THE ROAD LESS TRAVELLED: UTILIZATION OF FORMATE IN TWO BIOCHEMICAL REACTIONS IN GRAM-NEGATIVE BACTERIA

By

©2019
Nikola Kenjić

B.S., University of Florida, 2009
M.S.,University of Arizona, 2014

Submitted to the Department of Molecular Biosciences and the
Faculty of the Graduate School of the University of Kansas
In partial fulfillment of the requirements for the degree of
Doctor of Philosophy

_____
Chairperson – Audrey Lamb

_____
Committee members: Kristi Neufeld

_____
Joanna Slusky

_____
Mark Richter

_____
Timothy Jackson

Date defended: 07/02/2019

The Dissertation Committee for Nikola Kenjić certifies
that this is the approved version of the following dissertation:

THE ROAD LESS TRAVELLED: UTILIZATION OF FORMATE IN TWO BIOCHEMICAL REACTIONS IN GRAM-NEGATIVE BACTERIA

Committee:

_____
Audrey Lamb

_____
Kristi Neufeld

_____
Joanna Slusky

_____
Mark Richter

_____
Timothy Jackson

Date defended: 07/02/2019

ABSTRACT

Formate is an important intermediate in a number of metabolic reactions. Most of the formate pool in cells is generated from the degradation of primary metabolites (glucose, pyruvate, amino acids) and further degraded into CO_2 and CH_4. Remaining formate is recycled and utilized in one-carbon metabolism where a one-carbon formyl group is integrated into nucleotide metabolism, protein synthesis, and generation of secondary metabolites such as siderophores. Here I provide two examples of distinct and unique enzymes involved in formate generation and metabolism: PvdF from *Pseudomonas aeruginosa* and RibB from *Vibrio Cholerae*.

PvdF is one of two enzymes involved in generation of the formyl-hydroxyornithine (fOHOrn) moiety responsible for iron chelation in pyoverdin, a siderophore in *Pseudomonas aeruginosa.* Biochemical and structural studies suggest that PvdF is a unique new class of transformylase enzyme. PvdF catalyzes the movement of the formyl group from an N^{10} formyl-THF analogue to the substrate following the bireactant random substrate binding model. Structurally, PvdF has a transformylase fold with secondary structural element insertion so far unprecedented in the literature.

RibB is involved in the biosynthesis of riboflavin, vitamin B2. This is a magnesium dependent enzyme that catalyzes the conversion of the sugar ribulose 5-phosphate (Ru5P), a product of the pentose phosphate pathway, into 3,4-dihydroxy-2-butanone 4-phosphate (DHBP). The reaction catalyzed by RibB is an unusual deformylation reaction in which the fourth carbon of the five-carbon sugar is removed as formate. According to the literature, RibB catalyzes this reaction in the presence of di-metal Mg^{2+} centers following 1,2-methyl shift, called a skeletal rearrangement mechanism. Our evidence,

both biochemical and structural, suggests that RibB requires only one Mg^{2+} for catalysis. Furthermore, NMR and X-ray crystallography data point toward the formation of a 2-phosphoglycolic acid intermediate during RibB catalyzed reaction. These data suggest that a fragmentation mechanism, not a skeletal rearrangement, is the preferred mechanism of RibB catalysis.

ACKNOWLEDGEMENTS

This dissertation is dedicated to my family. My lovely wife, Dr Keely E. Brown and Lady Diva for believing in me and supporting me even when I have lost faith in myself. Without you none of this would be possible. To my mother Nada, father Mićo and sister Milica for endless support and courage. Thank you for your bravery and letting me go on this journey to accomplish my goal. To my American parents, Scott Brown and Cheryl Gray, thank you for providing me with a loving home away from home and parental advice away from my parents.

Sincere thank you to my mentor Dr Audrey Lamb for endless scientific advice and guidance, direction and mentorship over past couple of years. To Dr Chris Gamblin for allowing me to use instrumentation in his lab and for endless supply of coffee. To Dr Kathleen Meneely, Dr Trey Ronnebaum and Dr Jeff Mcfarlane for all scientific discussions, revisions and comments, you helped me learn something new every day. To Annemarie Chilton and all undergraduate students in Dr Lamb's lab over last couple of years for your continues help.

TABLE OF CONTENTS

List of Figures

List of Tables

CHAPTER 1:

Introduction

Structurally formate is one of the simplest carboxylate anions in the cell. This molecule is directly or indirectly involved in number of metabolic reactions [1-3]. For example, in one-carbon metabolism, formate is incorporated with tetrahydrofolate (THF) cofactors and further integrated into purine metabolism, used in iron chelation (siderophore biosynthesis) or as a recognition tool for first amino acids (fMet-tRNA) of protein biosynthesis. Current research on formate anabolic and catabolic reactions is extensive. How this molecule is going to be utilized will depend on bacterial extracellular environmental conditions [2, 4]. What has been proposed is that biggest proportion of generated formate is made during anaerobic degradation of carbohydrates such as glucose. Product of the degradation, pyruvate, is converted to formate by pyruvate formate lyase to make acetyl-CoA and formate. Formate is also made as the end product of degradation of amino acids (serine, glycine, threonine, histidine) as well as biproduct in generation of metabolites (such as riboflavin, formyl Met tRNA) [1, 5-7]. From the cellular formate pool this anion can either be reused, further oxidized into carbon dioxide or reduced into methane. For example, in purine metabolism of aerobic bacteria, when the essential cofactor N^{10}-formyl THF cannot be made, a purine intermediate can still be generated using intercellular formate and ATP. This has been demonstrated to be an alternative mechanism that would allow for bacterial growth and nucleotide generation even in the absence of essential (THF) cofactors [8, 9]. Formate is oxidized to carbon dioxide and hydrogen by formate dehydrogenase. These

dehydrogenases are complex and have been reported to be either NAD^+ dependent or NAD^+ independent. The NAD^+ independent reaction generates CO_2 in the presence of ferrocytochrome b1 and an iron sulfur cluster. Electrons generated during this process have been shown to contribute to bacterial respiration (proton motive force) as well as nitrate reduction. NAD^+ dependent reaction catalyzes the generation of CO_2 in typical NAD^+ dehydrogenase manner, by reduction of NAD^+ into NADH [2, 4, 10-12]. These enzymes have been studied extensively, they do not require metals and are usually not membrane bound. On the other hand, conversion of formate to methane is rare and the perfomed by class of organisms called methanogens. Methanogens are archaeal organisms that depend on simple carbon molecules such as carbon dioxide, formate, acetate and methanol. In short, in order for these organisms to survive they need to live in a proximity to bacteria, organisms that are able to convert complex molecules into their vital metabolites. Conversion of formate to methane is complex and it involves multiple enzymatic steps and presence of a pterin cofactor as well as F420 cofactors [2, 13].

The focus of this dissertation will be two examples of formate utilization. One is PvdF enzyme from *Pseudomonas aeruginosa.* This enzyme is crucial in pyoverdin biosynthesis using N^{10}-formyl-THF as a cofactor. The other is in riboflavin biosynthesis where enzyme RibB releases formate after conversion of ribulose 5-phosphate into product 3,4-dihydroxy-2-butanone 4-phosphate.

1.1 Role of formate in one-carbon metabolism.

One-carbon metabolism is the transfer of one-carbon groups such as methyl groups on folate cofactor to make *S*-adenosylmethionine. Herein, we will consider the transfer of one-carbon groups attached to folate, which is central to biosynthetic pathways involved in the formation of purine nucleotides and amino acids such as glycine, serine and methionine[14, 15]. Cancer cells require both a constant supply of nucleotides and a mechanism for the regulation of gene expression by methylation [16, 17]. Inhibition of enzymes in one-carbon metabolism provides an avenue to design antifolate inhibitors that are effective against cancer cell proliferation. Enzyme-specific inhibitors that are folate mimics have been developed that target dihydrofolate reductase [18, 19] and thymidylate synthase [20, 21], among others [22-28]. Some of these inhibitors, such as methotrexate, are also effective against other diseases such as arthritis and psoriasis [29, 30]. Indeed, many antifolate inhibitors are nonspecific [31].

Folates are comprised of three parts: a pterin ring (made from GTP), a *p*-aminobenzoic acid (pABP, derived from chorismate) and a glutamate tail [15]. Whereas enzymes for the biosynthesis of the folate core can be found in lower eukaryotes, protozoa, bacteria and plants [32, 33], mammals do not have genes for folate synthesis and must obtain folic acid, vitamin B1, from their diet. For this reason, one-carbon metabolism is also a target for the synthesis of antibacterial and anti-parasitic compounds. One of these inhibitors, trimethoprim-sulfamethoxazole, inhibits dihydrofolate reductase (**Figure 1.1**) [34].

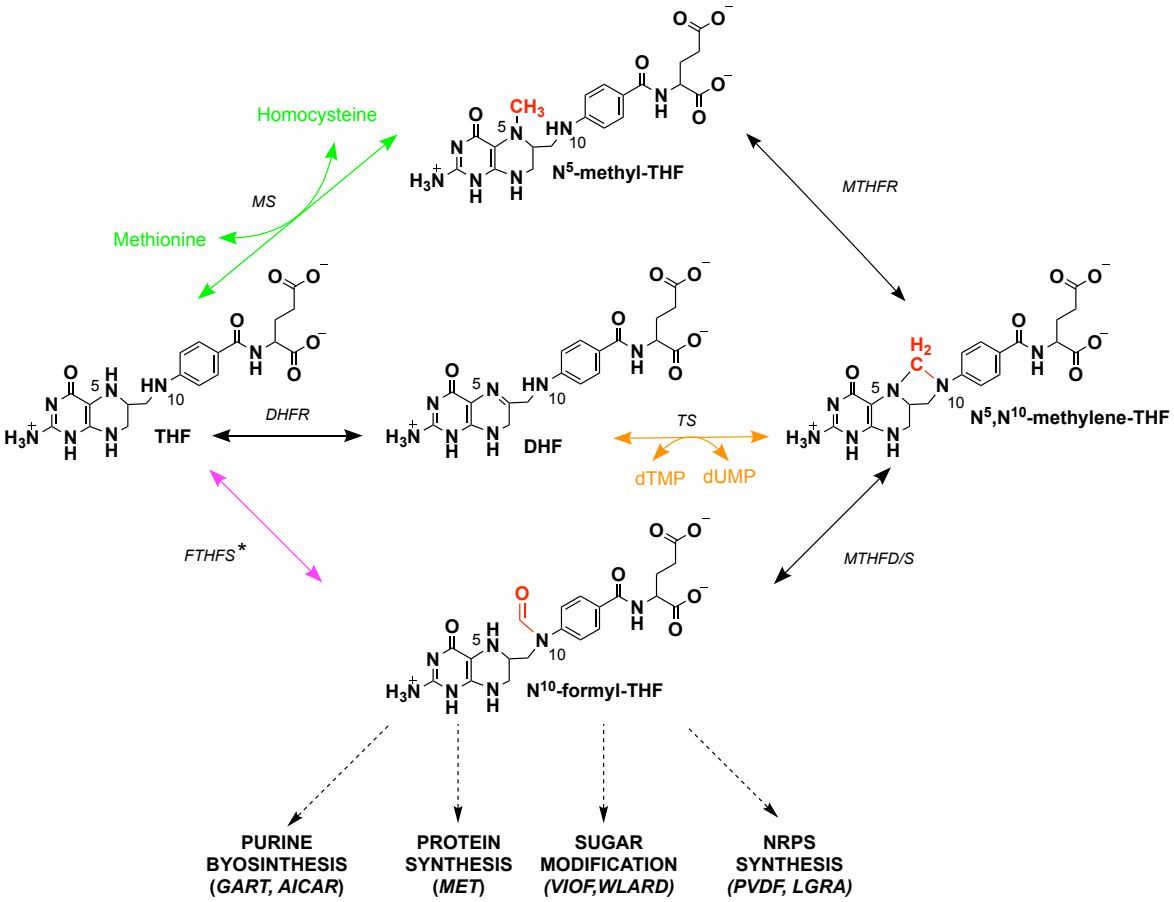


Figure 1.1. One carbon metabolism and utilization of N^{10}-formyl-THF in bacteria. Folate cofactors are indicated in bold letters. DHF stands for dihydrofolate, THF for tetrahydrofolate. Enzymes involved in one carbon metabolism designated as: DHFR (dihydrofolate reductase), MS (methionine synthase), MTHFR (5,10-methylenetetrahydrofolate reductase), MTHFD/S (5,10-methylenetetrahydrofolate dehydrogenase/cyclohydrolase), FTHFS (formyltetrahydrofolate synthetase), TS (thymidylate synthase). Reaction of formyltetrahydrofolate synthetase is predominantly present in anaerobic bacteria.

Tetrahydrofolate (THF) is the basic folate core (**Figure 1.1**) that is converted to other folate cofactors in one-carbon metabolism. Folates are generated to transfer methyl (N^5-methyl-THF), methylene (N^5,N^{10}-methylene-THF) or formyl (N^{10}-formyl-THF) groups. N^{10}-formyl-THF is required for de novo purine biosynthesis (becoming carbons C2 and C8), pathways that are well studied [35, 36]. In protein biosynthesis, a formyl group derived from N^{10}-formyl-THF and attached to a methionyl tRNA, such that $tRNA^{fMet}$ initiates translation [37-39]. While the utilization of N^{10}-formyl-THF in purine metabolism and protein synthesis has been investigated in detail, recent work has identified N^{10}-formyl-THF as a required cofactor in other biochemical reactions in pathogenic bacteria, such as siderophore biosynthesis and N-sugar formylation on the bacterial surface.

1.2 PurH – A bifunctional AICAR transformylase/cyclohydrolase.

PurH performs one of the last steps of purine biosynthesis and is a bifunctional enzyme containing aminoimidazole-4-carboxamide ribonucleotide (AICAR) transformylase and inosine monophosphate (IMP) cyclohydrolase activities. AICAR transformylase catalyzes the production of formyl-AICAR (FAICAR) from AICAR (**Figure 1.2, B**). FAICAR is transferred to the second domain, IMP cyclohydrolase, to produce inosine monophosphate. Both in solution and in the crystal structure, PurH is a homodimer with the monomers heavily intertwined around each other with 5000 Å of buried surface area [40]. The N-terminal domain of the protein is IMP cyclohydrolase while the C-terminal domain (the remaining 400 residues) comprises the AICAR transformylase. The AICAR

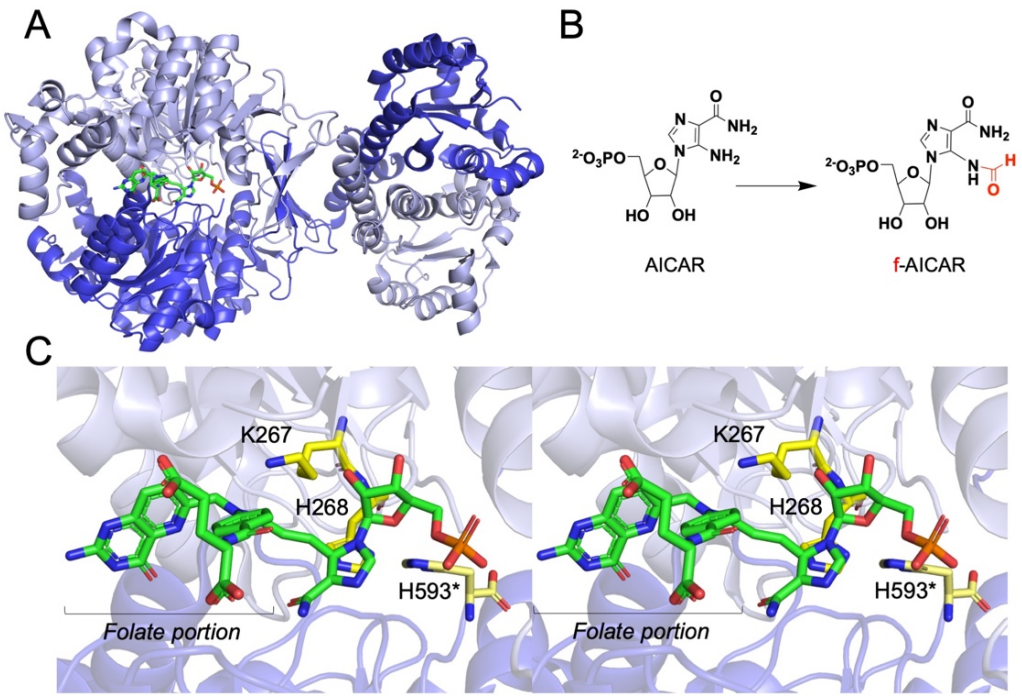


Figure 1.2. AICAR transformylase. (A) This bifunctional protein is a homodimer with two monomers intertwined around each other. The active site is located in the interface between the monomers. AICAR transformylase is larger of two subunits with transition state inhibitor (green) in the active site. (B) Reaction catalyzed by AICAR transformylase (C) AICAR (PDBID:1OZ0) has been crystallized in presence of transition state inhibitor with substrate and cofactor molecules joined together (green). Catalytic Lys267 (yellow) is hydrogen bonding distance away from the location of formate group on the cofactor. This residue stabilizes the transition state and provides the proton necessary for the formyl transfer. Other catalytic residues His268 (yellow) and His593 (from the opposing monomer, pale yellow) are proposed to be responsible for the proton shuttling form the solvent to Lys267. Detailed mechanism on Figure 1.5.

transformylase domain is comprised of three structural domains [41]. The first two domains have a very similar fold containing a five-stranded β-sheet core surrounded by helices. Both of these domains from each protein dimer contribute one β-sheet to make a 4 β-sheet bridge between IMP cyclohydrolase and AICAR transformylase. This barrier between the domains has been used as an argument against tunneling or movement of the intermediate FAICAR between the domains [40]. The third domain is represented as a small insertion into one of these domains with still unknown function. The active site of the AICAR transformylase is surface exposed and located in the dimer interface (**Figure 1.2, A**). Interestingly enough, a monomer of AICAR is not functional and the dimer is required for catalysis. Functional and structural studies of AICAR transformylase with a multi-substrate adduct inhibitor comprised of AICAR covalently linked to a THF analogue, were undertaken. A holo structure with this inhibitor bound implicated Lys267, His268 and His593 as residues important in catalysis (**Figure 1.2, C**) [42]. The proposed mechanism suggests a catalytic triad of histidine residues are used as a proton shuttle from the solvent to activate the primary amine of the substrate. The lysine stabilizes the oxyanion hole created by the formate in the transition state, protonation of the lysine regenerates the catalytic triad and the protons are returned back to the solvent [41, 42] (**Figure 1.2, C and Figure 1.5**)

1.3 Examples of GART enzymes.

Glycinamide ribonucleotide transformylases (GART) catalyze the formylation of glycinamide ribonucleotide with N^{10}-formyl-THF used as a formyl group donor. In purine biosynthesis, this enzyme is PurN and has been well studied both structurally and

functionally as it was considered a potential drug target. GART is composed of a Rossmann fold with 6-7 β-strands in the center of the molecule surrounded by α-helices with two subdomains. The N-terminal subdomain is responsible for binding of the THF cofactor and contains a folate binding loop while the C-terminal subdomain contains the active site loop and the catalytic His, Asn and Asp residues (**Figure 1.3**). The active site of the enzyme is surface exposed and is located at the interface of the two domains. The active site loop contains a conserved Asp residue as part of C-terminal domain. The importance of this loop has been studied in detail in human PurN where loop movement is directly correlated to the pH of the solution. For example at high pH (8.5) loop is ordered and readily binds the substrate, while at low pH this loop is disordered and prevents substrate binding [43]. In the *Mycobacterium tuberculosis* PurN, these studies could not be replicated, likely due to the different packing of the *M. tubeclulosis* PurN dimer (**Figure 1.3)** [44].

Unlike the THF cofactor that has a distinct binding domain, the substrate is stabilized by folate binding loop of the C-terminal domain with the nitrogen to be modified in the substrate orientated toward the catalytic residues and the formate of the cofactors. However, substrate binding and stabilization appears to be the main difference between GART-like proteins in the literature. Recently, structures of three other PurN homologues have been solved in which the transformylase activity is merely a domain of a larger protein with the second domain being responsible for substrate binding. The homolog, methionyl tRNA transformylase (MTF) is important in protein production and responsible for synthesis of f-Met bound to the tRNA [45]. A group of N-sugar

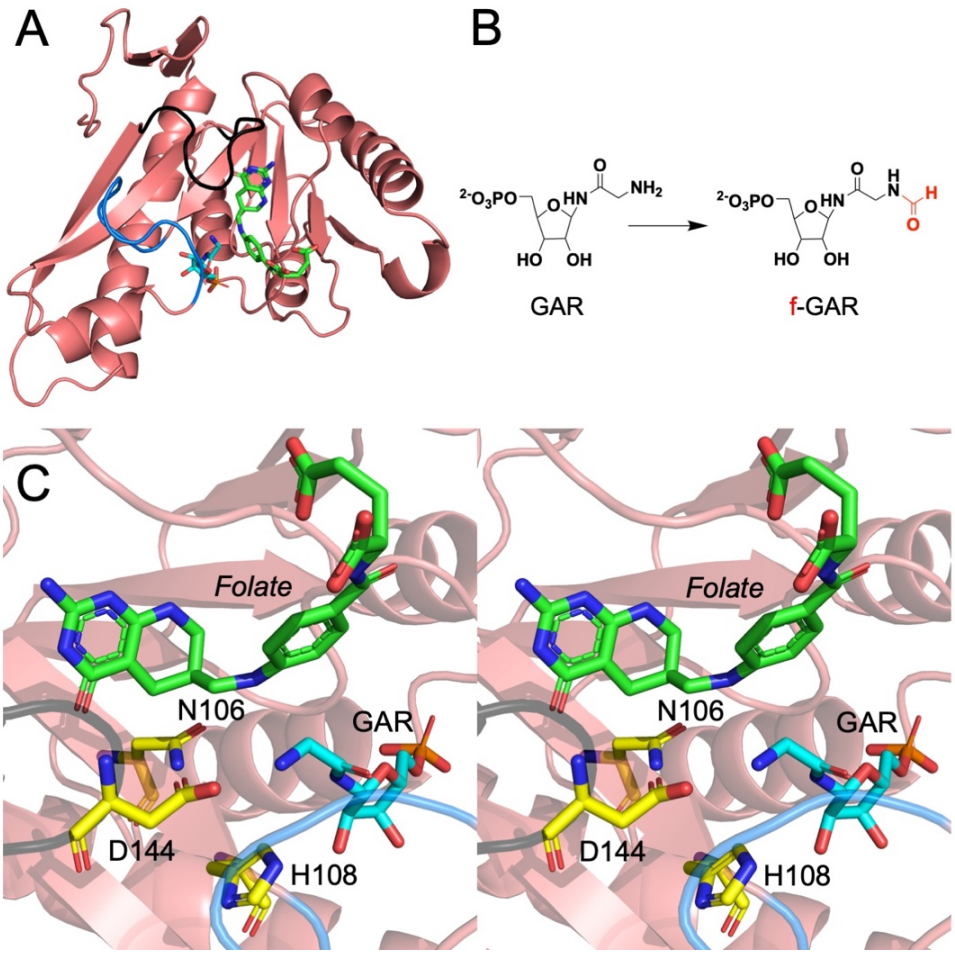


Figure 1.3. **GAR transformylase**. The enzyme is composed of 6 β-strands surrounded by helices. The active site is surface exposed surrounded by the active site loop (black) and the folate binding loop (blue). The cofactor (green) is orientated so that the formyl group and nitrogen of the substrate (cyan) are in binding distance of the GAR conserved catalytic triad Asn106, His 108 and Asp 144 (cyan).

transformylases are important in the LPS of Gram-negative bacteria where the growing sugar moiety is formylated on the S peptide in bacteria [46]. LgrA is a large nonribosomal peptide synthetase (NRPS) involved in gramicidin biosynthesis which contains a transformylation domain used as to modify the growing NRPS peptide (**Figure 1.4**) [47, 48]. Structurally, the transformylase domains in all of these enzymes are similar in fold to PurN (RMSD range 0.5 – 2.5 [49]) with the active site loop and conserved catalytic residues.

Mechanistic and kinetic investigation of the transformylases have been done by Benkovic and colleagues with studies predominantly oriented toward the *E. coli* PurN enzyme. Detailed turnover kinetic experiments suggest that the transformylation occurs via the bi-reactant systems in which either the cofactor or the substrate can form an initial complex with the enzyme [50]. However, binding of the cofactor is faster and preferred as the initial binding step. These data provide he basis for the current mechanism for transformylation, which is initiated by a nucleophilic attack of the terminal nitrogen of the substrate on the carbonyl of the formyl group. This event leads to the formation of a tetrahedral intermediate and formation of the oxyanion stabilized by an asparagine in the active site. A proton transfer leads to the formation of the final product. Even though this last step is simple, it is still debated. In one of the first presented structures of PurN [51], the active site contained a coordinated water suggesting water as the means of proton movement (**Figure 1.5, B**). However, in other structures including the PurN homologues, this water molecule was absent leading to the hypothesis that this proton transfer has to occur by other means. In order to provide

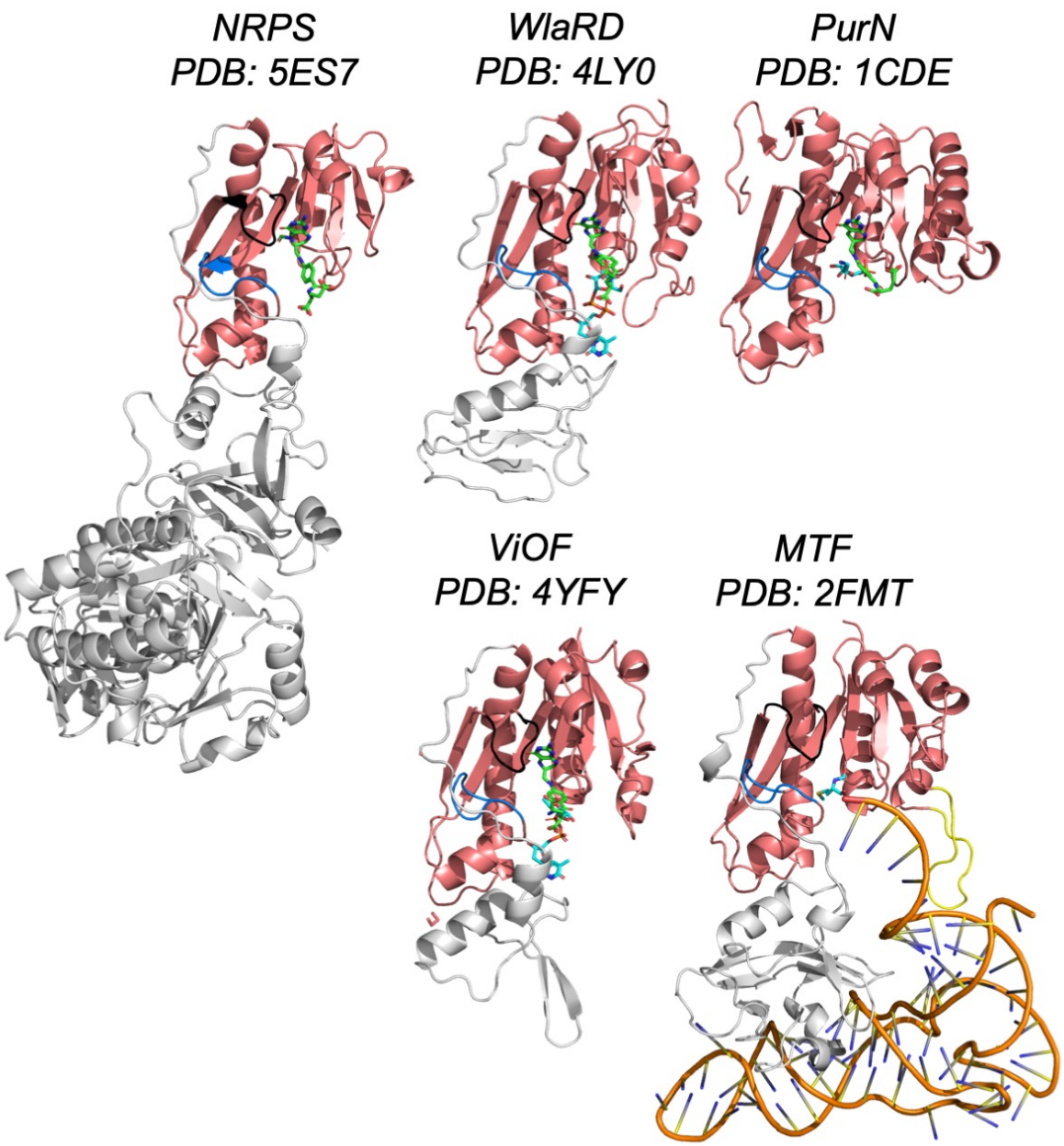


Figure 1.4. Examples of GAR transformylases. Presented are three classes of GART transformylases. NRPS, MTF and N-sugar transformylase (WlaRD and VioF). The N-terminus of these proteins (red) are structurally similar (RMSD = 2.2) with a conserved catalytic triad and crystallized bound with the folate cofactor (green). Active site is located in same region between folate loop (blue) and active site loop (black). The C-terminal domain of the proteins are very different and corresponds to the distinct functional characteristics of each enzyme. This C-terminal domain as well as the loop insertion in MTF (yellow) stabilize the substrate and insures its proper orientation prior to catalysis.

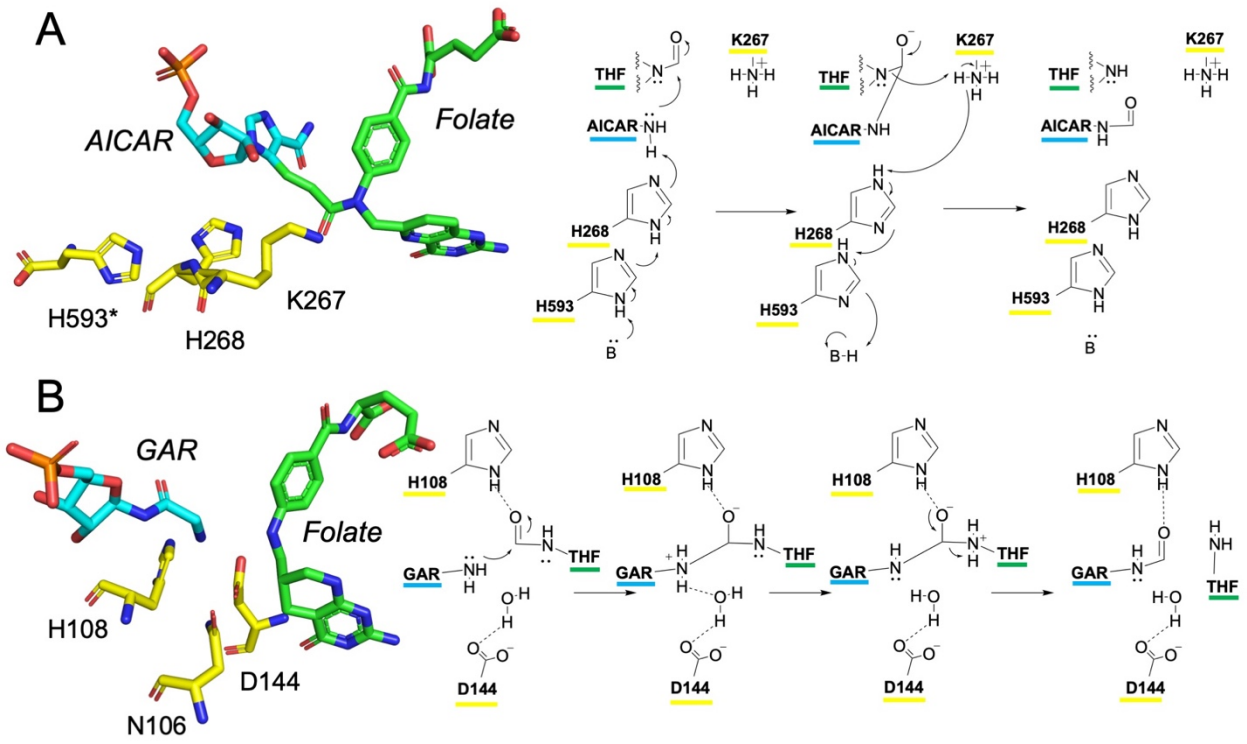


Figure 1.5 Active site comparison between AICAR transformylase and GART enzymes. (A) AICAR transformylase active site. Catalytic residues (yellow) are involved in proton transfer between the solvent and Lys267 that is directly supplying protons to the transition state intermediate (green and cyan, cyan portion of the compound is the substrate and green is the cofactor). (B) GAR transformylase active site. Represented is the active site with substrate GAR (cyan), folate cofactor (green) and catalytic residues (yellow) with one of proposed reaction mechanism. His108 and Asn106 (not shown) stabilize oxyanion and transition state intermediate while Asp144 coordinates the water molecules involved in proton transfer between substrate and cofactor.

a better overview, a closer description of the other PurN homologues is presented below.

1.4 Methionyl $tRNA_f$ Met formyltransferase (MTF).

The function of methionyl tRNA ($tRNA_f$ Met) formyltransferase is of utmost importance in protein translation. $tRNA_f$ Met in prokaryotes is recognized as the initial amino acid of a growing peptide chain. The $tRNA_f$ Met allows binding with elongation factor 2 and elongation factor 3 to signal the 30S subunit of the ribosome to start translation. Work by Shmitt *et al*. investigated the structural and functional aspects of MTF, the enzyme that formylates the methionine, in *E. coli* [52]. The authors provided a detailed comparison between MTF and GART. MTF is composed of two domains, the N-terminal domain that closely resembles the GART enzymes and a C-terminal domain composed of five antiparallel β-strands surrounded by α helices on either end. The N-terminal domain of MTF, while very similar to the GART enzymes (RMSD 1.4 Å), has several important differences that appear to be of functional importance. There is a loop insertion between the second β-strand and the second α-helix (yellow, **Figure 1.4, MTF**) at residues 37-45 composed of basic amino acid residues. The function of this insertion was unknown until the co-crystal structure of MTF with tRNA Met bound, was solved. The basic side chains of this insertion loop are involved in the binding of the phosphate backbone of the tRNA arm. The formylation efficiency of the enzyme has been shown to decrease four times in the absence of this insertion loop. The shape of the C-terminal domain was previously described as an oligonucleotide binding fold. This fold conforms to an orientation that allows the positively charged residues on its surface to interact with the L-shaped phosphate backbone of the substrate. In this way, the C-terminal loop

binds the substrate and brings it closer to the active site and provides specificity, locking the substrate in place for formylation by the N-terminal domain (**Figure 1.4, MTF**) [45, 52]

1.5 N-Sugar transformylases.

Transformylase activity has been shown to be of importance in the pathogenicity and extracellular recognition in Gram-negative bacteria. Lipopolysaccharides (LPS) are composed of three distinct regions: lipid A, core polysaccharide, and O-linked polysaccharide [53]. Diversity within this polysaccharide region is exhibited by the activity of N-sugar transformylases. The biosynthetic pathway describing the modification of sugars starts with glucose 1-phospate and involves the addition of a nucleotide, dehydration, and amination. N-sugar transformylases catalyze the reaction in which the formyl group from N^{10}-formyl-THF is transferred to the amine of the newly modified sugar [53].

Holden *et al.* have done extensive work with N-sugar transformylases and provided a couple of great examples of substrate specificities in this class of enzymes. One of the first examined structures was that of WlaRD, an N-sugar transformylase from *Campylobacter jejuni* [54]. While the substrate for this enzyme is not known, WlaRD was shown to be promiscuous, turning over both dTDP-Fuc3N and dTDP-Qui3N. The enzyme contains an N-terminal domain that is similar to the GART enzymes with a conserved catalytic triad and only one amino acid substitution in the folate binding loop (.S<u>A</u>LP.). The C-terminal domain is composed of 4 antiparallel β-strands. The function

of this domain is hypothesized to be for docking of the substrate such that Gln 223 and Tyr 222 are interacting with the thiamine ring of the substrate (**Figure 1.4**) [54].

Other types of enzymes from this class presented by the Holden lab include the characterization of VioF (*Providencia alcalifaciens O30*) and WbtJ (*Francisella tularensis*) [46]. While in WlaRD the transformylation occurs on the C3 amine, in VioF and WbtJ the transformylation is on the C4 amine. The N-terminal domain of these two enzymes are similar to GART and demonstrate similar structural and functional features. However, the C-terminal domain is very different and is composed of single α-helix and a β-hairpin. Genthe *et al.* described structural differences that allow for the differences in substrate specificity between the two types of sugar transformylase. First, VioF has a Trp104, residue absent in WlaRD, that prevents the same substrate conformation of the two enzymes. In VioF, residues Lys77 and Asp9 bind phosphoryl oxygen and the pyranosyl group of the substrate. In WlaRD, however, these residues are substituted with Asp79 and Lys9 and are unable to make these contacts [46, 54].

The last example of N-sugar transformylases was identified as QdtF from *Providencia alicalifaciens* O40. Like WlaRD, this enzyme catalyzed formylation on the C3 carbon of the sugar. What makes this example unusual is that it contains an additional third domain characterized as an ankyrin repeat. Composed of 33 amino acid residues in a helix-loop-helix motif, this domain has been demonstrated to bind the dTDP-Qui3N substrate. Mutation of the substrate binding residues within this repeat has been shown to considerably decrease the activity of the enzyme. While the actual *in vivo* substrate of this enzyme is not known, the presence of this typical eukaryotic regulatory mechanism

in a prokaryotic protein makes QdtF an intriguing addition to transformylase class of proteins [53, 55].

