# BIOSYNTHESIS OF THE AMINOPROPANOL MOIETY OF COBAMIDES BY BACTERIAL AND ARCHAEAL ENZYMES

by

NORBERT KEITH TAVARES

(Under the Direction of Jorge C. Escalante-Semerena)

## ABSTRACT

Adenosylcobalamin (Coenzyme $B_{12}$, AdoCbl) is a cofactor that is essential for organisms in all three domains of life, but is only synthesized by bacteria and archaea. While much is known about the biosynthesis of AdoCbl, many gaps remain. There is a separate aerobic and anaerobic pathway for the synthesis of the corrinoid ring of AdoCbl. These pathways involve a different set of enzymes and exist separately in different prokaryotic organisms. The pathways converge with the late steps where homologous enzymes carry out several, but not all steps. The late steps of the anaerobic pathway has been well studies in *Salmonella enterica* sv Typhimurium LT2. Homologous enzymes have not been identified for some of the late steps of the aerobic pathway, or within archaea or hyperthermophiles, which use the anaerobic pathway. Among these yet-to-be identified enzymes are the Adenosylcobalamin-*5'*-phosphate phosphatase (EC 3.1.3.73) (CobC in *S.* Typhimurium, CobZ in archaea), which catalyze the final step, and the L-threonine kinase (EC 2.7.1.177) (PduX in *S.* Typhimurium), which catalyzes the first step in the synthesis of the 1-amino-2-propanol-*O*-2-phosphate (AP-P) nucleotide linker moiety of AdoCbl.

Here we identify an enzyme named BluE, a L-threonine (L-Thr) kinase in *Rhodobacter sphaeroides,* an organism that synthesizes AdoCbl via the aerobic pathway. We show that *Rs*BluE has L-Thr kinase activity *in vitro* and complements a *S.* Typhimurium *pduX* strain *in vivo*. Phylogenic analysis

of *bluE* shows it is restricted to a few Rhodobacterales that have a strict requirement for the cobalamin with 5,6-dimethylbenzimidazole as the lower ligand and AP-P as the nucleotide linker.

To date no archaeal genomes have been found to encode *pduX* or *bluE* homologues. Investigations of the unusual archaeal L-threonine-*O*-3-phosphate (L-Thr-P) decarboxylase (CobD, EC 4.1.1.81) from *Methanosarcina mazei* revealed that this enzyme is bifunctional, conferring L-Thr kinase activity in addition to the expected L-Thr-P decarboxylation. *Mm*CobD has a $[4Fe\text{-}4S]^{2+}$ cluster domain that appears to have a regulatory function for both enzymatic activities. Bioinformatics analyses suggest that many organisms that lack *pduX* or *bluE* homologues encode a CobD with a slightly longer *N*-terminus similar to that of *Mm*CobD. This region might encode all or part of a motif for the L-Thr kinase activity.

INDEX WORDS: Cobamide, Adenosylcobalamin, $B_{12}$, *Salmonella*, *Rhodobacter,* 1-amono-2-propanol-*O*-2-phosphate

# BIOSYNTHESIS OF THE AMINOPROPANOL MOIETY OF COBAMIDES BY BACTERIAL AND ARCHAEAL ENZYMES

by

NORBERT KEITH TAVARES

BS, Southern Connecticut State University, 2006

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2016

© 2016

Norbert Keith Tavares

All Rights Reserved

BIOSYNTHESIS OF THE AMINOPROPANOL MOIETY OF COBAMIDES BY BACTERIAL AND ARCHAEAL ENZYMES

by

NORBERT KEITH TAVARES

Major Professor: Jorge Escalante-Semerena

Committee: Diana Downs
Robert Maier
Timothy Hoover

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
August 2016

DEDICATION

This dissertation is dedicated to you.  Thanks for having a look.

ACKNOWLEDGEMENTS

I would like to acknowledge the institutions of United States of America for the quality academic training I've received, and my country of birth, Jamaica, where the soil was tilled for the seeds of my values, morality, and work ethic to be planted by my mother and grandmother. The influence, support, love, and guidance of my sisters Eleanor and Nicola, and my mother Andrea Angela Ruthann Mitchell Tavares have been of great importance in my life and during my graduate work. Without them I could not have accomplished this and it would have no value or meaning.

I was set on the path to a PhD by the guidance of my undergraduate advisor Dr. Nicholas Edgington at Southern Connecticut State University. I received my first training in research as a result of a NSF funded REU programs at the University of Wisconsin with Dr. Cameron Currie and at the University of Minnesota with Dr. Michal Sadowsky. These positive experiences exposed me to research and made graduate school possible.

Special thanks goes to my research advisor Dr. Jorge Escalante-Semerana for the years of support. He has contributed significantly to my development as a scientist, and a moral member of the scientific community. He has given me the freedom to explore scientific ideas and to think about science in creative ways. I owe much more to Dr. Escalante than can be expressed here.

I must acknowledge the University of Wisconsin, where I started my PhD and received the bulk of my training, and the University of Georgia where my degree was completed. The feedback from my committee members (Katherine McMahon, Paul Weimer, Douglas Weibel, Diana Downs, Robert Maier, Timothy Hoover, and Michael Adams) at both institutions was valuable.

I am privileged to have met many good friends in and outside of the lab. Their support and simply their presence have made life and work pleasant. Special thanks to labmates and friends Sabra Katz-Wise, Teresa Curtis, Alex Tucker, Chi Ho Chan, Ted Moore, Michael Gray, Heidi Crosby, Kristy Hentchel, Chelsey VanDrisse and members of the Downs lab for technical help, support and most

importantly making the lab a pleasant place to work. Special acknowledgement goes to my wife and partner Paula Pappalardo for all her support, encouragement, love, and useful skills with R programming.

I would like to acknowledge the role played by chance and simple luck. It was by chance that I was born in a particular place, at a particular time in history, to my particular parents, and this has made all the difference in the world. Many other aspects of my life are also a result of chance and luck, which has afforded me the opportunity to accomplish the things I have, including the completion of this PhD. Though much hard work was involved, I must acknowledge that luck too helped in the creation of the work described in this document.

TABLE OF CONTENTS

Page

CHAPTER 1

# INTRODUCTION AND LITERATURE REVIEW

## OVERVIEW OF CORRINOIDS AND COBAMIDES

Corrinoids are members of the cyclic tetrapyrrole family of molecules that include Factor $F_{430}$, chlorophyll, and heme. Corrinoids contain cobalt as the central metal ion, while heme, chlorophyll, and factor $F_{430}$ contain iron, magnesium and nickel, respectively (1, 2). The cyclic tetrapyrrole of corrinoids is referred to as the corrin ring. Corrinoids are unique among the tetrapyrroles for having two axial ligands coordinated to the central cobalt ion. The upper ligand (*Co*β) can be a methyl (MeCba), hydroxyl (OHCba), cynao (CNCba, vitamin $B_{12}$), or *5'*-deoxyadenosine (AdoCba) group. The lower ligand (*Co*α) varies a great deal among organisms (3-5). Naturally occurring corrinoids have a nucleotide base that is a benzimidazole, purine, or phenol, linked to the corrin ring via an α-*N*-glycosidic bond. When a corrinoid lacks the lower ligand it is referred to as an incomplete corrinoid. A corrinoid with a lower ligand is known as a complete corrinoid or a complete cobamide (Cba) (6). Cobalamin (Cbl, vitamin-$B_{12}$) is a Cba with the purine analog *5,6*-dimethylbenzamidazole (DMB) as the base of the nucleotide in the lower ligand. The coenzymatic form (AdoCbl) of Cbl has a *5'*-deoxyadenosine as the upper axial ligand. Figure 1.1 shows the structure of AdoCbl.

