9PPs is consistent with its importance in ligand binding, as seen in the BT-KDN9PP/VO$_3$/KDN structure. All KDN9PPs identified in the Actinobacteria phylum are found as fusion proteins to a CMP transferase domain, that catalyzes the transfer of KDN to CMP [36]. The Actinobacteria phylum consists mainly of *Streptomyces* (the largest

162

genus, 21 genomes out of 25 identified in our bioinformatic analysis), which utilize anionic cell-wall polymers containing KDN [37-40]. As expected, no other bacterial species contained the additional KDN9PP gene, with few exceptions (see below). The absence of KDN9PP and other KDN synthesis genes correlates well with the rare appearance of KDN in bacteria.

Surprisingly, further genome searching indicated an alternative KDO8PP present in all other gram-negative bacteria, with Arg present at the first specificity residue and a variety of residues (Asp, Glu, Thr, Ser, etc) at the second specificity position. Thus, the enzyme observed over these bacteria is dissimilar to the *B. thetaiotaomicron* KDO8PP enzyme under study here.

The presence of a KDO8PP bearing Arg at this position indicates that the Bacteroides phylum divergently evolved with respect to KDO8PP from all other gram-negative bacteria. The divergent evolution possibly occurred through selective pressure to retain KDO8P phosphatase activity, while still allowing for variation of substrate binding residues to limit activity against other sugars present in the Bacteroides metabolome such as KDN9P.

Exceptions to the Rule

The class of Betaproteobacteria within the Proteobacteria phylum consists of aerobic or facultative anaerobic bacteria. In general, the bacterial species contain one C0 phosphatase, which is annotated as a KDO8PP with Arg at the first specificity position. A subset of the *Burkholderia* species contains two C0 phosphatases, both with sequences

consistent with assignment as KDO8PPs, the first enzyme bearing Arg at the first specificity position, and the second enzyme bearing Gly at the first specificity position. It appears that the *Burkholderia* species have evolved to contain two KDO8PP enzymes, which we cannot explain in terms of physiological function. Burkholderia are known to have a very potent LPS causing an upregulation of the innate immune response [41].

The *Francisella* species in the Gammaproteobacteria contain a single C0 phosphatase with Lys at the first specificity position. We propose that they are KDO8PPs because the Francisella species are shown to have the typical gram-negative Lipid A moiety [42]. The Fusobacteria phylum also contains KDO8PPs with Lys at this position. Fusobacteria are most similar to Bacteroides species, with a typical Lipid A moiety, but are noted to possess an exceptionally virulent LPS [43-44]. We propose that these Francisella and Fusobacteria enzymes identified are KDO8PPs, as the specificity residue was conservatively replaces from Arg to Lys , and no KDN9P synthase genes were identified within these genomes. The Acidobacteria class of Fibrobacteres phylum contains both KDO8PP (with Arg) and KDN9PP (with Glu). Acidobacteria are a highly acidic soil bacterium, and have only 4 total genome sequences available, so it is unclear the role the KDO or KDN plays in cell survival or defense.

Additional KDO8PP and KDN9PP Kinetic Analysis

We chose to functionally characterize KDO8PP enzymes with alternate residues present at the first specificity residue (these were additionally targeted for functional characterization by the Enzyme Function Initiative, http://www.enzymefunction.org).

The structures of *E. coli*, and *H. influenzae* KDO8PPs have previously been reported [15, 22] as well as the kinetics constants for the *E. coli* enzyme against KDO8P and are used for comparison. *L. pneumophila*, *P. syringae* and *V. cholera* are found in the gammaproteobacteria class of proteobacteria, the largest phylum of bacteria. *L. pneumophila subsp. pneumophila str. Philadelphia 1* KDO8PP contains an Ala at the first specificity position, Gln at the second specificity position, and is one of the exceptions to the rule we have observed. Other *L. pneumophilia* strains contain an Arg at the first specificity position.

The k_{cat} and K_m of each enzyme confirms its assignment as a KDO8PP orthologue. *L. pneumophila subsp. pneumophila str. Philadelphia 1* KDO8PP (LP-KDO8PP) is specific for KDO8P as substrate with $k_{cat} / K_m = 1.7 \times 10^3$ $M^{-1}s^{-1}$ (compared to $k_{cat} / K_m = 1.8 \times 10^1$ $M^{-1}s^{-1}$ for KDN9P), with a similar activity to BT-KDO8PP (**Table A.5**). *E. coli, P. syringae* and *V. cholera* KDO8PPs all contain an Arg at the first specificity position and a variety of residues at the second specificity position (Thr, Ser. Ala respectively).The kinetics show that each enzyme is specific for KDO8P (**Table A.5**), with k_{cat} / K_m ranging from 10^3 to 10^5 $M^{-1}s^{-1}$ and selectivity over KDN9P of 100 to 4,500 fold. Steady-state kinetics also confirm that *H. influenzae* KDO8PP (HI-KDO8PP) is a KDO8PP with a $k_{cat}/K_m = 4.8 \times 10^3$ M^{-1} s^{-1} towards KDO8P (**Table A.5**) and contains specificity residues Arg and Asp. HI-KDO8PP with KDN9P provided as substrate yields a $k_{cat}/K_m = 1.3$ M^{-1} s^{-1}, thus HI-KDO8PP is 370 fold more specific for KDO8P over KDN9P. This is a significant difference from the BT-KDO8PP (22 fold more specific for KDO8P over KDN9P) and indicates a significant role of Arg in the specificity of HI-