1.6 Formylating Domain in NRPS enzymes.

Nonribosomal peptide synthetases (NRPSs) are a class of multi domain enzymes that generate a variety of peptide molecules of diverse biological functions. NRPS genes are composed of modules in which each module is responsible for the addition of a single amino acid into a growing peptide chain in an assembly line fashion. Each module contains at least three domains: adenylation (A), peptidyl carrier protein (PCP) and condensation (C). The adenylation domain activates the substrate amino acids by the addition of adenosine monophosphate (AMP), transfers the activated amino acid to a phosphopantethionyl tail attached to the PCP domain. The PCP domain transfers the amino acid to the condensation domain for the formation of a peptide bond with the upstream nascent peptide. However, in some cases the modules have additional tailoring domains to decorate the peptide side chain allowing for the key functional features of the product molecules. An example of a tailoring domain is LgrA, the initial NRPS involved in the biosynthesis of the antibiotic gramicidin from *Bacillus brevis*. This module formylates the N-terminal nitrogen of the initial valine residue, using N^{10}-formyl-THF as the formyl donor. The structure of LgrA has been described in detail by Reimer *et al.* with the N-terminal domain of the protein containing a typical formyltransferase domain with an active site loop and catalytic triad that is required for the transfer of the formyl group from THF to the substrate. The C-terminal end of protein contains a

module for the incorporation of the initial formylated valine and the downstream glycine (**Figure 1.4**) [47, 48].

1.7 PvdF from Pseudomonas aeruginosa, a new class of transformylase.

While AICAR transformylase and GART represent the most studied examples of transformylases in nature, a new class of transformylases has recently been presented by our lab. The PvdF enzyme from pyoverdin biosynthesis in *Pseudomonas aeruginosa* catalyzes the formylation of hydroxyornithine to make N^5-formyl-hydroxyornithine. Identification of PvdF as a novel class of transformylases lies in both structural and functional evidence. Structurally, PvdF is similar to PurN, a GART, with an RMSD of 2.2 Å over 165 Cα residues. PvdF has the conserved catalytic residues of Asp, His, Asn, and the active site loop. The fold of PvdF is also conserved with 6 central β strands and a Rossman-like fold. However, PvdF is 100 amino acids longer than PurN, the GART from *E. coli*. These additional amino acids compose secondary structural elements that have been inserted in both the N and the C-terminal domains of the transformylase. Our current hypothesis suggests that these insertions are used for stabilization of both the N^{10}-formyl-THF and the substrate. Functionally, the PvdF substrate is unique in comparison to other transformylases. GART enzymes formylate primary amines of their substrates, while PvdF formylates the secondary amine of its substrate. This chemical property makes the terminal amine of hydroxyornithine more nucleophilic and more labile providing more reasons for its stabilization within the active site cavity. The PvdF substrate is significantly smaller in comparison to the substrates of other GART enzymes leading to a smaller active site cavity (**Figure 1.6**).

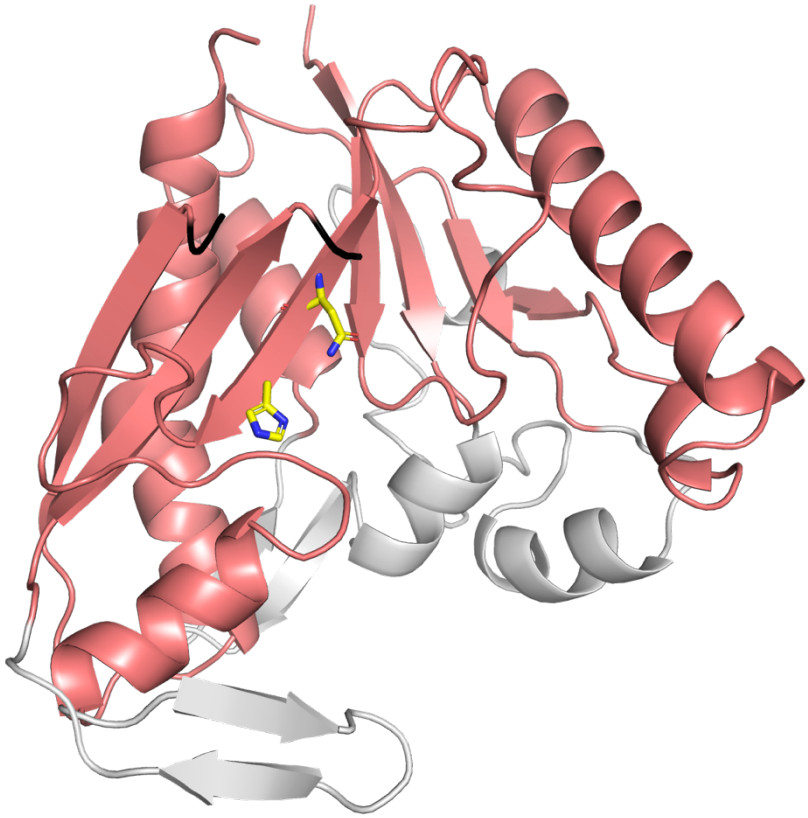


Figure 1.6. Structure of PvdF, hydroxyornithine transformylase from *Pseudomonas aeruginosa*. Majority of structures in Figure 1.4 have transformylase as a separate domain (red, Figure 1.4). PvdF however, has 100 amino acid insertions throughout the protein (gray). Regions of the protein that are structurally similar to GART are shown in red. Active site loop (black) is disordered, two catalytic residues Asn 168, His 170 represented in yellow.

Currently research is underway to solve the structure of PvdF with cofactor in the active site and to also identify the function of newly identified additional secondary structural elements within this structure.

1.8 AICAR and GART transformylases – structural and mechanistic comparison.

While they catalyze similar reactions, AICAR transformylase and GART are very distinct both structurally and functionally. The enzymes share no significant similarity in protein sequence or fold. A protein sequence comparison of human AICAR transformylase and GART enzymes in one example, has been used to hypothesize a conservation of GART active residues in AICAR transformylases [56]. However, mutagenesis studies indicated these residues are not required in AICAR transformylase activity [41]. Structurally, GART has a distinct Rossmann fold with the active site located between the two domains where one domain is responsible for cofactor binding and other for substrate binding and catalysis. AICAR transformylase has an alpha beta structure composed of three domains with an active site located in the protein dimer interface formed from domains 2 and 4 of opposing dimers. Functionally, AICAR transformylase is part of the bifunctional enzyme where the transformylase product is used by the IMP cyclohydrolase to make the final product. In GART, there are no examples of bifunctionality. Nevertheless, GART domains have been part of other transformylases where additional C-terminal domains were associated with substrate stabilization, or allosteric regulation of enzyme function (**Figure 1.4**).

It is also interesting to note that in purine metabolism both GART and AICAR transformylase have enzyme alternatives to make their respective products. For GART

reaction this is PurT and for AICAR transformylase this is the PurP gene. Both PurP and PurT use ATP to incorporate intercellular formate instead of using formate from the folate cofactor. PurT and PurP belong to the enzyme superfamily known as ATP-grasp enzymes [57]. These enzymes have a proposed acylphosphate intermediates that allow the transfer of an activated carbon unit via the nucleophilic attack of the amine of the substrate [58].

Mechanistic aspects of these two proteins are also different where composition of the substrate and the catalytic residues play key roles in formyl group transfer (**Figure 1.5**). The nucleophilicity of the formylated amines on the substrates of two proteins is different [42]. Positioning the formylated amine of AICAR close to the aromatic residues, due to resonance make this amine very unlikely to make the first nucleophilic attack to the formyl group of the cofactor without assistance of the residues in the active site. Two histidine residues, using the proton from the solvent, allow the lowering of the pKa of this amine [42]. In the case of GART, the initial step can proceed without direct involvement of active site residues. The active site residues in this transformylase seem to be used for stabilization of the transition state and proton transfer from the substrate to the cofactor. Differences in the catalytic proton movement in the two enzymes were outlined in mutagenetic studies of PurN and PurH. In mutagenetic studies of PurN, the conclusion suggest that no one residue is absolutely necessary for the activity of this enzyme. Indeed, one amino acid substitution of any conserved residues decreases the activity, while two or more mutations abolishes the activity of the enzyme [59]. Unlike PurN, PurH active site mutagenesis indicated an absolute requirement of all the residues in catalysis. These data demonstrate not only that the PurN active site is very

amenable to amino acid substitution but also the utilization of solvent in catalysis of these two enzymes. In GART, solvent or a water molecule is used to move protons in the transition state while in AICAR transformylase, activation of the formylated amine of the substrate is through the His residues that act as a proton shuttle between the solvent and the active site (**Figure 1.5**).

1.9 RibB enzyme from riboflavin biosynthesis.

All B-group vitamins have very important metabolic functions [60]. One of these vitamins, B2, otherwise known as riboflavin, is a crucial precursor for the production of the cofactors flavin adenine dinuclotide (FAD) and flavin adenine mononucleotide (FMN). These compounds are essential in redox and electron transport reactions in all forms of life [60, 61]. Besides this, riboflavin has been associated with a multitude of other functions such as: folate synthesis [60], iron absorption [62], carcinogen associated DNA damage [63] and inflammatory and immune response [64, 65].

Biosynthetic production of riboflavin is reserved for plants and bacteria. The gene cluster responsible for riboflavin synthesis in mammals is absent and riboflavin is mostly obtained from the diet. Due to this, research on riboflavin biosynthesis has drawn a lot of attention as a method for designing new antibacterial/ antimicrobial agents [66-68]

Riboflavin is made by the *rib* gene cluster which is composed of five genes, *ribABCDE.* As proposed by Bacher [61, 69], riboflavin biosynthesis is divided into two branches (**Figure 1.7**). In brief, biosynthesis starts with a molecule of GTP and the enzyme RibA, which is responsible for the removal of carbon C8 of the GTP nucleotide as formate.

The RibA product is transformed by RibD, which performs deamination of the pyrimidine ring as well as reduction of the ribose ring. In the other branch, a molecule of ribulose 5-phosphate (Ru5P) is converted by RibB to 3,4 dihydroxy-2-butanone phosphate (DHBP) releasing carbon C4 of the substrate as formate. RibE uses the products from the two branches to make ribityl dimethyllumazine. In the final step RibC catalyzes an interesting reaction where the enzyme uses two molecules of ribityl dimethyllumazine to make one molecule of riboflavin and reforms the RibD product that is recycled in the biosynthetic pathway [70].

RibB is a magnesium dependent enzyme. Research presented here suggests that this enzyme requires a single metal for catalysis. Also, I have provided evidence that support a new mechanism of formate release by this enzyme.

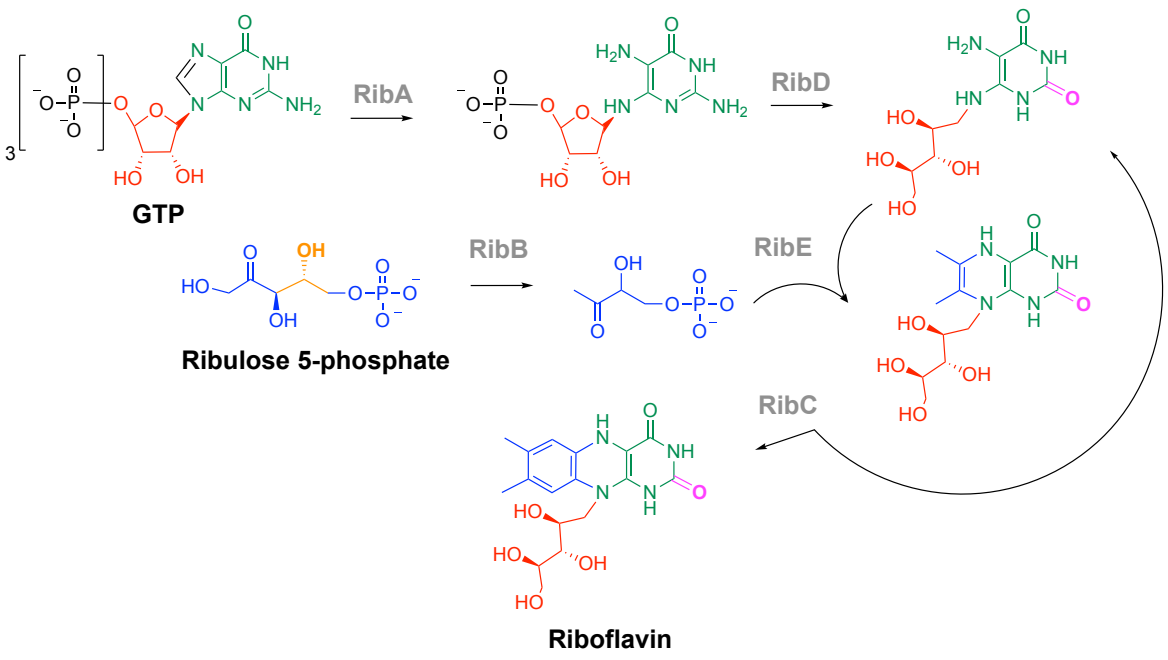


Figure 1.7. Biosynthetic pathway of riboflavin. Riboflavin is made from the gene cluster composed of five enzymes RibA-RibE. Pathway is divided into two branches. First branch starts with a molecule of GTP and action of RibA enzyme and second with a molecule of ribulose 5-phosphate and enzyme RibB.

1.10 REFERENCES

1. Brosnan, M.E. and J.T. Brosnan, *Formate: The Neglected Member of One-Carbon Metabolism.* Annu Rev Nutr, 2016. **36**: p. 369-88.
2. Crable, B.R., et al., *Formate formation and formate conversion in biological fuels production.* Enzyme Res, 2011. **2011**: p. 532536.
3. Lamarre, S.G., et al., *Formate: an essential metabolite, a biomarker, or more?* Clin Chem Lab Med, 2013. **51**(3): p. 571-8.
4. Leonhartsberger, S., I. Korsa, and A. Bock, *The molecular biology of formate metabolism in enterobacteria.* J Mol Microbiol Biotechnol, 2002. **4**(3): p. 269-76.
5. Tabor, H., et al., *Urocanic acid as an intermediate in the enzymatic conversion of histidine to glutamic and formic acids.* J Biol Chem, 1952. **196**(1): p. 121-8.
6. Gal, E.M. and A.D. Sherman, *L-kynurenine: its synthesis and possible regulatory function in brain.* Neurochem Res, 1980. **5**(3): p. 223-39.
7. Knappe, J., et al., *Pyruvate formate-lyase of Escherichia coli: the acetyl-enzyme intermediate.* Eur J Biochem, 1974. **50**(1): p. 253-63.
8. Sah, S., et al., *One-carbon metabolic pathway rewiring in Escherichia coli reveals an evolutionary advantage of 10-formyltetrahydrofolate synthetase (Fhs) in survival under hypoxia.* J Bacteriol, 2015. **197**(4): p. 717-26.
9. Nagy, P.L., et al., *Formyltetrahydrofolate hydrolase, a regulatory enzyme that functions to balance pools of tetrahydrofolate and one-carbon tetrahydrofolate adducts in Escherichia coli.* J Bacteriol, 1995. **177**(5): p. 1292-8.
10. Ferry, J.G., *Formate dehydrogenase.* FEMS Microbiol Rev, 1990. **7**(3-4): p. 377-82.
11. Kane, A.L., et al., *Formate Metabolism in Shewanella oneidensis Generates Proton Motive Force and Prevents Growth without an Electron Acceptor.* J Bacteriol, 2016. **198**(8): p. 1337-46.
12. Popov, V.O. and V.S. Lamzin, *NAD(+)-dependent formate dehydrogenase.* Biochem J, 1994. **301 (Pt 3)**: p. 625-43.
13. Stams, A.J. and C.M. Plugge, *Electron transfer in syntrophic communities of anaerobic bacteria and archaea.* Nat Rev Microbiol, 2009. **7**(8): p. 568-77.
14. Stover, P.J. and M.S. Field, *Trafficking of intracellular folates.* Adv Nutr, 2011. **2**(4): p. 325-31.
15. Suh, J.R., A.K. Herbig, and P.J. Stover, *New perspectives on folate catabolism.* Annu Rev Nutr, 2001. **21**: p. 255-82.
16. Newman, A.C. and O.D.K. Maddocks, *One-carbon metabolism in cancer.* Br J Cancer, 2017. **116**(12): p. 1499-1504.
17. Mentch, S.J., et al., *Histone Methylation Dynamics and Gene Regulation Occur through the Sensing of One-Carbon Metabolism.* Cell Metab, 2015. **22**(5): p. 861-73.
18. Allegra, C.J., et al., *Enhanced inhibition of thymidylate synthase by methotrexate polyglutamates.* J Biol Chem, 1985. **260**(17): p. 9720-6.
19. Allegra, C.J., et al., *The effect of methotrexate on intracellular folate pools in human MCF-7 breast cancer cells. Evidence for direct inhibition of purine synthesis.* J Biol Chem, 1986. **261**(14): p. 6478-85.
20. Hughes L.R., S.T.C., Boyle F.T., Jackman A.L, ed. *Raltitrexed (TomudexTM), a Highly Polyglutamatable Antifolate Thymidylate Synthase Inhibitor*. Antifolate Drugs in Cancer Therapy. Cancer Drug Discovery and Development, ed. J. A.L. 1999, Humana Press: Totowa, NJ. 147-165.

21. Grem, J.L., et al., *A Phase I study of raltitrexed, an antifolate thymidylate synthase inhibitor, in adult patients with advanced solid tumors.* Clin Cancer Res, 1999. **5**(9): p. 2381-91.
22. Erba, E., et al., *Mechanism of cytotoxicity of 5,10-dideazatetrahydrofolic acid in human ovarian carcinoma cells in vitro and modulation of the drug activity by folic or folinic acid.* Br J Cancer, 1994. **69**(2): p. 205-11.
23. Beardsley, G.P., et al., *A new folate antimetabolite, 5,10-dideaza-5,6,7,8-tetrahydrofolate is a potent inhibitor of de novo purine synthesis.* J Biol Chem, 1989. **264**(1): p. 328-33.
24. Shih, C., et al., *LY231514, a pyrrolo[2,3-d]pyrimidine-based antifolate that inhibits multiple folate-requiring enzymes.* Cancer Res, 1997. **57**(6): p. 1116-23.
25. Chen, V.J., et al., *Preclinical cellular pharmacology of LY231514 (MTA): a comparison with methotrexate, LY309887 and raltitrexed for their effects on intracellular folate and nucleoside triphosphate pools in CCRF-CEM cells.* Br J Cancer, 1998. **78 Suppl 3**: p. 27-34.
26. Mendelsohn, L.G., et al., *Enzyme inhibition, polyglutamation, and the effect of LY231514 (MTA) on purine biosynthesis.* Semin Oncol, 1999. **26**(2 Suppl 6): p. 42-7.
27. Morgan, J., et al., *Disrupting folate metabolism reduces the capacity of bacteria in exponential growth to develop persisters to antibiotics.* Microbiology, 2018. **164**(11): p. 1432-1445.
28. Salcedo-Sora, J.E. and S.A. Ward, *The folate metabolic network of Falciparum malaria.* Mol Biochem Parasitol, 2013. **188**(1): p. 51-62.
29. Said, S., E.W. Jeffes, and G.D. Weinstein, *Methotrexate.* Clin Dermatol, 1997. **15**(5): p. 781-97.
30. Feely, M.G., A. Erickson, and J.R. O'Dell, *Therapeutic options for rheumatoid arthritis.* Expert Opin Pharmacother, 2009. **10**(13): p. 2095-106.
31. Walling, J., *From methotrexate to pemetrexed and beyond. A review of the pharmacodynamic and clinical properties of antifolates.* Invest New Drugs, 2006. **24**(1): p. 37-77.
32. Bermingham, A. and J.P. Derrick, *The folic acid biosynthesis pathway in bacteria: evaluation of potential for antibacterial drug discovery.* Bioessays, 2002. **24**(7): p. 637-48.
33. Ravanel, S., et al., *Tetrahydrofolate biosynthesis in plants: molecular and functional characterization of dihydrofolate synthetase and three isoforms of folylpolyglutamate synthetase in Arabidopsis thaliana.* Proc Natl Acad Sci U S A, 2001. **98**(26): p. 15360-5.
34. Masters, P.A., et al., *Trimethoprim-sulfamethoxazole revisited.* Arch Intern Med, 2003. **163**(4): p. 402-10.
35. Baggott, J.E. and T. Tamura, *Folate-Dependent Purine Nucleotide Biosynthesis in Humans.* Adv Nutr, 2015. **6**(5): p. 564-71.
36. Wall, M., J.H. Shim, and S.J. Benkovic, *Human AICAR transformylase: role of the 4-carboxamide of AICAR in binding and catalysis.* Biochemistry, 2000. **39**(37): p. 11303-11.
37. Guillon, J.M., et al., *Importance of formylability and anticodon stem sequence to give a tRNA(Met) an initiator identity in Escherichia coli.* J Bacteriol, 1993. **175**(14): p. 4507-14.
38. Guillon, J.M., et al., *Disruption of the gene for Met-tRNA(fMet) formyltransferase severely impairs growth of Escherichia coli.* J Bacteriol, 1992. **174**(13): p. 4294-301.
39. Guillon, J.M., et al., *Nucleotides of tRNA governing the specificity of Escherichia coli methionyl-tRNA(fMet) formyltransferase.* J Mol Biol, 1992. **224**(2): p. 359-67.
40. Greasley, S.E., et al., *Crystal structure of a bifunctional transformylase and cyclohydrolase enzyme in purine biosynthesis.* Nat Struct Biol, 2001. **8**(5): p. 402-6.

41. Wolan, D.W., et al., *Structural insights into the avian AICAR transformylase mechanism.* Biochemistry, 2002. **41**(52): p. 15505-13.
42. Wolan, D.W., et al., *Structure of avian AICAR transformylase with a multisubstrate adduct inhibitor beta-DADF identifies the folate binding site.* Biochemistry, 2003. **42**(37): p. 10904-14.
43. Zhang, Y., et al., *Crystal structures of human GAR Tfase at low and high pH and with substrate beta-GAR.* Biochemistry, 2002. **41**(48): p. 14206-15.
44. Zhang, Z., et al., *Structures of glycinamide ribonucleotide transformylase (PurN) from Mycobacterium tuberculosis reveal a novel dimer with relevance to drug discovery.* J Mol Biol, 2009. **389**(4): p. 722-33.
45. Schmitt, E., et al., *Crystal structure of methionyl-tRNAfMet transformylase complexed with the initiator formyl-methionyl-tRNAfMet.* EMBO J, 1998. **17**(23): p. 6819-26.
46. Genthe, N.A., et al., *Molecular structure of an N-formyltransferase from Providencia alcalifaciens O30.* Protein Sci, 2015. **24**(6): p. 976-86.
47. Schoenafinger, G., et al., *Formylation domain: an essential modifying enzyme for the nonribosomal biosynthesis of linear gramicidin.* J Am Chem Soc, 2006. **128**(23): p. 7406-7.
48. Reimer, J.M., et al., *Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase.* Nature, 2016. **529**(7585): p. 239-42.
49. Krissinel, E. and K. Henrick, *Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2256-68.
50. Shim, J.H. and S.J. Benkovic, *Evaluation of the kinetic mechanism of Escherichia coli glycinamide ribonucleotide transformylase.* Biochemistry, 1998. **37**(24): p. 8776-82.
51. Klein, C., et al., *Towards structure-based drug design: crystal structure of a multisubstrate adduct complex of glycinamide ribonucleotide transformylase at 1.96 A resolution.* J Mol Biol, 1995. **249**(1): p. 153-75.
52. Schmitt, E., S. Blanquet, and Y. Mechulam, *Structure of crystalline Escherichia coli methionyl-tRNA(f)Met formyltransferase: comparison with glycinamide ribonucleotide formyltransferase.* EMBO J, 1996. **15**(17): p. 4749-58.
53. Holden, H.M., J.B. Thoden, and M. Gilbert, *Enzymes required for the biosynthesis of N-formylated sugars.* Curr Opin Struct Biol, 2016. **41**: p. 1-9.
54. Thoden, J.B., et al., *Structure of a sugar N-formyltransferase from Campylobacter jejuni.* Biochemistry, 2013. **52**(35): p. 6114-26.
55. Woodford, C.R., J.B. Thoden, and H.M. Holden, *New role for the ankyrin repeat revealed by a study of the N-formyltransferase from Providencia alcalifaciens.* Biochemistry, 2015. **54**(3): p. 631-8.
56. Rayl, E.A., B.A. Moroson, and G.P. Beardsley, *The human purH gene product, 5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase. Cloning, sequencing, expression, purification, kinetic analysis, and domain mapping.* J Biol Chem, 1996. **271**(4): p. 2225-33.
57. Galperin, M.Y. and E.V. Koonin, *A diverse superfamily of enzymes with ATP-dependent carboxylate-amine/thiol ligase activity.* Protein Sci, 1997. **6**(12): p. 2639-43.
58. Zhang, Y., M. Morar, and S.E. Ealick, *Structural biology of the purine biosynthetic pathway.* Cell Mol Life Sci, 2008. **65**(23): p. 3699-724.
59. Warren, M.S., A.E. Marolewski, and S.J. Benkovic, *A rapid screen of active site mutants in glycinamide ribonucleotide transformylase.* Biochemistry, 1996. **35**(27): p. 8855-62.
60. Thakur, K., et al., *Riboflavin and health: A review of recent human research.* Crit Rev Food Sci Nutr, 2017. **57**(17): p. 3650-3660.
61. Bacher, A., et al., *Biosynthesis of vitamin B2 (riboflavin).* Annu Rev Nutr, 2000. **20**: p. 153-67.

62. Powers, H., et al., *A proposed intestinal mechanism for the effect of riboflavin deficiency on iron loss in the rat.* British Journal of Nutrition, 1993. **69**(2): p. 553 - 561.
63. Webster, R.P., M.D. Gawde, and R.K. Bhattacharya, *Modulation of carcinogen-induced DNA damage and repair enzyme activity by dietary riboflavin.* Cancer Lett, 1996. **98**(2): p. 129-35.
64. Lakshmi, R., et al., *Effect of riboflavin or pyridoxine deficiency on inflammatory response.* Indian J Biochem Biophys, 1991. **28**(5-6): p. 481-4.
65. Schramm, M., et al., *Riboflavin (vitamin B2) deficiency impairs NADPH oxidase 2 (Nox2) priming and defense against Listeria monocytogenes.* Eur J Immunol, 2014. **44**(3): p. 728-41.
66. Long, Q., et al., *Riboflavin biosynthetic and regulatory factors as potential novel anti-infective drug targets.* Chem Biol Drug Des, 2010. **75**(4): p. 339-47.
67. Sassetti, C.M., D.H. Boyd, and E.J. Rubin, *Genes required for mycobacterial growth defined by high density mutagenesis.* Mol Microbiol, 2003. **48**(1): p. 77-84.
68. Mack, M. and S. Grill, *Riboflavin analogs and inhibitors of riboflavin biosynthesis.* Appl Microbiol Biotechnol, 2006. **71**(3): p. 265-75.
69. Bacher, A., et al., *Biosynthesis of riboflavin: GTP cyclohydrolase II, deaminase, and reductase.* Methods Enzymol, 1997. **280**: p. 382-9.
70. Fassbinder, F., M. Kist, and S. Bereswill, *Structural and functional analysis of the riboflavin synthesis genes encoding GTP cyclohydrolase II (ribA), DHBP synthase (ribBA), riboflavin synthase (ribC), and riboflavin deaminase/reductase (ribD) from Helicobacter pylori strain P1.* FEMS Microbiol Lett, 2000. **191**(2): p. 191-7.

CHAPTER 2:

PvdF, hydroxyornithine formyltransferase from *Pseudomonas aeruginosa*

The work discussed in this chapter has been published in the following research article:

Kenjić N. , Hoag M.R., Moraski G.C., Caperelli C.A., Moran G.R., Lamb A.L., *PvdF of pyoverdin biosynthesis is a structurally unique N^{10}-formyltetrahydrofolate-dependent formyltransferase*, Arch Biochem Biophys (2019) 664: 40-50[1]

2.1 INTRODUCTION

Pseudomonas aeruginosa is a gram-negative, opportunistic pathogen. It causes nosocomial infections in individuals with compromised immune systems (like burn victims and HIV patients) [1, 2] and pneumonia in individuals with cystic fibrosis [3]. According to the CDC, 51,000 patients are diagnosed with *Pseudomonas*-related infection per year, with 13% of these attributed to antibiotic resistant infections [4]. Due to an increase in antibiotic resistance of *P. aeruginosa,* it has been included in a group known as ESKAPE pathogens, group of pathogenic bacteria, both Gram-positive and Gram-negative, leading causes of nosocomial infections [5-7]. Antibiotic resistance of *P. aeruginosa* can be attributed to its ability to live in a biofilm form [8] and a plethora of different virulence factors [1]. Virulence factors can be proteins involved in development of acute infections (exoenzyme S, phospholipase C, immunoglobulins G), natural

[1] Contributions of the co-authors: Kenjić - performed the kinetic analysis and solved crystal structure, wrote the draft of the manuscript; Hoag – performed mass spectrometry analysis and provided respective data; Moraski – prepared OHOrn; Caperelli – prepared fDDF and DDF; Moran – mentor of Hoag, provided access and training for the mass spectrometry experiment, consulted on kinetic experimental design and interpretation, edited manuscript; Lamb – mentor of Kenjic, provided resources and training for experimentation, assisted with data analysis and manuscript preparation

products linked to chronic infections (siderophores such as pyoverdine or pyochelin or quorum sensing molecules such as N-acylhomoserine lactone [9]), or transcription regulatory systems [10]. In order to explore new approaches to control infection by this pathogen, researchers have turned to studies of some of these virulence factors.

One virulence factor of particular interest for chronic infections is involved in iron acquisition. Iron is critical for the function of enzymes in redox reactions, secondary metabolism, and DNA repair [11]. Therefore, during an infection, a pathogen needs to ensure a constant supply of iron to survive. For example, opportunistic pathogens such as *Pseudomonas aeruginosa* require 1μM of Fe for development. However, Fe(III) is insoluble and very limited in an infection environment, with concentration of about 10^{-19}M in human serum [12]. Iron can be acquired by bacteria through either direct or indirect means [13]. Pathogens can directly obtain iron from the environment (host or other bacteria) in the form of ferrous iron or by using iron acquisition proteins and cofactors such as lactoferrin [14], transferrin [15], ferritin and heme [16]. However, when iron in the environment is scarce, bacteria acquire iron by secretion of low-molecular weight, high-iron affinity molecules known as siderophores.

To ensure appropriate chelation that is ferric iron-specific, pathogens can vary siderophore functional groups and size. Depending on environmental conditions such as pH, pathogens secrete siderophores that allow for optimum iron chelation. As such, siderophores can be divided into five categories by functional group: hydroxamate, catecholate, phenolate, carboxylate, or mixed (**Figure 2.1**) [13].

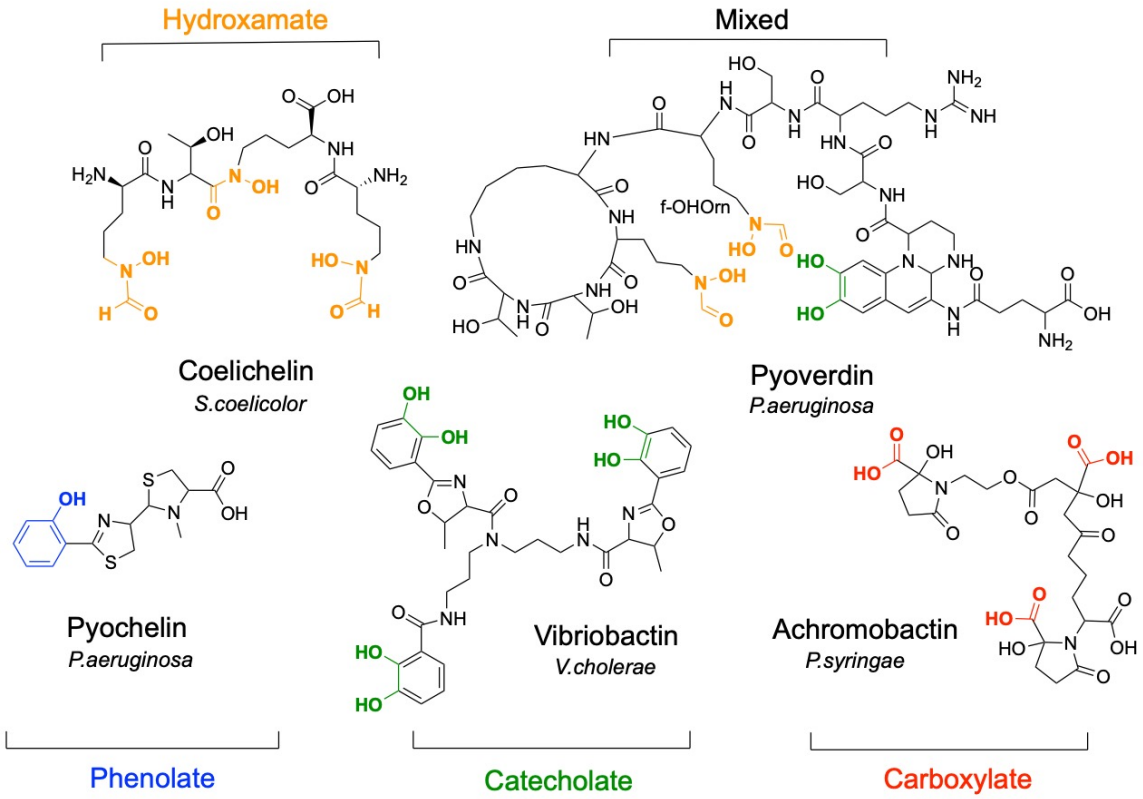


Figure 2.1. Classes of siderophores. Siderophores can be divided into five classes depending on functional groups used in iron chelation

P.aeruginosa primarily makes siderophores called pyoverdin and pyochelin. Previous studies have shown that pyoverdin is vital in virulence as well as biofilm development (ref) and as a signaling molecules in quorum sensing [8]. Structurally, pyoverdins are composed of a chromophore, an 8-12 amino acid linker peptide and a cyclic peptide portion, which is assembled by nonribosomal peptide synthethase (NRPS) enzymes. The length and backbone composition of the linker peptide varies between species of *Pseudomonas* [7], which has led to the characterization of over 50 pyoverdins divided into 3 different types [7]. Amino acids used in pyoverdin synthesis can be either proteogenic or non-proteogenic in origin and can include: D-amino acids, D-hydroxyornithine, cyclized-dN-hydroxyornithine, β-hydroxyaspartic acid, β-hydroxyhistidine [7]. *Pseudomonas aeruginosa* is one of two *Pseudomonas* species that contains the genes for synthesis of N^5-formyl-N^5-hydroxyornithine (fOHOrn), a modified amino acid used as an accessory for iron chelation in all three pyoverdine groups made by this pathogen. N^5-formyl-N^5-hydroxyornithine is made by two enzymes, PvdA and PvdF, and subsequently incorporated into pyoverdine via NRPS enzymes PvdJ and PvdI (**Figure 2.2**) [8]. PvdA, ornithine hydroxylase, was previously structurally and biochemically characterized, and shown to catalyze the formation of hydroxyornithine from ornithine using molecular oxygen, FAD, and NADPH [17]. It was hypothesized that PvdF utilizes hydroxyornithine to make formyl hydroxyornithine; however, the structure and mechanism had not been determined before this work (**Figure 2.2**). Previous studies had shown that a deletion strain of PvdF does not produce pyoverdine and is avirulant [18]. Furthermore, cell extracts of the PvdF deletion strain were not able to convert hydroxylamine to hydroxamate [18].