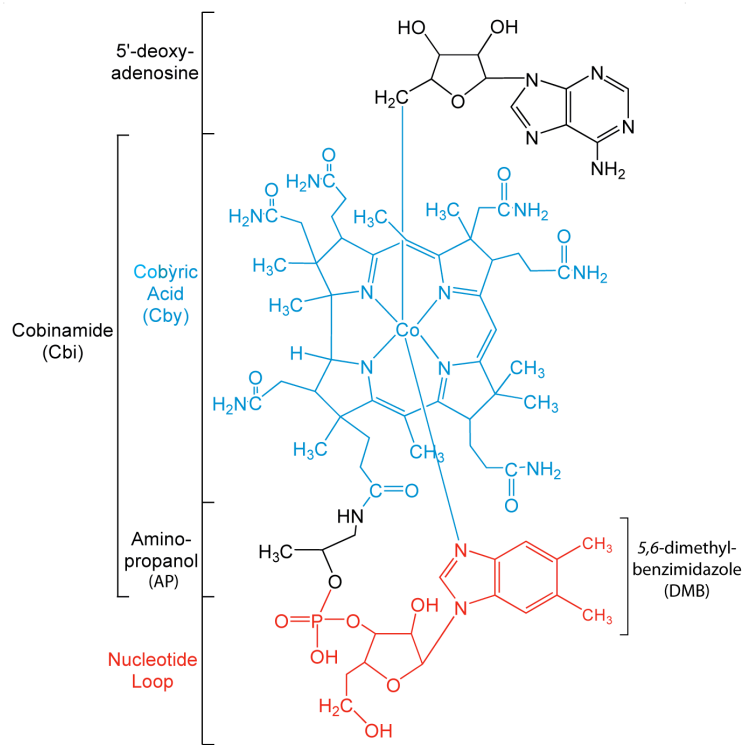



**Figure 1.1 Structure of Adenosylcobalamin (AdoCbl).** Relevant components of the molecule are bracketed and labeled. The coenzymatic for of AdoCbl has 5'deoxyadenosine (black) as the upper axial ligand. The aminopropanol (AP) moiety is assembled by PduX and CobD and attached to the Cby (blue) moiety by CbiB to form Cbi-P. The α-riboside nucleotide loop component in red is assembled by CobUSTC.

AdoCbl is the largest coenzyme known (1579 amu). It is biosynthesized exclusively by bacteria and archaea (6), and is used in many different metabolic processes including the synthesis of methionine (7, 8), methanogenisis (9), acetogenisis (4, 10, 11), catabolism of ethanolamine (12-16), 1,2-propaindiol (17-21), and short chain fatty acids (22-24), anaerobic glycerol (17, 25) and glutamate catabolism (26-28), organohalide respiration (29-34), and ribonucleotide metabolism (35-40). It is an essential nutrient for a number of eukaryotes, including mammals, which must acquire Cbl from their diet for use by Cbl-dependent methionine synthase (MTR, MetH) and methylmalonyl-CoA mutase (MCM) enzymes (35, 41, 42). Eukaryotes can only use Cbas with DMB as the lower ligand base (43-46). As a coenzyme, AdoCbl catalyzes carbon skeleton rearrangements using radical chemistry (42, 47). In the reductive dehalogenation reactions the Cba forms a cobalt-halogen bond to reduce the organohalide (48), but the

complete mechanism of catalysis is not yet fully understood. MeCbl functions as a cosubstrate in methyltransferase reactions (41).

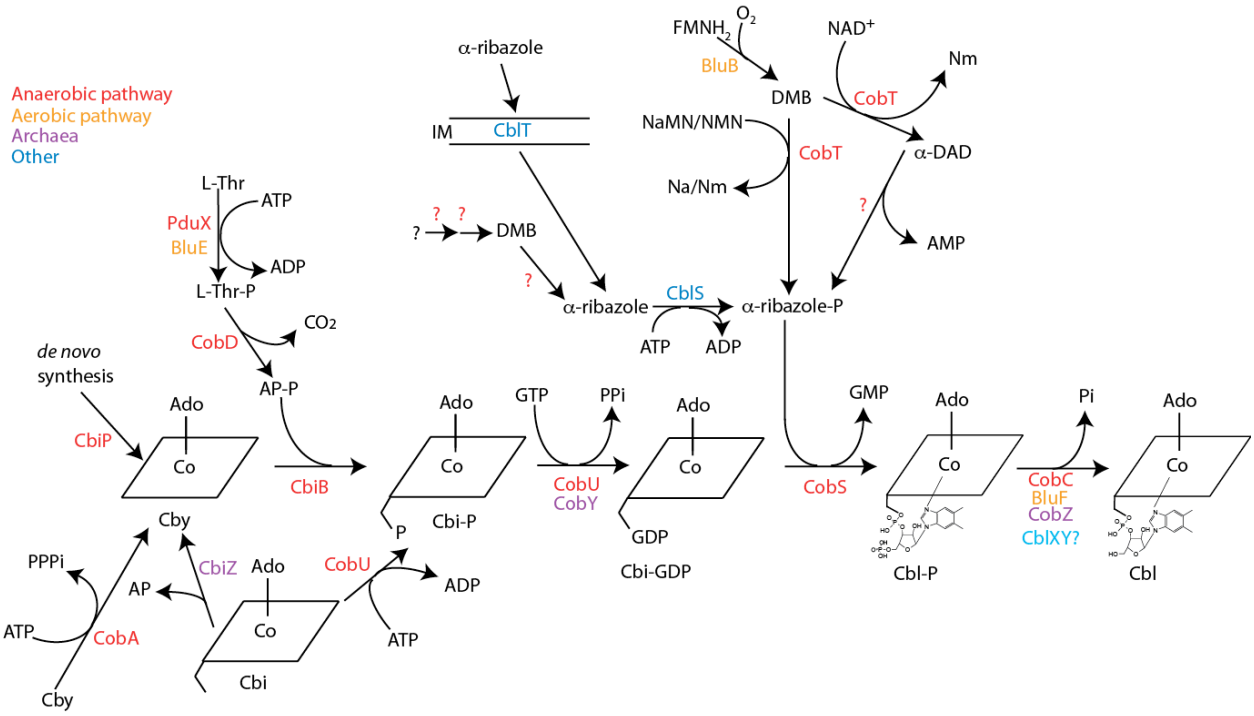



**Figure 1.2 Combined Nucleotide Loop Assembly (NLA) or late step of AdoCbl biosynthesis.** Enzymes labeled red are found in organism with use the early cobalt insertion /anaerobic pathway utilized by *S.* Typhimurium and are named using the *S.* Typhimurium enzyme naming convention. Enzymes labeled in yellow are found in only in organism that use the late cobalt insertion/aerobic pathway and are named using the *R. sphaeroides*/*R. capsulatus* naming convention. Enzymes labeled in purple are found only in archaea and are non-orthologous replacements for the equivalent enzyme function in *S. Typhimurium*. CbiZ is found in both bacteria and archaea. Enzymes colored in blue are found in other Cba producing bacteria and appear to be restricted to specific orders but not to the late or early cobalt insertion pathways. Ado, 5'-deoxyadenosine; Cby, cobyric acid; Cbi, cobinamide; Cbi-P, cobinamide-phosphate; Cbi-DGP, cobinamide- guanosine diphosphate; Cbl-P, cobalamin-phosphate; Ado-Cbl, adenlsylcobalamin; GMP, guanosine monophosphate; DMB, *5,6*-dimethylbenzimadazole; Na, nicotinate; Nm, nicotinamide; NaMN, nicotinate mononucleotide; NAD+, nicotinate dinucleotide; NMN, nicotinamide mononucleotide; α-DAD, dimethylbenzimadazole adenine dinucleotide; $FMNH_2$, reduced flavin mononucleotide; AP, aminopropanol; AP-P aminopropanol phosphate; L-Thr, L-threonine; L-Thr-P L-threonine phosphate; α-ribozole, α-DMB-riboside; α-ribozole-P; α-DMB-riboside monophosphate.