KDO8PP. The variability of the second specificity position in KDO8PPs indicates it plays a minor role in enzyme specificity.

Table A.15: Steady-state kinetics to assign KDO8PP and KDN9PP enzymes. Steady-state kinetic constants for enzyme catalyzed hydrolysis of KDN9P or KDO8P in 50 mM K^+HEPES containing 2 mM $MgCl_2$ (pH 7.0, 25°C).

Phylum /	Substra	k_{cat}	K_m	k_{cat}/K_m	ratio	Spec	PDB
Actinobacteria /	KDO8P	9.4 ±	1.2 ±	7.7 x 10^2	1.6	Glu /	3MMZ
	KDN9P	20 ±	1.7±	1.2 x 10^3			
Gammaproteobac	KDO8P	1300 ±	1.7 ±	7.7 x 10^5	4530	Arg	2R8E
	KDN9P	7.7 ±	4.6 ±	1.7 x 10^2		/ Th	
Gammaproteobac	KDO8P	78 ± 4	1.6 ±	4.8 x 10^3	370	Arg	1K1E
	KDN9P	2.1 ±	16 ± 5	1.3 x 10^1		/	
Gammaproteobac	KDO8P	8 ± 0.4	0.5 ±	1.7 x 10^3	100	Ala /	3N1U
	KDN9P	1.8 ±	10 ± 1	1.8 x 10^1		Gl	
Gammaproteobac	KDO8P	60 ±	0.5 ±0.	1.2 x 10^4	670	Arg	3MN1
	KDN9P	1.7 ±	9.6 ±	1.8 x 10^1		/ S	
Gammaproteobacteria / V. *cholerae* / Q9KP52	KDO8P	8.9 ±	0.19	4.7 x 10^3	200	Arg / Ala	3N07
	KDN9P	2.5 ± 0.1	11 ± 1	2.3 x 10^1			

An enzyme from *S. avermitillis* mis-annotated as KDO8P was found to possess the signature determinants for KDN9PP. Kinetic characterization demonstrates the SA-KDN9PP is specific for KDN9P with a $k_{cat} / K_m = 1.2 \times 10^3 \text{ M}^{-1}\text{s}^{-1}$. Notably, SA-KDN9PP occurs as a fusion protein with CMP transferase. Overall, then, the identity of KDN9PP enzymes can be based on Glu and Lys markers, and KDO8PP enzymes based on Gly and Ala or Arg and Asp/Ser/Ala/Gln/Thr markers.

Additional Structural Data

To correlate the bioinformatic and kinetic data, with structural information of KDO8PP and KDN9PP enzymes were compared (PDB ID 2R8E (*E. coli, EC-KDO8PP*), 1K1E (*H. influenzae, HI-KDO8PP*), 3MN1 (*P. syringae, PS-KDO8PP*), 3N1U (*L. pneumophila, LP-KDO8PP*), 3N07 (*V. cholera, VC-KDO8PP*), 3MMZ (*S. avermittillis, SA-KDN9PP*).