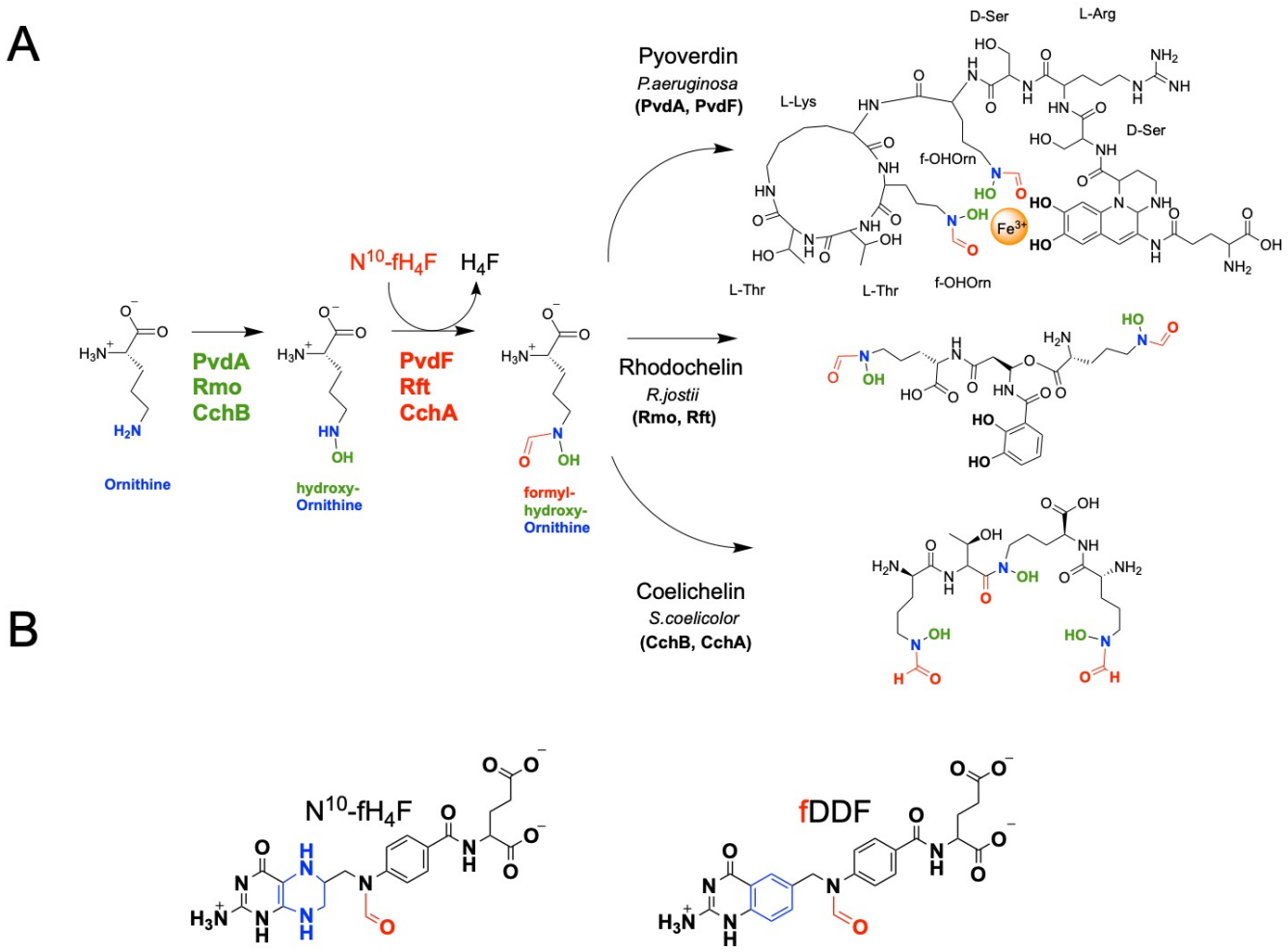


Figure 2.2. Role of PvdF in pyoverdin biosynthesis. (A) PvdA and PvdF derivatize ornithine to generate the precursor N^5-formyl-N^5-hydroxyornithine (fOHOrn). This molecule is subsequently incorporated into pyoverdin by nonribosomal peptide synthetases PvdF catalyzes formyl group transfer from N^{10}-formyltetrahydrofolate (N^{10}-fTHF) to N^5 of N^5-hydroxyornithine. In literature two other sets of enzymes have been show to catalyze reactions identical to PvdA/PvdF, Rmo/Rft form rhodochelin biosynthesis and CchB/CchA from coelichelin biosynthesis (B) Chemical representation of N^{10}-formyl-THF (right) and cofactor analogue fDDF (left). The ring altered between the cofactor and the analogue is highlighted in blue.

Genes that produce proteins that are functionally similar to PvdA and PvdF have been reported in clusters responsible for the synthesis of other siderophores like rhodochelin (*Rhodococcuss jostii RHA1*) [19], cholichelin (*Streptomyces coelicolor*) [20] and amychelin (*Amycolatopsis sp.AA4*) [21]. These siderophores are made by NPRS enzymes and the hydroxamate group on formyl-hydroxyornithine is used to chelate iron. Of these, only the rhodochelin enzymes Rmo and Rft and cholichelin enzyme CchB have been enzymatically characterized. While Rmo and CchB perform the same reaction as PvdA, Rft catalyzes the formylation reaction using N^{10}-formyl THF as a substrate. While PvdF, Rft and CChA share 12.8% with PvdA, Rmo and CchB share 35.4% sequence identity. [19] (**Figure 2.2**).

Here I report the structural characterization of a hydroxyornithine transformylase, PvdF from *Pseudomonas aeruginosa*. The structure reveals a core fold common among N^{10}-fTHF-dependent transformylases, including the glycinamide ribonucleotide transformylases (GART) [22-26], the methionyl-tRNA transformylase (MTF) [27, 28], and N-sugar transformylases of O-antigen formation [29-33]. However, the structure reveals large, unique insertions that we propose are important for binding the substrate OHOrn, and that place PvdF as the first documented member of a new structural subclass. This work includes a steady state kinetic analysis that indicates a partially ordered, formally random-sequential bireactant system that favors folate binding.

2.2 MATERIALS AND METHODS

2.2.1 Preparation of PvdF overexpression plasmid.

The *pvdF* gene was cloned from *Pseudomonas aeruginosa (PAO1)* genomic DNA, using polymerase chain reaction (PCR) with Herculase polymerase (Stratagene). The reaction was supplemented with 8% (v/v) DMSO as per manufacturer instructions due to high G-C content (61%). The forward primer (5'-AAT TAT ATA CAT ATG ACG AAA AGG AAA CTG GCC TA -3') contains an *NdeI* restriction site (underlined), and the reverse primer (5'-AAT ATA ATA CAG ATC TGG GAG CTT CTC GGC GAG CAG C-3') contains an *BglII* restriction site (underlined). The amplified DNA fragment was ligated into the correspondingly digested pET29b vector (Novagen) with T4 DNA ligase (New England BioLabs). The overexpression plasmid generates the PvdF protein with a C-terminal thrombin cleavage site followed by a histidine tag. This construct was further modified by site-directed mutagenesis to incorporate two stop codons at the C-terminus of the PvdF sequence so that the native PvdF protein, without purification tags, could be expressed. The Quik-Change® site-directed mutagenesis kit (Stratagene) was used with the forward primer (5'-CTG CTG GCC GAG AAG CTC TGA TGA CTG GGT ACC CTG GTG-3') and reverse complement primer (stop codons underlined).

2.2.2 Preparation of K72A,K74A-PvdF expression plasmid.

The PvdF K72A,K74A expression plasmid was prepared by the Genscript plasmid preparation and mutagenesis services. The *pvdF* gene was cloned into the pET29b expression plasmid at the *HindIII* restriction site on the 3' end and the *NdeI* restriction

site on the 5' end following the stop codon. The gene was synthesized such that the codon for K72 (AAA) was changed to encode alanine (GCA) and the codon for K74 (AAG) to encode alanine (GCG).

2.2.3 Wildtype PvdF expression and purification.

The PvdF plasmid was transformed into BL21(DE3) *E.coli* (New England BioLabs) for expression. Baffled flasks containing 1 L of LB Miller media containing 50 µg/ml of kanamycin were inoculated with 10 mL of overnight culture and grown at 37 ^{o}C in a shaker incubator (225rpm). When the OD_{600} reached 0.5, the temperature was lowered to 25 ^{o}C and allowed to equilibrate for 15 min. Expression was induced with isopropyl-β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.2 mM with shaking incubation for 16 hours. The cells were harvested by centrifugation (6000 x g, 5 min, 4 ^{o}C). The cell pellets were resuspended in 20 mL of 50 mM Tris-HCl pH 8.5 and lysed by three passes through a French press apparatus (35,000 psi). The lysate was centrifuged (12000 x g, 30 min, 4 ^{o}C) and the supernatant was injected onto a Source 30Q affinity column (GE Healthcare) pre-equilibrated with 50 mM Tris-HCl pH 8.5. The protein was eluted with a linear gradient of increasing NaCl to 500 mM. Protein fractions containing PvdF were confirmed by 15% SDS-PAGE and pooled. The salt concentration was adjusted to 1 M final concentration by slow addition of solid NaCl with gentle mixing. The protein was injected onto a Source Phenyl Sepharose (GE Healthcare) column pre-equilibrated in 50 mM Tris-HCl pH 8.5, 1 M NaCl. The PvdF protein was eluted from the column using a gradient to a buffer with no NaCl (50 mM Tris-HCl pH 8.5). Fractions containing PvdF were concentrated and injected onto a Superdex 200

gel filtration column (GE Healthcare) pre-equilibrated in 50 mM potassium phosphate pH 7.4. PvdF eluted at a molecular weight consistent with monomeric protein. The protein was concentrated with an Amicon® Ultracell® 30K centrifugal filter to 70 mg/mL as determined by Bradford assay, and stored at -80 °C. The purification protocol yielded 148 mg per liter of culture.

2.2.4 Expression and purification of K72A,K74A-PvdF.

The K72A,K74A-PvdF expression plasmid was transformed into BL21(DE3) *E.coli* (New England BioLabs). The variant protein was expressed and purified in a similar manner to wildtype PvdF, except that the phenyl sepharose column was not required to attain high purity. Therefore, the protein eluted from the Source 30Q affinity column was directly injected onto the Superdex 75 gel filtration column. This preparation yielded 55 mg of protein per liter of cell culture.

2.2.5 Selenomethionine substituted PvdF expression and purification.

Se-Met PvdF was produced according to the protocol by Van Duyne et al.[34] with some modifications. M9 minimal media was augmented with 2 mM $MgCl_2$, 0.1 mM $CaCl_2$, 0.4% (w/v) glucose and 50 μg/mL kanamycin. Growth cultures (1L) were inoculated with 10 mL of overnight culture and incubated at 37 °C in a shaker incubator (225 rpm) until an OD_{600} of 0.5 was reached. The temperature was lowered to 25 °C and an amino acid mixture was added to inhibit methionine production and allow for selenomethionine incorporation (the amino acid mixture included: 100 mg each of lysine, phenylalanine, threonine; 50 mg each of isoleucine, leucine, valine; 60 mg of

selenomethionine, per liter of culture). When the OD_{600} of the culture reached 1.0, IPTG was added to a final concentration of 0.2 mM and the culture was incubated for a further 16 hours with shaking. The SeMet protein purification was performed as for the native protein, with the exception that all buffers were supplemented with 2 mM dithiothreitol (DTT). The purified protein was concentrated with an Amicon® Ultracell® 30K centrifugal filter to 80 mg/mL as determined by Bradford assay and stored in -80 ^{o}C. The purification protocol yielded 100 mg per liter of culture.

2.2.6 PvdA protein expression and purification.

The PvdA enzyme was expressed and purified as previously reported [17, 35].

2.2.7 Preparation of hydroxyornithine (OHOrn).

N^{5}-hydroxyornithine was prepared by Garrett Moraski (Montana State University) according to the published protocol. [36]

2.2.8 Preparation of 10-formyl-5,8 dideazafolate (fDDF) and 5,8 dideazafolate (DDF).

Both fDDF and DDF were a generous gift from Dr. Carol Caperelli (University of Cincinnati). Both substrates were prepared according to the published protocols. [37]

2.2.9 *Steady state activity assays.*

The assay buffer contained 50 mM potassium phosphate buffer pH 7.4. The deformylation of fDDF is followed by the change in extinction coefficient at 295 nm ($\Delta\varepsilon$=18.9 $mM^{-1}cm^{-1}$)[38]. The assay was performed using a TgK scientific stopped-flow

instrument at 25 °C equipped with a mercury-xenon lamp. Enzyme (200 nM) with 11 mM N^5-hydroxyornithine was mixed at 1:1 ratio with varied fDDF concentration (5 μM to 642 μM). The rate was measured for 30 sec and the rate dependence was fit to the Michaelis-Menten equation. When hydoxyornithine was varied, the reaction was performed in 96 well flat–bottom plate (Corning cat # 9107) using a Varian 50MPR Microplate Reader, with a total reaction mixture per well of 300 μL. Each well contained a final enzyme concentration of 100 nM. The hydroxyornithine concentration was varied from 5.8 mM to 100 mM. The reaction was initiated with addition of 150 μM fDDF; the highest concentration possible at the fDDF λ_{max}, within the linear range of the instrument. Reaction progress was monitored at 295 nm in 1 sec cycle reads for 90 seconds. Data were fit to the Michaelis-Menten equation.

2.2.10 Coupled steady state activity assay.

Pseudomonas aeruginosa ornithine hydroxylase (PvdA) was used to generate the substrate OHOrn. The standard assay buffer contained 50 mM potassium phosphate buffer pH 7.4. The reaction was performed in a 1.5 mL quartz cuvette using a Cary 50 Bio UV-visible spectrophotometer. The initial reaction (600 μL) contained 1 μM PvdA, 150 μM FAD, 10 mM ornithine. Varying concentrations of NADPH were added (133 μM to 1 mM) for PvdA to generate defined OHOrn concentrations. The progress of the reaction was monitored at 300 nm as a measure of NADPH turnover. When the change in absorbance at 300 nm ceased, 1 mM fDDF (final concentration) was added to the cuvette and the spectrometer was blanked. The transformylase reaction was initiated by the addition of 200 nM PvdF and monitored at 295 nm for 30 sec using a 0.5 cm

pathlength. The initial rates were plotted versus [OHOrn], assuming that each NADPH consumed by PvdA produced one OHOrn molecule. The plot showed a non-hyperbolic velocity curve, with a decreasing rate at concentrations of OHOrn above 400 μM. Points lower than 133 μM were not obtained due to insufficient signal-to-noise.

2.2.11 Wildtype and K72A,K74A-PvdF progress curves.

Progress curves were measured using the same buffer as the steady state assays, generating OHOrn with 1 μM PvdA, 150 μM FAD, 10 mM ornithine, 500 μM NADPH. After the reaction ceased to change at 300 nm, the PvdF reaction was initiated by the addition of fDDF (46-183 μM) and 200 nM PvdF. The reaction was monitored at 295 nm for 60 min using 0.5 cm pathlength quartz cuvette.

2.2.12 Mass spectrometry.

Samples from the PvdA-PvdF reaction were diluted 1000-fold with LC-MS grade water (Sigma-Aldrich), and 10 μL of each dilution was analyzed by LC-MS over 65 minutes on an LCMS-IT-TOF (Shimadzu Scientific Instruments) with a Shim-pack XR-ODS column. The mobile phase consisted of 95% of an aqueous 0.1% formic acid solution and 5% acetonitrile (Sigma Aldrich), with a total flow rate of 0.2 mL/min. An ESI source was used, and acquisition was performed in scan mode from 120-550 m/z for both positive and negative ion modes. A 10 msec ion accumulation time was used, and event time was set to 100 msec. A three stage gradient was run as follows: 5% acetonitrile for 5 minutes, a linear gradient from 5% to 95% acetonitrile over 20 minutes, and 95% acetonitrile for another 20 minutes.

2.2.13 PvdF crystallization.

Purified SeMet protein was exchanged into 50 mM potassium phosphate buffer pH 7.4, 2 mM DTT and diluted to 40 mg/mL. A few flakes of powdered DDF were added to the protein solution, and the mixture was incubated on ice for 15 min. The protein solution was centrifuged (12000 x g, 30 sec, 4 ^{o}C). The protein was crystallized using the hanging-drop vapor diffusion method. Crystallization drops were prepared by mixing 1.5 µL protein solution with 1.5 µL precipitant solution containing 0.55 M sodium citrate, 0.1 M Tris-HCl pH 8.5. Rectangular-prism shaped crystals with dimensions 0.15 µm x 0.15 µm x 0.04 µm grew within 2 weeks. For data collection, crystals were soaked in a cryoprotectant solution containing the precipitant solution augmented with 20% ethylene glycol and flash cooled in liquid nitrogen.

2.2.14 Single crystal X-ray diffraction data collection and processing.

A single wavelength anomalous dispersion (SAD) dataset was collected at the Stanford Synchrotron Radiation Laboratory (SSRL, Stanford, CA) beamline 12-2 using a wavelength of 0.9795 at 100 K. This wavelength was based on a selenium fluorescence scan which showed a strong signal, with an inflection point at 0.9795 Å. The software package Blu-Ice [39] was used to collect 847 oscillation images (0.15 ° per image). The exposure time per frame was 0.2 sec with a transmission of 3%, and the crystal-to-detector distance set at 400 mm. Diffraction data were processed using XDS[40] to 2.34 Å with anomalous signal to 2.73 Å in the space group $P2_1$ with cell dimensions of a=128 Å, b=92.7 Å, c=128 Å, β=90.1 °. While the crystals appeared single, the diffraction pattern showed twinning (overlapping lattices) making space group determination

problematic. Data were frequently auto-processed as P422, but had to be manually re-processed in $P2_1$ in order to obtain the solution. Despite a strong anomalous signal, the SAD data did not lead to a solution using PHENIX. Crank2 [41] in the CCP4 program suite[42] was used to determine the location of 24 Se atoms, providing initial phases to build eight monomers in the asymmetric unit. This solution had a figure of merit (FOM) of 0.782 and R_{comb} of 0.351. XTRIAGE[43] identified the twin fractions (-l, k, h; -h, -k, l; l, -k, h) with (-h, -k, l) showing the highest twin fraction of 0.49. This twin fraction was applied in subsequent rounds of model building and refinement using Coot [44] and Phenix.Refine [45]. Water molecules were added automatically and inspected manually using Coot. Citrate molecules, derived from the crystallization conditions, were modeled manually using Coot. DDF molecules were built using LigandFit [46, 47] with restraints generated using eLBOW [48] and REEL [49]. Statistics for data refinement and analysis can be found in **Table 2.1**.

2.2.15 PvdF crystallographic model.

The final PvdF model contains eight monomers; however, the model is discontinuous with several chain breaks per monomer due to disorder. The amino acid summary can be found in **Table 2.2**. The model contains 336 waters, seven DDF molecules and eight citrate molecules. The DDF molecules are located at an interface between monomers. Four DDF molecules are present in 100% occupancy, whereas the remaining three were refined to 66-73% occupancy. Final Ramachandran analysis has been calculated with MolProbity [50] with 96.5% in the favored regions and one outlier in an area of poor

Table 2.1. PvdF SeMet data collection and refinement statistics[2]

	SeMet
Data collection	
Beamline	SSRL 12-2
Wavelength (Å)	0.9795
Space group	$P2_1$
Cell dimensions; *a, b, c* (Å), (°)	127.9 92.7 127.9 90 90.11 90
Resolution (Å)	38.74 – 2.3 (2.38 – 2.3)
R_{merge}[a]	0.089 (0.381)
Total observations	324216 (14816)
Total unique observations	121662 (5715)
Mean ((*I*) / sd(*I*))	9.2 (2.2)
Completeness (%)	91.8 (87.4)
CC(1/2)	0.909 (0.856)
Redundancy	2.7 (2.6)
Refinement	
Resolution (Å)	38.67 – 2.3 (2.38 – 2.3)
R_{cryst}[b]	0.2270 (0.2930)
R_{free}[c]	0.2494 (0.3019)
Total unique observations	121938 (12169)
No. of non-hydrogen atoms	17617
Protein	16963
Ligand	328
Water	326
rms deviation bonds (Å)	0.037
rms deviation angles (°)	2.31
Overall mean B-factor (Å^2)	39.3
Ramachandran plot analysis[e]	
Favored region	96.53
Allowed region	3.47
DDF	
Chain A; occupancy, B-factor (Å^2)	1.0, 30.5
Chain B; occupancy, B-factor (Å^2)	none
Chain C; occupancy, B-factor (Å^2)	0.69, 25.8
Chain D; occupancy, B-factor (Å^2)	1.0, 31.5
Chain E; occupancy, B-factor (Å^2)	0.72, 25.8
Chain F; occupancy, B-factor (Å^2)	0.66, 25.3
Chain G; occupancy, B-factor (Å^2)	1.0, 21.5
Chain H; occupancy, B-factor (Å^2)	1.0, 29.6

[2] $R_{merge} = \Sigma_h|I_h - \langle I \rangle|/\Sigma_h I_h$, where I_h is the intensity of reflection *h*, and $\langle I \rangle$ is the mean intensity of all symmetry-related reflections [b]$R_{cryst} = \Sigma||F_o| - |F_c||/\Sigma|F_o|$, F_o and F_c are observed and calculated structure factor amplitudes. [c]Five percent of the reflections were initially reserved to create an R_{free} test set used during each subsequent round of refinement.

Table 2.2. Ordered amino acids PvdF monomers

	Residues
Monomer A	3 – 225, 231 – 274
Monomer B	3 – 225, 231 – 274
Monomer C	3 – 26, 30 – 80, 85 – 224, 231 - 272
Monomer D	2 – 225, 232 – 274
Monomer E	3 – 224, 231 – 272
Monomer F	3 – 158, 163 – 225, 230 – 273
Monomer G	2 – 27, 32 – 225, 229 – 275
Monomer H	3 – 80, 84 – 225, 230 – 274

density. Root mean square deviation values were calculated using PDBeFold[51] and protein interaction interfaces were calculated using PDBePISA[52]. Structures figures were prepared using Pymol[53]. Atomic coordinates and structure factors for SeMet PvdF were deposited into the Protein Data Bank, with the accession code 6CUL.

2.3 RESULTS

2.3.1 Preparation of PvdF.

The PvdF protein was heterologously produced in *E. coli* and the purification was completed in three steps, using anion exchange, phenyl sepharose and gel filtration chromatography. The 31 kDa protein eluted from the gel filtration column at a molecular weight consistent with monomeric protein in solution (**Figure 2.3**). The SeMet protein was purified using a similar protocol, with the addition of a reducing agent to all buffers. The SeMet protein was crystallized using citrate as the precipitant, and crystals only formed in the presence of the product analogue 5,8-dideazafolate (DDF). As defined in the Materials and Methods section, the structure determination was complicated by twinning. Despite the technical difficulties, initial phase estimates for PvdF were determined by single wavelength anomalous dispersion phasing using the selenomethionine-substituted form of the protein to 2.34 Å (**Table 2.1**). A representative electron density map of the refined structure is depicted in **Figure 2.4 A,** with example density for the DDF and a citrate molecule in **Figure 2.4 B**. The asymmetric unit contains eight monomers, arranged in two rings with 4-fold rotational

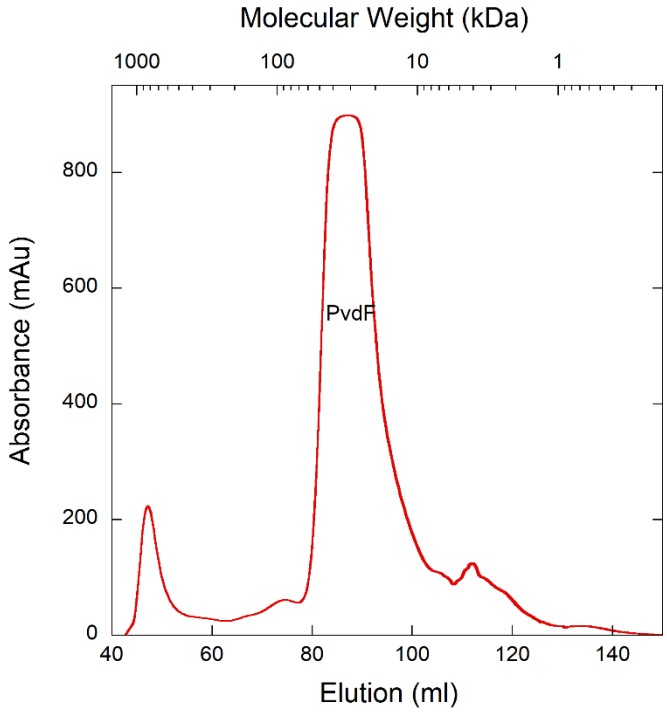


Figure 2.3. Size exclusion chromatography elution profile from a Superdex 200 column (GE Healthcare). PvdF elutes only in the peak at 89 ml, consistent with the molecular weight of a monomer, 31 kDa.

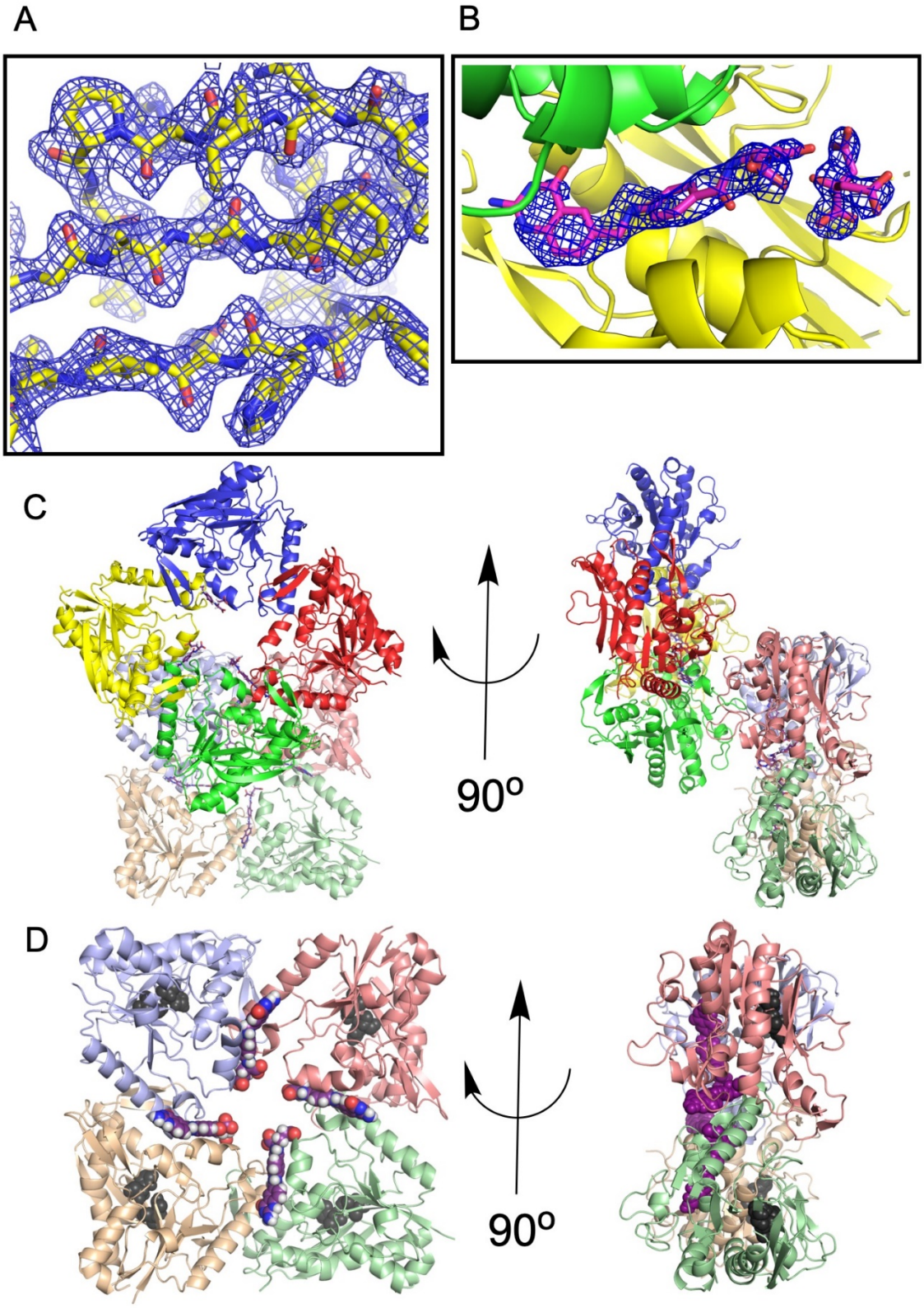

A
B
C
90º
D
90º

Figure 2.4. The asymmetric unit. (A) PvdF crystals appeared single in the drop; however, they showed evident twinning in the diffraction images. Refinement required implementation of the twin law (-h,-k,l). Representative electron density for a $2F_o$-F_c simulated annealing map (residues 166-174; 215-243) contoured at 1.5σ is shown. (B) PvdF crystals did not form without the cofactor analogue DDF. Electron density at the interfaces between seven of eight monomers is assigned to DDF. Electron density is displayed as a $2F_o$-F_c simulated annealing omit map contoured at 2σ. (C) PvdF crystallized with eight monomers in the asymmetric unit, as two rings with four-fold symmetry. Each monomer is a distinct color. (D) DDF (magenta) was observed at an interface between PvdF monomers. PvdF monomers E,F,G,H are shown and the location of the active site is highlighted with two of the three residues of the catalytic triad shown in black (the remaining residue is part of a disordered loop). If this were a productive catalytic binding mode, the formate would have to travel ~22 Å. In all other transformylases, the folate binds with the formate directly adjacent to the catalytic triad (within 5 Å).

symmetry (**Figure 2.4 C**). Consistent with the gel filtration data for the protein in solution, the average interface

area between monomers calculated by PDBePISA[52] was 719 $Å^2$, indicating that the four-fold symmetry is the result of the arrangement of the monomers in the crystal lattice and not indicative of an oligomeric state. Unexpectedly contacts within this monomer-monomer interface are mediated by the bound product analogue, 5,8 dideazafolate (DDF) which is not observed to bind in the putative active site, defined by the conserved catalytic triad (**Figure 2.4 D**).

2.3.2 Monomer architecture.

The core of the PvdF monomer shows the standard formyltransferase fold found in the N^{10}-formyltetrahydrofolate dependent enzymes, which is Rossmann-like, with a central 7-stranded sheet surrounded by helices and loops (**Figure 2.5 A**). The fold has been previously divided into two subdomains: an N-terminal subdomain for binding the folate substrate and a C-terminal subdomain for binding the substrate to be formylated [25, 26]. Glycinamide ribonucleotide transformylase (GART) is one of the best studied formyltransferases [22, 24], and *E. coli* GART (EcGART) will serve as a frame of reference for our discussion (**Figure 2.5 B**). The core of PvdF and EcGART are structurally conserved; however, PvdF has three major structural insertions. In the N-terminal subdomain, EcGART has a short loop between strand 1 and helix A. In PvdF, strand 1 is followed by a short helix, which is labeled helix a, and then an antiparallel β-sheet (strands i and ii), before helix A (grey in **Figure 2.5**). This is a total insertion of 23 amino acids (residues 12 to 34 in PvdF). In EcGART, the connection between the

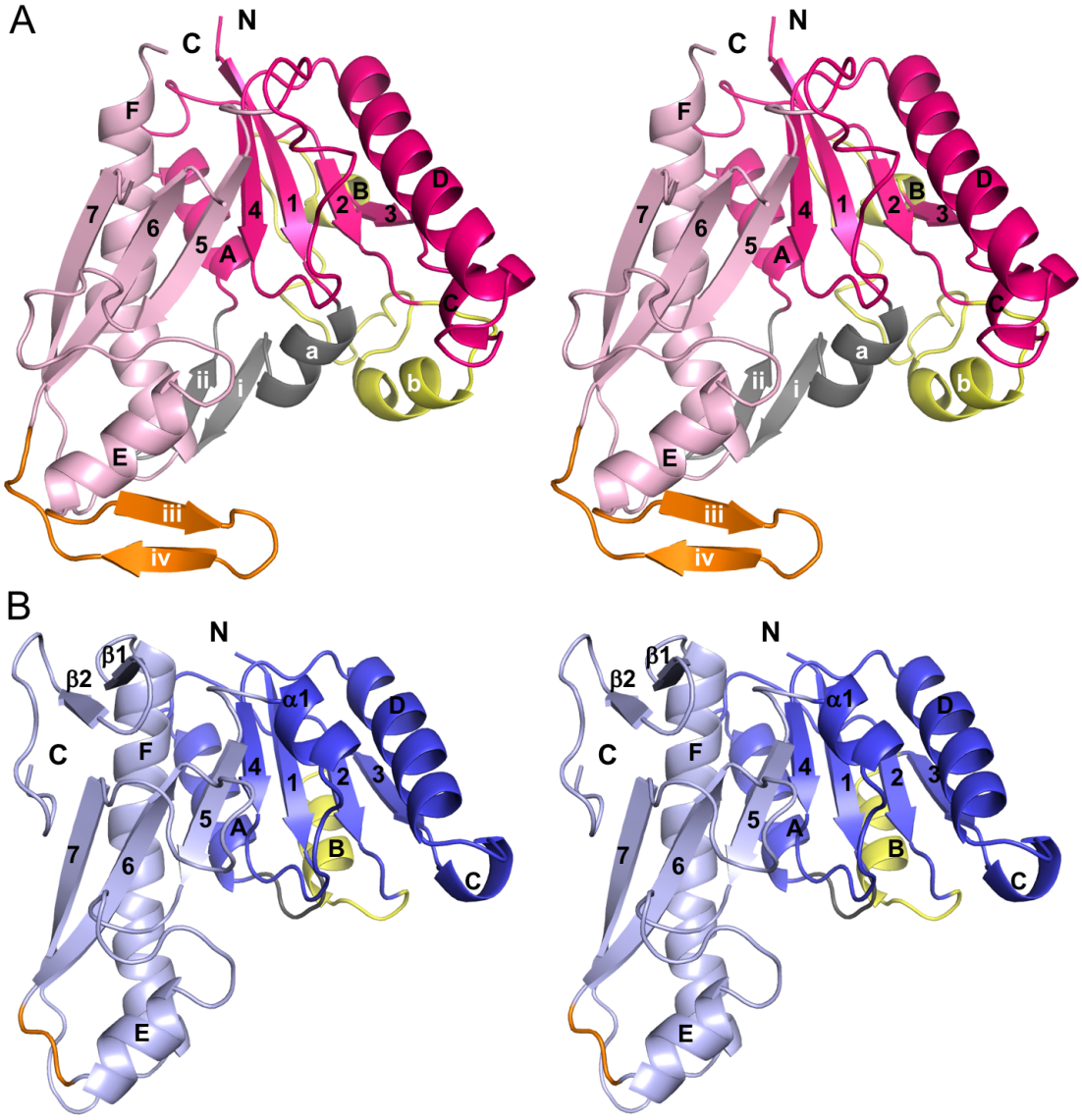


Figure 2.5. Walleye stereoimages showing monomer topology comparison between PvdF and EcGART. (A) While the formyltransferase fold and core of these proteins is similar, PvdF contains three major insertions sites respectively colored in gray (residues 12-34, helix a, strands i and ii), yellow (residues 63-102, loop and helix b) and orange (197-214, strands iii and iv). The N-terminal domain of PvdF (residues 1-162) in shown in magenta and the C terminal domain is in light pink (residues 163-275). (B) The N-terminal domain of EcGART (residues 1-100) is shown in dark blue and the C terminal domain is in light blue (residues 101-209). Insertion sites are represented in same colors as in panel A. EcGART contains structural elements that are not seen in PvdF, helix α1 and terminal sheet β1-β2. Helices are labeled with letters and sheets are indicated in numbers.

second β-strand and helix B is also a short loop. Helix B is followed by another short loop that connects to strand 3. PvdF has a large insertion at this connection (40 amino acids), beginning at residue 63. Helix B of the Rossman fold is replaced with helix b, which is roughly parallel to helix a (not packed against the central sheet). Helix b is followed by a long loop that does pack against the central sheet, structurally replacing helix B of EcGART. PvdF does have a short turn of a helix B before it rejoins the standard fold at residue 102 in strand 3. These changes are highlighted in yellow in **Figure 2.5**. Finally, in the C-terminal subdomain, PvdF has an insertion between helix E and strand 6, shown in orange in **Figure 2.5**. This insertion, residues 197 to 214 (18 amino acids), forms an antiparallel β-sheet with strands labeled iii and iv. EcGART continues after the F helix, with 2 additional strands forming a small sheet, structural elements not within the Rossmann-like fold and not found in PvdF.