## *S. TYPHIMURIUM* AS A MODEL ORGANISM FOR COBAMIDE BIOSYNTHESIS

*Salmonella enterica* sv Typhimurium LT2 has been used as a model organism for the study of AdoCba biosynthesis for many years. There are several advantages to using *S.* Typhimurium as a model organism for studying Cba biosynthesis, such as the ease and rapid growth when cultured in minimal medium, good genetic systems, and several Cba-dependent enzymes with which phenotyping can be

evaluated. The work presented here focuses on the last steps of the AdoCba biosynthesis pathway, known as the nucleotide loop assembly (NLA) (Fig. 1.2). In *S.* Typhimurium*, de novo* synthesis of the corrin ring (early steps) only occurs under strictly anaerobic conditions (49, 50). This allows the study of the NLA branch of the pathway under aerobic conditions without the need for the introduction of additional mutations in the genes of corrin ring synthesis pathway. Transport and salvaging of incomplete corrinoids (*e.g.* cobinamide (Cbi) and cobyric acid (Cby)) and complete cobamides (*e.g.* Cbl) occurs under both aerobic and anaerobic conditions (6), which allows feeding of intermediates for the phenotypic evaluation of specific mutations in the pathway. This approach along with mutagenesis and other genetic and biochemical techniques has enabled the identification of all the major enzymes involved in the NLA steps of the pathway in *S.* Typhimurium (51-64). As a result, *S.* Typhimurium is an excellent tool for heterologous expression of putative AdoCba biosynthetic genes from other organism. The enzymatic functions of several AdoCba biosynthetic genes from difficult-to-culture organisms have been identified in this manner (65-70).

*S.* Typhimurium possessed three cobalamin-dependent enzymes. Ethanolamine ammonia-lyase (EutBC, EC 4.3.1.7) (71) is required for the utilization ethanolamine (EA) (72-74), and 1,2-propanediol dehydratase (PduCD, EC 4.2.1.28 (75, 76) is necessary for growth on 1,2-propanediol (PD) (75, 77). Utilization of these two carbon sources requires the synthesis or supplementation of high concentrations of Cbl (78). *S.* Typhimurium also utilizes the Cba-dependent methionine synthase (MetH, EC 2.1.1.13), which synthesizes methionine from homocysteine and methyl-tetrahydrofolate (79, 80). *S.* Typhimurium also encodes a Cba-independent methionine synthase (MetE, 2.1.1.14) (81-83). An advantage of MetH is in the efficiency of the enzyme, which has turnover number 122 times greater than MetE (82). Picomolar concentrations of Cba are sufficient to meet the methionine synthesis needs of *S.* Typhimurium under laboratory conditions (84). The low (MetH) and high (EutBC, PduCD) Cba demands of these enzymes create a set of growth conditions which makes the assessment of Cba-dependent phenotypes simple and versatile in *S.* Typhimurium.

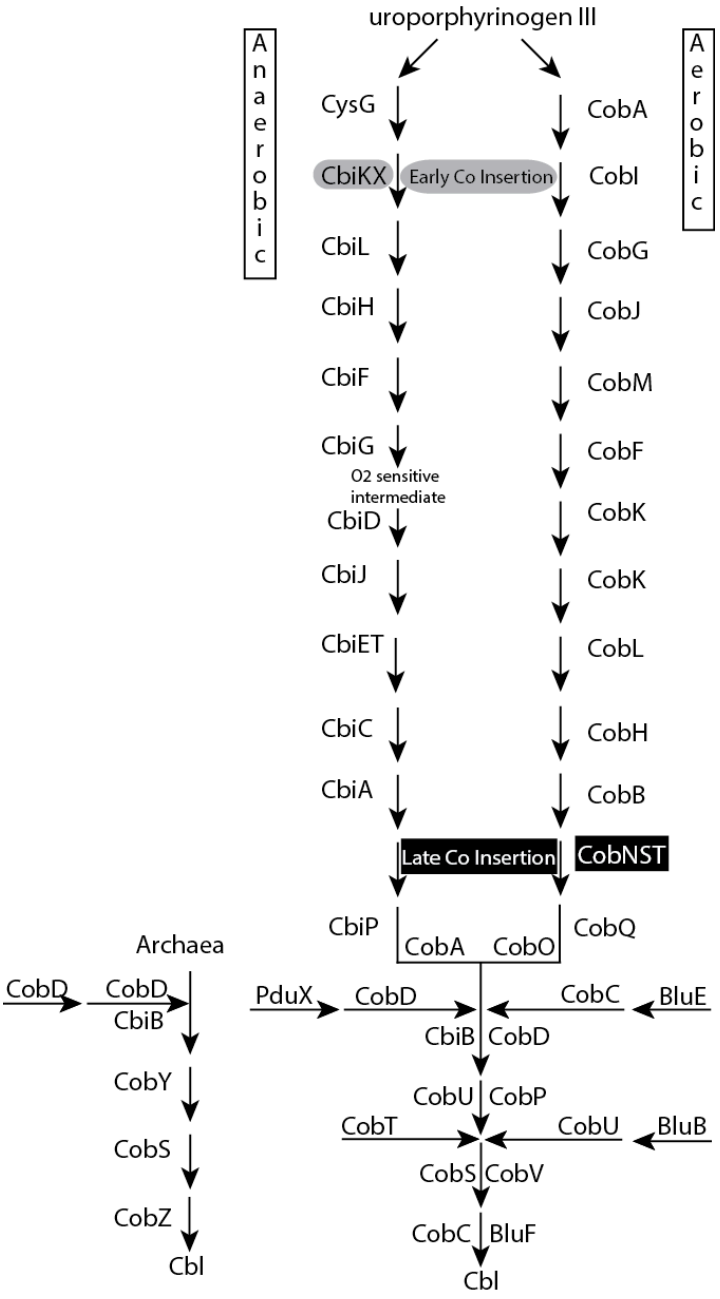



**Figure 1.3 Late and early cobalt insertion pathways for AdoCbl biosynthesis.** The late (aerobic) and early (anaerobic) cobalt insertion pathways starting with uroporphyrinogen III and converging into the shared nucleotide loop assembly (NLA) pathway to produce $B_{12}$ (Cbl). Archaea use the early insertion pathway but use several non-orthologues replacement enzymes for several steps of the NLA pathway. Only enzyme names are shown, substrates other start and end product are omitted. The oxygen sensitive intermediate step of the anaerobic pathway is shown. Cobaltochelatase responsible for insertion of Co in the early (gray oval) and late (black square) cobalt insertion pathways are highlighted. *S.* Typhimurium nomenclature is used for the early Co and NLA enzymes and *P. denitrificans*/*R. capsulatus* nomenclature is used for the late Co and corresponding NLA enzymes. *M. mazei* nomenclature is used for the archaeal genes.