Structural overlays of all available KDO8PP and KDN9PP enzymes show the three-dimensional location of the putative specificity residues in the active site is retained, (**Figure A.7**). Additionally the location always maps in primary structure with the first specificity residue positioned two residues after HADSF motif two (Ser/Thr), and the second specificity residue positioned 11 amino acids after the first. Overall, specificity residues in KDO8PP and KDN9PP enzymes throughout bacteria are conserved in primary and tertiary structure.
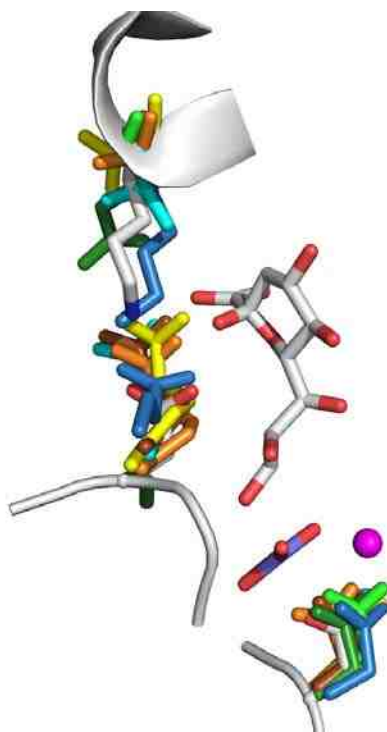
Figure A.51: Overlay of all KDO8PP and KDN9PP specificity residues. Side chains of motif one Asp and both specificity residues (indicated in parentheses) shown for EC-KDO8PP (Arg78, Thr89, cyan), HI-KDO8PP (Arg60, Asp71, yellow), PS-KDO8PP (Arg69, Ser80, orange), LP-KDO8PP (Ala71, Gln82, forest green), VC-KDO8PP (Arg75, Ala86, brown), SA-KDN9PP (Glu64, Lys75,blue), BT-KDO8PP (Gly63, Ala74, green), BT-KDN9PP (Glu56, Lys67, gray with oxygen colored red and nitrogen colored blue) active overlayed. KDN (gray), VO_3^- (slate), Mg^{2+} (magenta sphere), and ribbon diagram of BT-KDN9PP shown for reference. Figure generated using PyMol

168

Structural Analysis of HI-KDO8PP with KDO and VO$_3^-$

The functional importance of Arg at the first specificity position is highlighted by the structure of the complex of HI-KDO8PP with KDO and VO$_3^-$ as a transition state mimic, which was solved to 1.8 Å (**Figure A.8**). Two conformations of KDO were found in the crystal structure, each modeled with 50% occupancy (**Figure A.8**). The second conformer is not indicative of KDO8P binding, as the terminal OH forms a hydrogen bond to oxygen of VO$_3^-$, and was thus not included in analysis. Structural analysis shows three charge-charge interactions and two hydrogen bonds to the leaving-group sugar. Arg60 makes a mono-dentate salt bridge to an oxygen of the substrate carboxyl (2.8 Å, **Figure A.8**) revealing the basis of the functional importance of this residue in the binding of substrate. The second specificity residue, Asp71, does not interact with ligand, which correlates with the variability of that residue found in our bioinformatic data. The other important ligand-protein interactions include a bi-dentate salt bridge from Arg 68 to the carboxyl group of KDO (2.8 and 2.9 Å), and a hydrogen bond between the carbonyl of Val38 and Met178 to O2 (2.5 Å) and O4 (3.2Å), respectively. Also, a water molecule forms a bridging hydrogen bond to O7 of KDO (3 Å) and Asp108 (2.8 Å) of HI-KDO8PP. The C6-C8 carbon chain of KDO is packed neatly against the side chain methylene carbons of Arg60. The C-terminus of HI-KDO8PP (from the 'cap' subunit) is found adjacent to the active site, effectively blocking solvent from the binding pocket (the final residue, Gln180 could not be modeled into density). The backbone oxygen of Met178 forms a hydrogen bond with O3 of KDO. This structure, combined with C-terminal deletion mutant kinetics for BT-KDO8PP and the *E. coli* KDO8PP (*vide infra*