2.3.3 Structural homologues.

The closest structural homologues to PvdF are the GART proteins, which are found in the pathway for the de novo biosynthesis of purines. The root mean square deviation (rmsd) calculated for the comparison of PvdF to EcGART is 2.2 Å for 165 Cα residues (**Figure 2.6 A and 2.6 B**). PvdF is 275 amino acids in length, whereas EcGART is 209 amino acids. The three insertions listed above account for most of the differences in Cα comparison, with other more subtle changes found within other loops. The methionyl-tRNA formyltransferase (MTF), which formylates the primary amine of the methionine attached to the initiator methionyl-tRNA, also shares a similar fold with PvdF and EcGART. The enzyme from *Yersinia pestis* is shown in **Figure 2.6 C**, which when

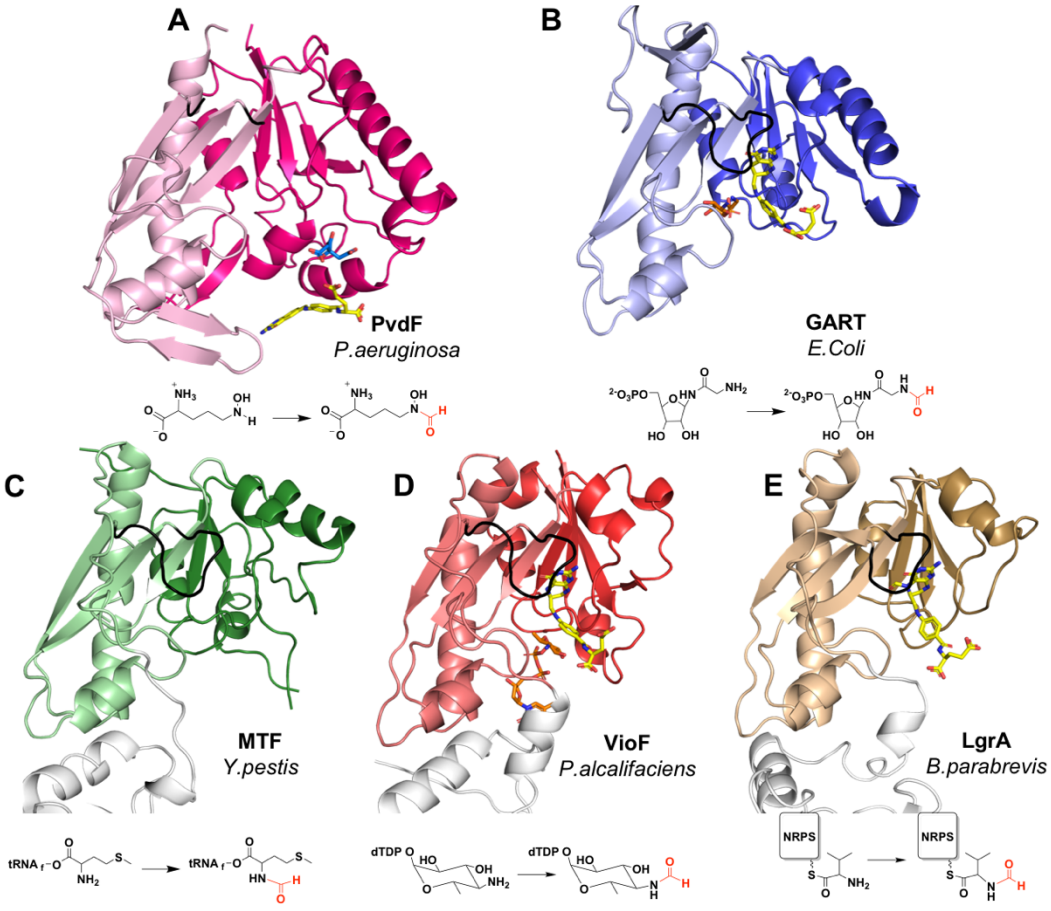


Figure 2.6. Comparison of PvdF with structural and functional homologues. (A) PvdF (B) GART from *E.Coli* (PDB:1CDE; sequence identity when compared to PvdF: 27%, calculated in LALIGN); (C) methionyl t-RNA transformylase from *Y.pestis* (PDB:3R8X; seq id: 26%); (D) VioF, sugar N-transformylase from *P.alcalifaciens* (PDB:4YFY; seq id: 28%); and (E) the NRPS formyltransferase domain of LgrA from *B.parabrevis* (PDB:5ES7; seq id: 25%). In each case the darker shade is the N-terminal and lighter shade is the C-terminal domain. In the bottom row the transformylase is part of a large multi-functional enzyme and domains without transformylase activity are white. The active site loop is shown in black. Folate analogues are shown in yellow with respective substrates in orange. The citrate molecule from crystallization in PvdF is shown in cyan. The reaction catalyzed by each enzyme is represented under the structure.

compared to PvdF has an rmsd of 2.3 Å over 150 Cα residues. The sugar N-transformylases involved in production of modified sugars for incorporation into O-antigens also fall into this structural and functional class. Of these, VioF from *P. alcalifaciens* O30 showed the closest structural similarity to PvdF with an rmsd of 2.5 Å over 145 Cα residues (**Figure 2.6 D**). Many of the sugar N-transformylases include C-terminal domains with other catalytic activities or regulatory roles [29-33]. Recently, the structure of the gramicidin initiation module (LgrA) was determined, which includes a formyltransferase domain (**Figure 2.6 E**). Interestingly, this formyltransferase domain has been proposed to be incorporated into the NRPS assembly line as the result of a gene duplication and horizontal transfer of a sugar N-transformylase [54], potentially an evolutionary precursor of VioF. At the C-terminal end of the LgrA transformylase domain is a new structural element, a loop that includes an α-helix, that serves as a linker to the adenylation domain of the NRPS module [54]. Note that the secondary structure insertions of PvdF are unique among these enzymes (**Figure 2.6**), placing PvdF in a new structural subclass of N^{10}-fTHF dependent transformylase enzymes.

2.3.4 Folate binding pocket.

Enzymes dependent on N^{10}-formyltetrahydrofolate are sometimes identified by a folate binding motif, HxSLLPxxxG where x is any residue, in the C-terminal subdomain (despite the N-terminal domain being labeled the folate-binding domain) (**Figure 2.7**) [29]. The histidine in this sequence is one of three residues in the catalytic triad, discussed later. This sequence starts in strand 5 and continues through the loop that connects to helix E, forming a portion of the folate binding pocket. For many of the N-

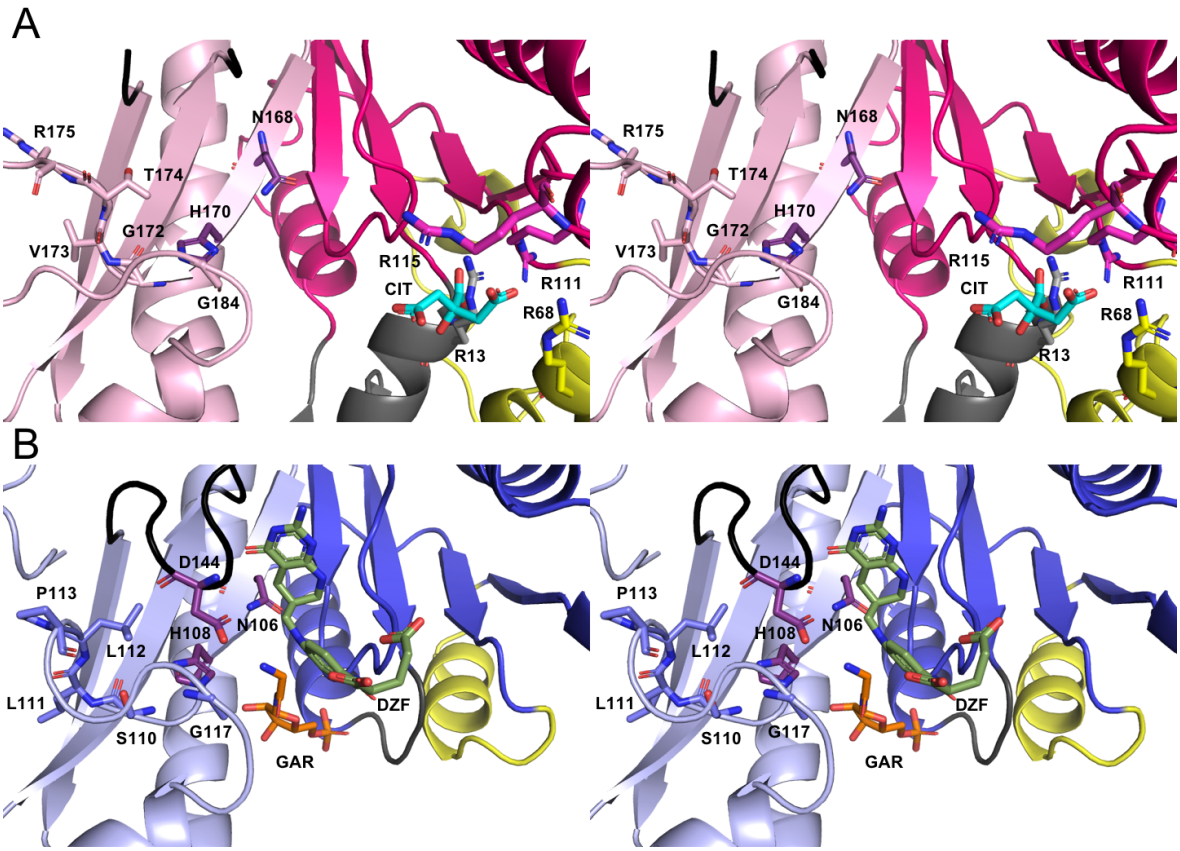


Figure 2.7. Walleye stereo comparison of the active sites of PvdF (A) and EcGART (B). EcGART contains the folate binding motif (S110-P113, light blue), that is not conserved in PvdF (G172-R175, light pink). The catalytic triad of Asn, His and Asp (shown in purple) is conserved in both PvdF and EcGART with Asp being part of the folate binding loop (shown in black). This loop was mobile in PvdF and not resolved. Because of unique insertions (gray and yellow, Panel A), PvdF contains an arginine binding pocket that contains a citrate molecule (cyan) derived from the crystallization mother liquor. By comparison to the binding orientation of the THF analogue in GART (dark green, Panel B), we propose that this arginine pocket may stabilize the mobile glutamate tail of the folate analogue in PvdF during catalysis.

sugar transformylases and for PvdF, this sequence is not conserved. The initial histidine and final glycine residues are conserved in three dimensions in PvdF. However, the remaining residues are not conserved, and the loop is 5 amino acids longer. The resulting sequence is: 170-HxGVTRyyyyyxxxG-184, where y is any residue in the extended PvdF loop. The primary contacts with the folate as determined in the GART and N-sugar transformylases are through hydrophobic interactions and hydrogen bonds between the methylpterin rings and the protein backbone found in the loop connecting strands 4 and 5, and in the loop connecting strands 6 and 7. The loop connecting strands 4 and 5 in EcGART and most of the other transformylases includes a helix (labeled α1 in **Figure 2.5 B**). In PvdF, this loop is similarly coiled, but does not make the hydrogen bonds requisite to define this as a helix. While the residues that interact with the methylpterin of the folates are not conserved between the GART and N-sugar transformylases, the shape, hydrophobicity of the pocket, and hydrogen bonding interactions are. In PvdF, the loop connecting strands 4 and 5 maintains the proper shape for interaction with the methylpterin rings; however, the loop connecting strands 6 and 7 is disordered, potentially because no folate is bound in the active site. This loop has been noted in other transformylases to be mobile [25, 55, 56]. We hypothesize that for catalysis, the folate binds in the analogous location in PvdF when compared to both the GART and N-sugar transformylases. PvdF crystals would not grow without inclusion of a folate analogue. However, the folates evident in the electron density map are found forming crystal contacts, as was described previously, rather than in the active site.

2.3.5 OHOrn binding.

The substrates differ widely between N-transformylase groups, and include nucleotide precursors, sugar-nucleotides, tRNA, amino acids, and amino acids attached to a NRPS carrier domain through a phosphopantetheinyl linker, as documented in **Figure 2.6**. It is not surprising that the individual substrate binding interactions are specific, with the ultimate goal of presenting the amine group undergoing formylation within range of the formyl group of the N^{10}-formyltetrahydrofolate. While there is no substrate bound in the PvdF structure, comparison to the holo-structures of transformylases previously determined suggests that the OHOrn substrate will bind in an analogous location (**Figure 2.7 A**). The loop between strand 1 and helix A forms important hydrogen bonding and ionic interactions with the phosphates of the substrates in the GART and N-sugar transformylases (grey in **Figure 2.7 B**) [22, 25, 30, 31]. In PvdF, the loop that contains these amino acids is not present, and this is instead the location of the first major insertion. Indeed, helix a, which is in the same three-dimensional space, has two sidechains, from Asn14 and Asp18, that point into the cavity and may form interactions with the backbone of the OHOrn substrate. The PvdF helix a is placed more interior in the active site than the GART or N-sugar transformylase 1-A loop, and the Asn14 and Asp18 sidechains would place the substrate deeper in the active site, potentially accounting for the considerably shorter OHOrn substrate.

The loop in the tRNA transformylase connecting β-strand 2 and helix B has been considered important for binding of the tRNA substrate [27, 28]. This loop corresponds with the second major insertion in PvdF; however helix b is in the comparable three dimensional location (yellow in **Figure 2.7**). Helix b is unlikely to play a direct role in

hydroxyornithine binding in PvdF, being too distant from the putative substrate binding site. Finally, helix F has a proline forming a kink in these homologues, not necessarily at the same turn in the helix, but still suitable to promote a conformation in which the N-terminus of the helix is bent toward the substrate binding cavity. In the GART and N-sugar transformylases, charged and polar residues of helix F form hydrogen bonds with the substrates. The analogous residue in PvdF is Arg252, which may serve a similar role. In all, the location for substrate binding is likely analogous to that seen in the GART and sugar transformylases, but the residues that promote binding may be contributed at least in part by the structural features that are unique to PvdF.

2.3.6 Citrate.

The model of PvdF has a well-ordered citrate bound by three arginine residues (13, 68, and 111) from helix a, helix b, and helix C in every monomer (**Figure 2.7 A**). The citrate molecule is derived from the mother liquor, which required greater than 0.5 M citrate for protein crystal formation. The binding of citrate in this site is undoubtedly a crystallization artifact. Nevertheless, the citrate is in close proximity of N^{10}-fTHF binding site, and is bound with a free carboxylate less than 3 Å from the putative location for the folate glutamate tail (**Figure 2.7 B**). This suggests that the citrate from the crystallization conditions, in large molar excess, prevented binding of folate in the active site. If this is correct, Arg111 or potentially Arg115 (nearby but not bound to citrate) may be involved in binding the glutamate tail of N^{10}-fTHF; however, a new crystal form with folate bound in the active site would be necessary to establish this. It is important to note that there are no comparable binding interactions for binding the glutamate tail

within the previously determined N^{10}-fTHF dependent transformylase domains: the tail is frequently found to be disordered or having high B-factors in structures where folate is bound [22, 26, 30, 31].

2.3.7 Catalytic Triad.

The loop connecting strands 6 and 7 (black in **Figure 2.6** and **Figure 2.7**) has been named both the active site loop [25] and the folate binding loop [26]. As noted before, this loop has been documented in other transformylases to be mobile, so it is not surprising that this loop is disordered in PvdF in the absence of folate. This loop harbors an aspartic acid that is one of three residues in a proposed catalytic triad. The other two residues, a histidine and an asparagine, are located in strand 5. These three residues are conserved in PvdF: Asn168, His170, and Asp229 (purple in **Figure 2.7**). In the GART and sugar transformylases, the binding pose of the folate is such that the formyl group on N^{10}-fTHF is positioned at the center of the triad [24, 29, 30]. The proposed mechanism for the GART enzymes, and by extension all enzymes of this class, suggests that the amino group of the substrate performs a nucleophilic attack on the carbonyl of the formyl group of N^{10}-fTHF, generating a tetrahedral intermediate. The catalytic triad residues are proposed to serve as general acid-general base residues to promote intermediate formation and resolution of the catalytic cycle [24, 29].

2.3.8 PvdF steady state kinetics.

N^{10}-fTHF dependent transformylases, such as those from purine biosynthesis, have been successfully assayed using the analogue 10-formyl-5,8-dideazafolate (fDDF) [38,

57]. When the formyl group is removed from fDDF, there is an increase in absorbance at 295 nm, allowing for a convenient continuous spectroscopic assay. When OHOrn is held in excess, and the varied substrate is fDDF, kinetic parameters are readily determined: K_m = 60 ± 10 μM, k_{cat} = 1.7 ± 0.1 sec^{-1} (**Figure 2.8**, red curve). However, the converse reaction, with fDDF in excess and OHOrn as the varied substrate, yielded a physiologically improbable K_m value: K_m = 30 ± 10 mM, k_{cat} = 1.0 ± 0.2 sec^{-1} (**Figure 2.9**). Furthermore, the experiment was not repeatable with each subsequent experiment showing an increase in K_m and a decrease in k_{cat}. The Km effect can be rationalized as OHOrn is known to be unstable [58-60], and so the effective substrate concentration was diminishing with time. We hypothesized that a solution to this problem was to have the preceding enzyme of the biosynthetic pathway, the ornithine hydroxylase PvdA, generate the necessary substrate *in situ*. PvdA is a flavin-dependent enzyme that must be reduced by NADH with each catalytic cycle [17, 35, 61]. The appropriate concentration of OHOrn was produced from PvdA by varying and limiting the concentration of NADH. The steady state plot produced in this manner, showed a non-hyperbolic velocity curve (blue in **Figure 2.8**). The curve does not fit well to a Michealis-Menten or a substrate inhibition model. Instead, the data suggest a random-sequential bireactant mechanism in which the pathway for fDDF binding first is preferred for product generation, a model that as has been previously described for other enzymes [62, 63].

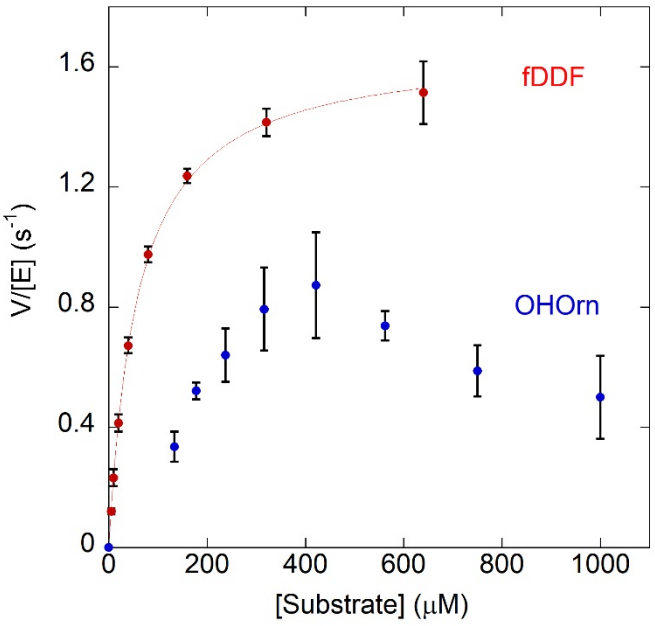


Figure 2.8. In the presence of fDDF as the varied substrate PvdF shows typical Michaelis-Menten kinetics (red trace, K_m = 60 ± 10 μM, k_{cat} = 1.7 ± 0.1 sec^{-1}). In this experiment, the synthesized OHOrn was used. Higher concentrations of fDDF were not possible due to the highly absorbant nature of the compound exceeding the linear range of the spectrometer. Using hydroxyornithine as the varied substrate generated by PvdA in a coupled assay, PvdF initially exhibits a sigmoidal curve (blue dotted trace) until 400 uM, after which the rate decreases. The model that best describes this behavior is a random bireactant mechanism in which the binding of fDDF is preferred first step.

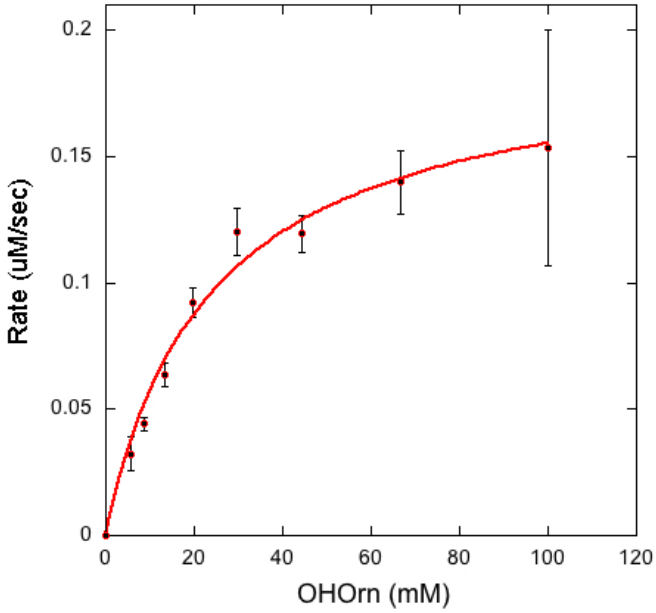


Figure 2.9. Steady state kinetics in the presence of synthesized hydroxyornithine as a substrate. The high value of K_M is rationalized by the instability of the OHOrn substrate. This plot was generated in triplicate using a plate reader on a single day. The kinetic constants derived from the above plot are K_m = 30 ± 10 mM, k_{cat} = 1.0 ± 0.2 sec^{-1}. On subsequent days, the K_M increased and k_{cat} decreased, and the experiment could not be reliably replicated.

2.3.9 Product detection.

The absorbance assay described above indicates the loss of the formyl group from fDDF, but not necessarily for formation of the product fOHOrn. To confirm fOHOrn production, the PvdA-PvdF reaction was analyzed by LCMS, monitoring for fOHOrn (m/z=177.1), fDDF (m/z=466.2) and DDF (m/z=438.1) (**Figure 2.10**). fOHOrn and DDF were observed, and the fDDF decreased in reactions containing both enzymes and all necessary substrates (flavin, NADH, Orn, fDDF), whereas controls that did not contain one of the enzymes (PvdA or PvdF) did not show production of the fOHOrn or DDF products.

2.3.10 The observed DDF binding mode is a crystallization artifact.

We hypothesized that the observed binding site for DDF, outside the active site and 22 Å distant from the catalytic triad, is a crystallization artifact (**Figure 2.4 D**). As mentioned previously, the crystals only grew in the presence of DDF, so potentially this binding promoted the formation of an oligomerization interface that promoted crystallization. Despite >10 years of effort, these twinned crystals were the best to date and the only ones that produced a refined structure. However, we now have the benefit of a refined structure in which we can analyze crystallization contacts. We generated a

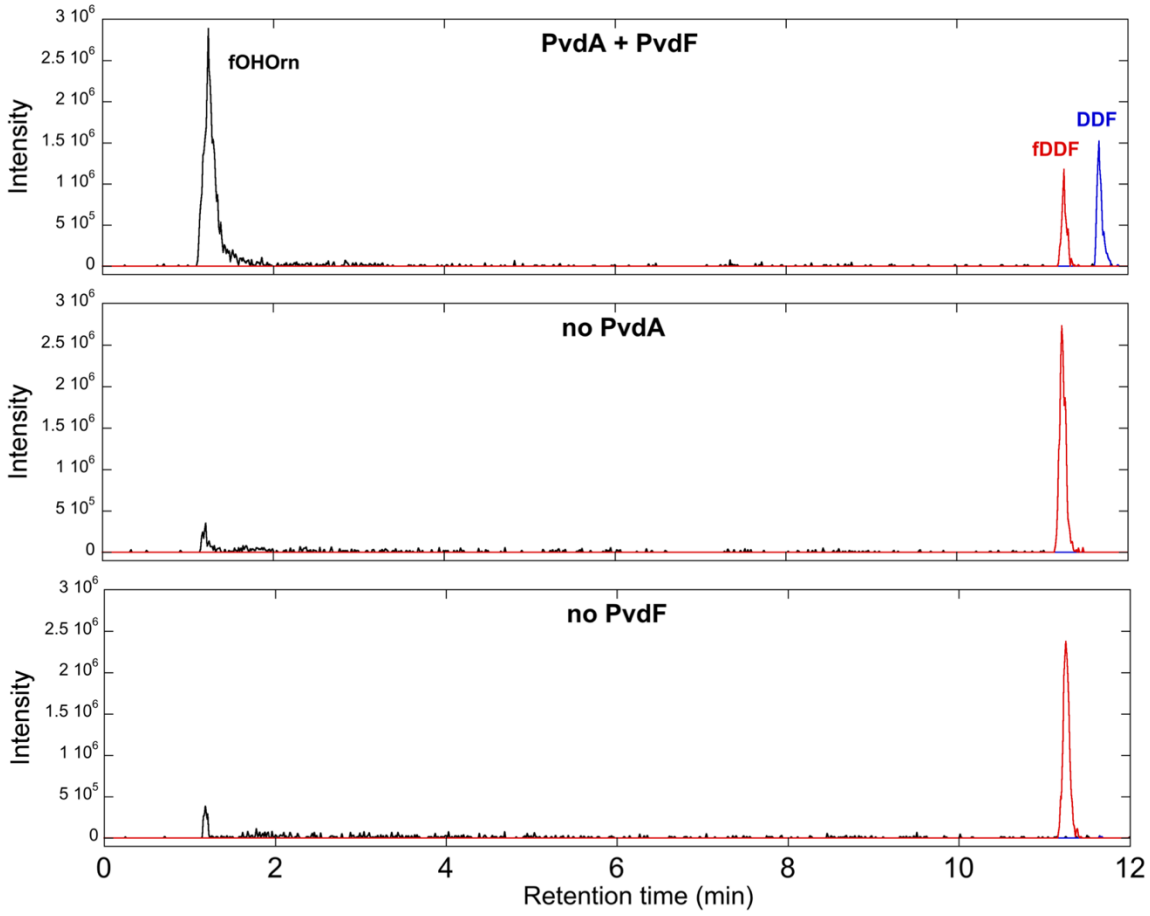


Figure 2.10. The formation of the product fOHOrn was monitored by LCMS (black trace, m/z=177.0), along with DDF (blue trace, m/z=438.1) and fDDF (red trace, m/z=466.2). Unlike in the control samples (lower two panels), the complete reaction (top panel) shows clear formation of the product by catalytic formyl group transfer.

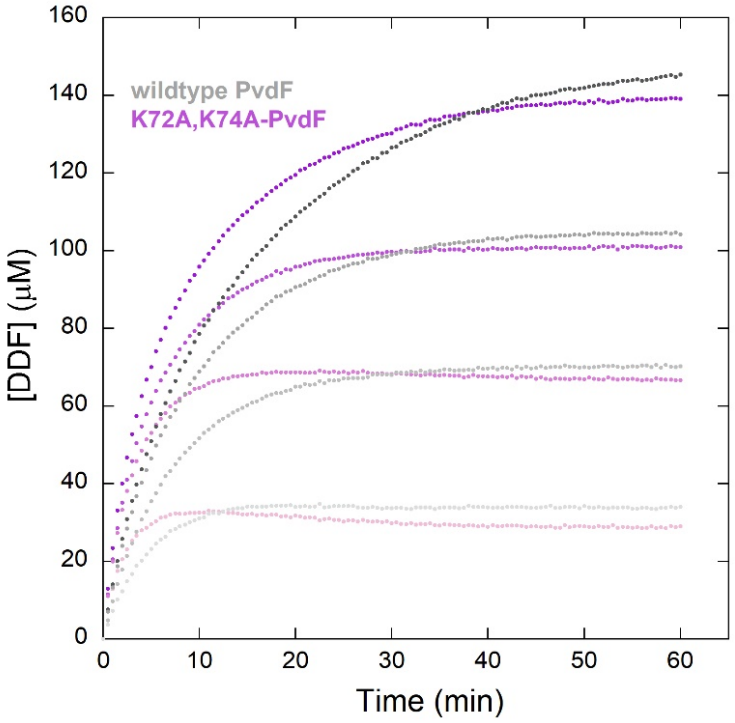


Figure 2.11. Progress curve comparison between the wild type (shade of grey) and K72A, K74A-PvdF variant (shades of purple/pink). The curves represent four concentrations of fDDF (from light to dark shades). The K72A,K74A-PvdF does not crystallize, suggesting that the crystallographic binding site for DDF is disrupted. The variant is more active than the wildtype, suggesting that the catalytic binding site for the folate is not what is observed in the structure.

variant, K72A,K74A-PvdF. These two lysine residues flank the DDF binding site. In monomers C and D, K72 directly hydrogen bond with the glutamate tail of DDF. K74 of one monomer is proximity of E65 of the next monomer in the ring, and in two of the eight cases, these residues form a hydrogen bond. The double K→A variant did not crystallize, but the protein was active as shown in full progress curves (**Figure 2. 11**).

Indeed, the slope of the curves for the K72A,K74A-PvdF is slightly steeper than those of the WT proteins, indicating that the protein is marginally more active with this variant. Taken together, we suggest that the K72A,K74A-PvdF variant no longer has the crystallographic DDF binding site, and yet is catalytically active. Therefore, the catalytically relevant binding of the folate is not seen in this PvdF structure. Instead, we hypothesize that in the catalytic complex, the folate will bind such that the formate to be transferred (attached to N^{10}) will be adjacent to the catalytic triad, as seen in all other homologues of this family.

2.4 DISCUSSION

PvdF is the formyltransferase that converts N^5-hydroxyornithine (OHOrn) to N^5-formyl-N^5-hydroxyornithine (fOHOrn) so that fOHOrn can be incorporated into the siderophore pyoverdin by a nonribosomal peptide synthetase assembly line (**Figure 2.2**). An N^{10}-fTHF-dependent hydroxyornithine transformylase involved in siderophore biosynthesis has been functionally characterized previously [19]. The enzyme, rhodochelin formyltransferase, or Rft, is involved in the biosynthesis of the mixed

catecholate-hydroxamate siderophore rhodochelin by *Rhodococcus jostii* RHA1, a gram positive soil bacteria. Like pyoverdin, this siderophore includes two formylhydroxyornithine residues for iron chelation, and is assembled by a nonribosomal peptide synthetase. Using an HPLC-MS assay, the authors showed conversion of OHOrn to fOHOrn, but there are no structural data for Rft. Sequence comparisons indicate that Rft is not a close structural homologue of PvdF. Instead, Rft is likely to be structurally similar to either the tRNA transformylase (FMT) or the N-sugar transformylase ArnA, an *E. coli* enzyme involved in lipid A modification that transformylates UDP-4-amino-4-deoxy-L-arabinose [19]. Rft, FMT, and ArnA all have the conserved HxSLLPxxxG motif for binding the folate co-substrate that PvdF lacks, and they all lack the major insertions highlighted in **Figure 2.5** that are specific to PvdF. Also unlike PvdF, they all have a C-terminal domain that provides additional functionality (to enhance substrate binding or to provide an additional catalytic activity). Finally, Rft is an allosteric enzyme showing positive cooperativity, and is proposed to be a tetramer in solution [19]. PvdF is a monomer in solution and in the crystals. The 4-fold ring structure seen in **Figure 2.4 B** is the result of crystal packing. Therefore, Rft is more functionally and likely structurally similar to the N-sugar transformylases and the tRNA transformylases than to PvdF.

The initiation module of the NRPS for the biosynthesis of the antibiotic gramicidin (LgrA) includes a formyltransferase that has been structurally characterized. The LgrA N^{10}-fTHF-dependent formyltransferase domain transfers a formate to the backbone amine of a valine while the amino acid is covalently attached to the peptidyl carrier domain by a phosphopantethienyl tether [54]. This is in contrast to PvdF, which is a

stand-alone accessory enzyme that formylates the sidechain amine of the free OHOrn substrate before the product fOHOrn is activated by an NRPS adenylation domain and attached to the carrier domain of the PvdI or PvdJ proteins [8]. LgrA has been hypothesized to be an evolutionary descendent of an N-sugar transformylase [54]. Both LgrA and *N*-sugar transformylases are considerably shorter in length, and lack the secondary structure insertions of PvdF, again suggesting that PvdF belongs to a distinct structural class of N^{10}-fTHF dependent formyltransferase.

PvdF maintains the 7-stranded β-sheet core in a Rossmann-like fold common to the N^{10}-fTHF dependent transformylases, and the catalytic triad characteristic for this class of enzymes is conserved. The structurally unique features of PvdF are likely responsible for interaction with the smaller amino acid substrate. The mechanism previously defined for the GART enzymes and hypothesized to be conserved in the class likely holds for PvdF. A detailed steady state kinetic analysis of human GART demonstrated an ordered-sequential kinetic mechanism in which the folate binds first [64]. Subsequent, pre-steady state kinetic experiments for the *E. coli* GART defined a random sequential kinetic mechanism in which folate and GAR bind in no obligatory order, but for which the apoenzyme has higher affinity for fDDF than GAR [57]. For PvdF, varying the concentration of the substrate analogue fDDF (OHOrn in excess) generated Michaelis-Menten kinetics with k_{cat} and K_m values in keeping with an enzyme from secondary metabolism. N^5-hydroxyornithine was a difficult substrate with which to work, and when used as the varied substrate, the data were not reproducible and provided kinetic values that were not physiologically relevant. However, when the ornithine hydroxylase (PvdA) of the same biosynthetic pathway was used to generate

the OHOrn *in situ*, reproducible data could be generated. Interestingly, the curve was nonhyperbolic and not well fit by Michealis-Menten nor a substrate inhibition model (**Figure 2.8**). Instead, these data represent a model defined by Ferdinand [62] and later echoed by Segel [63] in which the bireactant system shows random binding of the two substrates, but favors binding of the folate over the OHOrn, very similar to the kinetic models for the GART proteins.

2.5 REFERENCES

1. Moradali, M.F., S. Ghods, and B.H. Rehm, *Pseudomonas aeruginosa Lifestyle: A Paradigm for Adaptation, Survival, and Persistence.* Front Cell Infect Microbiol, 2017. **7**: p. 39.
2. Gellatly, S.L. and R.E. Hancock, *Pseudomonas aeruginosa: new insights into pathogenesis and host defenses.* Pathog Dis, 2013. **67**(3): p. 159-73.
3. Khan, T.Z., et al., *Early pulmonary inflammation in infants with cystic fibrosis.* Am J Respir Crit Care Med, 1995. **151**(4): p. 1075-82.
4. CDC, *Antibiotic Resistance threats in the United States*, D.o.H.a.H. Serices, Editor. 2013 p. 114.
5. Gulick, A.M., *Nonribosomal peptide synthetase biosynthetic clusters of ESKAPE pathogens.* Natural Product Reports, 2017. **34**(8): p. 981-1009.
6. Sadikot, R.T., et al., *Pathogen-host interactions in Pseudomonas aeruginosa pneumonia.* Am J Respir Crit Care Med, 2005. **171**(11): p. 1209-23.
7. Meyer, J.M., *Pyoverdines: pigments, siderophores and potential taxonomic markers of fluorescent Pseudomonas species.* Arch Microbiol, 2000. **174**(3): p. 135-42.
8. Visca, P., F. Imperi, and I.L. Lamont, *Pyoverdine siderophores: from biogenesis to biosignificance.* Trends Microbiol, 2007. **15**(1): p. 22-30.
9. Antunes, L.C., et al., *Quorum sensing in bacterial virulence.* Microbiology, 2010. **156**(Pt 8): p. 2271-82.
10. Ben Haj Khalifa, A., et al., *[Virulence factors in Pseudomonas aeruginosa: mechanisms and modes of regulation].* Ann Biol Clin (Paris), 2011. **69**(4): p. 393-403.
11. Neilands, J.B., *Siderophores: structure and function of microbial iron transport compounds.* J Biol Chem, 1995. **270**(45): p. 26723-6.
12. Vasil, M.L. and U.A. Ochsner, *The response of Pseudomonas aeruginosa to iron: genetics, biochemistry and virulence.* Mol Microbiol, 1999. **34**(3): p. 399-413.
13. Miethke, M. and M.A. Marahiel, *Siderophore-based iron acquisition and pathogen control.* Microbiol Mol Biol Rev, 2007. **71**(3): p. 413-51.
14. Ellison, R.T., 3rd, *The effects of lactoferrin on gram-negative bacteria.* Adv Exp Med Biol, 1994. **357**: p. 71-90.
15. Cornelissen, C.N., *Transferrin-iron uptake by Gram-negative bacteria.* Front Biosci, 2003. **8**: p. d836-47.
16. Braun, V., *Iron uptake mechanisms and their regulation in pathogenic bacteria.* Int J Med Microbiol, 2001. **291**(2): p. 67-79.
17. Meneely, K.M. and A.L. Lamb, *Biochemical characterization of a flavin adenine dinucleotide-dependent monooxygenase, ornithine hydroxylase from Pseudomonas aeruginosa, suggests a novel reaction mechanism.* Biochemistry, 2007. **46**(42): p. 11930-7.
18. McMorran, B.J., et al., *Involvement of a transformylase enzyme in siderophore synthesis in Pseudomonas aeruginosa.* Microbiology, 2001. **147**(Pt 6): p. 1517-24.
19. Bosello, M., et al., *An enzymatic pathway for the biosynthesis of the formylhydroxyornithine required for rhodochelin iron coordination.* Biochemistry, 2012. **51**(14): p. 3059-66.
20. Pohlmann, V. and M.A. Marahiel, *Delta-amino group hydroxylation of L-ornithine during coelichelin biosynthesis.* Org Biomol Chem, 2008. **6**(10): p. 1843-8.

21. Seyedsayamdost, M.R., et al., *Structure and biosynthesis of amychelin, an unusual mixed-ligand siderophore from Amycolatopsis sp. AA4.* J Am Chem Soc, 2011. **133**(30): p. 11434-7.
22. Almassy, R.J., et al., *Structures of apo and complexed Escherichia coli glycinamide ribonucleotide transformylase.* Proc Natl Acad Sci U S A, 1992. **89**(13): p. 6114-8.
23. Deis, S.M., et al., *Structural and Enzymatic Analysis of Tumor-Targeted Antifolates That Inhibit Glycinamide Ribonucleotide Formyltransferase.* Biochemistry, 2016. **55**(32): p. 4574-82.
24. Klein, C., et al., *Towards structure-based drug design: crystal structure of a multisubstrate adduct complex of glycinamide ribonucleotide transformylase at 1.96 A resolution.* J Mol Biol, 1995. **249**(1): p. 153-75.
25. Zhang, Y., et al., *Crystal structures of human GAR Tfase at low and high pH and with substrate beta-GAR.* Biochemistry, 2002. **41**(48): p. 14206-15.
26. Zhang, Z., et al., *Structures of glycinamide ribonucleotide transformylase (PurN) from Mycobacterium tuberculosis reveal a novel dimer with relevance to drug discovery.* J Mol Biol, 2009. **389**(4): p. 722-33.
27. Schmitt, E., S. Blanquet, and Y. Mechulam, *Structure of crystalline Escherichia coli methionyl-tRNA(f)Met formyltransferase: comparison with glycinamide ribonucleotide formyltransferase.* EMBO J, 1996. **15**(17): p. 4749-58.
28. Schmitt, E., et al., *Crystal structure of methionyl-tRNAfMet transformylase complexed with the initiator formyl-methionyl-tRNAfMet.* EMBO J, 1998. **17**(23): p. 6819-26.
29. Gatzeva-Topalova, P.Z., A.P. May, and M.C. Sousa, *Crystal structure and mechanism of the Escherichia coli ArnA (PmrI) transformylase domain. An enzyme for lipid A modification with 4-amino-4-deoxy-L-arabinose and polymyxin resistance.* Biochemistry, 2005. **44**(14): p. 5328-38.
30. Genthe, N.A., et al., *Molecular structure of an N-formyltransferase from Providencia alcalifaciens O30.* Protein Sci, 2015. **24**(6): p. 976-86.
31. Thoden, J.B., et al., *Structure of a sugar N-formyltransferase from Campylobacter jejuni.* Biochemistry, 2013. **52**(35): p. 6114-26.
32. Woodford, C.R., J.B. Thoden, and H.M. Holden, *New role for the ankyrin repeat revealed by a study of the N-formyltransferase from Providencia alcalifaciens.* Biochemistry, 2015. **54**(3): p. 631-8.
33. Zimmer, A.L., J.B. Thoden, and H.M. Holden, *Three-dimensional structure of a sugar N-formyltransferase from Francisella tularensis.* Protein Sci, 2014. **23**(3): p. 273-83.
34. Van Duyne, G.D., et al., *Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin.* J Mol Biol, 1993. **229**(1): p. 105-24.
35. Olucha, J., et al., *Two structures of an N-hydroxylating flavoprotein monooxygenase: ornithine hydroxylase from Pseudomonas aeruginosa.* J Biol Chem, 2011. **286**(36): p. 31789-98.
36. Mayfield, J.A., et al., *Comprehensive spectroscopic, steady state, and transient kinetic studies of a representative siderophore-associated flavin monooxygenase.* J Biol Chem, 2010. **285**(40): p. 30375-88.
37. Caperelli, C.A. and J. Conigliaro, *Synthesis of 10-Acetyl-5,8-Dideazafolic Acid - a Potent Inhibitor of Glycinamide Ribonucleotide Transformylase.* Journal of Medicinal Chemistry, 1986. **29**(10): p. 2117-2119.
38. Smith, G.K., et al., *On the cofactor specificity of glycinamide ribonucleotide and 5-aminoimidazole-4-carboxamide ribonucleotide transformylase from chicken liver.* Biochemistry, 1981. **20**(5): p. 1241-5.
39. McPhillips, T.M., et al., *Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines.* Journal of Synchrotron Radiation, 2002. **9**: p. 401-406.