**BIOSYNTHESIS OF COBAMIDES - MULTIPLE PATHWAYS**

The early steps of AdoCba biosynthesis are divided into two main pathways that are utilized independently in different AdoCba-producing prokaryotes. The main difference between the two pathways is the timing of the insertion of the central cobalt ion into the corrinoid ring (50, 85, 86) (Fig. 1.3). The early cobalt insertion pathway has an oxygen sensitive intermediate (87), and as a result, this pathway is referred to as the 'anaerobic pathway'. The anaerobic pathway is found in a diverse array of bacteria and archaea, but is best characterized in *S.* Typhimurium (51-64). The late cobalt insertion pathway is found primarily among the α-Proteobacteria, and while it is functional under aerobic or anaerobic conditions it is often referred to as the 'aerobic pathway'. The aerobic pathway is best characterized in *Pseudomonas denitrificans* (88-101) and *Rhodobacter capsulatus* (102-108). These two pathways contain many homologues but vary with regard to the cobaltochelatases responsible for cobalt insertion, CobS, CobT, and CobN (101) in the aerobic pathway, and CbiK, CbiX or CysG (109-111) in the anaerobic pathway, as well as the ring contraction enzymes CbiH (anaerobic), CobG and CobZ (aerobic) (106, 112, 113). These two pathways converge in that late steps referred to as the nucleotide loop assemble (NLA) (Fig. 1.2), which begins with the substrate cobyric acid (Cby) produced by the enzyme CbiB. For the most part, organisms that use either the late or early cobalt insertion pathway appeared to possess homologues of almost all 7 enzymes involved in the late steps of the pathway. With the availability of more sequenced genomes, we now know that AdoCba biosynthesis across many species of prokaryotes is much more complex than the simple two pathway model. Many AdoCba producers do not have homologues of proteins within either pathway, which suggest the existence of yet to be discovered enzymes and or branch pathways for AdoCba biosynthesis. However, several non-orthologous replacements for gene products characterized in the model Ado producers *S.* Typhimurium*, P. denitrificans*, and *R. capsulatus*, have been identified. For example, CobY is a Cbi GTP:adenosylcobinamide-phosphate guanylyltransferase in archaea which replaces CobU/CobP (Fig. 1.2) (65). The AdoCba-phosphate phosphatase (CobC in *S.* Typhimurium) has also been replaced by a non-orthologous enzyme named CobZ in archaea (66, 67). ArsAB is a heterodimeric protein found in strict

anaerobe *Sporomusa ovata* (114), which replaces the homodimeric CobT from *S.* Typhimurium, and specifically activates phenolic bases (114, 115), an activity that CobT cannot perform. Within some Firmicutes, CobT has been supplemented with or completely replaced by CblS, an α-ribazole (α-R) kinase (70). The presence of this enzyme along with an α-ribazole transporter (CblT) (70), suggests these organisms are capable of savaging sufficient α-R from the environment to satisfy their AdoCbl needs. This in turn suggests that α-R is being generated and released into the environment in sufficient quantities by other organisms that share the same ecological niches as these Firmicutes. This kind of metabolic interaction, cross-feeding, or potential symbiosis involving AdoCba biosynthesis has not been rigorously explored, but these findings suggest the likely existence of such relationships (reviewed (116)). The most likely enzymatic source of α-R in the environment is CbiZ. CbiZ from *R. sphaeroides* and *M. mazei* are amidohydrolases that recognizes Cba precursors such as, Cbi, or complete cobamides such as pseudo-Cba. CbiZ from these organisms cleave incomplete corrinoids and Cbas without DMB as the base to form the precursor Cby (69, 117-119). Cby then enters the AdoCba synthesis pathway further upstream to ensure the correct base, in this case DMB, is attached to the corrin ring. PseudoCbl is a Cba with an α-linked adenine base in place of DMB (120, 121). While the Cba-dependent enzymes in organisms like *S.* Typhimurium can use pseudoCbl, other Cba-dependent enzymes, such as methylmalonyl CoA mutase (MCM) found in humans and *R. sphaeroides,* can only use Cbas containing DMB as the base. CbiZ allows *R. sphaeroides* to salvage any available Cba in the environment and remodel it into a Cba with the correct base to suit its needs. The α-linked base is likely recycled internally or released back into the environment where it can be picked up by organisms containing a version of the α-ribazole transporter CblT, which may have specific affinities for particular bases (*e.g*. DMB, Ade, phenol, etc). The CbiZ from *R. sphaeroides* and *M. mazei* specifically cleave non-DMB bases, while CblT and CblS proteins studied to date have specific affinities for DMB-containing α-R. This would suggest the likely existence of a variant of CbiZ that cleaves Cbas with DMB as the base to generate the α-R, which is then salvaged by organisms with CblTS. Such a variant of CbiZ has yet to be identified. Figure 1.4 shows a model for the combined global AdoCba biosynthetic and salvaging pathways as of the date of this publication.