169

and [18]) demonstrates that the C-terminus of KDO8PP plays a role in positioning ligand, allowing effective catalysis. Previously, the only ligand bound KDO8PP structure available was with a C-terminal truncation [15].
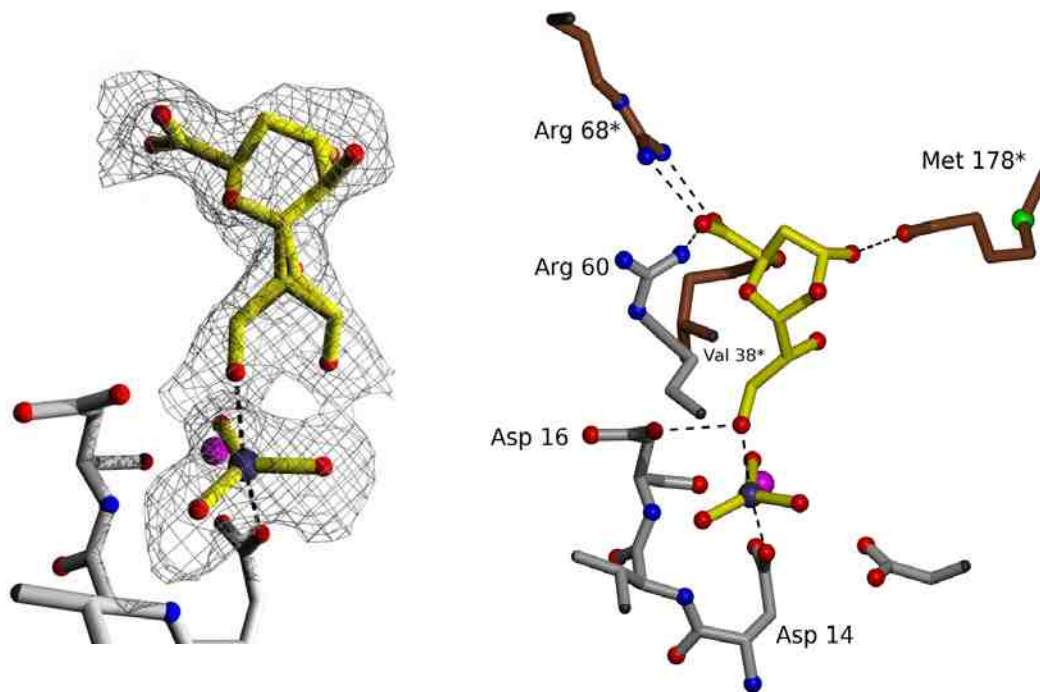
Figure A.52: Structure of HI-KDO8PP in complex with KDO and VO$_3^-$(yellow) is shown and Asp14, Val15 and Asp16 (motif one, gray, left panel). Hydrogen bonds (dashed line) between KDO the catalytic domain (gray residues) and 'cap domain' (brown residues with *) are shown. Magnesium is represented as a magenta sphere and vanadium is represented as a slate blue sphere. Figures generated using MOLSCRIPT and POVS CRIPT [54-55].

Both structure and kinetics of KDO8PP from *H. influenzae, E. coli* and *B. thetaiotaomicron* demonstrates the function of the C-terminus in binding substrate and

170

providing solvent exclusion. In contrast, the C-terminal tail of BT-KDN9PP does not

stabilize substrate; structural analysis shows it is positioned far from the active site. The

C-terminal tail of KDN9PPs and KDO8PPs are functionally disparate and represent a key

difference not observable via sequence alignment alone.

The other contribution to shape of the binding pocket is the specificity determinant

residues. HI-KDO8PP has Arg60 (Asp71, the second specificity marker does not interact

with substrate) and BT-KDN9PP has Glu56 and Lys67, which directly form hydrogen

bonds to the substrate. BT-KDO8PP has Gly63 and Ala74 at these positions, significantly

altering the shape of the active site compared to HI-KDO8PP and BT-KDN9PP. With

respect to the size and shape of the binding cleft, HI-KDO8PP is more similar to BT-

KDN9PP than BT-KDO8PP. This provides explanation as to why HI-KDN9PP can

accommodate KDN9P as a substrate (**Table A.5**).

Our data indicate that KDO8PP has evolved to retain functionality while altering

substrate binding. This is confirmed by our kinetic and structural analysis. KDO8PP

either provides Arg as a direct ligand for KDO8P, or Gly to provide an altered binding

pocket that is catalytically efficient toward KDO8P.

Variability Analysis

To identify any other possible conserved specificity residues, the sequence entropy

of each set of KDN9PP and KDO8PP enzyme sequences was determined. A

171

multiple sequence alignment of all KDN9PP sequences (44 available sequences from Actinobacteria, Bacteroides and Chlorobi phyla) was utilized to calculate entropy which was mapped our structure of BT-. For KDO8PP with the Gly marker (KDO8PP-Gly), sequences from Bacteroides (47) and Cyanobacteria (11) were aligned, with results mapped onto the BT-KDO8PP structure. For KDO8PP with the Arg marker (KDO8PP-Arg), selected sequences from all other phyla were aligned with results mapped onto the HI-KDO8PP structure.