40. Kabsch, W., *Integration, scaling, space-group assignment and post-refinement.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 133-44.
41. Pannu, N.S., et al., *Recent advances in the CRANK software suite for experimental phasing.* Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 331-7.
42. Winn, M.D., et al., *Overview of the CCP4 suite and current developments.* Acta Crystallographica Section D-Biological Crystallography, 2011. **67**: p. 235-242.
43. 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.
44. Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.
45. Afonine, P.V., et al., *Towards automated crystallographic structure refinement with phenix.refine.* Acta Crystallogr D Biol Crystallogr, 2012. **68**(Pt 4): p. 352-67.
46. Terwilliger, T.C., et al., *Ligand identification using electron-density map correlations.* Acta Crystallogr D Biol Crystallogr, 2007. **63**(Pt 1): p. 101-7.
47. Terwilliger, T.C., et al., *Automated ligand fitting by core-fragment fitting and extension into density.* Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 8): p. 915-22.
48. Moriarty, N.W., R.W. Grosse-Kunstleve, and P.D. Adams, *electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation.* Acta Crystallogr D Biol Crystallogr, 2009. **65**(Pt 10): p. 1074-80.
49. Moriarty, N.W., E.J. Draizen, and P.D. Adams, *An editor for the generation and customization of geometry restraints.* Acta Crystallogr D Struct Biol, 2017. **73**(Pt 2): p. 123-130.
50. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 1): p. 12-21.
51. Krissinel, E. and K. Henrick, *Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2256-68.
52. Krissinel, E. and K. Henrick, *Inference of macromolecular assemblies from crystalline state.* Journal of Molecular Biology, 2007. **372**(3): p. 774-797.
53. *The PyMOL Molecular Graphics Systems, Version 2.0 Schrödinger, LLC.*
54. Reimer, J.M., et al., *Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase.* Nature, 2016. **529**(7585): p. 239-42.
55. Chen, P., et al., *Crystal structure of glycinamide ribonucleotide transformylase from Escherichia coli at 3.0 A resolution. A target enzyme for chemotherapy.* J Mol Biol, 1992. **227**(1): p. 283-92.
56. Su, Y., et al., *A pH-dependent stabilization of an active site loop observed from low and high pH crystal structures of mutant monomeric glycinamide ribonucleotide transformylase at 1.8 to 1.9 A.* J Mol Biol, 1998. **281**(3): p. 485-99.
57. Shim, J.H. and S.J. Benkovic, *Evaluation of the kinetic mechanism of Escherichia coli glycinamide ribonucleotide transformylase.* Biochemistry, 1998. **37**(24): p. 8776-82.
58. Franke, J., et al., *Nitro versus hydroxamate in siderophores of pathogenic bacteria: effect of missing hydroxylamine protection in malleobactin biosynthesis.* Angew Chem Int Ed Engl, 2013. **52**(32): p. 8271-5.
59. Neumann, C.S., et al., *Biosynthesis of Piperazic Acid via N5-Hydroxy-Ornithine in Kutzneria spp. 744.* Chembiochem, 2012. **13**(7): p. 972-976.
60. Tomlinson, G. and T. Viswanatha, *Synthesis and properties of -N-hydroxyornithine.* Can J Biochem, 1973. **51**(6): p. 754-63.
61. Olucha, J. and A.L. Lamb, *Mechanistic and structural studies of the N-hydroxylating flavoprotein monooxygenases.* Bioorg Chem, 2011. **39**(5-6): p. 171-7.

62. Ferdinand, W., *The interpretation of non-hyperbolic rate curves for two-substrate enzymes. A possible mechanism for phosphofructokinase.* Biochem J, 1966. **98**(1): p. 278-83.
63. H.Segel, I., *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady - State Enzyme Systems* 1975: John Wiley & Sons, Inc. 957.
64. Caperelli, C.A., *Mammalian glycinamide ribonucleotide transformylase. Kinetic mechanism and associated de novo purine biosynthetic activities.* J Biol Chem, 1989. **264**(9): p. 5053-7.

CHAPTER 3:

3,4-dihydroxy-2-butanone 4-phosphate synthase (RibB) of riboflavin biosynthesis has a mononuclear magnesium active site

3.1 INTRODUCTION

The RibB enzyme is functionally and structurally unique. Functionally, RibB is a magnesium-dependent enzyme that removes carbon C4 of Ru5P, a five-carbon sugar, to make the product DHBP, a four-carbon compound (**Figure 3.1**). Current evidence for this is mainly based on ^{13}C feeding studies as well as ^{13}C NMR endpoint experiments [1-4]. Due to the complexity of the reaction it is not surprising that the RibB reaction is considered to be one of the rate-limiting steps in riboflavin biosynthesis [5]. RibB enzymes have turnover numbers reported to be within a couple of turnovers per minute [6]. Structural examination of RibB testifies to its unique nature as well. At first glance, RibB enzymes have $\alpha+\beta$ fold (**Figure 3.2**). However, comparison of RibB to proteins deposited in the Protein Data Bank (PDB) results only in other RibB proteins. In an analysis using PDBeFold [7], the results yield 47 RibB proteins with best match corresponding to RibB from *Salmonella enterica* (RMSD of 0.54 Å over 203 Cα of 203 residues) and lowest matched RibB protein from *Methanocaldococcus jannaschii* (RMSD of 1.59 Å over 191 Cα residues of 219. While the RibB fold has an $\alpha+\beta$ structure arranged in a common global arrangement (sheets

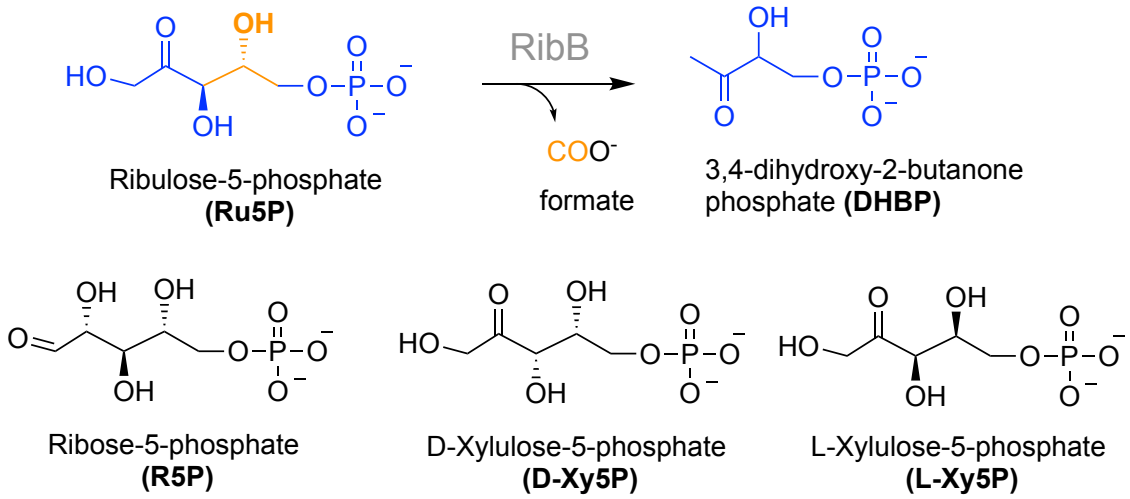


Figure 3.1. Reaction catalyzed by RibB from riboflavin biosynthesis. (A) RibB removes carbon C4 of ribulose 5-phosphate (Ru5P) as formate (orange), generating one molecule of 3,4-dihydroxy-2-butanone phosphate (DHBP) (blue). (B) Isomers used in this study. Ribose 5-phosphate (R5P) is an aldose sugar equivalent to Ru5P. Xylulose 5-phosphate is an epimer at carbon C3.

surrounded by helices) the connectivity of the secondary structural elements and protein topology are unique. In Gram-negative bacteria RibB enzymes have been reported to be dimeric in solution and in crystal structures. In Gram-positive bacteria, however, the RibB enzyme is found in complex with RibA. Isolation of only the RibB domain leads to formation of a dimer as in the Gram-negative homologues [8, 9]. The active site of the enzyme is surface exposed and in proximity to the dimer interface (**Figure 3.2**). This site is also surrounded with two mobile loops. The shorter loop, loop 1, is composed of acidic residues and is important for binding of the substrate and metal. The longer loop, loop 2, has not been functionally characterized. However, major conformational changes are observed in this loop upon substrate and metal binding [10-12].

A multitude of RibB structures have been reported. The majority of these structures are apo structures, structures with Mn^{2+}, Zn^{2+} and Ca^{2+}, structures with sulfate or phosphate in the active site, and few are holo-structures with metal and substrate bound [9-17]. From this structural evidence it was concluded that the enzyme requires two metal ions for catalysis. Here we report intrinsic tyrosine fluorescence and activity assay profiles that show RibB binds and requires a single metal ion for full activity. We also report high-resolution structures of RibB with an inactive substrate analogue ribose 5-phosphate (R5P), stereoisomers D and L-xylulose 5-phoshate (DXy5P and LXy5P) and 1.08 Å apo-RibB structure with loop 1 in an open conformation. The structures reveal the location of a single metal binding site that is highly coordinated by both substrate and amino acid residues in the active site.

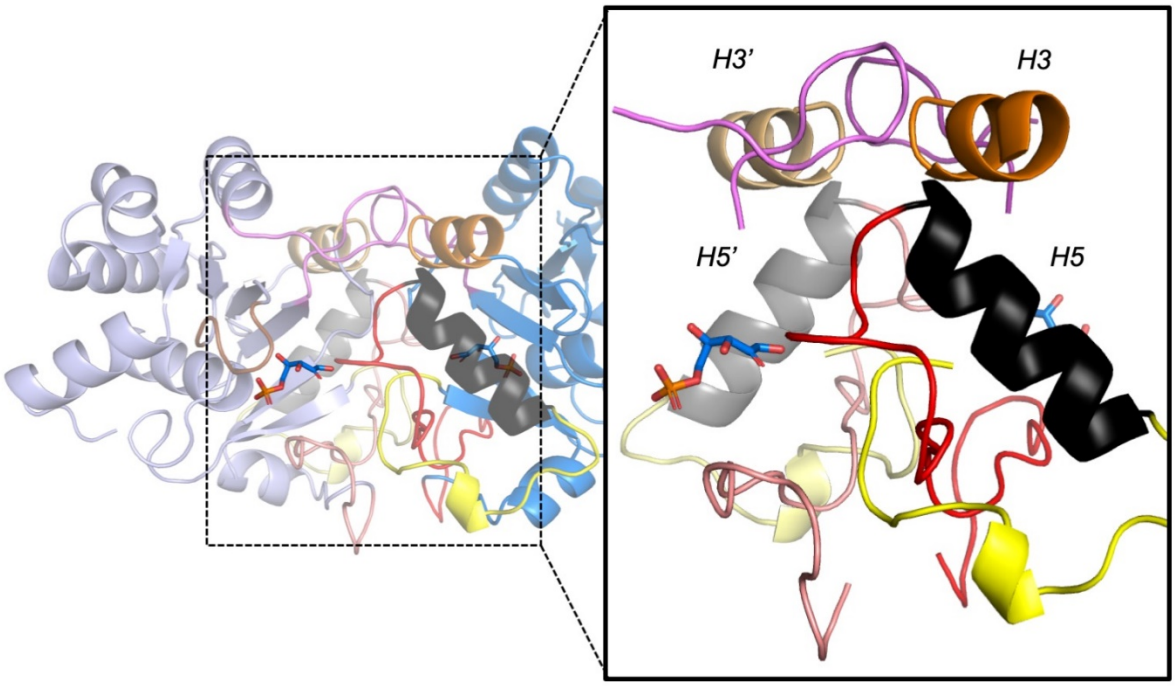


Figure 3.2. RibB dimer and dimer interface. RibB is a homodimer composed of two identical monomers. Each monomer is represented with light and darker shade of same color (light blue and dark blue). Interface is composed of 10 hydrogen bonding interaction, 150 non-bonding interaction and one salt-bridge over 48 amino acid residues. Hydrogen bonds come predominantly from helix 5 (black,gray) and helix 3 (orange, light orange) interaction with residues 175 -185 (purple,violet). Nonbonding interactions come from polypeptide chain originating on either end on helix 5. Chain 1 (red, light red) start on the N-terminal end of helix 5 and passes the interface to surround the enzyme active site of the same monomer. Location of the active site is represented by molecule of R5P (blue sticks). End of the chain that covers the active site (residues 87-96) has been represented as catalytically important loop 2. Chain 2 (yellow, light yellow) starts on the C-terminal end of helix 5 in opposing monomer. Besides non-bonding interaction with chain 1, it is partially buried in active site of opposing monomer where His137 is making a salt bridge with Glu175 and hydrogen bonding distance away from terminal hydroxyls of the substrate. Active site is surface exposed and covered with loop 2 on one side opposing loop 1 (34-41).

3.2 MATERIALS AND METHODS

3.2.1 VcRibB cloning.

An overexpression construct for the *Vibrio cholerae* RibB gene was prepared by Genscript plasmid preparation and subcloning services. The *ribB* gene from *Vibrio cholerae* (sequence ID: AE003853.1, strain: N16961, taxid: 243277) was initially synthesized in the pUC15 vector. The final construct was obtained by sub cloning this gene into pET28a+ (restriction sites: *NdeI*/*XhoI*), to yield the protein with an N-terminal 6 His tag.

3.2.2 VcRibB purification.

Purification of 3,4-dihydroxy-2-butanone 4-phosphate synthase, RibB from *Vibrio cholerae* was done as described previously [12] with a few changes. The RibB construct was transformed into BL21(DE3) *E.Coli* (New England Biolabs) and grown overnight at 37 ^{o}C in 50 ml of LB Miller media (150 ml conical flask) in a shaker incubator (250 rpm) (containing 50 µl of 50 mg/ml of kanamycin). 1L of LB Miller broth in a baffled flask containing 1 ml of 50 mg/ml kanamycin was inoculated with 10 ml of the overnight culture and grown at 37 ^{o}C in a shaker incubator (250 rpm). Protein expression was induced when the OD_{600nm} reached 0.8 with a final concentration of 0.5 mM IPTG (isopropyl β-D-1-thiogalactopyranoside) and further incubated at 37 ^{o}C for 4 hours. Cells were harvested by centrifugation (6000 x g,10min, 4 ^{o}C). The cell pellet was resuspended in 10 ml of 50 mM Tris-HCl pH 8.0, 50 mM imidazole and 500 mM NaCl per liter of culture broth. Resuspended cells were lysed by passing them through French

press (35000 psi) three times, on ice. The cell lysate was centrifuged at 12000 x g for 40 min at 4 ^{o}C. The supernatant was injected onto a 25 ml Nickel Superdex fast flow (GE Healthcare) column pre-equilibrated with 50 mM Tris:HCl, 50 mM imidazole and 500 mM NaCl pH 8.0. The protein was eluted with a 10 column volume linear gradient of 50 mM Tris-HCl, 500 mM imidazole, 500 mM NaCl pH 8.0. RibB eluted at 200 mM imidazole. The protein (100 ml) was concentrated to 30 ml using an Amicon® nitrogen concentrator with a 10kDa cutoff filter and injected onto a Superdex 200 gel-filtration column (GE Healthcare), pre-equilibrated with 25 mM Tris-HCl pH 8.0. RibB eluted as a dimer and was isolated and concentrated using an Amicon® Ultracell® 30K centrifugal concentrators to 40 mg/ml by Bradford and stored at -80 ^{o}C for later use. The purification yield was 40 mg of protein per liter of culture.

3.3.3 VcRibB purification in presence of EDTA.

In order to remove as much of residual divalent ions, prior to size exclusion step, the protein has been incubated with final concentration of 2 mM EDTA and left on ice for 10 min. The buffer for the size exclusion column contained 100 µM EDTA. Before start of the experiment, the protein was exchanged again into freshly made 25 mM Tris-HCl pH 8.0, 100 µM EDTA prior to use in metal binding and activity studies (section 3.3.7 and 3.3.8).

3.3.4 Steady state kinetics of RibB with varied substrate, ribulose 5-phosphate (Ru5P).

Ru5P (Sigma Aldrich) was dissolved in 50 mM Tris-HCl pH 7.5 to a stock concentration of 90 mM. Serial dilutions of substrate, by factor of a half, were prepared with a final

concentration range from 4 mM to 0.5 µM. In a reaction vial of 200 µl, diluted Ru5P, 10 µM enzyme, and 10 mM $MgCl_2$ were mixed and incubated for 30 min at room temperature. After 30 min, the reaction was quenched and the color developed by addition of freshly made 175 µL of saturated naphthol (dissolved in 1N NaOH to 30 mg/ml) and 250 µL of saturated creatine (dissolved in water, 35mg/ml) [6]. The color was left to develop for 30 min after which it was scanned between 450 nm and 650 nm using Cary 50 Bio UV-visible spectrophotometer. The absorbance at λ_{max} 525 nm was corrected by subtracting with the absorbance at 650nm and then converted to the amount of 3,4-dihydroxy-2-butanone phosphate (DHBP) (in nmol) using a 3,4-butadione standard curve. As the concentration of the substrate was not saturating in this assay, steady state parameters were obtained by using the equation:

$$v = V_{max}\frac{([E_T] + [S_T] + K_M) - \sqrt{([E_T] + [S_T] + K_M)^2 - 4[E_T][S_T]}}{2[E_T]}$$

Values for V_{max} and K_M are averages of three trials and errors are reported as standard deviation between values. Error propagation was used in the determination of error values for k_{cat} and k_{cat}/K_M.

3.3.5 Steady state kinetics with varied Mg^{2+}.

The steady state kinetics of RibB with varied Mg^{2+} were determined as with Ru5P with one change. The standard reaction mixture (200 µL) contained diluted $MgCl_2$ (99.9% pure, Fisher, 1.6 mM to 12.2 µM), 1.8 mM Ru5P and 10 µM enzyme.

3.3.6 Steady state kinetics with varied Mn^{2+}.

Steady state parameters were obtained in the same manner as with Mg^{2+} except 1 μM enzyme and pure $MnCl_2$ (250 μM to 1.9 μM, 99% pure, Fisher) was used.

3.3.7 Determination of number of metals required for activity.

RibB protein purified in the presence of EDTA was exchanged into 25 mM Tris-HCl, 100 μM EDTA pH 8.0 and concentrated to 37.7 mg/ml using an Amicon® Ultracell® 30K centrifugal concentrator. In each reaction of 200 μL, the final concentration of 50 mM Tris-HCl pH 7.5, 100 μM EDTA, 1.8 mM Ru5P and 0 to 220 μM of $MgCl_2$ in increments of 20 μM were combined. The reaction was initiated by addition of 60 μM enzyme and incubated for 1 hour at room temperature. Steady state data were calculated using the above procedure. The experiment was repeated three times, and values reported are the averages of three trials with errors reported as standard deviations.

3.3.8 Determination of number of metals required for binding.

Intrinsic tyrosine fluorescence was measured using a Cary 50 Eclipse fluorometer with an excitation wavelength of 280 nm and emission recorded from 290 nm to 400 nm (excitation slit set to 10 nm and emission slit set to 5 nm). The reaction mix, 200 μL, contained 50 mM Tris-HCl pH 7.5, 100 μM EDTA, 1.8mM Ru5P and 60 μM RibB (50mM Tris:HCl, 100 μM EDTA pH 8), was titrated in 1 μL increments of 2 mM $MgCl_2$, 100 μM EDTA solution. The reported Mg^{2+} concentration account for the dilution. Tyrosine fluorescence at 302 nm was corrected with the florescence at 302 nm for the control sample (titration with 100 μM EDTA). Experiments were repeated three times and data

points reported are averages of three trials with error reported as the standard deviation of the respective trials.

3.3.9 EPR studies of VcRibB metal binding

Depending on the sample, EPR samples were prepared by mixing millimolar ratio of VcRibB, $MnCl_2$ and Ru5P (LXy5P) in 50 mM Tris-HCl pH 7.5 and 10% glycerol. In 300 µL final volume the reaction was performed at 4 ^{o}C and initiated by direct addition of enzyme into the reaction mixture placed in a 4 mm quartz EPR tube. After an allotted time period the reaction was quenched by submerging the EPR tube in liquid nitrogen. X-band EPR data were collected on a 9 GHz Bruker EMXPlus spectrometer. Experiments were run at 10 K with the use of an Oxford ESR900 continuous-flow liquid helium cryostat equipped with an Oxford ITC503 temperature system. Perpendicular-mode data were collected in a dual-mode Bruker ER4116DM cavity. Spectra were recorded using the following non-saturating conditions: 9.64 GHz microwave frequency, 2.0 mW microwave power, 4 G modulation amplitude, 100 kHz modulation frequency, and 40.96 ms time constant.

3.3.10 VcRibB crystallization

All crystals were grown using the hanging-drop vapor diffusion method. Each drop (3 µl) was prepared by mixing protein and the precipitant solution in equal amounts.

3.3.11 VcRibB apo crystallization

Protein was crystallized at 40 mg/ml. The precipitant solution contained final concentration of 0.1 M Na_2HPO_4/NaH_2PO_4 pH9.3, 16% PEG 3350, 0.3M glycine. Cube shaped crystal grew to their full size in three weeks. For data collection crystal was soaked in precipitant solution supplemented with 20% PEG3350, 5.2mM Ru5P and 0.2mM $MgCl_2$.

3.3.12 VcRibB crystallization with D-ribulose 5-phosphate (D-Ru5P).

Purified protein at 32.9 mg/ml has been mixed with 15 times molar excess of D-ribulose 5-phosphate in 50 mM Tris-HCl pH 7.5 (Sigma Aldrich) and incubated on ice for 15 min. The precipitant protein solution was 0.1 M Lithium Acetate and 18% PEG 3350. In preparation for data collection rod shaped crystals were first soaked in precipitant solution mixed with a final concentration of 40 mM D-ribulose 5-phosphate overnight, then soaked in cryoprotect containing precipitant supplemented with 30% ethylene glycol.

3.3.13 VcRibB crystallization with D-ribose 5-phosphate (D-R5P) and Mn^{2+}.

Purified protein at 32.9 mg/ml was mixed with a few flakes of D-ribose 5-phosphate (Sigma Aldrich) and incubated on ice for 15 min. The precipitant solution contained a final concentration of 0.1 M Lithium acetate and 16% PEG 3350. Rod shaped crystals grew to their full size in 48 hours. For data collection, crystals were soaked in precipitant solution supplemented with 30% ethylene glycol and 4 mM $MnCl_2$ (99% pure, Fisher).

3.3.14 VcRibB crystallization with D-xylulose 5-phosphate (D-Xy5P) with Mn^{2+}.

Purified protein at 32.9 mg/ml has been mixed with 15 times molar excess of D-xylulose 5-phosphate in 50 mM Tris-HCl pH 7.5 (Sigma Aldrich) and incubated on ice for 15 min. The precipitant protein solution was 0.1 M Lithium Acetate and 12% PEG 3350. In preparation for data collection rod shaped crystals were first soaked in precipitant solution mixed with a final concentration of 40 mM D-xylulose 5-phosphate overnight, then soaked in cryoprotect containing precipitant supplemented with 4mM $MnCl_2$ and 30% ethylene glycol.

3.3.15 VcRibB crystallization, L-xylulose 5-phosphate (L-Xy5P) with Mn^{2+}.

Purified protein at 32.9 mg/ml has been mixed with 15 times molar excess of L-xylulose 5-phosphate in 50 mM Tris-HCl pH 7.5 (Sigma Aldrich) and incubated on ice for 15 min. Protein was crystallized using the hanging-drop vapor diffusion method. Precipitant protein solution was 0.1 M Lithium Acetate and 14% PEG 3350. In preparation for data collection rod shaped crystals were first soaked in precipitant solution augmented with a final concentration of 40 mM L-xylulose 5-phosphate then soaked in cryoprotect containing precipitant supplemented with 4mM $MnCl_2$ and 30% ethylene glycol.

3.3.16 X-ray diffraction data collection.

Data for all RibB crystal variants were collected at the Stanford Synchrotron Radiation Laboratory (SSRL, Stanford, CA) beamline 12-2 using a wavelength of 0.9795 Å at 100 K (except for apo-RibB). The software package Blu-Ice [18, 19] was used to collect 1200 oscillation images (0.15 ° per image) with an exposure time of 0.2 sec.

3.3.17 X-ray diffraction data collection and processing of apo-RibB

Data were collected at the wavelength of 0.8526 Å at 100K. Diffraction data were collected using a beam size of 50 µm x15 µm with transmission set to 1.5 % and a detector distance of 188.1 mm. Diffraction data were processed using XDS [20] to 1.08 Å in the space group P $2_12_12_1$ with cell dimensions of a=56.84 Å, b=73.90 Å, c=95.82 Å, $\alpha = \beta = \gamma$ =90.0 °. A phasing solution was obtained by molecular replacement using PHENIX, Phaser-MR [21] with the 4P8J as a model [12]. One molecular replacement solution was identified with an LLG of 28265 and a TFZ of 156. The solution was subjected to alternating cycles of model building and refinement using Coot [22] and Phenix.Refine [23, 24]. Water molecules were added automatically and inspected manually using Coot. Due to the high resolution, all atoms were refined with anisotropic factors. Statistics for data refinement and analysis can be found in **Table 3.1**.

The final model has two monomers in the asymmetric unit with residues 7 to 83 and 97 to 216 fully resolved in one monomer and residues 5 to 83 and 97 to 216 fully resolved in the other. The final model contained 490 water molecules. Ramachandran analysis was calculated with MolProbity [25] showing 97.94% in the favored regions and no outliers.

3.3.18 X-ray diffraction data collection and processing of RibB with D-ribulose 5-phosphate.

Diffraction data were collected using a beam size of 100 µm x 60 µm with transmission set to 5.0 % and a detector distance of 300.4 mm. Diffraction data were processed

using XDS [20] to 1.80 Å in the space group C 2 2 2_1 with cell dimensions of a=60.67 Å, b=78.63 Å, c=90.32 Å, α = β = γ =90.0 °. A phasing solution was obtained via molecular replacement using PHENIX, Phaser-MR [21] with the 4P8E as a model [12]. One molecular replacement solution was identified with an LLG of 5323 and a TFZ of 51. The solution was subjected to alternating cycles of model building and refinement using Coot [22] and Phenix.Refine [23, 24]. Water molecules were added automatically and inspected manually using Coot. The ribulose 5-phosphate molecule was added manually. Restraints for ribulose 5-phosphate was generated using eLBOW [26] and REEL [27]. Statistics for data refinement and analysis can be found in **Table 3.1**.
The final model has one monomer with residues 2 to 217 fully resolved. The final model contained 148 water molecules and molecule of ribulose 5-phosphate in the active site. Ramachandran analysis has been calculated with MolProbity [25] with 97.2% in the favored regions and no outliers. Structures figures were prepared using Pymol [28].

3.3.19 X-ray diffraction data collection and processing of RibB with D-ribose 5-phosphate and Mn^{2+}.

Diffraction data were collected using a beam size of 100 µm x 60 µm with transmission set to 17.5 % and a detector distance of 351.6 mm. Diffraction data were processed using XDS [20] to 2.0 Å in the space group C 2 2 2_1 with cell dimensions of a=60.25 Å, b=78.58 Å, c=89.93 Å, α = β = γ =90.0 °. A phasing solution was obtained by molecular replacement using PHENIX, Phaser-MR [21] with 4P8E as a model [12]. One molecular replacement solution was identified with an LLG of 3913 and a TFZ of 43. The solution was subjected to alternating cycles of model building and refinement using Coot [22]

and Phenix.Refine [23, 24]. Water molecules were added automatically and inspected manually using Coot. The ribose 5-phosphate molecule and the manganese were added manually. Restraints for ribose 5-phosphate were generated using eLBOW [26] and REEL [27]. Statistics for data refinement and analysis can be found in **Table 3.1**. The final model has one monomer in the asymmetric unit with residues 3 to 217 fully resolved. The final model contained 68 water molecules, one molecule of ribose 5-phosphate and one Mn^{2+} ion in the active site. Ramachandran analysis was calculated with MolProbity [25] showing 95.8% in the favored regions and 0.94% of outliers. Outliers belong to residues Ser 5 and Asp 214 that are in areas of poor electron density.

3.3.20 X-ray diffraction data collection and processing of RibB with D-xylulose 5-phosphate and Mn^{2+}.

Diffraction data were collected using a beam size of 150 µm x 80 µm with transmission set to 35.0 % and a detector distance of 390.0 mm. Diffraction data were processed using XDS [20] to 2.0 Å in the space group C 2 2 2_1 with cell dimensions of a=61.03 Å, b=79.11 Å, c=91.34 Å, $\alpha = \beta = \gamma = 90.0$ °. A phasing solution was obtained by molecular replacement using PHENIX, Phaser-MR [21] with 4P8E as a model [12]. One molecular replacement solution was identified with an LLG of 4225 and a TFZ of 48.3. The solution was subjected to alternating cycles of model building and refinement using Coot [22] and Phenix.Refine [23, 24]. Water molecules were added automatically and inspected manually using Coot. D-xylulose 5-phosphate molecule and the manganese were added manually. Restraints for D-xylulose 5-phosphate were generated using

eLBOW [26] and REEL [27]. Statistics for data refinement and analysis can be found in **Table 3.1**.

The final model has one monomer in the asymmetric unit with residues 3 to 86 and 94 to 217 fully resolved. The final model contained 106 water molecules one molecule of D-xylulose 5-phospahte one Mn^{2+} ion in the active site and two ethylene glycol molecules. Ramachandran analysis was calculated with MolProbity [25] showing 98.54% in the favored regions and no outliers.

3.3.21 X-ray diffraction data collection and processing of RibB with L-xylulose 5-phosphate and two Mn^{2+}metal ions.

Diffraction data were collected using a beam size of 100 µm x 60 µm with transmission set to 59.0 % and a detector distance of 347.1 mm. Diffraction data were processed using XDS [20] to 2.0 Å in the space group C 2 2 2_1 with cell dimensions of a=60.37 Å, b=79.02 Å, c=91.35 Å, α = β = γ =90.0 °. A phasing solution was obtained via molecular replacement using PHENIX, Phaser-MR [21] with 4P8E as a model [12]. One molecular replacement solution was identified with an LLG of 3586 and a TFZ of 45. The solution was subjected to alternating cycles of model building and refinement using Coot [22] and Phenix.Refine [23, 24]. Water molecules were added automatically and inspected manually using Coot. L-xylulose 5-phosphate molecule and two manganese ions were added manually. Restraints for L-xylulose 5-phosphate were generated using eLBOW [26] and REEL [27]. Statistics for data refinement and analysis can be found in **Table 3.1**.

The final model has one monomer with residues 2 to 217 fully resolved. The final model contained 120 water molecules, a molecule of L-xylulose 5-phospahte and two Mn^{2+} ions in the active site. Ramachandran analysis has been calculated with MolProbity [25] with 97.66% in the favored regions and no outliers. Structures figures were prepared using Pymol [28].

Table 3.1: RibB variant data collection and refinement data statistics[3]

	RibB-apo	RibB-Ru5P	RibB-R5P	RibB-DXy5P	RibB-LXy5P
Data collection					
Beamline	SSRL 12-2	SSRL 12-2	SSRL 12-2	SSRL 12-2	SSRL 12-2
Wavelength (Å)	0.9795	0.9795	0.9795	0.9795	0.9795
Space group	$P2_12_12_1$	$C222_1$	$C222_1$	$C222_1$	$C222_1$
Cell *a*, *b*, *c* (Å)	56.84 73.90 95.82	60.67 78.63 90.32	60.25 78.58 89.93	61.03 79.11 91.34	60.37 79.02 91.35
Resolution (Å)	36.95 - 1.08 (1.10 - 1.08)	39.31 - 1.80 (1.84 -1.80)	39.29 - 1.95 (2.0 - 1.95)	39.55 – 2.0 (2.05 – 2.0)	39.51 - 2.0 (2.05 - 2.0)
R_{merge}[a]	0.066 (0.865)	0.084 (0.600)	0.085 (0.868)	0.069 (0.418)	0.075 (0.418)
Total observations	1128679 (48565)	135338 (6988)	104193 (6753)	99522 (7084)	99288 (6060)
Total unique observations	172713 (8395)	20407 (1209)	15760 (1092)	15176 (1092)	15092 (1060)
Mean ((*I*) / sd(*I*))	13.0 (2.3)	13.0 (2.3)	12.4 (2.0)	16.7 (4.2)	16.9 (3.8)
Completeness (%)	99.8 (98.7)	99.9 (99.6)	98.8 (97.0)	99.3 (99.3)	99.6 (97.0)
Redundancy	6.5 (5.8)	6.6 (5.8)	6.6 (6.2)	6.6 (6.5)	6.6 (5.7)
CC(1/2)	0.945 (0.028)	0.998 (0.869)	0.999 (0.778)	0.998 (0.904)	0.999 (0.903)
Refinement					
Resolution (Å)	36.95 - 1.08 (1.12 - 1.08)	39.31 -1.8 (1.9 – 1.8)	36.01 – 1.95 (2.0 – 1.95)	36.3 -2.0 (2.1 - 2.0)	36.26 - 2.0 (2.1 - 2.0)
R_{cryst}[b]	0.1306 (0.1875)	0.1633 (0.2081)	0.1739 (0.2381)	0.1603 (0.1792)	0.1578 (0.1715)
R_{free}[c]	0.1479 (0.2148)	0.1985 (0.2628)	0.2333 (0.2949)	0.2171 (0.2429)	0.2116 (0.2393)
Total unique observations	172514 (17007)	20371 (2005)	15737 (1530)	15141 (1493)	15067 (1486)
No. of non-hydrogen atoms	3569	1811	1706	1741	1768
Protein	3079	1649	1623	1612	1631
Ligand	0	14	15	23	17
Water	490	148	68	106	120
rms deviation bonds (Å)	0.006	0.011	0.011	0.011	0.012
rms deviation angles (°)	0.97	1.20	1.16	1.18	1.05
Overall mean B-factor (Å^2)	13.74	21.80	31.05	28.74	26.68
Ramachandran plot analysis[c]					
Favored region	97.94	97.20	95.77	98.54	97.66
Allowed region	2.06	2.80	3.29	1.46	2.34
Outlier region	0.0	0.0	0.94	0.0	0.0

[3] $R_{merge} = \Sigma_h|I_h - \langle I \rangle|/\Sigma_h I_h$, where I_h is the intensity of reflection *h*, and $\langle I \rangle$ is the mean intensity of all symmetry-related reflections [b]$R_{cryst} = \Sigma||F_o| - |F_c||/\Sigma|F_o|$, F_o and F_c are observed and calculated structure factor amplitudes. [c]Five percent of the reflections were initially reserved to create an R_{free} test set used during each subsequent round of refinement.

3.3 RESULTS

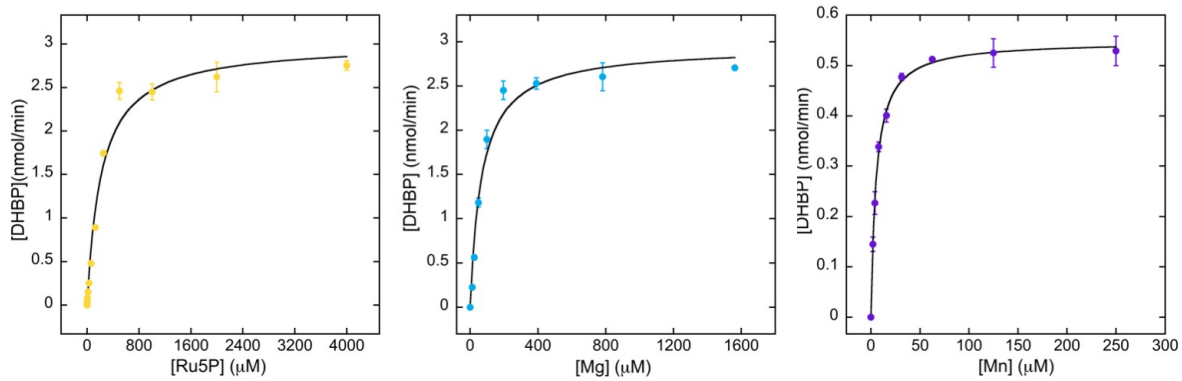
3.3.1 RibB protein purification and dimer architecture.