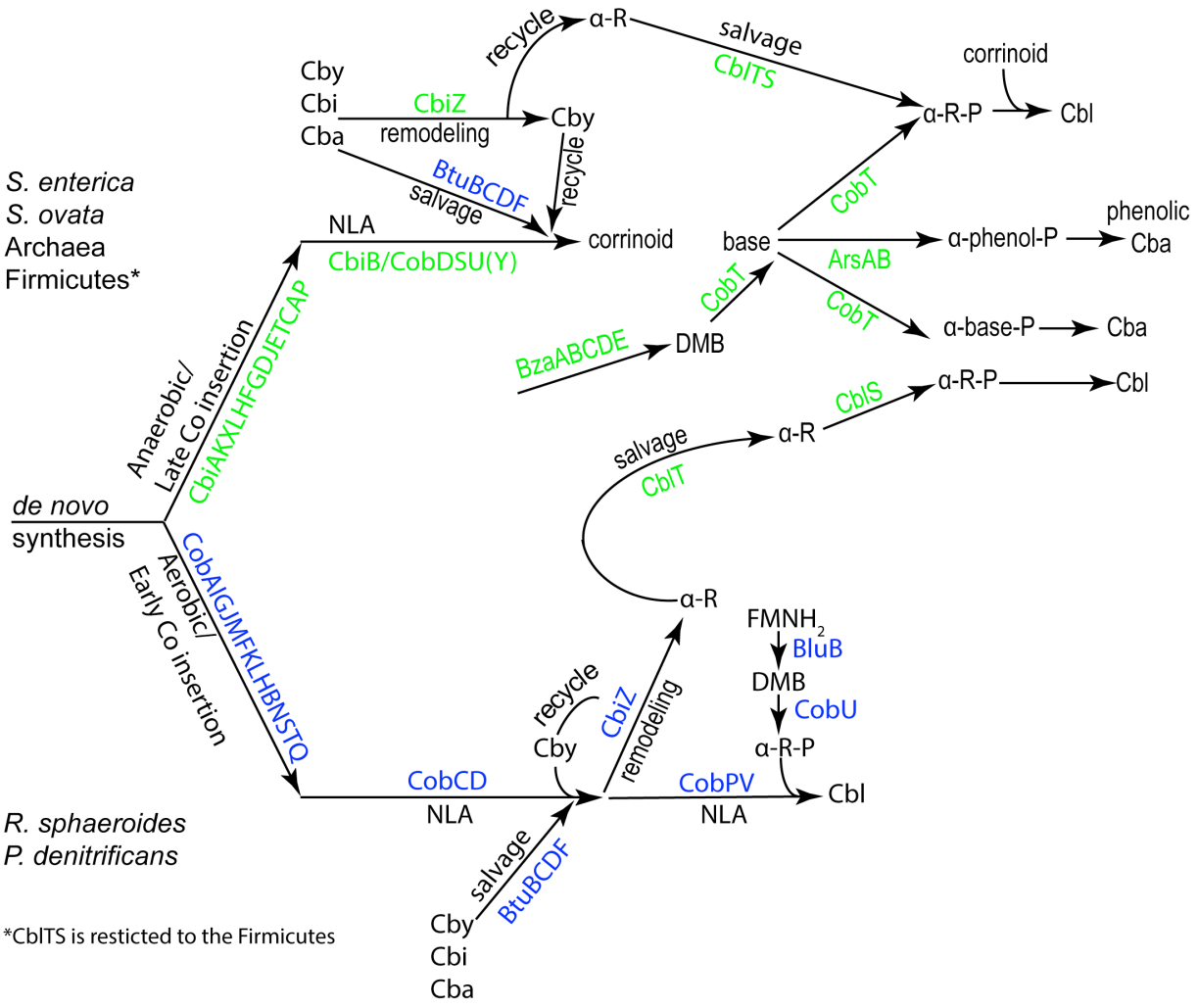



**Figure 1.4. Model of cobamide salvaging, recycling, remodeling, and base attachment pathways across various species of Cba producers.** Enzymes are colored blue. The anaerobic and aerobic *de novo* synthesis pathways are shown. Cross feeding occur with the salvage and recycling of cobamides by CbiZ to produce Cby, which is remodeled with a new α-ribotide containing the preferred base of the organism. The cleaved α-ribotide (e.g. α-R) is presumable released and salvaged by Firmicutes which posses the transporter CblT and α-R kinase CblS. *De novo* synthesis of α-ribotide from various bases is determined by the which NaMN:base phosphotransferase (EC 2.4.2.21) the organism possess; CobT, ArsAB, or CobU. NLA, nucleotide loop assembly; Cby, cobyric acid; Cbi, cobinamide; Cbl, cobalamin; Cba, cobamide; DMB, 5,6-dimethylbenzimadazole; $FMNH_2$, reduced flavin mononucleotide; α-ribazole, α-DMB-riboside; α-ribazole-P; α-DMB-riboside monophosphate, BzaABCDE; 5-hydroxybenzimidazole synthase, BtuBCDF; corrinoid transporter.

The enzymology of the early Co insertion pathway has been more fully characterized than the late insertion pathway because it has been extensity studied in the model organism *S.* Typhimurium. As a result, all of the primary enzymes involved in AdoCba biosynthesis have been identified in this organism.

While AdoCba biosynthesis has been studied in late Co insertion organisms such as *P. denitrificans* (88-101) and *R. sphaeroides* (117-119, 122), several major enzymes within the pathway are unknown. There are even greater gaps in our knowledge of AdoCba biosynthesis pathways of other bacterial phyla and archaea. Within this body of work is described the characterization of the several of these missing enzymes. Chapter 2 describes the characterization of BluE, the missing L-threonine (L-Thr) kinase of some Rhodobacterales. Chapter 3 and 4 presents data that show CobD in the methanogenic archaea *Methanosarcina mazei* is a bifunctional metalloenzyme with both L-Thr kinase and L-Thr-P decarboxylase activities.

**SYNTHESIS OF THE AMINOPROPANOL PHOSPHATE LINKER**

The focus of this thesis is on the synthesis if (*R*)-1-amino-2-propanol-*O*-2-phosphate (AP-P) and ethanolamine-*O*-phosphate (EA-P), which function as linkers in the attachment of the $N^I$-(5-phospho-α-D-ribosyl)-base (α-ribotide) to the corrinoid ring, Cby. The structural difference between AP-P and EA-P is the absence of the methyl group attached at carbon 176 (Fig. 1.5) (34). If EA-P is used as the linker the prefix "nor" is added to the name of the resulting cobamide, *e.g*. norCbl, when DMB is the base, norpseudoCbl (34) for Adenine, (Fig. 1.5) or norFactor A (3, 123-125) when 2-methyladenine is the base. In *S.* Typhimurium AP-P is generated by the activities of two enzymes. First the L-threonine kinase (EC 2.7.1.177), PduX, phosphorylates L-Thr to produce L-Thr-P (51). The product of this reaction is then decarboxylated by L-threonine-*O*-3-phosphate decarboxylase (EC 4.1.1.81), CobD, to produce (*R*)-1-amino-2-propanol *O*-2-phosphate (AP-P) (52). AP-P is then attached to Cby by cobinamide-phosphate synthase, CbiB (58) (Fig. 1.2). The L-threonine kinase in a number of organisms including archaea, thermophilic bacteria, Actinobacteria, and for organisms that use the late Co insertion pathway has not been identified.

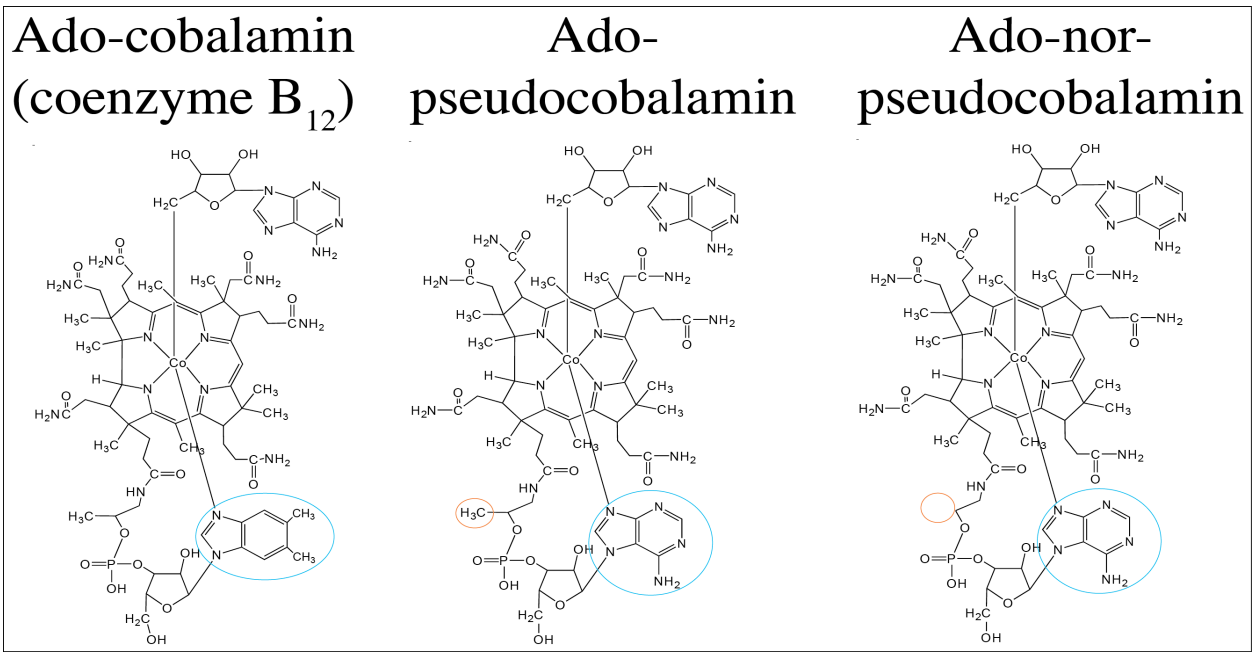



**Figure 1.5. Structural comparison of Cbl, pseudoCbl, and norpseduoCbl.** Cbl contains DMB and the lower ligand base where as pseudoCbl contains adenine (blue circles). Norpsedudo is missing the methyl group at carbon 176 (orange circles).