Variability analysis confirms the identity of KDN9PP specificity residues, as expected, and identifies Ser 37 (BT-KDN9PP numbering) as absolutely conserved (**Figure A.9**). Ser37 does not contribute to ligand binding in the structure presented above (4.2 Å from KDN), it may affect the polarity of the active site.
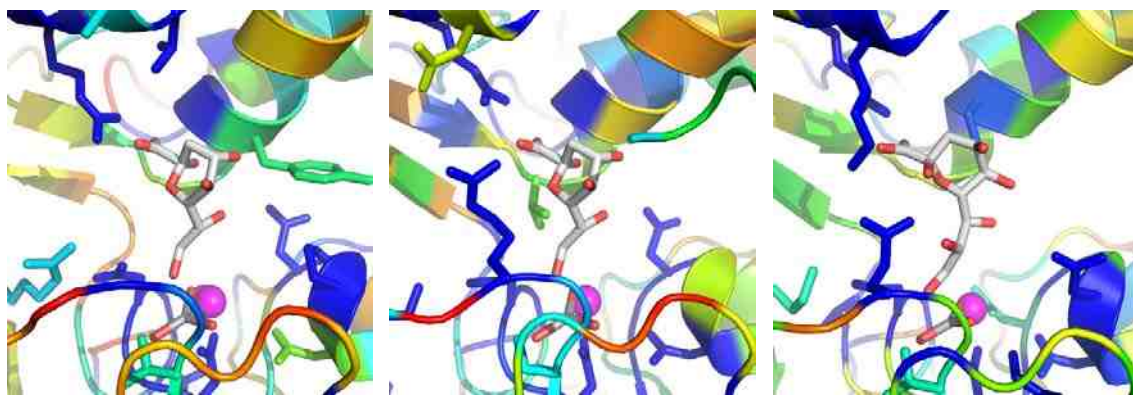
Figure A.53: Variability of KDO8PP-Gly, KDO8PP-Arg and KDN9PP. Ribbon diagram surrounding the active site of each structure: KDO8PP-Gly variability mapped onto BT-KDO8PP (left panel, KDO (yellow sticks) is docked), KDO8PP-Arg variability mapped onto HI-KDO8PP (middle panel), and KDN9PP variability mapped onto BT-KDN9PP (right panel). HAD motifs and specificity residues are shown as sticks.

The variability analysis of KDO8PP-Gly shows that Arg38 (BT-KDO8PP numbering) is conserved throughout Bacteroides and is conservatively replaced in the Cyanobacteria (6 / 12 Arg, 5 / 12 Lys). Likewise, in KDO8PP-Arg the orthologous position is absolutely conserved as Lys throughout all phyla. The functionality of this Lys/Arg residue is to position the acid/base Asp residue motif one). In KDO8PP-Arg proteins, Lys is excluded from solvent by the specificity residue Arg, in comparison to KDO8PP-Gly proteins, Arg/Lys is exposed to solvent due to the presence of Gly at the specificity position. This can be seen by comparing electrostatic surfaces of BT-KDO8PP with HI-KDO8PP calculated by GRASP [25] (**Figure A.10**). The insertion of a positive charge into the KDO8P binding site provides a potential ligand for the KDO8P carboxylate. Another mechanism could be to alter the electrostatic environment of the active site to favor negatively charged ligand. Docking models of KDO8P into BT-KDO8PP show that it is plausible to suggest KDO8P can form a weak charge charge interaction with Arg38 and Arg71 (**Figure A.10**, 4 Å and 4.4 Å respectively). Arg38 is present behind KDO8P carboxyl, and Arg71 is found to the left of Arg37 in **Figure A.8**. cannot visibly be docked into the same position. We hypothesize that Arg/Lys 38 is contributing to ligand binding or altering the electrostatic environment of the active site to make it more favorable for KDO8P binding, which has a negative charge due to the carboxylate. This mechanism feasibly allows for selectivity of KDO8P over KDN9P. Thus KDO8PP-Gly enzymes have possibly evolved a secondary mechanism for ligand binding, alternative to KDN9PP and KDO8PP-Arg.
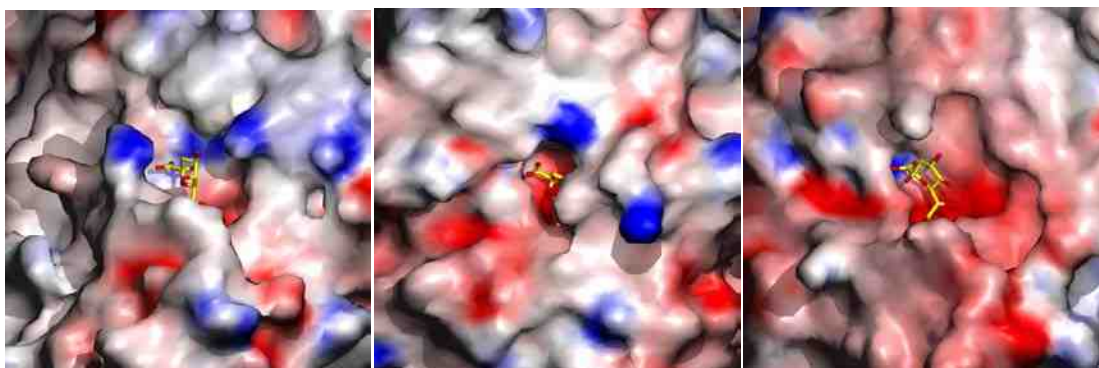
Figure A.54: Electrostatic potential of BT-KDN9PP, BT-KDO8PP and HI-KDO8PP active site. Electrostatic surface potential generated using GRASP [30]. BT-KDO8PP surface is shown with KDO8P docked (yellow sticks, left panel), HI-KDO8PP surface is shown with KDO for positioning reference (middle panel), and BT-KDN9PP surface shown with KDN (right panel).