RibB was heterologously produced in *E.coli* using a purification protocol with two steps: nickel affinity and size exclusion. The protein eluted predominantly as a dimer. The apo-RibB is dimeric. In presence of substrate and metal, the protein was solved in a higher symmetry space group ($C222_1$) with one monomer in asymmetric unit. The protein is still dimeric, with a dimer interface in line with a crystallographic 2-fold axis. PISA analysis was performed, identifying the dimeric surface interface of 1400 $Å^2$ [29]. Indeed, this is indicative that one dimer is composed of two identical, stable monomers without any major conformational changes or disorder in protein structure. The RibB monomer has a $\alpha + \beta$ fold with no recognizable binding motifs. The center of the protein is composed of 6 β-strands that are surrounded by 9 α-helixes. Overall 48 residues are involved in formation of dimer interface accounting for the third of all amino acid residues in the protein. Out of 153 non-bonded contacts, 10 are hydrogen bonds and 2 contacts (His137 to Glu175) are salt bridges. The hydrogen bonds originate from an interaction between the N-terminal portion of helix 5 (residues 112-123, brown, **Figure 3.2**) and the C-terminal-end of helix 3 (residues 53-62, orange, **Figure 3.2**) with the loop 175-185 of the opposing monomer (green, **Figure 3.2**). Other hydrophobic interactions come from interactions of polypeptide chains originating on the ends of helix 5. N-terminal of helix 5 there is a span of 26 amino acids that goes through the dimer interface reaching the active site on the opposite

end of the monomer. A portion of this chain covers the active site and is referred to as loop 2 (residues 85-97). C-terminal to helix 5 of the opposing monomer (residues 124-138, yellow, **Figure 3.2**). Besides interaction with loop 2, one residue of this chain, His137, is buried in the active site of the opposing monomer making a salt bridge with Glu175. A nitrogen of the imidazole ring of this side chain is also within hydrogen bonding distance of hydroxyls of O1 and O2 of the substrate and has been hypothesized to be involved in initial catalytic steps of the mechanism [11].

3.3.2 Steady state kinetics.

A steady state kinetic assay was developed based on a previously reported assay that measures the formation of terminal ketones. A colorimetric change indicative of this moiety is developed in the presence of saturating amounts of creatine and naphthanol [30, 31]. This assay was used to determine kinetic parameters when Ru5P, Mg^{2+} or Mn^{2+} was the varied substrate (**Figure 3.3**). The turnover numbers for Ru5P and Mg^{2+} are similar with values of 1.50 ± 0.05 min^{-1} and 1.50 ± 0.03 min^{-1}, respectively. Reported, K_M values for Ru5P, Mg^{2+} and Mn^{2+} are 219 μM, 68 μM and 5.3 μM. The value for Ru5P is comparable to the value reported in the literature [12]. However, when Mn^{2+} was substituted for Mg^{2+}, the K_M was 13-fold lower (5.3 ± 0.5 μM) and the k_{cat} was 2-fold higher (2.73 ± 0.03 min^{-1}). This represent the 24-fold increase in the catalytic efficiency (360 ± 20 $M^{-1}sec^{-1}$ for Mg^{2+} and 8600 ± 800 $M^{-1}sec^{-1}$ for Mn^{2+}).



Substrate	K_M (μM)	k_{cat} (min^{-1})	k_{cat}/K_M ($M^{-1}sec^{-1}$)
Ru5P	219 ± 6	1.50 ± 0.05	115 ± 5
Mg^{2+}	68 ± 3	1.50 ± 0.03	360 ± 20
Mn^{2+}	5.3 ± 0.5	2.73 ± 0.03	8600 ± 800

Figure 3.3. Steady state kinetics. Steady state kinetic parameters with three varied substrates: Ru5P (yellow), $MgCl_2$ (cyan), $MnCl_2$ (purple).

3.3.3 RibB requires one metal in binding and activity.

Historically, it is known that RibB is a Mg-dependent enzyme, but the stoichiometry was not discussed. Crystal structures with Zn^{2+} show two metal ions in the active site (**Figure 3.6**). Metals were titrated against a specific concentration of enzyme (60 μM) and activity and metal binding were measured. To ensure no metal is present before the start of the titration, enzyme was purified in buffer containing 100 μM EDTA. Therefore, assuming EDTA has a stronger binding of metal than RibB, a change in activity is expected once the metal binding exceeds 100 μM.

The titration experiments show that one equivalent of metal binds (**Figure 3.4**) and is required for the activity (**Figure 3.5**). The activity assay confirms that RibB is inactive in the presence of Zn^{2+} but active in the presence of Mg^{2+} and Mn^{2+}. One would be tempted to propose that additional Mn^{2+} ions over 1:1 ratio were inhibitory. However, after the addition of one equivalent of Mn^{2+}, enzyme precipitation is observed, which is one explanation for the decrease in the activity over 160 μM Mn^{2+} (**Figure 3.5**). Mn^{2+} has also been previously reported to interfere with the color development in the assay [6].

Here I have reported five structures of the RibB enzyme: apo RibB, RibB with substrate bound and no metal, RibB in complex with ribose 5-phosphate and Mn^{2+} and RibB in complex with stereoisomers L-xylulose 5-phosphate and D-xylulose 5-phosphate. Orientation of structural elements is very close with RMSD of 0.2 Å (over 201 out of 217 Cα residues), and in comparison with other *Vibrio cholerae* RibB structures [12] with RMSD of 0.26 Å (over 200 out of 217 Cα residues) [7].

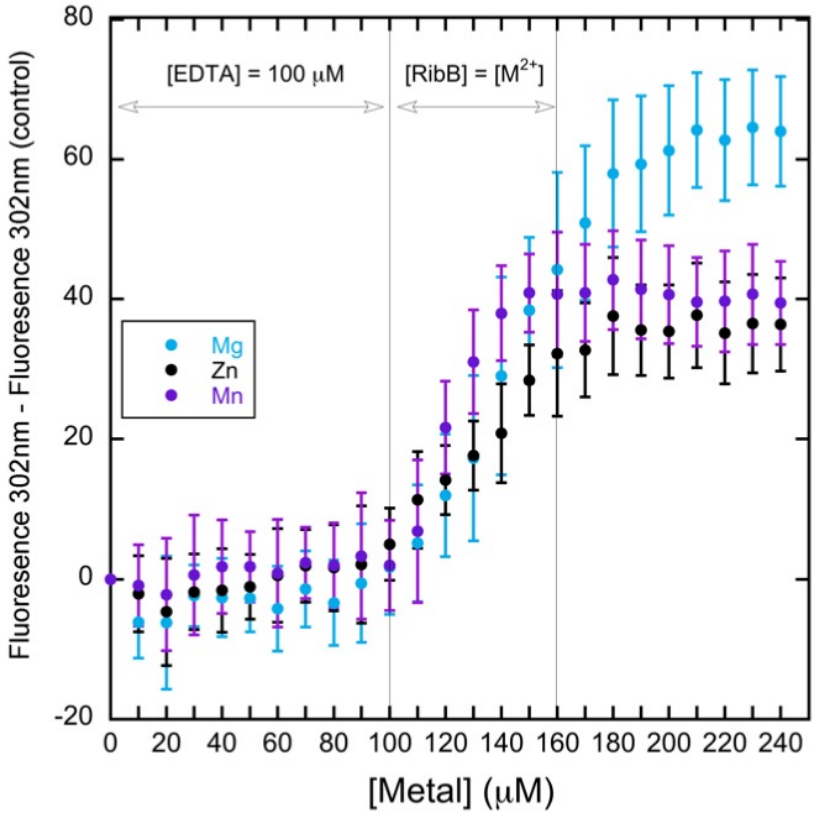


Figure 3.4. One metal equivalent is binds to RibB. Intrinsic tyrosine fluorescence indicates binding of metal to RibB in the ratio of 1:1. 60 μM RibB purified in the presence of 100 μM EDTA was titrated with Mg^{2+} (cyan), Zn^{2+} (black) and Mn^{2+} (purple). Fluorescence increases after 100 μM EDTA concentration has been overcome. There is no significant change in fluorescence after an additional 60 μM metal, is added. This concentration is equivalent to the concentration of enzyme used in the experiment.

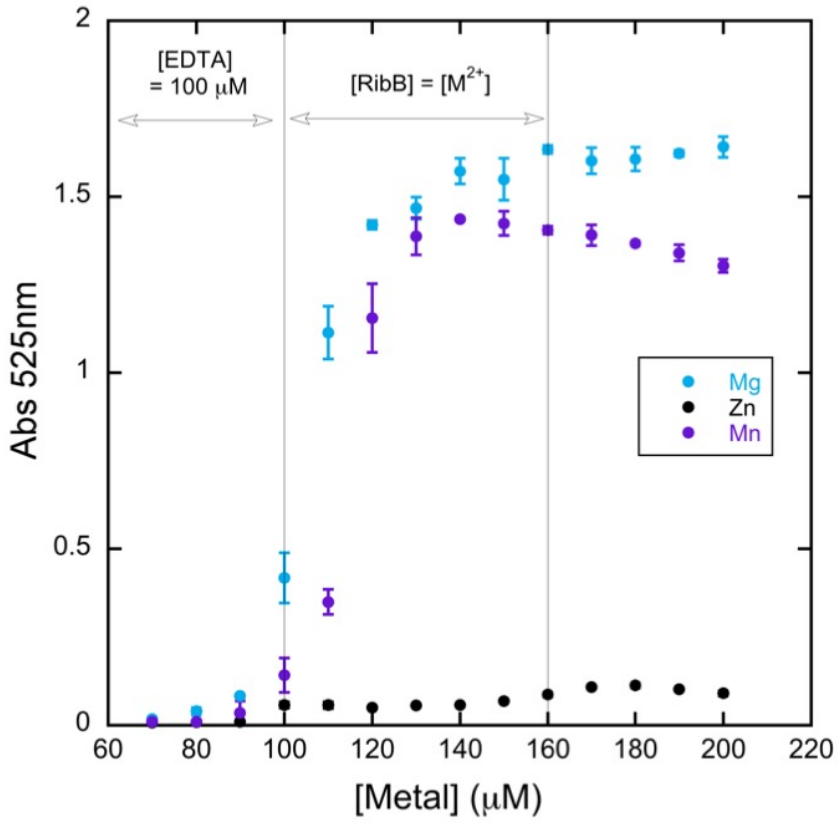


Figure 3.5. One metal equivalent is required for RibB activity. Activity of 60 μM RibB in the presence of 100 μM EDTA was measured after titration with Mg^{2+} (cyan), Zn^{2+} (black) and Mn^{2+} (purple). Activity increases upon addition of 100 μM metal and no longer increases at the 160 μM. This shows 1:1 ratio of metal to RibB is required to activity.

3.3.4 Structural analysis of apo-RibB structure

The structure of apo-RibB has been solved as a dimer with one dimer in the asymmetric unit. Even though this structure has been cryoprotected in the presence of substrate and Mg^{2+} there was no electron density for either one of these. As mentioned previously, the active site is composed of two loops, loop 1 (34-41) and loop 2 (85-97) located opposite to each other, surrounding the substrate in the active site with loop 1 orientated away from the protein dimer interface (**Figure 3.2**). Loop 1 (residues 34-41) has been previously reported to be involved substrate and metal recognition and binding [10, 12]. This loop is composed of acidic amino acid residues. Out of 8 amino acids residues, only 2 residues Arg38 and Asn40 are not acidic and others are equally divided between Glu and Asp. In this apo-RibB structure we find loop 1 in the open conformation with important Glu39 residues pointing away from the active site (**Figure 3.6**). Function of loop 2 based on our structural data is unclear. This loop in known to be mobile and well-ordered only in examples of structures with inactive complexes. Considering that this is an apo structure it comes as no surprise that loop 2 in both monomers was disordered.

3.3.5 Structural analysis of Ru5P (D-ribulose 5-phosphate) RibB structure

Structure of Ru5P bound to RibB was solved by co-crystallizing the substrate with enzyme and soaking 15 molar excess of the substrate prior to cryoprotection. The final structure was well ordered with B-factors of 21.8 $Å^2$ for 1.8 Å structure. Both catalytic loops in this structure are in closed conformations. Loop 1 is pointed toward the active site were Glu39 is expected to form the metal binding pocket with the substrate. In

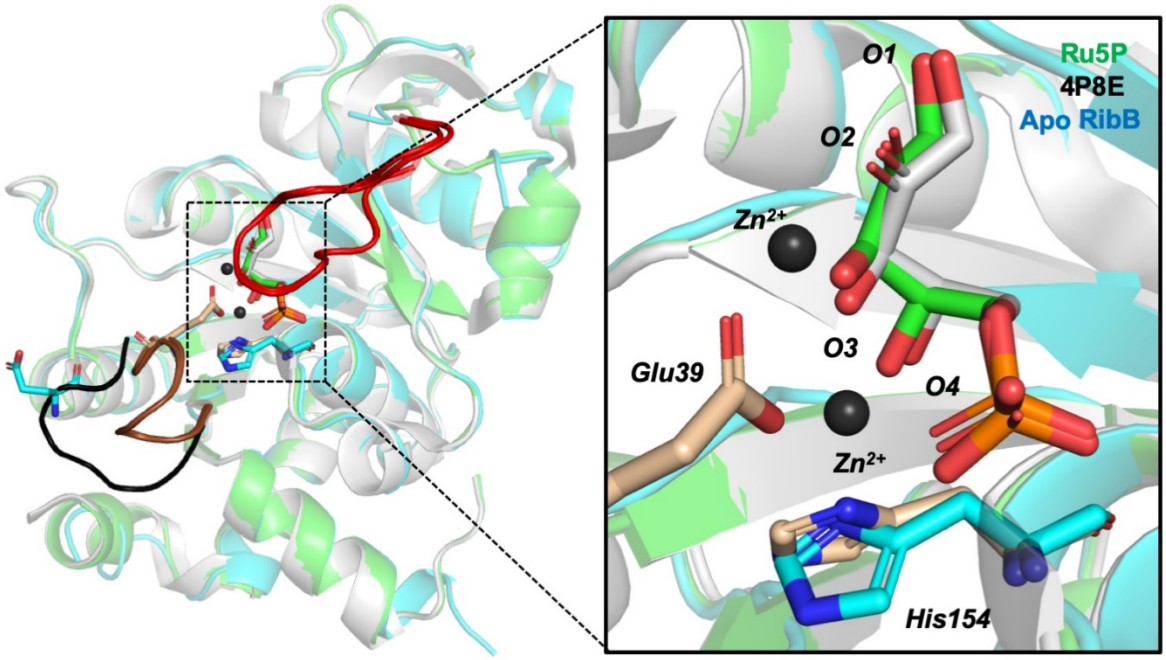


Figure 3.6. Comparison of RibB Ru5P structure, apo-RibB and published structure of RibB with Zn^{2+} (4P8E). Active site is surrounded by two loops. Loop 1 in the closed form (brown) coordinates the metal and stabilizes the substrate in active site. Metal is coordinated by the substrate and Glu39 and His154. In apo-RibB (cyan) loop 1 is in open conformation (black) with Glu39 (cyan) pointing away from active site and plane of imidazole ring of His154 rotated by 90 degrees (cyan). While in apo-RibB loop 2 (red) is disordered, in Ru5P it is modelled over active site. Also, RibB has usually been modelled with two metals in the active site. Ru5P solved structure, shown in green, has been super-imposed with published di-Zn^{2+} RibB structure (PDB:4P8E, gray), Zn^{2+} ions shown in black spheres. Modeled Ru5P molecules in Zn^{2+} structure (gray) have a different binding pose with a shift of 0.4 Å away from the active site. Zn^{2+} ions (black spheres) are in same binding location as Mn^{2+} ions in L-Xy5P structure.

comparison to the apo-RibB structure, the closing of the loop 1 also induces the 90 degrees rotation in the plane of the imidazole ring of His154 (**Figure 3.6**). The only interaction between substrate and loop 2 is weak interaction between Thr94 and hydroxyl of O3 (**Figure 3.7**).

The residues responsible in the substrate binding in the RibB active site can be divided in three sets of residues: 1) Phosphate binding, 2) metal binding and 3) terminal hydroxyl binding residues. The residues responsible for binding of the phosphate group on the substrate are: Arg38, Arg151 and Thr155. Note that one of these residues Arg38 belongs to loop 1 defining the importance of this chain in substrate recognition. Mn^{2+} is coordinated by His154 and Glu39 and the substrate, hydroxyls at carbons C3 and C4 (O1, O2), phosphate and water molecule hydrogen bonded by other acidic residues of loop 1, Asp43 and Glu41 (**Figure 3.7**). This binding mode we will refer to, from this point forward, as first metal binding site. While Ru5P structure was not crystallized in the presence of metal this binding site is occupied by a water molecule. On the other side of the molecule, the hydroxyl at the carbon C1 is part of hydrogen bonding network formed between this hydroxyl, water molecule and three residues Glu175, Cys68, Ser64. The hydroxyl is equidistant between GluO2ε and Cys68 (3.4 Å). A water molecule, that is 2.4 Å away from this hydroxyl is also in the center, hydrogen bonding distance away between Glu175, Cys68 and Ser64. The water molecule in this network we will refer to as W2 in our further discussion. In the analysis of dimer interfaces, we have mentioned that polypeptide chain 2 partially intrudes into the active site of the opposing monomer with one of the residues His137 being hydrogen bonding distance away form Glu175

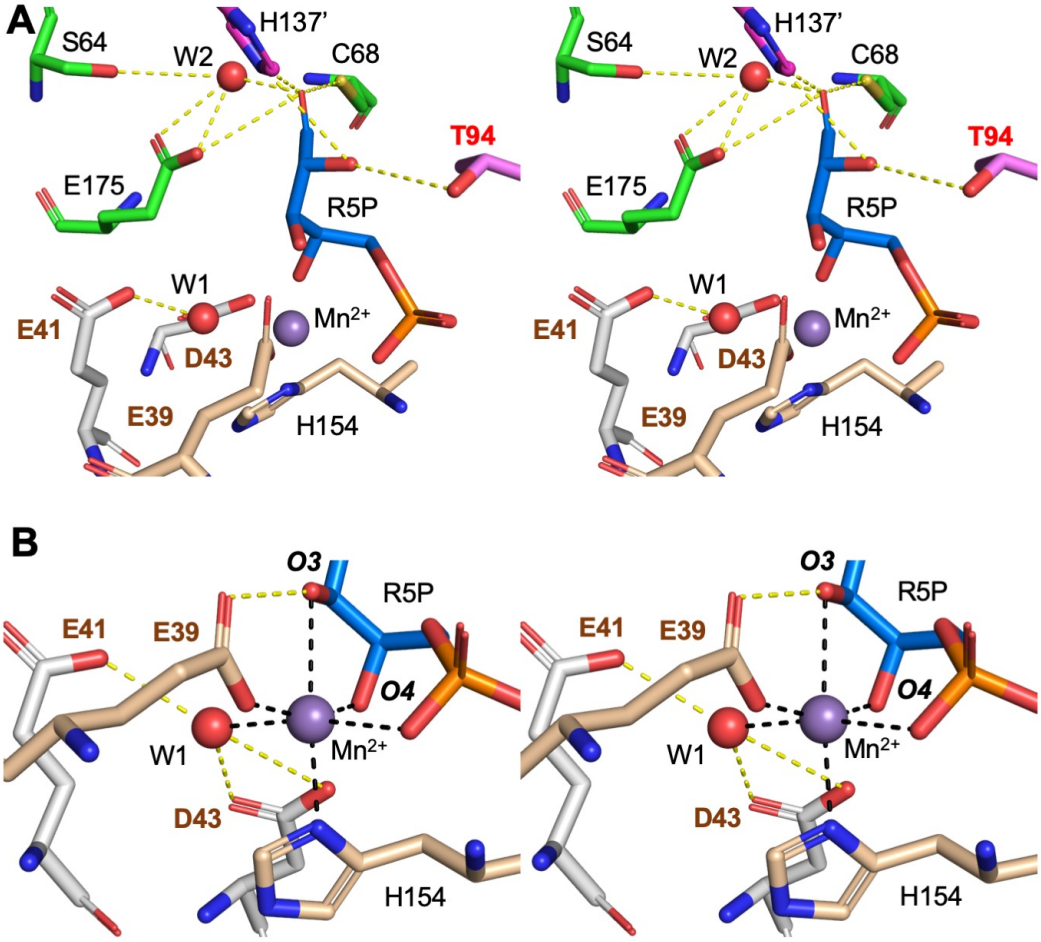


Figure 3.7. Active site composition and metal binding. (A) Stereo view representation of residues involved in substrate binding in active site. Ribose 5-phospate is shown in blue sticks. Substrate is stabilized by phosphate binding (not represented here), Mn^{2+} residues (residues directly involved in binding shown in wheat and indirectly shown in gray) and binding to hydroxyls at carbons C1 and C2 (shown in green). Active site has two important water molecules. Water molecule in first metal binding site (W1) is bonding distance away from Asp43 and Glu41. Second water molecule (W2) is binding distance away from hydroxyl at carbon C1 and Glu175, Ser64 and Cys68. Residues that belong to the catalytically important loop 1 are highlighted in brown. Thr94, residue of loop2 (highlighted in red) has a weak hydrogen bonding interaction with hydroxyl of carbon C2. His137 (purple) is a residue of opposing monomer hydrogen-bonding distance away from hydroxyls at C1 and C2. (B) Detailed stereo-view representation of Mn^{2+} (shown in purple sphere) binding in respective first metal binding pocket. Mn^{2+} metal ion is bound in the octahedral manner electrostatically coordinated by functional groups on substrate and on active site residues. Substrate coordinates Mn^{2+} by hydroxyls at C3 (O3) and C4 (O4) and phosphate. Other coordination sites are from His154, Glu39 and a water molecule.

(**Figure 3.7**). Also, Nε2 of His137 is equidistant from the hydroxyls of carbon C1 and C2.

In the comparison to the published Ru5P structure with two Zn^{2+} ions, the substrates have a small difference in the binding orientation in the active site with Ru5P from Zn^{2+} shifted by 0.4 Å. This small shift could be attributed to the need to accommodate the Zn^{2+} metal ions in the active site. Furthermore, Zn^{2+} structure has a second metal binding site coordinated by Glu39 and hydroxyls at O2 and O3. This site is also in a very electronegative binding pocket surrounded with acidic residues Glu39, Asp41 and Glu175 (**Figure 3.6**).

3.3.6 Structural analysis of R5P (ribose 5-phosphate) RibB structure

The metal binding is dependent on both the substrate and protein residues. Therefore, investigation of the metal binding requires the use of the substrate that has the same structural features as Ru5P. Also, this compound had to be inactive with addition of metal in order to aid in structure solution. Ribose 5-phosphate (R5P) was a great candidate compound, as an aldose sugar to Ru5P it was inactive and structurally similar to the substrate.

The R5P and Ru5P bound RibB structures do not show any significant differences in orientation of structural elements, both loops are in closed orientation with RMSD of 0.134 Å between the two structures. The most noticeable differences are in the active site. R5P was solved with one Mn^{2+} ion in the active site. This metal ion is located in

previously proposed first metal binding site. There was no evident electron density in the second metal site.

While the orientation of the Ru5P ligand is similar to R5P, one difference is a rotation of the hydroxyl at the position of the carbon C2 by 42 degrees (as measures by dihedral angles O2, C2, C3, O3 in Pymol [28]). The most likely explanation is difference in carbon hybridization at carbon C2 between two substrates (**Figure 3.8, A**).

3.3.7 Structural analysis of D-Xy5P (D-xylulose 5-phosphate) RibB structure

D-Xy5P is the only Ru5P isomer with significant activity (**Figure 3.8, D**). In the presence of D-Xy5P, VcRibB is 2.6 times less active in comparison to the substrate D-Ru5P. Therefore, in the presence of metal, D-Xy5P will make a catalytically active complex. This can be used as a justification for the absence of the mobile loop 2. In the active site D-Xy5P occupies the same binding cavity as the substrate. Observing along its carbon backbone, D-Xy5P is positioned as a mirror image of the substrate. This seems obvious as D-Xy5P is an epimer of the substrate at carbon C3. This also means that while hydroxyls at carbon C3 are oriented in the same direction, hydroxyl (O3*) and carbon C3 of D-Xy5P are moved 1.7 Å away from their respective locations in the substrate. Location of O3 of the substrate is now populated with hydroxyl at carbon C4 of D-Xy5P (O4*, **Figure 3.8, B**).

D-Xy5P has an unprecedented change in the conformation of the Glu39 side chain. Over the two states, A and B (**Figure 3.8, B**), state A (wheat, **Figure 3.8, B**) is the representative of location of Glu39 in other RibB structures. On the other hand, state B

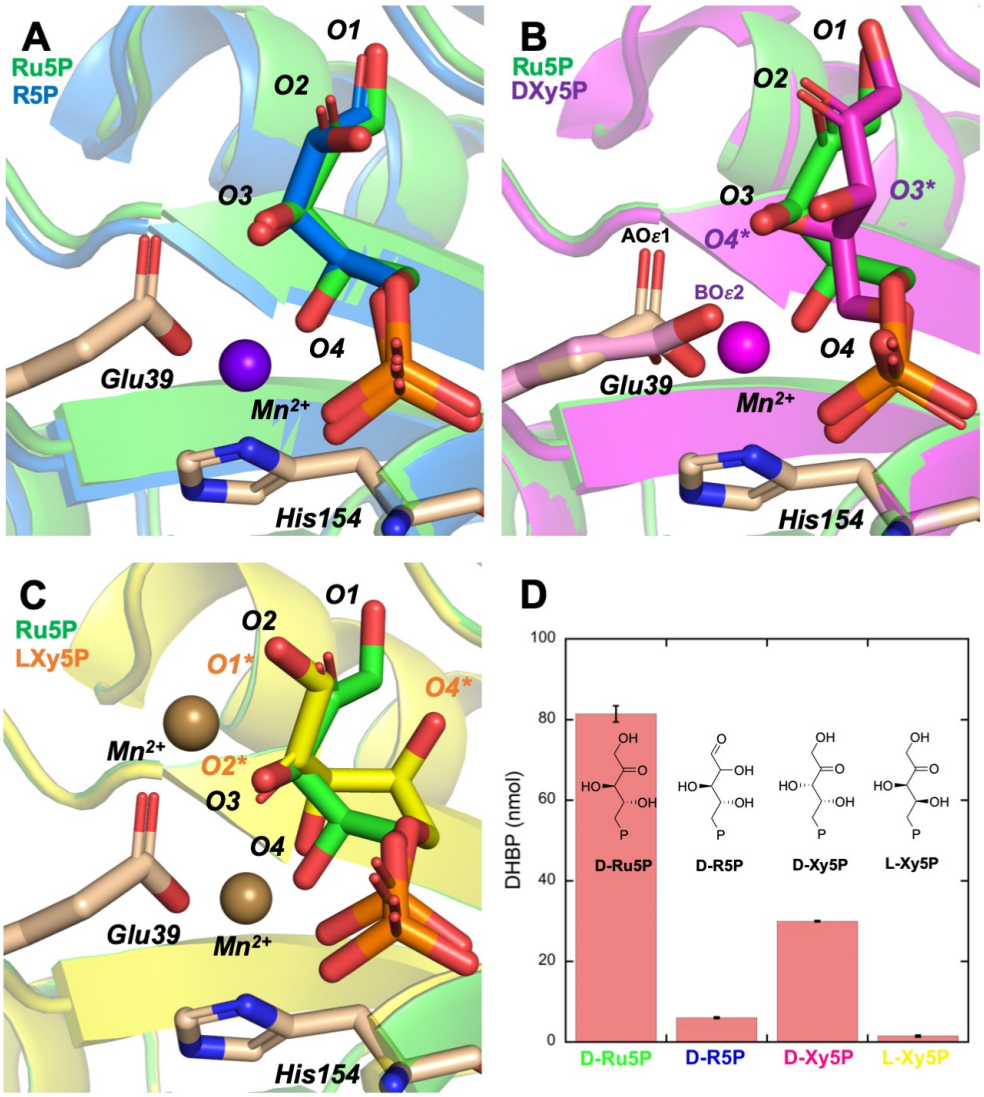


Figure 3.8. Comparison between active site binding and activity of RibB substrate isomers. (A) Ribulose 5-phosphate (green) and ribose 5-phosphate (blue) are in similar binding orientations in the active site. One difference is the orientation of the hydroxyl (position O2) of R5P of 42 degrees (dihedral angle) in comparison to carbonyl of Ru5P. (B) D-Xylulose 5-phospate (violet) is an epimer of ribulose 5-phosate (green) at carbon C3 leading to distinct orientation of hydroxyls (O3 vs O3*). Change in stereochemistry at this hydroxyl at leads to movements of hydroxy at O4* to a location occupied by substrate at O3. Also, this structure has two distinct orientation of Glu39 side chain. State A (wheat) is identical to Glu39 side chain in other structures. State B (pink), in a 72 degrees clockwise rotation of state A side chain. Reason behind the rotation could be attributed to hydrogen bonding of side chain to hydroxyls as O3 (C) Structure of L-Xy5P (yellow) shows distinct binding orientation in comparison to Ru5P (green). As expected by stereochemistry oxygen of carbon C4 in LXy5P(O4*) is rotated away from the Mn binding site. Oxygen atoms at position 1 (O1*) and 2 (O2*) are rotated away from their binding position closer to Glu 39. This change was sufficient enough to make an additional Mn^{2+} binding site (brown). (D) Bar diagram representation of activity comparison between substrate,

D-Ru5P, and isomers. While R5P and L-Xy5P are inactive, D-Xy5P is 2.5 times less active then the substrate.

(pink, **Figure 3.8, B**) is a rotation of the Glu39A side chain by 72 degrees clockwise. According to calculations of occupancy both states occur in equal measures with 51% occupancy is in A state and 49% in B state. While reason behind this flip in side chain remain unknown, one hypothesis could be related to the hydrogen bonding of Glu39 to hydroxyl at O3 (**Figure 3.8, B**). In state A (wheat, **Figure 3.8, B**) side chain of Glu39 is oriented in plane with hydroxyls O3 and O4 of the Ru5P with Oε1 being hydrogen bonding distance away, 2.5 Å, from O3 and Oε2 binding coordinating the Mn^{2+}. In state B (pink, **Figure 3.8, B**) it appears that the movement of side chain goes in line with the location of O3 in D-Xy5P (O3*, **Figure 3.8, B**). Oε1 of state B is coordinating the Mn^{2+}, while Oε2 is weak hydrogen bonding distance away, 3Å, from O3* of D-Xy5P. Furthermore, the distance between O3 and O3* is identical to the distance between hydrogen bonding oxygen atoms on Glu39 in the two states, of 1.7 Å. This distance is identical to the movement of C3* hydroxyl of D-Xy5P. It appears that the movement of the C3* of D-Xy5P corelated with the change in Glu39 side chain orientation. This could suggest that Glu39 could have more catalytic role rather than only be responsible for metal binding.

D-Xy5P has been refined with a Mn^{2+} ion in the first metal binding site. Changes in the orientation of the O3 and O4 hydroxyls cause coordination changes to Mn^{2+} in D-Xy5P. Mn^{2+} in now coordinated only with hydroxyl at O4 and phosphate because O3* is too far away from the metal ion to be considered a valid coordination ligand, 3.7 Å. This makes Mn^{2+} directly coordinated to 5 ligands instead of 6 ligands in octahedral manner as previously described. It is important to note that we have found some electron density suggestive of a water molecule at the position of O4 which would be a sixth ligand of the

metal coordination shell. However, during the refinement process that electron density disappeared which is suggestive of a very low occupancy of that water molecule and it is a reason why it was not modeled in the structure.

3.3.8 Structural analysis of L-Xy5P (L-xylulose 5-phosphate) RibB structure

L-Xy5P bound structure, in comparison, has significant changes in ligand orientation. In truth, this was to be expected as L-Xy5P differs from the substrate in stereochemistry at carbons C3 and C4. In the active site, the change of the stereochemistry at C4 changes the orientation of the hydroxyl of L-Xy5P: instead of being coordinated to Mn^{2+} ion, this moiety is within hydrogen bonding distance of Cys68. Due to orientation of L-Xy5P ligands are close to the first Mn^{2+} metal center with average ligand distance of 2.15 Å. Metal ion is now being coordinated by oxygens on carbon C2 and C3 and by the phosphate which is pushed away from the phosphate binding pocket by 0.7 Å. Changes in the coordination of Mn^{2+} have also drastic changes in the orientation of the atoms at the C1 end of L-Xy5P. In comparison to the substrate binding, the hydroxyl at carbon C1 has been rotated and moves 3.5 Å closer to Glu39 residue (**Figure 3.8, D**). This shift was enough to make a convenient electrostatic binding pocket for binding of another Mn^{2+} ion. This Mn^{2+} binding site, matches the location of second metal binding site previously reported.

3.3.9 EPR investigation of Mn^{2+} binding

Di-nuclear Mn^{2+} centers have been previously investigated by EPR in the dinitrogenase reductase enzyme [32]. According to this study, EPR spectra of di-nuclear Mn^{2+} have

characteristic features at 2800G and 3800G with hyperfine splitting of 45G. A high intensity six peak pattern at 3400G is suggestive of free Mn^{2+} metal ions in solution [32]. Following the similar methodology, we should be able to distinguish a two Mn^{2+} ion interaction at a specific point of time in the catalyzed reaction. First, in order to inspect Mn^{2+} binding in RibB an equimolar concentration of enzyme and Mn^{2+} (3mM) were mixed with substrate Ru5P (2mM) and reactions left to proceed for 1 min, 10 min and 60 min before freezing (**Figure 3.9, A**). Absence of hyperfine splitting in these spectra indicate absence of di-nuclear Mn^{2+} in excess of metal during the catalytic reaction. Moreover, subsequent decrease in intensity of six peaks at 3400G upon addition of enzyme and substrate is suggestive of metal binding within the enzyme. Second, structural data of L-Xy5P (**Figure 3.8, C**) provided direct evidence of VcRibB structure with di-nuclear Mn^{2+} metal ions separated by 4 Å. Using L-Xy5P in EPR studies provided a control experiment for Mn^{2+} binding in the enzyme active site. As expected, EPR studies are suggestive of di-nuclear Mn^{2+} center both in presence of metal excess (dark purple, **Figure 3.9, C**) after 10 min and equimolar solution after 10 sec (hyperfine splitting of 46G and 2800G). Lastly, lowering the ratio of Mn^{2+} and freezing the reaction mixture (in presence of Ru5P) within 10 sec lead to an ERP spectra indicative of di-nuclear Mn^{2+} center, predominant in the equimolar solution (**Figure 3.9, B**). These results would suggest that in the beginning of the reaction di metal center can bind in the active site. However, as hyperfine splitting is absent in samples with excess metal (dark blue, **Figure 3.9, B**) in later time point, it can be concluded that di-nuclear Mn^{2+} is not required for enzyme catalytic function.

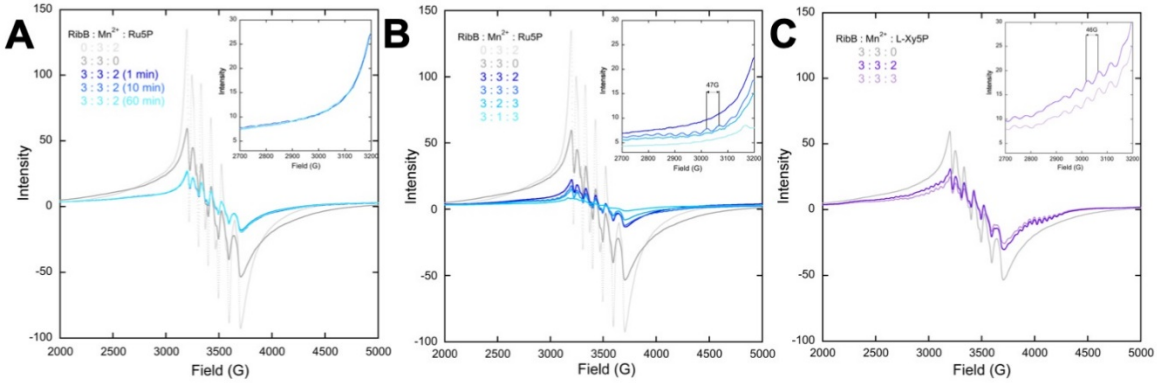


Figure 3.9. Examination of RibB catalyzed reaction using perpendicular-mode EPR spectroscopy. Free Mn^{2+} in solution exhibits a characteristic six-line signal centered at 3400G. The features at 2800G and 3800G with hyperfine splitting of 45G are suggestive a dinuclear Mn^{2+} species. (A) RibB reaction in presence of excess Mn^{2+} left to react for period of 1, 10 and 60 min and quenched by freezing the reaction in an EPR tube with liquid nitrogen. Decrease of intensity in the peak associated with free Mn^{2+} in solution is observed upon addition of enzyme and substrate, which is suggestive of metal binding within the enzyme. (B) Lowering the ratio of Mn^{2+} and quenching the reaction in shorter period of time (10sec) leads to similar hyperfine splitting, suggestive of a dinuclear Mn^{2+} center. This feature, however, disappears in presence of excess Mn^{2+} and extending the reaction to 20sec (dark blue line). (C) Substitution of D-Ru5P for L-Xy5P in presence of excess of Mn^{2+} (dark purple, 10min) leads to characteristic dinuclear Mn^{2+} features (hyperfine splitting = 46G). Equimolar solution of reaction components after 10sec produces a similar spectrum. [4]

[4] Data acquired by Melissa C. Denler, graduate student in Dr Timothy A. Jackson lab. Melissa and Dr Timothy A. Jackson helped in further processing and data analysis

3.4 DISCUSSION

The RibB enzyme in the riboflavin biosynthesis is structurally and functionally unique. Functionally this enzyme is magnesium dependent and catalyzes a reaction that abstracts carbon C4 of the substrate ribulose 5-phospate and removes the hydroxyl on carbon C1 to make a final product (**Figure 3.1**). Structurally, the topology of this enzyme is unparalleled and no other enzyme with similar chemistry has been documented. Structural data of RibB enzymes with Mn^{2+}, Zn^{2+}, Ca^{2+} have been presented as a paradigm that enzyme catalysis required two metals for full activity. However, no further stoichiometric evidence demonstrating this was presented. Here, we show steady state kinetics of the RibB enzyme in the presence of Mn^{2+} as well as crystal structures of apo-RibB, RibB in complex with inactive analogue, ribose 5-phosphate with one Mn^{2+}, stereoisomers L-xylulose 5-phosphate with two Mn^{2+} metal ions, D-xylulose 5-phosphate with one Mn^{2+} and structure of RibB with substrate ribulose 5-phosphate with no metal. Furthermore, we present binding, activity assay and EPR evidence that one metal binding site is required for the full enzyme activity.