**DISSERTATION OUTLINE**

This thesis will focus on the steps of the AdoCba biosynthetic pathway for which enzymes within the late Co insertion pathway have not been identified, and therefore appear to have the most enzymatic diversity. Figure 1.4 shows the diversity of enzymes for α-ribotide synthesis (CobT, ArsAB) and salvage (CblS). There is evidence that CblS from *Geobacillus kaustophilus* is a bifunctional enzyme that can synthesis α-ribazole phosphate (unpublished data) in addition to its characterized α-ribazole kinase activity (70). This enzymatic diversity is also found in the first step of the linker synthesis with the generation of L-Thr –P by PduX and BluE, and the final step in the pathway with the dephosphorylation of AdoCbl-P by CobC, BluF, or CobZ.

Chapter 2 will discuss the identification and characterization BluE a L-Thr kinase of the AdoCbl producers in the order Rhodobacterales that uses the late Co insertion pathway. Chapter 3 will expand on the number of known L-Thr kinases with the characterization of L-Thr kinase activity in CobD from methanogens, an activity that is absent from CobD from *S.* Typhimurium. Chapter 4 further characterizes

CobD from *M. mazei*. The work in this chapter is focused on the characterization of the metal binding domain of *Mm*CobD and demonstrates it is has one or more $[4Fe\text{-}4S]^{2+}$ cluster cofactor which is not required for, but influences the enzymatic activities of *Mm*CobD. Chapter 5 offers the initial characterization of the bacterial corrinoid amidohydrolase CbiZ from *R. sphaeroides* and offers phylogenetic data that draws the conclusion it is of bacterial origin and was acquired through horizontal transfer. Chapter 6 summarizes the significance of this work and suggests future directions for the expansion of this portion of the field.

**REFERENCES**


1. **Battersby AR, Fookes CJ, Matcham GW, McDonald E.** 1980. Biosynthesis of the pigments of life: formation of the macrocycle. Nature **285:**17-21.

2. **Battersby AR.** 2000. Tetrapyrroles: the pigments of life. Nat Prod Rep **17:**507-526.

3. **Renz P.** 1999. Biosynthesis of the 5,6-dimethylbenzimidazole moiety of cobalamin and of other bases found in natural corrinoids, p 557-575. *In* Banerjee R (ed), Chemistry and Biochemistry of B12. John Wiley & Sons, Inc., New York.

4. **Stupperich E, Eisinger HJ, Kräutler B.** 1988. Diversity of corrinoids in acetogenic bacteria. *p*-Cresolylcobamide from *Sporomusa ovata*, 5-methoxy-6-methylbenzimidazolylcobamide from *Clostridium formicoaceticum* and vitamin B12 from *Acetobacterium woodii*. Eur J Biochem **172:**459-464.

5. **Allen RH, Stabler SP.** 2008. Identification and quantitation of cobalamin and cobalamin analogues in human feces. Am J Clin Nutr **87:**1324-1335.

6. **Escalante-Semerena JC.** 2007. Conversion of cobinamide into adenosylcobamide in bacteria and archaea. J Bacteriol **189:**4555-4560.

7. **Weissbach H, Taylor R.** 1966. Role of vitamin B12 in methionine synthesis. Fed Proc **25:**1649-1656.

8. **Taylor RR.** 1982. B12-dependent methionine synthesis, p 307-356. *In* Dolphin D (ed), B12, vol 2. John Wiley & Sons, New York.

9. **Stadtman TC, Blaylock BA.** 1966. Role of B12 compounds in methane formation. Fed Proc **25:**1657-1661.

10. **Stupperich E, Eisinger HJ, Albracht SP.** 1990. Evidence for a super-reduced cobamide as the major corrinoid fraction in vivo and a histidine residue as a cobalt ligand of the *p*-cresolyl cobamide in the acetogenic bacterium *Sporomusa ovata*. Eur J Biochem **193:**105-109.

11. **Stupperich E, Eisinger HJ, Kräutler B.** 1989. Identification of phenolyl cobamide from the homoacetogenic bacterium *Sporomusa ovata*. Eur J Biochem **186:**657-661.

12. **Babior BM.** 1970. The mechanism of action of ethanolamine ammonia-lyase, a B12-dependent enzyme. VII. The mechanism of hydrogen transfer. J Biol Chem **245:**6125-6133.

13. **Chang GW, Chang JT.** 1975. Evidence for the B12-dependent enzyme ethanolamine deaminase in *Salmonella*. Nature **254:**150-151.

14. **Blackwell CM, Turner JM.** 1978. Microbial metabolism of amino alcohols. Formation of coenzyme B12-dependent ethanolamine ammonia-lyase and its concerted induction in *Escherichia coli*. Biochem J **176:**751-757.

15. **Babior BM.** 1982. Ethanolamine ammonia-lyase, p 263-288. *In* Dolphin D (ed), B12, vol 2. John Wiley & Sons, New York.

16. **Roof DM, Roth JR.** 1988. Ethanolamine utilization in *Salmonella typhimurium*. J Bacteriol **170:**3855-3863.

17. **Toraya T, Honda S, Kuno S, Fukui S.** 1978. Coenzyme B12-dependent diol dehydratase: regulation of apoenzyme synthesis in Klebsiella pneumoniae (Aerobacter aerogenes) ATCC 8724. J Bacteriol **135:**726-729.

18. **Poznanskaja AA, Tanizawa K, Soda K, Toraya T, Fukui S.** 1979. Coenzyme B12-dependent diol dehydrase: purification subunit heterogeneity, and reversible association. Arch Biochem Biophys **194:**379-386.

19. **Toraya T, Honda S, Fukui S.** 1979. Fermentation of 1,2-propanediol and 1,2-ethanediol by some genera of enterobacteriaceae involving coenzyme-B12-dependent diol dehydratase. J Bacteriol **139:**39-47.

20. **Rondon MR, Escalante-Semerena JC.** 1992. The *poc* locus is required for 1,2-propanediol-dependent transcription of the cobalamin biosynthetic (*cob*) and propanediol utilization (*pdu*) genes of *Salmonella typhimurium*. J Bacteriol **174:**2267-2272.

21. **Bobik TA, Ailion M, Roth JR.** 1992. A single regulatory gene integrates control of vitamin B12 synthesis and propanediol degradation. J Bacteriol **174:**2253-2266.

22. **Rétey J.** 1982. Methylmalonyl-CoA mutase, p 357-380. *In* Dolphin D (ed), B12, vol 2. John Wiley & Sons, New York.

23. **Halpern J.** 1985. Mechanisms of coenzyme $B_{12}$-dependent rearrangements. Science **227:**869-875.