The active site design does not eliminate KDN9P phosphatase activity in BT-KDO8PP, as confirmed by kinetics (**Table A.2**). This can be explained by the presence of Arg71 in BT-KDO8PP, which corresponds to Arg 64 in BT-KDN9PP, and Arg 68 in HI-KDO8PP, both of which form a bi-dentate salt bridge to ligand (**Figure A.5**, **Figure A.8**). Variability analysis shows that this residue is conserved in all three enzymes (KDN9PP – 0, KDO8PP-Gly – 0.112, KDO8PP-Arg – 0.094) and Arg is always found in the same position in primary structure, eight residues downstream from the first specificity residue. The presence of Arg at this position allows for KDN9P to weakly bind to BT-KDO8PP in the absence of the specificity residues Glu and Lys.

Phylogenetic Correlation

Recent phylogenetic proteome analysis provides a bacterial tree of life for comparison with our findings [43]. The Bacteroides phylum is shown as a divergence on the path to gram-negative bacteria, namely the proteobacteria phylum. Our data correlate well with this tree of life, indicating that at some point in time a divergence occurred, where the Bacteroides phylum evolved separately from other gram-negative bacteria, possibly due to co-habitation with humans. Our data indicate that bacteria which are symbiotic with humans contain both KDO8PP and KDN9PP genes (e.g. Bacteroides and Chlorobi), while pathogenic bacteria tend to not have a KDN9PP gene, nor the corresponding synthase. Although, some bacteria like *H. influenzae* have evolved a secondary mechanism for obtaining KDN and its derivatives by importing sialic acid from the extracellular environment [44], it appears that the chosen mechanism in Bacteroides is to mimic human cells by acquiring the ability to produce sialic acids used to decorate the cell wall in humans.

Evolutionary Divergence and Duplication

Enzyme Specificity

A literature search illustrates the complexity in identifying a single residue which induces substrate selectivity seen here, as very few examples can be found. One example is tyrosine ammonia-lyase (TAL) of the aromatic lyase family. TAL is selective for L-Tyr over L-Phe by 150 – 300 fold depending on the enzyme analyzed [45] and contains a His residue adjacent to the active site. Sequence alignments confirm that His is present

175

in the same position in other TALs, while phenylalanine ammonia-lyases contain Phe at the same position. Mutation of His89 to Phe in TAL modifies the substrate selectivity to Phe over Tyr (15,000 fold) [45]. A similar example is the alteration of dyhydroflavonol 4-reductase specificity in *Gerbera* by changing a single amino acid (Asn134Leu) [46]. This results in the switch of specificity to solely utilize dihydrokaempferol as substrate, excluding dihydroquercetin and dihydromyricetin, which only differ in the addition of one or two hydroxyls, respectively, present on the ring structure [46]. The lack of examples in the literature highlights the difficulty in uncovering the mechanisms of substrate selectivity. The work described herein represents a structure-function-bioinformatic approach which is feasible for teasing out subtleties in enzyme specificity.

The correlation of structural, kinetic and bioinformatic data confirms the identity of two KDO8P phosphatases within bacterial species. The vast majority of bacteria contain a KDO8PP enzyme with an Arg residue at the identified specificity position. The HI-KDO8PP liganded structure highlights the importance of Arg in positioning KDO8P for catalysis. Yet, some KDO8PPs (i.e. *B. thetaiotaomicron* and *L. pneumophila subsp. pneumophila str. Philadelphia 1*) are specific for KDO8P, but replaced Arg with Gly or Ala. This exemplifies the variability and adaptability of enzymes. The loss of a large charged residue changes the shape of the KDO8P binding pocket which allows KDO8P binding, while reducing the ability of the enzyme to take in alternative substrates. This is critical for bacteria like *B. thetaiotaomicron* which must produce KDO for normal growth, and KDN to mimic human cells to prevent the innate immune response.

References:

1. Hisano, T., et al., *Crystal structure of L-2-haloacid dehaloge-nase from Pseudomonas sp. YL. An alpha/beta hydrolase structure that is different from the alpha/beta hydrolase fold.* J. Biol. Chem., 1996. **271**: p. 20322-20330.

2. 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.

3. Lahiri, S.D., et al., *Analysis of the substrate specificity loop of the HAD superfamily cap domain.* Biochemistry, 2004. **43**(10): p. 2812-20.

4. 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.

5. Xu, J., et al., *A genomic view of the human-Bacteroides thetaiotaomicron symbiosis.* Science, 2003. **299**(5615): p. 2074-6.