Use of Mn^{2+} as a replacement for Mg^{2+} ions in magnesium depended protein is a common methodology [33, 34]. In enzyme complexes, both Mn^{2+} and Mg^{2+} ions have typical 6 coordinate geometry with magnesium ions having much higher probability of formation of hexahydrate complexes. Both are considered to be hard acids as breaking of their hexacoordinate shell requires significant thermodynamic contribution [34]. Their ionic radii is also very similar with reported values of 0.65 Å for magnesium and 0.74 Å for manganese [34]. Mn^{2+} ions have a big advantage when used as replacement to

Mg^{2+} ions in crystal structure studies. While Mg ions are isoelectronic with water and cannot be readily distinguished based on the electron density, Mn^{2+} is a transition metal with greater electron density. Use of Ca^{2+} or Zn^{2+} in these studies could be problematic due to differences in physical properties of these metals and metal of interest, Mg^{2+}. Ca^{2+} ion is also a hard acid, but the ionic radius is larger than Mg^{2+} (0.99 Å). Also, Ca^{2+} ions have been described to have coordination geometries of 6 or 8. Enzymatic sites made for Ca^{2+} ion are usually designed by enzymes that sequester it and readily bind to it, such as EF hand motif [34]. Zinc is different than either magnesium, manganese or calcium. While its ionic radius is similar to magnesium and manganese (0.71 Å), zinc like calcium, easily interchanges between different coordination states (coordination number of 4,5 or 6). As such zinc is not as strong of a Lewis acid which has been proposed as one of the reason for RibB inactivity with this metal [11]. Finally, Mn^{2+} is EPR active which provides better opportunities to study metal binding and coordination chemistry.

Use of Mn^{2+} as a metal of choice in RibB enzyme kinetics was beneficial. In comparison to Mg^{2+}, the efficiency of the enzyme in the presence of Mn^{2+} was 24 times higher (**Figure 3.3**). Chemically this could be explained by ligand preferences between these metals. Mg^{2+} has greater preference for oxygen ligands (hydroxyls, carboxylic acids) and is not commonly seen in complex with nitrogen ligands which is not true for Mn^{2+} complexes. In nature, enzymes have been shown to use this method to discriminate between Mg^{2+} or Mn^{2+} metal ion binding [33, 34]. It is clear however, based on the crystal structures, that the first binding site has a nitrogen binding ligand from His154 (**Figure 3.7, B**). This could explain the kinetics and stronger association between Mn^{2+}

the lower K_M and the active site ligands allowing for better stabilization of substrate in the active site and more efficient steady state kinetic profile (**Figure 3.3**). Nevertheless, RibB is most likely a magnesium-dependent enzyme. Because the availability of Mn^{2+} in the cell is low in comparison to Mg^{2+} [35] the likelihood of Mn^{2+} association with RibB in the cell is unlikely.

Earlier we have described two metal titration experiments that demonstrate that RibB binds one metal and requires one metal in the catalysis. Our structure of the inactive substrate, aldose sugar analogue, ribose 5-phosphate, supports that location of this metal site is in the first metal site coordinated by His154, Glu39 and the substrate. It is important to note that both ribose 5-phosphate and ribulose 5-phosphate have same orientation in the active site with no changes in protein structure or residue orientation (**Figure 3.8**). The previously determined structure of RibB with glycerol supports our hypothesis. Glycerol in this structure is oriented in same way as carbohydrate backbone of the substrate and a single Mn^{2+} is in the same orientation as in our structures (PDB:1K40) [11].

Proposed second metal binding site in the literature is coordinated by Glu39 and hydroxyls O3 and O2 of the substrate (**Figure 3.6**). Considering that other chelating ligands on this metal are water molecules this would suggest that this site is labile and might not have a catalytic purpose in the enzyme. The region of second metal binding is surface exposed, in proximity to acidic residues in the active site (Glu39, Glu41,

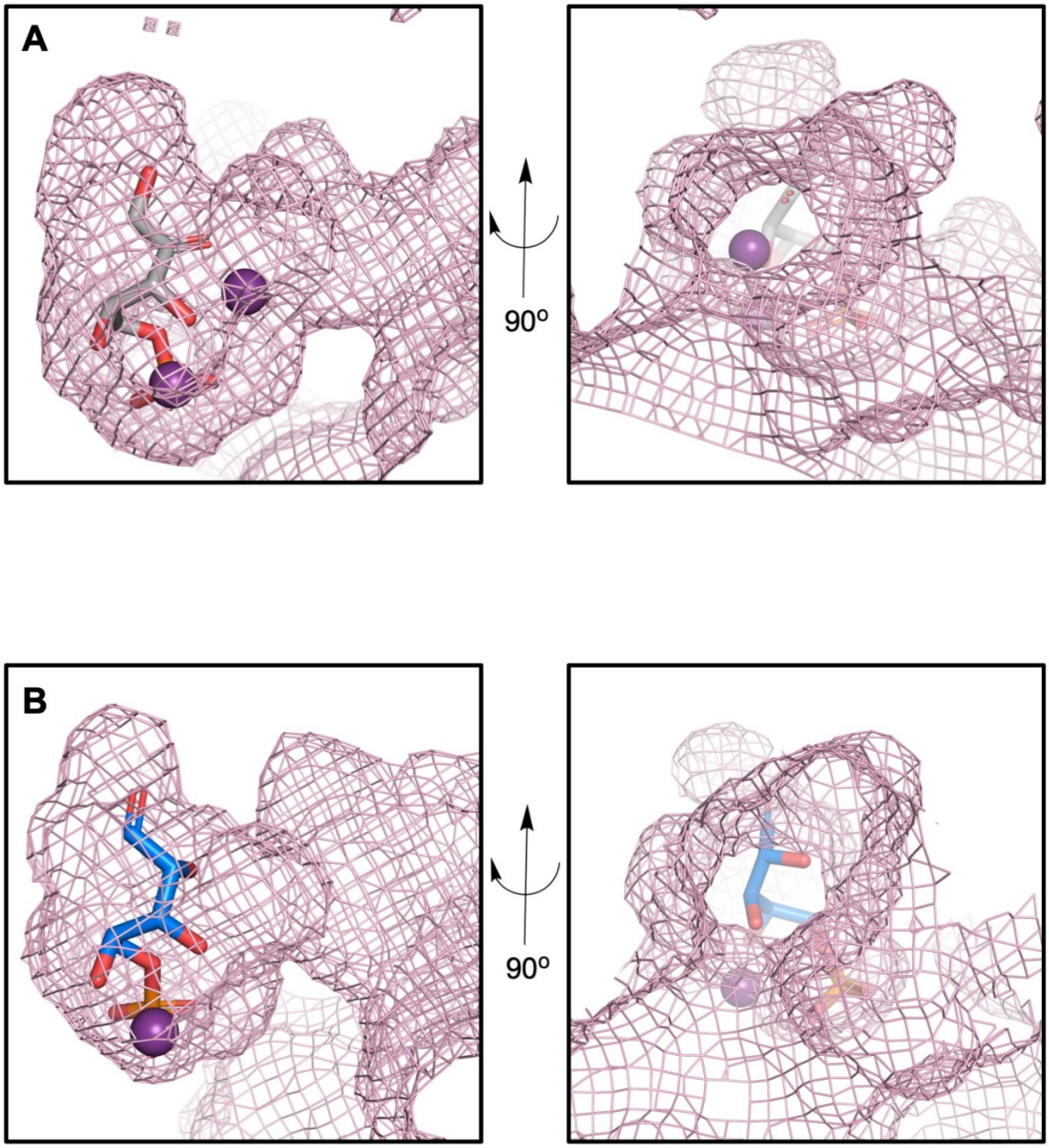


Figure 3.10. RibB active site binding cavity. Active site of the RibB occupies surface area of 237 $Å^2$ and volume of 95 $Å^3$ (generated by CASTp 3.0 [36]). In the closed conformation of active site loops substrate binding is buried deep in the pocket between two monomers. First metal binding site (purple) is coordinated by the protein and the substrate. In contrary proposed second metal binding site is open to the solvent allowing the diffusion of the metal between solvent and this ligand-rich binding environment. (A) Surfaces binding area of VcRibB structure with Zn^{2+} (PDB:4P8E). (B) Structure of VcRibB with R5P

Glu175, Asp43) making an ideal electrostatic area to attract the metal ions (**Figure 3.10**). Therefore, it is not surprising that in excess of metal ions this area of the active site can be a preferred binding area for the metal. As we have shown with the structure of the non-catalytic L-Xy5P, the proximity of the substrate hydroxyls (O1* and O2*, **Figure 3.8, C**) are brought closer to the Glu39 locking binding metal in place. EPR data does show some evidence of weak di-metal presence in first few seconds of the catalyzed reaction. As these features are absent in later stages of catalyzed reaction in presence of excess of metal, dinuclear Mn^{2+} center is not required for catalysis.
The evidence indicates a single metal binding site in VcRibB, at the location of first metal binding site. We therefore conclude that during catalysis, one metal is required near the C4 portion of the substrate. This doesn't necessarily eliminate the currently proposed mechanism of 1,2 methyl shift but certainly opens up possibilities for a new mechanistic alteration to the substrate by RibB.

3.5 REFERENCES:

1. Volk, R. and A. Bacher, *Biosynthesis of Riboflavin - the Structure of the 4-Carbon Precursor.* Journal of the American Chemical Society, 1988. **110**(11): p. 3651-3653.
2. Volk, R. and A. Bacher, *Biosynthesis of riboflavin. Studies on the mechanism of L-3,4-dihydroxy-2-butanone 4-phosphate synthase.* J Biol Chem, 1991. **266**(31): p. 20610-8.
3. Bacher, A., Q. Le Van, and M. Buhler, *Biosynthesis of riboflavin. Incorporation of D-[1-13C]RIbose* Journal of American Chemical Society, 1982. **104**: p. 3753 - 3755.
4. Bacher, A., et al., *Biosynthesis of riboflavin. Incorporation of 13C-labeled precursors into the xylene ring.* J Biol Chem, 1983. **258**(22): p. 13431-7.
5. Kelly, M.J., et al., *The NMR structure of the 47-kDa dimeric enzyme 3,4-dihydroxy-2-butanone-4-phosphate synthase and ligand binding studies reveal the location of the active site.* Proc Natl Acad Sci U S A, 2001. **98**(23): p. 13025-30.
6. Picollelli, M.A., P.V. Viitanen, and D.B. Jordan, *Spectrophotometric determination of 3, 4-dihydroxy-2-butanone-4-phosphate synthase activity.* Anal Biochem, 2000. **287**(2): p. 347-9.
7. Krissinel, E. and K. Henrick, *Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2256-68.
8. Li, J., et al., *The crystal structure and biochemical properties of DHBPS from Streptococcus pneumoniae, a potential anti-infective target for Gram-positive bacteria.* Protein Expr Purif, 2013. **91**(2): p. 161-8.
9. Singh, M., et al., *The crystal structure reveals the molecular mechanism of bifunctional 3,4-dihydroxy-2-butanone 4-phosphate synthase/GTP cyclohydrolase II (Rv1415) from Mycobacterium tuberculosis.* Acta Crystallogr D Biol Crystallogr, 2013. **69**(Pt 9): p. 1633-44.
10. Liao, D.I., et al., *Structural definition of the active site and catalytic mechanism of 3,4-dihydroxy-2-butanone-4-phosphate synthase.* Biochemistry, 2002. **41**(6): p. 1795-806.
11. Steinbacher, S., et al., *Structure of 3,4-dihydroxy-2-butanone 4-phosphate synthase from Methanococcus jannaschii in complex with divalent metal ions and the substrate ribulose 5-phosphate: implications for the catalytic mechanism.* J Biol Chem, 2003. **278**(43): p. 42256-65.
12. Islam, Z., et al., *Structural basis for competitive inhibition of 3,4-dihydroxy-2-butanone-4-phosphate synthase from Vibrio cholerae.* J Biol Chem, 2015. **290**(18): p. 11293-308.
13. Liao, D.I., et al., *Crystal structure of 3,4-dihydroxy-2-butanone 4-phosphate synthase of riboflavin biosynthesis.* Structure, 2001. **9**(1): p. 11-8.
14. Liao, D.I., P.V. Viitanen, and D.B. Jordan, *Cloning, expression, purification and crystallization of dihydroxybutanone phosphate synthase from Magnaporthe grisea.* Acta Crystallogr D Biol Crystallogr, 2000. **56**(Pt 11): p. 1495-7.
15. Steinbacher, S., et al., *Metal sites in 3,4-dihydroxy-2-butanone 4-phosphate synthase from Methanococcus jannaschii in complex with the substrate ribulose 5-phosphate.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 7): p. 1338-40.
16. Kumar, P., et al., *Potential anti-bacterial drug target: structural characterization of 3,4-dihydroxy-2-butanone-4-phosphate synthase from Salmonella typhimurium LT2.* Proteins, 2010. **78**(16): p. 3292-303.
17. Echt, S., et al., *Potential anti-infective targets in pathogenic yeasts: structure and properties of 3,4-dihydroxy-2-butanone 4-phosphate synthase of Candida albicans.* J Mol Biol, 2004. **341**(4): p. 1085-96.

18. McPhillips, T.M., et al., *Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines.* Journal of Synchrotron Radiation, 2002. **9**: p. 401-406.
19. Soltis, S.M., et al., *New paradigm for macromolecular crystallography experiments at SSRL: automated crystal screening and remote data collection.* Acta Crystallographica Section D-Biological Crystallography, 2008. **64**: p. 1210-1221.
20. Kabsch, W., *Integration, scaling, space-group assignment and post-refinement.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 133-44.
21. McCoy, A.J., et al., *Phaser crystallographic software.* J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.
22. Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.
23. 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.
24. Afonine, P.V., et al., *Towards automated crystallographic structure refinement with phenix.refine.* Acta Crystallogr D Biol Crystallogr, 2012. **68**(Pt 4): p. 352-67.
25. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 1): p. 12-21.
26. Moriarty, N.W., R.W. Grosse-Kunstleve, and P.D. Adams, *electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation.* Acta Crystallogr D Biol Crystallogr, 2009. **65**(Pt 10): p. 1074-80.
27. Moriarty, N.W., E.J. Draizen, and P.D. Adams, *An editor for the generation and customization of geometry restraints.* Acta Crystallogr D Struct Biol, 2017. **73**(Pt 2): p. 123-130.
28. *The PyMOL Molecular Graphics Systems, Version 2.0 Schrödinger, LLC.*
29. Krissinel, E. and K. Henrick, *Inference of macromolecular assemblies from crystalline state.* Journal of Molecular Biology, 2007. **372**(3): p. 774-797.
30. Westerfeld, W.W., *A Colorimetric Determination of Blood Acetion.* J. Biol. Chem, 1945 **161**: p. 495-502
31. Speckman, R.A. and E.B. Collins, *Specificity of the Westfield adaptation of the Voges - Proskauer test* Applied and Enviromental Microbiology 1982. **44**(1): p. 40-43.
32. Antharavally, B.S., R.R. Poyer, and P.W. Ludden, *EPR Spectral Evidence for Binuclear Mn(II) Center in Dintorgenase Reductase-Activating Glycohydrolase from Rhodospirillum rubrum.* J Am Chem Soc, 1998. **120**: p. 8897 - 8898
33. Bock, C.W., et al., *Manganese as a replacement for magnesium and zinc: Functional comparison of the divalent ions.* Journal of the American Chemical Society, 1999. **121**(32): p. 7360-7372.
34. Glusker, J.P., A.K. Katz, and C.W. Bock, *Metal ions in bilogical systems.* Rigaku Journal, 1999. **16**(2): p. 8 - 17.
35. Hohle, T.H. and M.R. O'Brian, *The mntH gene encodes the major Mn(2+) transporter in Bradyrhizobium japonicum and is regulated by manganese via the Fur protein.* Mol Microbiol, 2009. **72**(2): p. 399-409.
36. Wei Tian, et al., *CASTp 3.0: computed atlas of surface topology of protein* Nucleic Acid Research, 2018. **46**: p. W363 - W367.

CHAPTER 4:

Identification of intermediates in 3,4-dihydroxy-2-butanone 4-phosphate synthase (RibB) of riboflavin biosynthesis

4.1 INTRODUCTION

The RibB enzyme converts a molecule of the five-carbon sugar ribulose 5-phosphate into a four-carbon compound 3,4-dihydoxy-2 butanone 4-phosphate by removal of carbon C4 of the substrate as formate. The consensus in the literature is a 1,2-methyl shift mechanism, a model that has been related to mechanisms of D-xylose isomerase [1, 2], D-ribulose 5-phosphate isomerase [3, 4]) and 1-deoxy-D-xylulose 5-phosphate (DXP) reductoisomerase from terpenoid biosynthesis [5]). This mechanism involves generation of the intermediate in which methyl group is shifted to the neighboring carbon with high partial positive charge, such as carbon of the carbonyl (**Figure 4.1**). Implementation of this skeletal rearrangement in the reaction mechanism of RibB and DXP reductoisomerase are summarized in **Figure 4.1**. In the example of the DXP reductoisomerase, the methyl shifted is the carbon C1 (**Figure 4.1, A, red**). In RibB, this carbon is equivalent to phosphorylated carbon C5 of the substrate (**Figure 4.1, B, red**). For the DXP reductoisomerases an alternative mechanism, called fragmentation mechanism, has been proposed [5]. Unlike the canonical mechanism, this model involves the generation of the β-hydroxy ketone intermediate (**Figure 4.1**). This ketone intermediate would undergo retro-aldol reaction in which skeletal carbon-carbon bond of

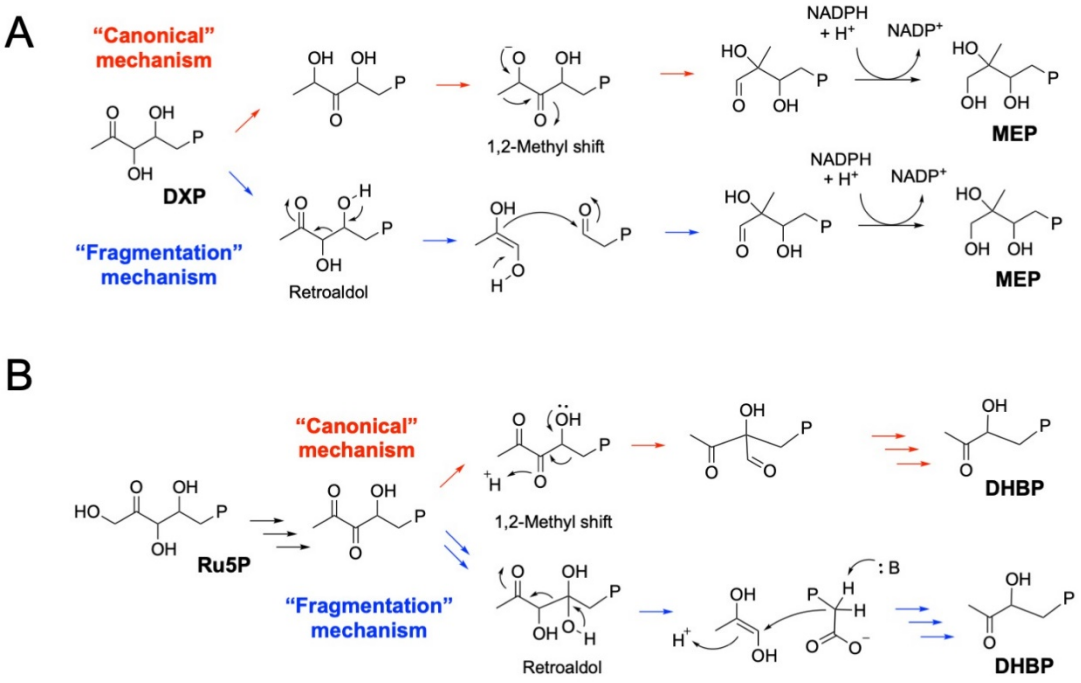


Figure 4.1. Examples of proposed enzymatic mechanisms, canonical (1,2-methyl shift) vs fragmentation (retro-aldol) mechanism. (A) Proposed enzymatic pathways for 1-deoxy-D-xylulose 5-phosphate (DXP) reductosiomerase (B) Proposed enzymatic pathways for RibB catalyzed reaction. While chemical logic of fragmentation and canonical mechanism is similar in these enzymes, reductoisomerases have a crucial NADPH oxidation step.

β-hydroxy is broken to generate two fragment pieces. A potential conformation change in one of the fragments and acid-base chemistry results in the fragments being rejoined to make a new intermediate.

The evidence supporting the canonical mechanism over the fragmentation mechanism is scarce. First argument supporting 1,2-methyl shift is based on isotopically labeled feeding studies. In these experiments, based in enzymes from terpenoid biosynthesis, stochiometric ratios of labeled fragments were measured during the catalyzed reaction. As fragments were not detected during the reaction, this led to the conclusion that fragmentation mechanism is improbable. Nevertheless, the design of the experiment assumed that the substrate fragments are released during catalysis [6]. In the case of RibB, evidence supporting the canonical mechanism is based on end point assays, disregarding that fragments are enzyme intermediates and can only be detected on a much shorter time scale [7-9]. In addition, Cleland and colleagues provided evidence for a fragmentation mechanism in ribulose 5-phosphate isomerase using kinetic isotope effects, in which a change in the rate constants is observed when atoms in a substrate are replaced by heavy isotopes [4]. Therefore, due to limited and inconclusive evidence, it is valuable to investigate the fragmentation mechanism in the RibB reaction.

The two proposed mechanisms for RibB are shown in **Figure 4.2**. In order to distinguish between these mechanisms, we will focus on identifying the pivotal intermediates in each proposed model. For example, in the fragmentation mechanism, a key step would be identification of the enediolate of hydroxyacetone and 2-phosphoglycolic acid (red box

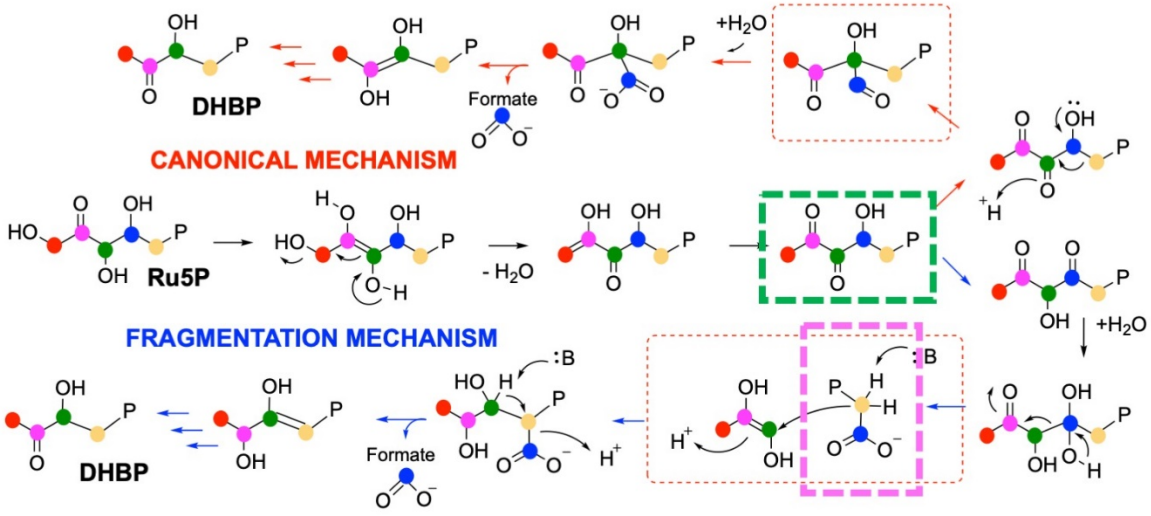


Figure 4.2. Proposed mechanisms for RibB reaction. Both proposed pathways start with molecule of Ru5P in the same manner (black arrows) by dehydration, removal of the hydroxyl at the carbon C1 (red dot) to make intermediate 1, 2-hydroxy-3,4-dioxopentyl phosphate (green box). Intermediate 1 can make product, DHBP via fragmentation mechanism (blue arrows) or canonical mechanism (red arrows). In upcoming experiments identification of crucial intermediates (red boxes) will help us discern between two mechanisms**.** Fragmentation mechanism leads to the formation of two fragments: enediolate of hydroxyacetone and 2-phosphoglycolic acid (pink box). In further discussion we will refer to 2-phosphoglycolic acid intermediate as intermediate 2.

blue arrow, **Figure 4.2**). Conversely, one would look for the formation of the 1,2-methyl shift product in support of the canonical mechanism (red box, red arrow, **Figure 4.2**). Here, I report evidence, using ^{13}C NMR and X-ray crystallography, that the mechanism of the RibB reaction might indeed follow a fragmentation mechanism model.

4.2 MATERIALS AND METHODS

4.2.1 VcRibB purification.

The protein was purified as described in chapter 3. For purposes of the NMR experiment, in the last step of the purification the enzyme was exchanged into 25mM Tris:HCl pH 8.0, 10% glycerol.

4.2.2 6-phosphogluconate dehydrogenase (Ec6PGDH) purification.

Purification of 6-phosphogluconate dehydrogenase from *E.Coli* (*Ec*6PGDH) was done as previously described [10]. The gene sequence from *E.Coli* K-12 (ATCC #47076) was synthesized and placed into a pET-28b(+) vector using *NdeI* and *HindIII* restriction sites (stop codon placed after *HindIII* 3' end) by Genscript. This overexpression construct yields a construct with N-terminal His tag. The purification protocol is similar to the RibB protein purification with a few changes. During cell growth, cells were induced with 1 mM IPTG (isopropyl β-D-1-thiogalactopyranoside) and further incubated at 20 ^{o}C overnight while shaking (220 RPM). Protein was separated out by Nickel Supedex fast flow (GE Healthcare), eluting at 300 mM imidazole, as determined by SDS-PAGE. The fractions containing the Ec6PGDH were dialyzed into 25 mM Tris:HCl pH 8.0 with two

buffer changes at 8 hours and 12 hours. The protein was concentrated and stored for later use at -80 ^{o}C. The final yield of the protein was 185 mg per liter of culture.

4.2.3 Preparation of ^{13}C labeled ribulose 5-phosphate.

Ribulose 5-phosphate was prepared by reconstitution of the pentose phosphate pathway following the previously published protocol [11]. In first vial of 4 ml, 50 mM Tris:HCl pH7.5, 30 mM of ATP, 30 mM of unifomly labeled ^{13}C glucose (Cambridge Isotope Laboratories) and 10 mM DTT (GoldBio) were added and the pH was adjusted to 7.5 using 1M NaOH. 60U of hexokinase (Sigma Aldrich, *Saccharomyces cervisiae*) was added and the solution incubated at 37 ^{o}C for 30 min. In a second vial of 4 ml, 50 mM Tris:HCl pH 7.5, 10 mM of NADPH, 50 mM of α-ketoglutarate and 75 mM of ammonium chloride were added and the pH was adjusted to 7.5 with 1N NaOH. The two vials were combined and 0.5 mg/ml final concentration of Ec6PGDH, 20 U of glutamate dehydrogenase (Sigma Aldrich, bovine liver) and 12 U of glucose 6-phosphate dehydrogenase (Sigma Aldrich, *Saccharomyces cervisiae*) were added. The reaction was incubated at room temperature overnight. Barium chloride, at a final concentration of 50 mM, was added to the solution, mixed and incubated at room temperature for 5 min. A white precipitate formed and was pelleted by centrifuging the solution in a swing bucket rotor centrifuge for 20 min at 3100 x g (4 ^{o}C). The white precipitate was resuspended in ethanol and incubated at -20 ^{o}C for 4 hours. The precipitate was washed in 90% ethanol three times and dried over anhydrous calcium chloride *in vacuo*. The dried white solid was resuspended in 30 ml of water and 50mM final concentration of barium sulfate was added. The white precipitate was removed by

centrifugation for 20 min, 3100 x g at 4 °C. The supernatant was lyophilized producing a white powder that was resuspended in 50 mM Tris:HCl pH 7.5. The concentration of Ru5P was determined by colorimetric assay [12] and standard curve using D-Ru5P (Sigma Aldrich) as previously described. The Ru5P was authenticated by ^{13}C NMR using previously published peak assignments [9].

4.2.4 Single-turnover NMR experiments.

RibB in 50mM Tris:HCl, 10% glycerol pH 8.0 was concentrated to 111 mg/ml (4.3 mM). The reaction mixture contained 50 mM Tris:HCl pH 7.5, 20% D_2O, 2 mM ^{13}C Ru5P and 2.8 mM enzyme in 700 µL final volume at 4 °C. The reaction was initiated by addition of enzyme and was quenched with 70 µl of 4 M H_2SO_4 at allocated time points. The reaction mixture was stored in -20 °C until NMR data acquisition. NMR data were collected using a Brucker AVIII 500MHz NMR equipped with cryogenically-cooled carbon observe probe. The ^{13}C NMR spectra were obtained after 1200 scans, at 25 °C, over 40 min per experiment. Spectra preliminarily referenced based on published spectra peaks[5] [9].

4.2.5 VcRibB crystallization with intermediate 1 and Mn^{2+}.

Purified protein at 32.9 mg/ml was mixed with 15 times molar excess of D-ribulose-5-phosphate in 50 mM Tris-HCl pH 7.5 (Sigma Aldrich) and incubated on ice for 15 min. Protein was crystallized using the hanging-drop vapor diffusion method. Each drop (3

[5] NMR data acquired with help from Justin T. Douglas and Sarah A. Neuenswander. Support for the NMR instrumentation was provided by NIH Shared Instrumentation Grant # S10RR024664 and NSF Major Research Instrumentation Award # 1625923

µl) was prepared by mixing protein and precipitant solution in equal amounts. The precipitant protein solution was 0.1 M Lithium Acetate and 16% PEG 3350. In preparation for data collection, rod shaped crystals were first soaked in precipitant solution mixed with a final concentration of 40 mM D-ribulose 5-phosphate overnight then soaked in cryoprotect containing precipitant supplemented with 30% ethylene glycol and 4mM $MnCl_2$ for 3 min.

4.2.6 VcRibB crystallization with intermediate 2 and Mn^{2+}.

Purified protein at 32.9 mg/ml was mixed with 15 times molar excess of D-ribulose-5-phosphate in 50 mM Tris-HCl pH 7.5 (Sigma Aldrich) and incubated on ice for 15 min. Protein was crystallized using the hanging-drop vapor diffusion method. Each drop (3 µl) was prepared by mixing protein and precipitant solution in equal amounts. The precipitant protein solution was 0.1 M Lithium Acetate and 16% PEG 3350. In preparation for data collection, rod shaped crystals were first soaked in precipitant solution mixed with a final concentration of 40 mM D-ribulose 5-phosphate overnight then soaked in cryoprotect containing precipitant supplemented with 30% ethylene glycol and 4mM $MnCl_2$ for 70 min.

4.2.7 X-ray diffraction data collection.

Data for both intermediate crystal variants were collected at the Stanford Synchrotron Radiation Laboratory (SSRL, Stanford, CA) [13] beamline 12-2 using a wavelength of 0.9795 Å at 100 K. The software package Blu-Ice [14] was used to collect 1200 oscillation images (0.15 ° per image) with an exposure time of 0.2 sec.

4.2.8 X-ray diffraction data collection and processing for RibB intermediate 1 structure with Mn^{2+}

Diffraction data were collected using a beam size of 100 µm x 60 µm with transmission set to 9.5% and a detector distance of 366.4mm. Diffraction data were processed using XDS [15] to 2.2 Å in the space group C 2 2 2_1 with cell dimensions of a=60.34 Å, b=78.56 Å, c=90.54 Å, α = β = γ =90.0°. A phasing solution was obtained via molecular replacement using PHENIX, Phaser-MR [16] with 4P8E(PDBID) [17] monomer as a model. The model was prepared by removing all water molecules, hydrogens and any previously modeled substrates. The search procedure was set to look for one copy of the model in an asymmetric unit and all possible space group enantiomers. One molecular replacement solution was identified with an LLG of 3268 and a TFZ of 41. The solution was subjected to alternating cycles of model building and refinement using Coot [18] and Phenix.Refine [19, 20]. Water molecules were added automatically and inspected manually using Coot. Intermediate 1 and manganese ion were added manually. Restraints for intermediate 1 were generated using eLBOW [21] and REEL [22]. Statistics for data refinement and analysis can be found in **Table 4.1**.

Table 4.1: Crystallographic data for VcRibB structures with intermediates[6]

	VcRibB – Intermediate 1	VcRibB – Intermediate 2
Data collection		
Beamline	SSRL 12-2	SSRL 12-2
Wavelength (Å)	0.9795	0.9795
Space group	$C222_1$	$C222_1$
Cell *a, b, c* (Å), (°)	60.34, 78.56, 90.54, 90	60.91, 78.99, 90.37, 90
Resolution (Å)	39.28 – 2.20 (2.27 – 2.20)	39.49 – 2.10 (2.16 – 2.10)
R_{merge}[a]	0.101 (0.593)	0.095 (0.611)
Total observations	73493 (6101)	86439 (6747)
Total unique observations	11272 (955)	13068 (1038)
Mean ($\langle I \rangle$ / sd(I))	11.5 (2.8)	11.7 (2.7)
Completeness (%)	99.9 (99.2)	99.9 (99.7)
Redundancy	6.5 (6.4)	6.6 (6.5)
CC(1/2)	0.998 (0.870)	0.998 (0.861)
Refinement		
Resolution (Å)	36.03 – 2.2 (2.27 – 2.2)	39.49 -2.1 (2.17 – 2.1)
R_{cryst}[b]	0.1684 (0.2002)	0.1798 (0.2224)
R_{free}[c]	0.2264 (0.2982)	0.2347 (0.3045)
Total unique observations	11250 (1089)	13042 (1285)
No. of non-hydrogen atoms	1707	1656
Protein	1631	1578
Ligand	23	11
Water	53	67
rms deviation bonds (Å)	0.013	0.012
rms deviation angles (°)	1.22	1.39
Overall mean B-factor (Å^2)	37.13	38.40
Ramachandran plot analysis:		
Favored region	98.13	97.06
Allowed region	1.87	2.45
Outlier region	0.0	0.49

[6] $R_{merge} = \Sigma_h|I_h - \langle I \rangle|/\Sigma_h I_h$, where I_h is the intensity of reflection h, and $\langle I \rangle$ is the mean intensity of all symmetry-related reflections [b]$R_{cryst} = \Sigma||F_o| - |F_c||/\Sigma|F_o|$, F_o and F_c are observed and calculated structure factor amplitudes. [c]Five percent of the reflections were initially reserved to create an R_{free} test set used during each subsequent round of refinement.

4.2.9 X-ray diffraction data collection and processing for RibB intermediate 2 structure with Mn^{2+}

Diffraction data were collected using a beam size of 100 µm x 60 µm with transmission set to 11.0% and a detector distance of 404.0 mm. Diffraction data were processed using XDS [15] to 2.1 Å in the space group C 2 2 2_1 with cell dimensions of a=60.91 Å, b=78.99 Å, c=90.37 Å, $\alpha = \beta = \gamma = 90.0°$. A phasing solution was obtained via molecular replacement using PHENIX, Phaser-MR [16] with 4P8E(PDBID) [17] monomer as a model. The model was prepared by removing all water molecules, hydrogens and any previously modeled substrate. The search procedure was set to look for one copy of the model in an asymmetric unit and all possible space group enantiomers. One molecular replacement solution was identified with an LLG of 3098 and a TFZ of 44.0. The solution was subjected to alternating cycles of model building and refinement using Coot [18] and Phenix.Refine [19, 20]. Water molecules were added automatically and inspected manually using Coot. Intermediate 2 and manganese ions were added manually. Restraints for intermediate 2 were generated using eLBOW [21] and REEL [22]. Statistics for data refinement and analysis can be found in **Table 4.1**.