24. **Mancia F, Keep NH, Nakagawa A, Leadlay PF, McSweeney S, Rasmussen B, Bosecke P, Diat O, Evans PR.** 1996. How coenzyme B12 radicals are generated: the crystal structure of methylmalonyl-coenzyme A mutase at 2Å resolution. Structure **4:**339-350.

25. **Toraya T, Fukui S.** 1977. Immunochemical evidence for the difference between coenzyme-B12-dependent diol dehydratase and glycerol dehydratase. Eur J Biochem **76:**285-289.

26. **Reitzer R, Gruber K, Jogl G, Wagner UG, Bothe H, Buckel W, Kratky C.** 1999. Glutamate mutase from *Clostridium cochlearium*: the structure of a coenzyme B12-dependent enzyme provides new mechanistic insights. Structure Fold Des **7:**891-902.

27. **Ohmori H, Ishitani H, Sato K, Shimizu S, Fukui S.** 1974. Metabolism of glutamate in purple nonsulfur bacteria: participation of vitamin $B_{12}$. Agric Biol Chem **38:**359-365.

28. **Switzer RL.** 1982. Glutamate mutase, p 289-306. *In* Dolphin D (ed), B12, vol 2. John Wiley & Sons, New York.

29. **Yan J, Im J, Yang Y, Loffler FE.** 2013. Guided cobalamin biosynthesis supports Dehalococcoides mccartyi reductive dechlorination activity. Philos Trans R Soc Lond B Biol Sci **368:**20120320.

30. **Bommer M, Kunze C, Fesseler J, Schubert T, Diekert G, Dobbek H.** 2014. Structural basis for organohalide respiration. Science **346:**455-458.

31. **Reinhold A, Westermann M, Seifert J, von Bergen M, Schubert T, Diekert G.** 2012. Impact of vitamin B12 on formation of the tetrachloroethene reductive dehalogenase in *Desulfitobacterium hafniense* strain Y51. Appl Environ Microbiol **78:**8025-8032.

32. **Smith MH, Woods SL.** 1994. Regiospecificity of chlorophenol reductive dechlorination by vitamin B12s. Appl Environ Microbiol **60:**4111-4115.

33. **Wohlfarth G, Diekert G.** 1999. Reductive dehalogenases., p 871-893. *In* Banerjee R (ed), Chemistry and Biochemistry of B12. John Wiley & Sons, Inc., New York.

34. **Kräutler B, Fieber W, Osterman S, Fasching M, Ongania K-H, Gruber K, Kratky C, Mikl C, Siebert A, Diekert G.** 2003. The cofactor of tetrachloroethene reductive dehalogenase of *Dehalospirillum multivorans* is Norpseudo-B12, a new type of natural corrinoid. Helvetica Chimica Acta **86:**3698-3716.

35. **Banerjee R, Ragsdale SW.** 2003. The many faces of vitamin B12: catalysis by cobalamin-dependent enzymes. Annu Rev Biochem **72:**209-247.

36. **Blakley RL.** 1982. Cobalamin-dependent ribonucleotide reductases, p 381-418. *In* Dolphin D (ed), B12, vol 2. John Wiley & Sons, New York.

37. **Booker S, Licht S, Broderick J, Stubbe J.** 1994. Coenzyme B12-dependent ribonucleotide reductase: evidence for the participation of five cysteine residues in ribonucleotide reduction. Biochemistry **33:**12676-12685.

38. **Jordan A, Torrents E, Jeanthon C, Eliasson R, Hellman U, Wernstedt C, Barbé J, Gibert I, Reichard P.** 1997. B12-dependent ribonucleotide reductases from deeply rooted eubacteria are structurally related to the aerobic enzyme from Escherichia coli. Proc Natl Acad Sci USA **94:**13487-13492.

39. **Jordan A, Reichard P.** 1998. Ribonucleotide reductases. Ann Rev Biochem **67:**71-98.

40. **Fontecave M, Mulliez E.** 1999. Ribonucleotide Reductases, p 731-756. *In* Banerjee R (ed), Chemistry and Biochemistry of B12. John Wiley & Sons, Inc., New York.

41. **Banerjee RV, Matthews RG.** 1990. Cobalamin-dependent methionine synthase. FASEB J **4:**1450-1459.

42. **Banerjee R.** 2003. Radical carbon skeleton rearrangements: catalysis by coenzyme B12-dependent mutases. Chem Rev **103:**2083-2094.

43. **Watanabe F, Tanioka Y, Miyamoto E, Fujita T, Takenaka H, Nakano Y.** 2007. Purification and characterization of corrinoid-compounds from the dried powder of an edible cyanobacterium, *Nostoc commune* (Ishikurage). J Nutr Sci Vitaminol (Tokyo) **53:**183-186.

44. **Watanabe F.** 2007. Vitamin B12 Sources and Bioavailability. Exp Biol Med **232:**1266-1274.

45. **Lengyel P, Mazumder R, Ochoa S.** 1960. Mammalian Methylmalonyl Isomerase and Vitamin B(12) Coenzymes. Proc Natl Acad Sci U S A **46:**1312-1318.

46. **Firth J, Johnson BC.** 1954. Pseudo-B12 activity in the baby pig. Science **120:**352-353.

47. **Ludwig ML, Matthews RG.** 1997. Structure-based perspectives on B12-dependent enzymes. Annu Rev Biochem **66:**269-313.

48. **Payne KAP, Quezada CP, Fisher K, Dunstan MS, Collins FA, Sjuts H, Levy C, Hay S, Rigby SEJ, Leys D.** 2015. Reductive dehalogenase structure suggests a mechanism for B12-dependent dehalogenation. Nature **517:**513-516.

49. **Vitreschak AG, Rodionov DA, Mironov AA, Gelfand MS.** 2003. Regulation of the vitamin B12 metabolism and transport in bacteria by a conserved RNA structural element. RNA **9:**1084-1097.

50. **Escalante-Semerena JC, Warren MJ.** 2008. Biosynthesis and Use of Cobalamin ($B_{12}$). *In* Böck A, Curtiss III R, Kaper JB, Karp PD, Neidhardt FC, Nyström T, Slauch JM, Squires CL (ed), *EcoSal - Escherichia coli* and *Salmonella*: cellular and molecular biology. ASM Press, Washington, D. C.

51. **Fan C, Bobik TA.** 2008. The PDUX enzyme of *Salmonella enterica* is an L-threonine kinase used for coenzyme $B_{12}$ synthesis. J Biol Chem **283:**11322-11329.

52. **Brushaber KR, O'Toole GA, Escalante-Semerena JC.** 1998. CobD, a novel enzyme with L-threonine-*O*-3-phosphate decarboxylase activity, is responsible for the synthesis of (*R*)-1-amino-2-propanol *O*-2-phosphate, a proposed new intermediate in cobalamin biosynthesis in *Salmonella typhimurium* LT2. J Biol Chem **273:**2684-2691.

53. **O'Toole GA, Trzebiatowski JR, Escalante-Semerena JC.** 1994. The *cobC* gene of *Salmonella typhimurium* codes for a novel phosphatase involved in the assembly of the nucleotide loop of cobalamin. J Biol Chem **269:**26503-26511.

54. **Anderson PJ, Lango J, Carkeet C, Britten A, Kräutler B, Hammock BD, Roth JR.** 2008. One pathway can incorporate either adenine or dimethylbenzimidazole as an alpha-axial ligand of $B_{12}$ cofactors in *Salmonella enterica*. J Bacteriol **190:**1160-1171.