6. Vuorio, R. and M. Vaara, *The lipid A biosynthesis mutation lpxA2 of Escherichia coli results in drastic antibiotic supersusceptibility.* Antimicrob Agents Chemother, 1992. **36**(4): p. 826-9.

7. Kim, M.J.H., W. J.; Sweers, H. M.; Wong, C. H.. *Enzymes in carbohydrate synthesis: N-acetylneuraminic acid aldolase catalyzed reactions and preparation of N-acetyl-2-deoxy-D-neuraminic acid derivatives. .* JACS, 1988. **110**(19): p. 6481-6.

8.   Onishi, H.R., et al., *Antibacterial agents that inhibit lipid A biosynthesis.* Science, 1996. **274**(5289): p. 980-2.

9.   Lu, Z., et al., *Structure-function analysis of 2-keto-3-deoxy-D-glycero-D-galactonononate-9-phosphate phosphatase defines specificity elements in type C0 haloalkanoate dehalogenase family members.* J Biol Chem, 2009. **284**(2): p. 1224-33.

10. Iwasaki, M., S. Inoue, and F.A. Troy, *A new sialic acid analogue, 9-O-acetyl-deaminated neuraminic acid, and alpha -2,8-linked O-acetylated poly(N-glycolylneuraminyl) chains in a novel polysialoglycoprotein from salmon eggs.* J Biol Chem, 1990. **265**(5): p. 2596-602.

11. Inoue, S., K. Kitajima, and Y. Inoue, *Identification of 2-keto-3-deoxy-D-glycero--galactonononic acid (KDN, deaminoneuraminic acid) residues in mammalian tissues and human lung carcinoma cells. Chemical evidence of the occurrence of KDN glycoconjugates in mammals.* J Biol Chem, 1996. **271**(40): p. 24341-4.

12. Angata, T., et al., *Biosynthesis of KDN (2-keto-3-deoxy-D-glycero-D-galacto-nononic acid). Identification and characterization of a KDN-9-phosphate synthetase activity from trout testis.* J Biol Chem, 1999. **274**(33): p. 22949-56.

13. Tanner, M.E., *The enzymes of sialic acid biosynthesis.* Bioorg Chem, 2005. **33**(3): p. 216-28.

14. Wang, L., et al., *Human symbiont Bacteroides thetaiotaomicron synthesizes 2-keto-3-deoxy-D-glycero-D- galacto-nononic acid (KDN).* Chem Biol, 2008. **15**(9): p. 893-7.

15. Radaev, S., et al., *Structure and mechanism of 3-deoxy-D-manno-octulosonate 8-phosphate synthase.* J Biol Chem, 2000. **275**(13): p. 9476-84.

16. Biswas, T., et al., *The tail of KdsC: conformational changes control the activity of a haloacid dehalogenase superfamily phosphatase.* J Biol Chem, 2009. **284**(44): p. 30594-603.

17. Potterton, E., et al., *A graphical user interface to the CCP4 program suite.* Acta Crystallogr D Biol Crystallogr, 2003. **59**(Pt 7): p. 1131-7.

18. Winn, M.D., *An overview of the CCP4 project in protein crystallography: an example of a collaborative project.* J Synchrotron Radiat, 2003. **10**(Pt 1): p. 23-5.

19. Brünger, A.T., et al., *Crystallography & NMR system: A new software suite for macromolecular structure determination.* Acta Crystallogr D Biol Crystallogr, 1998. **54 (Pt 5)**: p. 905-21.

20. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.

21. 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.

22. Parsons, J.F., et al., *From structure to function: YrbI from Haemophilus influenzae (HI1679) is a phosphatase.* Proteins, 2002. **46**(4): p. 393-404.

23. Murshudov, G.N., A.A. Vagin, and E.J. Dodson, *Refinement of macromolecular structures by the maximum-likelihood method.* Acta Crystallogr D Biol Crystallogr, 1997. **53**(Pt 3): p. 240-55.

24.   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.

25.   Garcia-Boronat, M., et al., *PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery.* Nucleic Acids Res, 2008. **36**(Web Server issue): p. W35-41.

26.   Shannon, C.E., *The mathematical theory of communication. 1963.* MD Comput, 1997. **14**(4): p. 306-17.

27.   Inoue, S. and K. Kitajima, *KDN (deaminated neuraminic acid): dreamful past and exciting future of the newest member of the sialic acid family.* Glycoconj J, 2006. **23**(5-6): p. 277-90.

28.   Gil-Serrano, A.M., et al., *Structural determination of a 5-O-methyl-deaminated neuraminic acid (Kdn)-containing polysaccharide isolated from Sinorhizobium fredii.* Biochem J, 1998. **334 (Pt 3)**: p. 585-94.