4.2.10 Intermediate 1 crystallographic model.

The final model has one monomer in the asymmetric unit with residues 2 to 217 fully resolved. The final model contained 53 water molecules and an intermediate 1, one Mn^{2+} ion in the active site and two molecule of ethylene glycol. Ramachandran analysis was calculated with MolProbity [23] with 98.13% in the allowed regions and no outliers.

4.2.11 Intermediate 2 crystallographic model.

The final model has one monomer in the asymmetric unit with residues 3 to 217 fully resolved. The final model contained 68 water molecules and one intermediate 2 and two Mn^{2+} ions in the active site. Ramachandran analysis was calculated with MolProbity [23] showing 97.06% in the favored regions and 0.49% of outliers. Asp214 was the outlier and falls in an area of poor electron density.

4.3 RESULTS

4.3.1 Structure of RibB intermediate 1 and consequences on catalysis

RibB crystals were grown in presence of substrate and subsequently soaked into $MnCl_2$ cryoprotectant solution. Structure refinement led to the generation of electron density that by our analysis corresponds to one of the first intermediates of the RibB catalytic cycle. In the first attempt to model the correct substrate we used the ribulose 5-phosphate substrate. After subsequent rounds of structure refinement and modeling, negative electron density indicated the absence of a portion of the molecule at the carbon C1. As the first step in RibB mechanism (green box, **Figure 4.2**) using intermediate 1 as an alternative modeling compound was justified. Indeed, modelling of intermediate 1 (2-hydroxy-3,4-dioxopentyl phosphate) in the active site of RibB provided a model with acceptable crystallographic statistics. After refinement, electron density of simulated annealing composite omit map had electron density of the substrate set at 1.5 sigma with B-factors of 42.24 $Å^2$.

In comparison to the model of substrate binding presented in chapter 3, significant changes in the substrate binding and metal coordination can be observed. As described in chapter 3, metal is coordinated by Glu39, His154, a water molecule (hydrogen bonded to Asp34), and to the substrate by the phosphate and hydroxyls O4 and O3. Additionally, the hydroxyl O1 is 3.4 Å away from Cys68 and Glu175. Intermediate 1 binding is more elongated in comparison to the substrate. The hydroxyl at O1 has been removed as a water molecule during dehydration reaction, therefore the position in the protein structure is now occupied by the hydroxyl of O2. Like its predecessor, the hydroxyl at O2 is hydrogen bonded to Cys68 and Glu175 (**Figure 4.3**). On the other end of the molecule, at the metal binding site, this 'elongation' leads to displacement of the hydroxyl at O3 form the Mn^{2+} coordination sphere and the changing of O4 coordination from equatorial to axial position. The original O4 location in the substrate structure has been replace by a water molecule (W2, **Figure 4.3**) that is 1.9 Å away from Mn^{2+} metal center. The intermediate 1 structure was determined with one metal ion in the active site with no density for a second metal ion. Therefore, this structural data can be used as further evidence for a single metal requirement for the enzyme binding and catalysis (**Chapter 3**).

4.3.2 Generation of the intermediate 2

Prolonged soaking of Ru5P containing crystals with $MnCl_2$ lead to a RibB crystal structure that has electron density for a compound much smaller than the substrate.

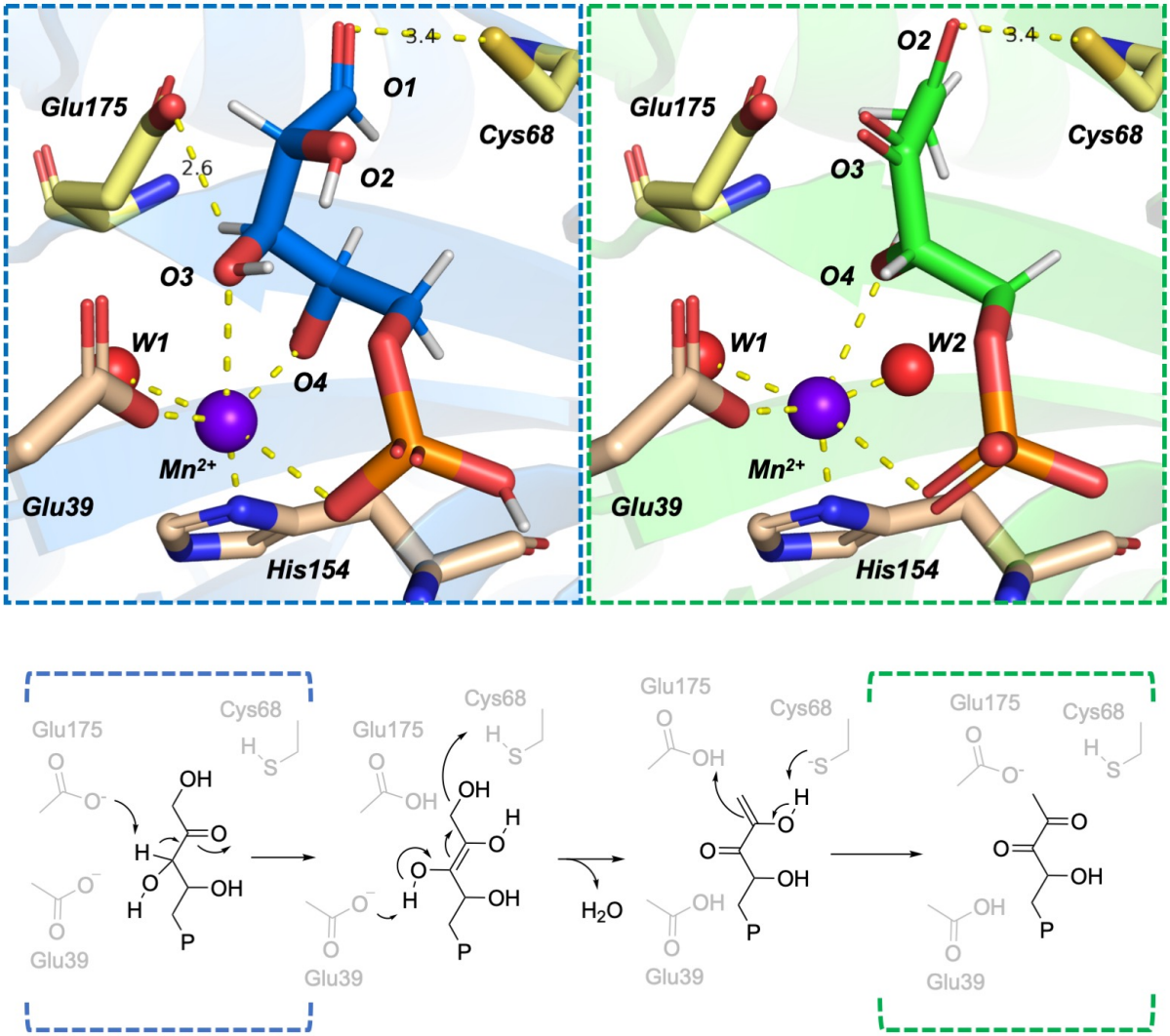


Figure 4.3. Detailed proposed mechanism of intermediate 1 formation. (A) Active site overview of R5P (blue) and intermediate 1 (green) binding. Note the changes in the coordination ligands of Mn^{2+}. In R5P structure Mn^{2+} is coordinated with oxygens O3, O4 and phosphate from the substrate, Glu39, His154 and water molecule (W1). With removal of hydroxyl (at O1) only hydroxyl at O4 and phosphate are coordinating the metal. Previous placement of O4 is now populated with the water molecule (W2) preserving octahedral geometry of Mn^{2+}. (B) Previous research and bonding distance of Glu175 and Cys68 would suggest that these residues have an important role in enolization and removal of hydroxyl on carbon C1 as water.

Upon inspection this electron density accommodates two Mn^{2+} ions and two carbon compound with the phosphate moiety. The only compound in either reaction mechanism that could satisfy these characteristics would be 2-phosphoglycolic acid, one of the fragments in the fragmentation model. Indeed, after refinement we have generated the simulated annealing composite omit map set to 2 sigma confirming successful compound modeling (**Figure 4.4, B**).

Unlike the already presented RibB structures, the structure with intermediate 2 had a disordered loop (residues 87-95). The absence of this loop could indicate a major conformational change in the active site during catalysis. Furthermore, modeling of 2-phosphoglycolic acid in the active site raises the important question of the location of other fragmentation piece, the enediolate. Absence of loop 2 and the enediolate fragment lead us to the hypothesis that loop 2 might be involved in stabilization of this fragment during the catalysis. In its ordered state, as we can see in the intermediate 1 structure, loop 2 sits above the location where the enediol fragment would reside.

The intermediate 2 structure also has two Mn^{2+} ions in the active site. The location of the first Mn^{2+} is identical to the first metal binding site of the substrate and intermediate 1 (**Chapter 3**). The second metal binding site is coordinated by the Glu39, the carboxylic acid of the fragment and 4 water molecules. The presence of the second metal in active site could be attributed to absence of loop 2. Moreover, coordination to extensive number of water molecule could suggest that this metal binding is factitious, frequently exchanging with the solvent, which contain excess Mn^{2+}. Location of the metals in this structure is also identical to the location of Zn^{2+} metal ions in 4P8E

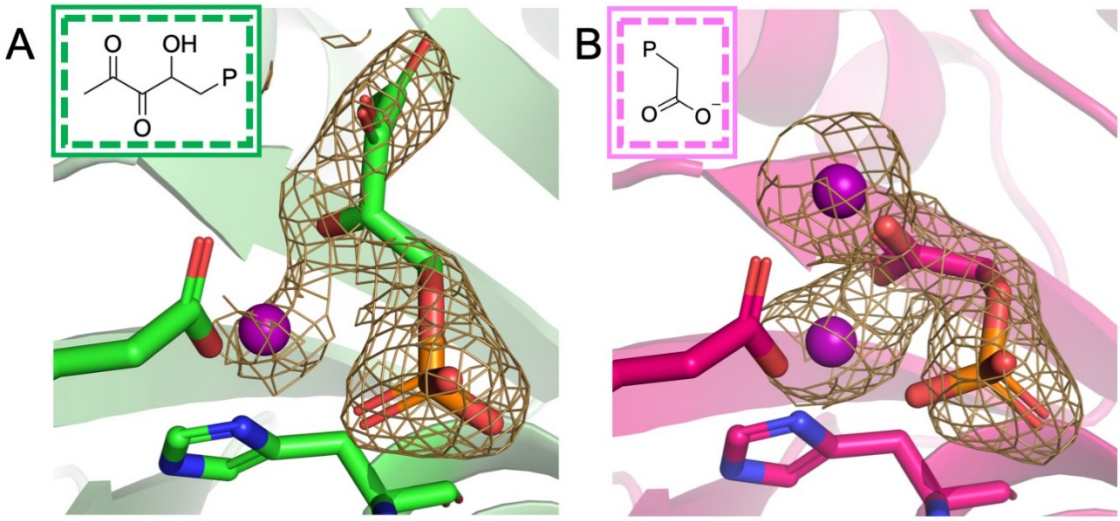


Figure 4.4. Simulated annealing composite omit map of RibB structures with intermediates. (A) Active site modeled with the intermediate 1 (green box), structure has one Mn^{2+} ion with map set to 1.5 sigma. (B) Site modeled with the 2-phosphoglycolic acid, one of two fragments in the fragmentation mechanism, map set to 2.0 sigma. Unlike in (A) this structure has disordered loop 2 (residues 87-96) and it has an additional Mn^{2+} binding in the previously proposed second metal binding site (see Chapter 3).

structure (**Figure 3.6**). This data can be used as another form of evidence for frequency of metal binding in the second binding site (**Figure 3.8**).

4.3.3 Single turnover NMR supports structural data

Uniformly ^{13}C labeled Ru5P was used as a substrate for single turnover experiments with RibB purified in 10% glycerol at 4 degrees. At specific time points the reactions were quenched with acid and analyzed using ^{13}C NMR. The synthesis of the ^{13}C labeled substrate, generated using pentose phosphate pathway enzymes as described in methods, and product formation by RibB were confirmed by comparison to previously published peak assignments [9]. The most prominent peaks in the spectrum are the peaks associated with carbon C2 ((t), 212ppm) represented by a purple dot in **Figure 4.5,** and the formate product peak ((s), 171 ppm), represented by a blue dot in **Figure 4.5**. As it has been previously reported, we have also observed that assignment and peak intensity of peaks at carbon C3 and C5 were dependent on percent of D_2O available in the solution. Peaks at other carbons can only be seen if the D_2O concentration is less than 5% [9].

With timepoints to five minutes, we did not observe the formation of the formate peak. Instead we saw appearance of an intermediate 1, in the range of 212 ppm. A peak assigned as a triplet peak corresponds to carbon C2, **Figure 4.5**. Over the course of 5 minutes, this peak is shifted by 0.7 ppm upfield with intervening quartets as intermediate is formed. The generation of this new triplet peak is due to removal of the hydroxyl at the carbon C1. The data also show the formation of doublet at 166.2 ppm

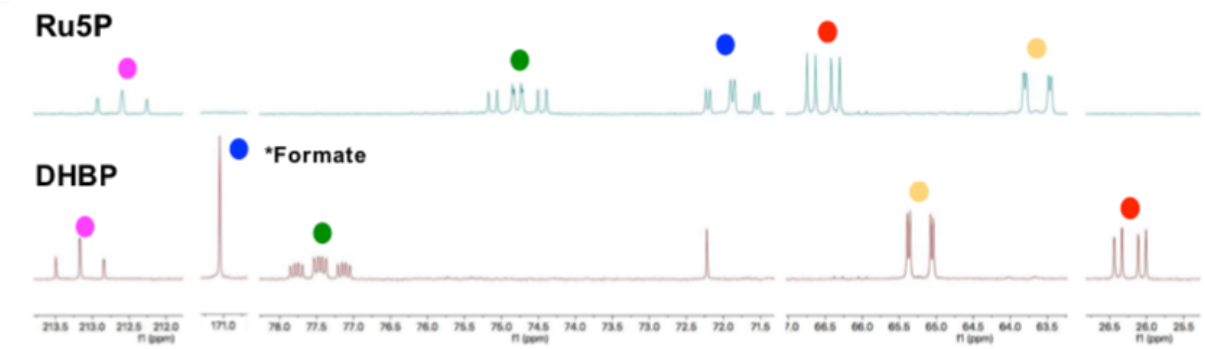


Figure 4.5. ^{13}C decoupled NMR spectra of Ru5P (top) and the product (bottom) in 2%D_2O.

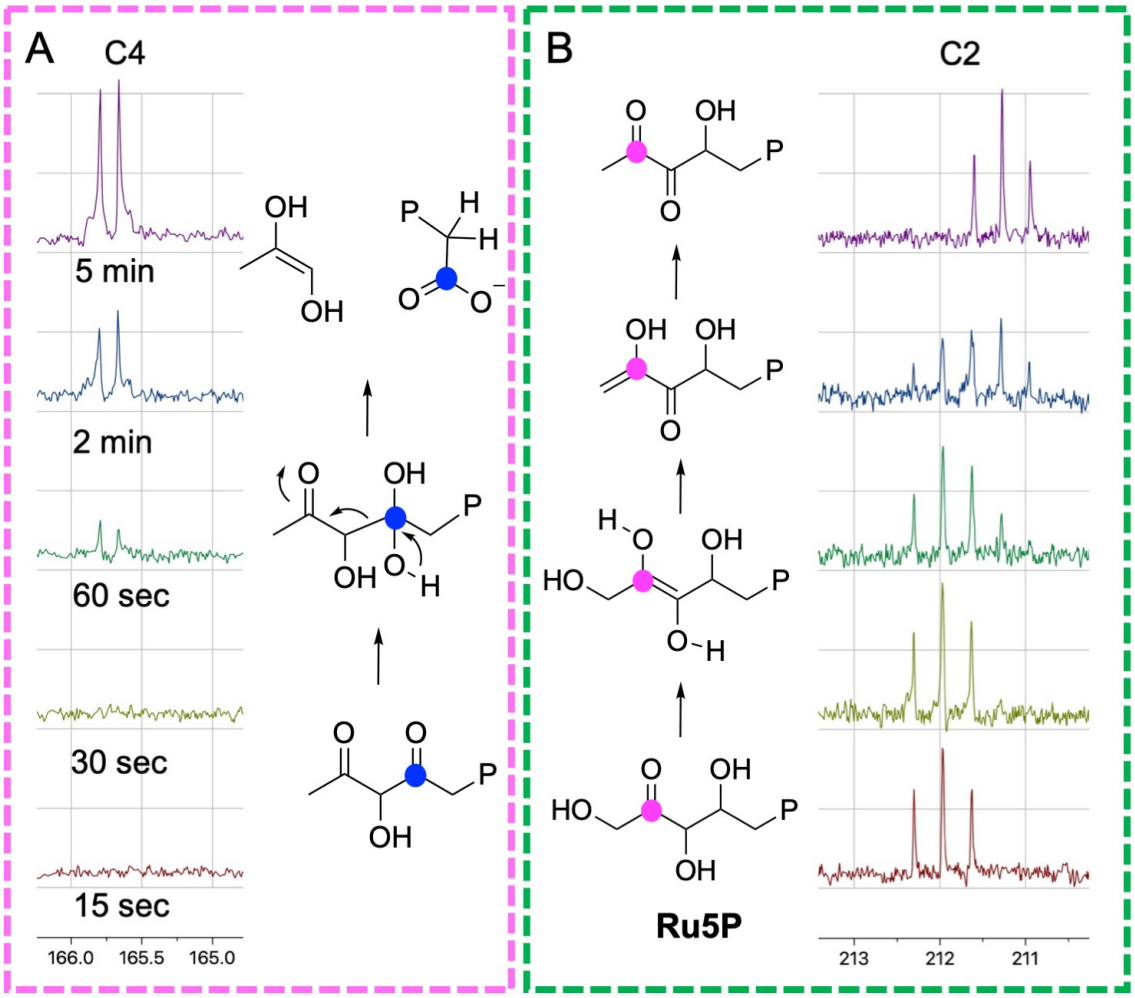


Figure 4.6. Single turnover NMR provides additional evidence of fragmentation mechanism. RibB catalyzed reaction in NMR tube, under single turnover conditions in presence of VcRibB in 10% glycerol at 4°C with acid quench. (A) Data shows the appearance of doublet after 60 sec in the two proposed reaction mechanisms. Only the possible intermediate with this NMR profile would be Intermediate 2. (B) Also, the dehydration reaction at the carbon C2 is demonstrated as the gradual shift of the triplet from 212 to 211 ppm.

(Figure 4.6). Based on the chemical properties we know that this peak has an adjacent carbon atom with carboxy group or equivalent (150 ppm or higher). Therefore, this doublet represents the 2-phoshoglycolic intermediate from the fragmentation mechanism also observed in the crystal structure (intermediate 2).

4.4 DISCUSSION

RibB catalyzes the removal of the carbon C4 of the substrate Ru5P as formate to make the DHBP product. There are competing models that have been described for the related enzymes such as DXP reductoisomerase enzyme in terpene biosynthesis [5, 19], and isomerase enzymes such as D-ribulose 5-phosphate isomerase [3, 4] or D-xylose isomerase[1, 2]. These models include a skeletal rearrangement model (mechanism involving 1,2-methyl shift) and a fragmentation model (fragment generation after retro-aldol reaction). These enzymes all catalyze a reaction on a structurally similar substrate. The substrates DXP and ribulose 5-phosphate differ in functional groups on C1 and the stereochemistry at C3, while D-Ru5P and D-Xy5P are epimers at carbon C3. However, the reactions catalyzed by all of these enzymes are different. While RibB reactions catalyzes the metal-dependent removal of formate, the reductoisomerases perform an NADPH assisted reduction of the intermediate in the final step, and the isomerases catalyze change of the stereochemistry of the substrate. Nevertheless, the skeletal rearrangement or canonical mechanism has been established in the literature as the model that describes the RibB catalysis. Structural and NMR data we present here shows the presence of 2-hydroxy-3,4-dioxopentyl phosphate (green box, **Figure**

4.2) and a 2-phosphoglycolate intermediates (pink box, **Figure 4.2**). Evidence provided suggests that fragmentation mechanism, not canonical mechanism is most probable mechanism followed by RibB.

Intermediate 1, 2-hydroxy-3,4-dioxopentyl phosphate, is the last common intermediate between the two proposed models (**Figure 4.2**). The structure of RibB with bound intermediate 1 identifies residues that are involved in the dehydration of the substrate at carbon C1 (red dot, **Figure 4.2**). Involvement of Glu175 and Cys68 has been previously discussed [24]. According to this mechanism, Liao *et al*. suggest these two residues (Glu174, Cys66) are involved in initial enolization and dehydration step of 1,2 skeletal rearrangement mechanism. Furthermore, this mechanism suggested that the imidazole side chain of opposing monomer, His136 (identical to His 137, Chapter 3), is used as a proton donor during the enolization step of the reaction. However, extensive mutagenesis studies on RibB active site residues suggest that any change to Glu39 or Glu175 completely abolished the activity of the enzyme. Also, mutation of Cys68 or His137 to serine lead to decrease in enzyme activity by 3 and 7 times, respectively [25, 26]. Here we provide an alternative method for formation of intermediate 1 (**Figure 4.3**). Catalysis starts with Glu175 abstracting a proton from carbon C3, 2.6 Å away. The loss of the proton promotes enolization and elimination of water generated at C1 with a proton donated by Cys68. In last two steps Glu175 and Cys68 are regenerated to make intermediate 1 (**Figure 4.3**). It is interesting to note that in the next step of the mechanism a water molecule is required for the hydroxylation at the carbon C4 (**Figure 4.2**). A water molecule (W2) in evident in intermediate 1 structure (**Figure 4.3**) located ideally beside the carbon C4. If the next step in the mechanism is hydration of C4, it

comes as no surprise that water W2 is coordinated to the Mn^{2+}. This provides additional evidence for the fragmentation mechanism, since hydration occurs after the skeletal rearrangement in the canonical mechanism. After hydration at carbon C4 the fragmentation mechanism proposes formation of a β-hydroxy ketone intermediate that, after retro-aldol reaction, produces two fragments. We have visualized one of the fragments, 2-phosphoglycolic acid, in the crystal structure. The single turnover ^{13}C NMR corroborate this hypothesis with the appearance of a doublet at 166ppm. In accordance with the fragmentation mechanism, the next step in the reaction would require the nucleophilic attack by carbon C5 (yellow, **Figure 4.2**) to join the fragments. However, in the binding pose in the structure, phosphoglycolate is oriented in catalytically non-covalent conformation. This would indicate that this fragment will have to go through conformational change or repositioning in order for reaction to proceed, which could be the results of the catalytically irrelevant Mn^{2+} ion altering the binding mode as concluded by the disordered loop.

We have presented considerable evidence that indicate a new mechanism is required to describe the chemistry of the RibB enzyme. We are currently optimizing the NMR experiment to identify all the intermediates in real time (not using quenching experiments). We are also derivatizing the 2-phosphoglycolic acid intermediate for detection using the mass spectrometer.

4.5 REFFERENCES

1. Allen, K.N., et al., *Isotopic exchange plus substrate and inhibition kinetics of D-xylose isomerase do not support a proton-transfer mechanism.* Biochemistry, 1994. **33**(6): p. 1481-7.
2. Allen, K.N., et al., *Role of the divalent metal ion in sugar binding, ring opening, and isomerization by D-xylose isomerase: replacement of a catalytic metal by an amino acid.* Biochemistry, 1994. **33**(6): p. 1488-94.
3. Lee, L.V., et al., *Role of metal ions in the reaction catalyzed by L-ribulose-5-phosphate 4-epimerase.* Biochemistry, 2000. **39**(16): p. 4821-30.
4. Lee, L.V., M.V. Vu, and W.W. Cleland, *13C and deuterium isotope effects suggest an aldol cleavage mechanism for L-ribulose-5-phosphate 4-epimerase.* Biochemistry, 2000. **39**(16): p. 4808-20.
5. Murkin, A.S., K.A. Manning, and S.A. Kholodar, *Mechanism and inhibition of 1-deoxy-D-xylulose-5-phosphate reductoisomerase.* Bioorg Chem, 2014. **57**: p. 171-85.
6. Arigoni, D., et al., *Terpenoid biosynthesis from 1-deoxy-D-xylulose in higher plants by intramolecular skeletal rearrangement.* Proc Natl Acad Sci U S A, 1997. **94**(20): p. 10600-5.
7. Bacher, A., Q. Le Van, and M. Buhler, *Biosynthesis of riboflavin. Incorporation of D-[1-13C]RIbose* Journal of American Chemical Society, 1982. **104**: p. 3753 - 3755.
8. Volk, R. and A. Bacher, *Biosynthesis of Riboflavin - the Structure of the 4-Carbon Precursor.* Journal of the American Chemical Society, 1988. **110**(11): p. 3651-3653.
9. Volk, R. and A. Bacher, *Biosynthesis of riboflavin. Studies on the mechanism of L-3,4-dihydroxy-2-butanone 4-phosphate synthase.* J Biol Chem, 1991. **266**(31): p. 20610-8.
10. Chen, Y.Y., et al., *Conformational changes associated with cofactor/substrate binding of 6-phosphogluconate dehydrogenase from Escherichia coli and Klebsiella pneumoniae: Implications for enzyme mechanism.* J Struct Biol, 2010. **169**(1): p. 25-35.
11. Richter, G., et al., *Biosynthesis of riboflavin: 3,4-dihydroxy-2-butanone-4-phosphate synthase.* Methods Enzymol, 1997. **280**: p. 374-82.
12. Picollelli, M.A., P.V. Viitanen, and D.B. Jordan, *Spectrophotometric determination of 3, 4-dihydroxy-2-butanone-4-phosphate synthase activity.* Anal Biochem, 2000. **287**(2): p. 347-9.
13. Soltis, S.M., et al., *New paradigm for macromolecular crystallography experiments at SSRL: automated crystal screening and remote data collection.* Acta Crystallographica Section D-Biological Crystallography, 2008. **64**: p. 1210-1221.
14. McPhillips, T.M., et al., *Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines.* Journal of Synchrotron Radiation, 2002. **9**: p. 401-406.
15. Kabsch, W., *Integration, scaling, space-group assignment and post-refinement.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 133-44.
16. McCoy, A.J., et al., *Phaser crystallographic software.* J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.
17. Islam, Z., et al., *Structural basis for competitive inhibition of 3,4-dihydroxy-2-butanone-4-phosphate synthase from Vibrio cholerae.* J Biol Chem, 2015. **290**(18): p. 11293-308.
18. Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.
19. Afonine, P.V., et al., *Towards automated crystallographic structure refinement with phenix.refine.* Acta Crystallogr D Biol Crystallogr, 2012. **68**(Pt 4): p. 352-67.

20. 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.
21. Moriarty, N.W., R.W. Grosse-Kunstleve, and P.D. Adams, *electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation.* Acta Crystallogr D Biol Crystallogr, 2009. **65**(Pt 10): p. 1074-80.
22. Moriarty, N.W., E.J. Draizen, and P.D. Adams, *An editor for the generation and customization of geometry restraints.* Acta Crystallogr D Struct Biol, 2017. **73**(Pt 2): p. 123-130.
23. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 1): p. 12-21.
24. Liao, D.I., et al., *Structural definition of the active site and catalytic mechanism of 3,4-dihydroxy-2-butanone-4-phosphate synthase.* Biochemistry, 2002. **41**(6): p. 1795-806.
25. Fischer, M., et al., *Biosynthesis of riboflavin in archaea studies on the mechanism of 3,4-dihydroxy-2-butanone-4-phosphate synthase of Methanococcus jannaschii.* J Biol Chem, 2002. **277**(44): p. 41410-6.
26. Steinbacher, S., et al., *Structure of 3,4-dihydroxy-2-butanone 4-phosphate synthase from Methanococcus jannaschii in complex with divalent metal ions and the substrate ribulose 5-phosphate: implications for the catalytic mechanism.* J Biol Chem, 2003. **278**(43): p. 42256-65.

CHAPTER 5:

5.1 CONCLUSIONS

In bacterial cells, formate can be recycled and reused by one-carbon metabolism or treated as a byproduct of primary metabolic pathways. In one-carbon metabolism, a folate cofactor, N^{10}-formyl-THF, is used to transport a formyl group and incorporate it into purine metabolism, protein synthesis, and biosynthesis of secondary metabolites. Enzymes that facilitate incorporation of formate from folate cofactors are known as transformylases. Here, I describe structural and functional studies of two formate-utilizing enzymes: PvdF, a new class of transformylase from Pseudomonas aeruginosa, and RibB enzyme from riboflavin biosynthesis.

5.1.1 AICAR and GAR transformylases.

AICAR transformylases and GART are prime examples of the utilization of the folate cofactor N^{10}-formyl-THF in the incorporation of one carbon unit into a variety of metabolites. In chapter 1, I have outlined differences between these two enzymes classes. Also, I provide examples of other kinds of GART-like enzymes and their function in protein synthesis (MTF) [1, 2], N-sugar transformylation [3], and in NRPS peptide assembly [4]. It is very evident that all of these enzymes share a common transformylase fold that is responsible for the movement of formate from the folate cofactor to the substrate. In all of these enzymes, the N-terminal domain “additions” are structurally different but serve a common purpose to stabilize and orient the substrate before the formylation occurs. It was therefore suggested these enzymes are distantly

related through gene transfer [4]. AICAR are significantly different from GART enzymes. These transformylases can only be found as part of nucleotide biosynthesis and follow a transformylation mechanism that is distinct from GART enzymes. These enzymes also do not share any structural resemblance as AICAR transformylases have a different fold and are part of larger two domain proteins.

5.1.2 PvdF hydroxyornithine transformylase from P.aeruginosa.

Iron acquisition is important for virulence in *Pseudomonas aeruginosa,* an opportunistic, ESKAPE pathogen. Secretion of siderophores is one method to obtain iron that is necessary for bacterial survival in an iron-limiting environment [5-7]. In the mixed-type siderophore pyoverdin, iron is chelated in an octahedral manner with two hydroxamate groups from fOHOrn and catecholate of the chromophore. In chapter 2, I explored structural and functional characteristics of the enzyme responsible for hydroxamate group formation in pyoverdin, PvdF. Here I have demonstrated that PvdF catalyzes the formate group transfer from N^{10}-formyl-THF analogue to the substrate. Structurally, however, PvdF has several long insertions that together account for over 100 of the 275 amino acids that we hypothesize are responsible for the substrate and cofactor stabilization. As these elements are not present in any other transformylase, PvdF represents a unique class of transformylase enzyme. In the future, it would be interesting to see if our hypothesis on function of structural insertions is correct. This can be further investigated in PvdF structural studies. Structure of PvdF with substrate and cofactor bound in the active site can elucidate more not just about structural insertions but also about enzyme mechanism. PvdF can now also be used as model to

solve other hydroxyornithine transformylases such as Rft from rhodochelin biosynthesis and CchA from coelichelin biosynthesis [8]. The knowledge gained in the study of PvdF, its unique structural and functional features, can be an important next step toward better rational drug design, especially considering the wealth of antifolate inhibitors at our disposal.

5.1.3 3,4-dihydroxy-2-butanone 4-phosphate synthase (RibB) of riboflavin biosynthesis. In mammals, riboflavin, or vitamin B2, serves numerous functions. In one of them riboflavin is utilized as the main precursor to FAD and FMN, cofactors in the electron transport chain [9-14]. Unfortunately, mammals cannot synthesize this vitamin and have to obtain it through external sources. Plants, bacteria and lower eukaryotes do synthesize riboflavin. This is the reason why further elucidation of enzymes in the riboflavin biosynthetic pathway can be used as potential antibacterial or antimicrobial targets [15-17]. In this study, we investigated a fascinating reaction catalyzed by the enzyme RibB, a rate-limiting enzyme in riboflavin biosynthesis. Literature and research presented over the past couple of decades described RibB as a magnesium-dependent enzyme that required two metal centers for catalysis. Consensus in the literature also states that the probable enzyme mechanism follows 1,2 methyl-shift, a skeletal rearrangement mechanism that facilitates carbon C4 removal by making a new carbon-carbon bond between carbon C3 and C5 of Ru5P [18-22]. Evidence shown in these studies are contrary to results I present in chapters 3 and 4. Instead of two metals, RibB binds one metal and requires only one metal for catalysis. Stoichiometry evidence, using activity and binding studies, EPR, and structural data clearly indicate that the

RibB catalytic cycle requires one metal for the activity by a reaction mechanism that is distinctly dissimilar to a 1,2-methyl shift (Chapter 3). This new mechanism, referred to as a fragmentation mechanism, yields two fragments of Ru5P during catalysis. One of these fragments, 2-phosphoglycolic acid, was captured during single turnover NMR studies as well as in the crystal structure. Positioning of this intermediate in the crystal structure does not provide any further clues on the next catalytic step of the reaction. This could mean that a significant change in the orientation of the fragment is necessary for further catalysis. Also, the absence of a second enediolate fragment and disordered loop 2 might suggest a potential function of this loop (Chapter 4). This enigma could be resolved with a crystal structure with both fragment intermediates bound, SAR by NMR, or further single turnover NMR studies with individually ^{13}C labeled substrates.

5.2 REFFERENCES

1. Schmitt, E., S. Blanquet, and Y. Mechulam, *Structure of crystalline Escherichia coli methionyl-tRNA(f)Met formyltransferase: comparison with glycinamide ribonucleotide formyltransferase.* EMBO J, 1996. **15**(17): p. 4749-58.
2. Schmitt, E., et al., *Crystal structure of methionyl-tRNAfMet transformylase complexed with the initiator formyl-methionyl-tRNAfMet.* EMBO J, 1998. **17**(23): p. 6819-26.
3. Holden, H.M., J.B. Thoden, and M. Gilbert, *Enzymes required for the biosynthesis of N-formylated sugars.* Curr Opin Struct Biol, 2016. **41**: p. 1-9.
4. Reimer, J.M., et al., *Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase.* Nature, 2016. **529**(7585): p. 239-42.
5. Meyer, J.M., *Pyoverdines: pigments, siderophores and potential taxonomic markers of fluorescent Pseudomonas species.* Arch Microbiol, 2000. **174**(3): p. 135-42.
6. Gulick, A.M., *Nonribosomal peptide synthetase biosynthetic clusters of ESKAPE pathogens.* Natural Product Reports, 2017. **34**(8): p. 981-1009.
7. Sadikot, R.T., et al., *Pathogen-host interactions in Pseudomonas aeruginosa pneumonia.* Am J Respir Crit Care Med, 2005. **171**(11): p. 1209-23.
8. Bosello, M., et al., *An enzymatic pathway for the biosynthesis of the formylhydroxyornithine required for rhodochelin iron coordination.* Biochemistry, 2012. **51**(14): p. 3059-66.
9. Bacher, A., et al., *Biosynthesis of vitamin B2 (riboflavin).* Annu Rev Nutr, 2000. **20**: p. 153-67.
10. Lakshmi, R., et al., *Effect of riboflavin or pyridoxine deficiency on inflammatory response.* Indian J Biochem Biophys, 1991. **28**(5-6): p. 481-4.
11. Powers, H., et al., *A proposed intestinal mechanism for the effect of riboflavin deficiency on iron loss in the rat.* British Journal of Nutrition, 1993. **69**(2): p. 553 - 561.
12. Schramm, M., et al., *Riboflavin (vitamin B2) deficiency impairs NADPH oxidase 2 (Nox2) priming and defense against Listeria monocytogenes.* Eur J Immunol, 2014. **44**(3): p. 728-41.
13. Thakur, K., et al., *Riboflavin and health: A review of recent human research.* Crit Rev Food Sci Nutr, 2017. **57**(17): p. 3650-3660.
14. Webster, R.P., M.D. Gawde, and R.K. Bhattacharya, *Modulation of carcinogen-induced DNA damage and repair enzyme activity by dietary riboflavin.* Cancer Lett, 1996. **98**(2): p. 129-35.
15. Long, Q., et al., *Riboflavin biosynthetic and regulatory factors as potential novel anti-infective drug targets.* Chem Biol Drug Des, 2010. **75**(4): p. 339-47.
16. Mack, M. and S. Grill, *Riboflavin analogs and inhibitors of riboflavin biosynthesis.* Appl Microbiol Biotechnol, 2006. **71**(3): p. 265-75.
17. Sassetti, C.M., D.H. Boyd, and E.J. Rubin, *Genes required for mycobacterial growth defined by high density mutagenesis.* Mol Microbiol, 2003. **48**(1): p. 77-84.
18. Liao, D.I., et al., *Crystal structure of 3,4-dihydroxy-2-butanone 4-phosphate synthase of riboflavin biosynthesis.* Structure, 2001. **9**(1): p. 11-8.
19. Liao, D.I., et al., *Structural definition of the active site and catalytic mechanism of 3,4-dihydroxy-2-butanone-4-phosphate synthase.* Biochemistry, 2002. **41**(6): p. 1795-806.
20. Steinbacher, S., et al., *Metal sites in 3,4-dihydroxy-2-butanone 4-phosphate synthase from Methanococcus jannaschii in complex with the substrate ribulose 5-phosphate.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 7): p. 1338-40.

21. Steinbacher, S., et al., *Structure of 3,4-dihydroxy-2-butanone 4-phosphate synthase from Methanococcus jannaschii in complex with divalent metal ions and the substrate ribulose 5-phosphate: implications for the catalytic mechanism.* J Biol Chem, 2003. **278**(43): p. 42256-65.
22. Islam, Z., et al., *Structural basis for competitive inhibition of 3,4-dihydroxy-2-butanone-4-phosphate synthase from Vibrio cholerae.* J Biol Chem, 2015. **290**(18): p. 11293-308.