29.   Knirel, Y.A., et al., *Structure of the capsular polysaccharide of Klebsiella ozaenae serotype K4 containing 3-deoxy-D-glycero-D-galacto-nonulosonic acid.* Carbohydr Res, 1989. **188**: p. 145-55.

30.   Xu, J., et al., *Evolution of symbiotic bacteria in the distal human intestine.* PLoS Biol, 2007. **5**(7): p. e156.

31.   Trent, M.S., et al., *Diversity of endotoxin and its impact on pathogenesis.* J Endotoxin Res, 2006. **12**(4): p. 205-23.

32.	Moran, A.P., M.M. Prendergast, and B.J. Appelmelk, *Molecular mimicry of host structures by bacterial lipopolysaccharides and its contribution to disease.* FEMS Immunol Med Microbiol, 1996. **16**(2): p. 105-15.

33.	Bugla-Ploskonska, G., et al., *Sialic acid-containing lipopolysaccharides of Salmonella O48 strains--potential role in camouflage and susceptibility to the bactericidal effect of normal human serum.* Microb Ecol, 2010. **59**(3): p. 601-13.

34.	Bugla-Ploskonska, G. and W. Doroszkiewicz, *Bactericidal activity of normal bovine serum (NBS) directed against some Enterobacteriaceae with sialic acid-containing lipopolysaccharides (LPS) as a component of cell wall.* Pol J Microbiol, 2006. **55**(3): p. 169-74.

35.	Bryant, D.A. and N.U. Frigaard, *Prokaryotic photosynthesis and phototrophy illuminated.* Trends Microbiol, 2006. **14**(11): p. 488-96.

36.	Terada, T., et al., *Synthesis of CMP-deaminoneuraminic acid (CMP-KDN) using the CTP:CMP-3-deoxynonulosonate cytidylyltransferase from rainbow trout testis. Identification and characterization of a CMP-KDN synthetase.* J Biol Chem, 1993. **268**(4): p. 2640-8.

37.	Shashkov, A.S., et al., *Cell wall anionic polymers of Streptomyces sp. MB-8, the causative agent of potato scab.* Carbohydr Res, 2002. **337**(21-23): p. 2255-61.

38.	Shashkov, A.S., et al., *A polymer with a backbone of 3-deoxy-D-glycero-D-galacto-non-2-ulopyranosonic acid, a teichuronic acid, and a beta-glucosylated ribitol teichoic acid in the cell wall of plant pathogenic Streptomyces sp. VKM Ac-2124.* Eur J Biochem, 2002. **269**(24): p. 6020-5.

39. Tul'skaia, E.M., et al., *[Anionic carbohydrate-containing cell wall polymers of Streptomyces melanosporofaciens and related species].* Mikrobiologiia, 2007. **76**(1): p. 48-54.

40. Naumova, I.B. and A.S. Shashkov, *Anionic polymers in cell walls of gram-positive bacteria.* Biochemistry (Mosc), 1997. **62**(8): p. 809-40.

41. De Soyza, A., et al., *Burkholderia cenocepacia lipopolysaccharide, lipid A, and proinflammatory activity.* Am J Respir Crit Care Med, 2004. **170**(1): p. 70-7.

42. Gunn, J.S. and R.K. Ernst, *The structure and function of Francisella lipopolysaccharide.* Ann N Y Acad Sci, 2007. **1105**: p. 202-18.

43. Jun, S.R., et al., *Whole-proteome phylogeny of prokaryotes by feature frequency profiles: An alignment-free method with optimal feature resolution.* Proc Natl Acad Sci U S A, 2010. **107**(1): p. 133-8.

44. Severi, E., et al., *Sialic acid transport in Haemophilus influenzae is essential for lipopolysaccharide sialylation and serum resistance and is dependent on a novel tripartite ATP-independent periplasmic transporter.* Mol Microbiol, 2005. **58**(4): p. 1173-85.

45. Watts, K.T., et al., *Discovery of a substrate selectivity switch in tyrosine ammonia-lyase, a member of the aromatic amino acid lyase family.* Chem Biol, 2006. **13**(12): p. 1317-26.

46. Johnson, E.T., et al., *Alteration of a single amino acid changes the substrate specificity of dihydroflavonol 4-reductase.* The Plant Journal, 2001. **25**(3): p. 325-333.

47.   Schrödinger, L., *The PyMOL Molecular Graphics System, Version 1.2r3pre*.

48.   Kraulis, P., *MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures.* J. Appl. Crystallogr., 1991. **24**: p. 946-950.

49.   Fenn, T.D., D. Ringe, and G.A. Petsko, *POVScript+: a program for model and data visualization using persistence of vision ray-tracing.* Journal of Applied Crystallography, 2003. **36**: p. 944-947.