55. **Maggio-Hall LA, Escalante-Semerena JC.** 1999. In vitro synthesis of the nucleotide loop of cobalamin by *Salmonella typhimurium* enzymes. Proc Natl Acad Sci U S A **96:**11798-11803.

56. **O'Toole GA, Escalante-Semerena JC.** 1995. Purification and characterization of the bifunctional CobU enzyme of *Salmonella typhimurium* LT2. Evidence for a CobU-GMP intermediate. J Biol Chem **270:**23560-23569.

57. **Zayas CL, Escalante-Semerena JC.** 2007. Reassessment of the late steps of coenzyme $B_{12}$ synthesis in *Salmonella enterica*: Evidence that dephosphorylation of adenosylcobalamin-5'-phosphate by the CobC phosphatase is the last step of the pathway. J Bacteriol **189:**2210-2218.

58. **Zayas CL, Claas K, Escalante-Semerena JC.** 2007. The CbiB protein of *Salmonella enterica* is an integral membrane protein involved in the last step of the de novo corrin ring biosynthetic pathway. J Bacteriol **189:**7697-7708.

59. **Escalante-Semerena JC, Johnson MG, Roth JR.** 1992. The CobII and CobIII regions of the cobalamin (vitamin B12) biosynthetic operon of Salmonella typhimurium. J Bacteriol **174:**24-29.

60. **Trzebiatowski JR, O'Toole GA, Escalante-Semerena JC.** 1994. The *cobT* gene of *Salmonella typhimurium* encodes the NaMN: 5,6-dimethylbenzimidazole phosphoribosyltransferase responsible for the synthesis of $N^1$-(5-phospho-alpha-D-ribosyl)-5,6-dimethylbenzimidazole, an intermediate in the synthesis of the nucleotide loop of cobalamin. J Bacteriol **176:**3568-3575.

61. **Roth JR, Lawrence JG, Rubenfield M, Kieffer-Higgins S, Church GM.** 1993. Characterization of the cobalamin (vitamin $B_{12}$) biosynthetic genes of *Salmonella typhimurium*. J Bacteriol **175:**3303-3316.

62. **Escalante-Semerena JC, Roth JR.** 1987. Regulation of cobalamin biosynthetic operons in *Salmonella typhimurium*. J Bacteriol **169:**2251-2258.

63. **Jeter RM, Roth JR.** 1987. Cobalamin (vitamin $B_{12}$) biosynthetic genes of S*almonella typhimurium*. J Bacteriol **169:**3189-3198.

64. **Raux E, Lanois A, Levillayer F, Warren MJ, Brody E, Rambach A, Thermes C.** 1996. *Salmonella typhimurium* cobalamin (vitamin $B_{12}$) biosynthetic genes: functional studies in *S. typhimurium* and *Escherichia coli*. J Bacteriol **178:**753-t767.

65. **Otte MM, Escalante-Semerena JC.** 2009. Biochemical characterization of the GTP:adenosylcobinamide-phosphate guanylyltransferase (CobY) enzyme of the hyperthermophilic archaeon *Methanocaldococcus jannaschii*. Biochemistry **48:**5882-5889.

66. **Zayas CL, Woodson JD, Escalante-Semerena JC.** 2006. The *cobZ* gene of *Methanosarcina mazei* Gö1 encodes the nonorthologous replacement of the α-ribazole-5'-phosphate phosphatase (CobC) enzyme of *Salmonella enterica*. J Bacteriol **188:**2740-2743.

67. **Woodson JD, Escalante-Semerena JC.** 2006. The *cbiS* gene of the archaeon *Methanopyrus kandleri* AV19 encodes a bifunctional enzyme with adenosylcobinamide amidohydrolase and alpha-ribazole-phosphate phosphatase activities. J Bacteriol **188:**4227-4235.

68. **Gray MJ, Escalante-Semerena JC.** 2007. Single-enzyme conversion of $FMNH_2$ to 5,6-dimethylbenzimidazole, the lower ligand of $B_{12}$. Proc Natl Acad Sci U S A **104:**2921-2926.

69. **Woodson JD, Escalante-Semerena JC.** 2004. CbiZ, an amidohydrolase enzyme required for salvaging the coenzyme $B_{12}$ precursor cobinamide in archaea. Proc Natl Acad Sci USA **101:**3591-3596.

70. **Gray MJ, Escalante-Semerena JC.** 2010. A new pathway for the synthesis of alpha-ribazole-phosphate in *Listeria innocua*. Mol Microbiol **77:**1429-1438.

71. **Faust LR, Connor JA, Roof DM, Hoch JA, Babior BM.** 1990. Cloning, sequencing, and expression of the genes encoding the adenosylcobalamin-dependent ethanolamine ammonia-lyase of *Salmonella typhimurium*. J Biol Chem **265:**12462-12466.

72. **Stojiljkovic I, Baumler AJ, Heffron F.** 1995. Ethanolamine utilization in *Salmonella typhimurium*: nucleotide sequence, protein expression, and mutational analysis of the *cchA cchB eutE eutJ eutG eutH* gene cluster. J Bacteriol **177:**1357-1366.

73. **Kofoid E, Rappleye C, Stojiljkovic I, Roth J.** 1999. The 17-gene ethanolamine (*eut*) operon of *Salmonella typhimurium* encodes five homologues of carboxysome shell proteins. J Bacteriol **181:**5317-5329.

74. **Roof DM, Roth JR.** 1989. Functions required for vitamin B12-dependent ethanolamine utilization in *Salmonella typhimurium*. J Bacteriol **171:**3316-3323.

75. **Bobik TA, Xu Y, Jeter RM, Otto KE, Roth JR.** 1997. Propanediol utilization genes (*pdu*) of *Salmonella typhimurium*: three genes for the propanediol dehydratase. J Bacteriol **179:**6633-6639.

76. **Toraya T.** 2000. Radical catalysis of $B_{12}$ enzymes: structure, mechanism, inactivation, and reactivation of diol and glycerol dehydratases. Cell Mol Life Sci **57:**106-127.

77. **Jeter RM.** 1990. Cobalamin-dependent 1,2-propanediol utilization by *Salmonella typhimurium*. J Gen Microbiol **136:**887-896.

78. **Bradbeer C.** 1965. The clostridial fermentation of choline and ethanolamine. II. Requirement for a cobamide coenzyme by an ethanolamine deaminase. J Biol Chem **240:**4675-4681.

79. **Drennan CL, Huang S, Drummond JT, Matthews RG, Ludwig ML.** 1994. How a protein binds $B_{12}$: A 3.0Å X-ray structure of $B_{12}$-binding domains of methionine synthase. Science **266:**1669-1674.

80. **Bandarian V, Pattridge KA, Lennon BW, Huddler DP, Matthews RG, Ludwig ML.** 2002. Domain alternation switches B(12)-dependent methionine synthase to the activation conformation. Nat Struct Biol **9:**53-56.

81. **Peariso K, Zhou ZS, Smith AE, Matthews RG, Penner-Hahn JE.** 2001. Characterization of the zinc sites in cobalamin-independent and cobalamin-dependent methionine synthase using zinc and selenium X-ray absorption spectroscopy. Biochemistry **40:**987-993.

82. **Gonzalez JC, Banerjee RV, Huang S, Sumner JS, Matthews RG.** 1992. Comparison of cobalamin-independent and cobalamin-dependent methionine synthases from *Escherichia coli*: two solutions to the same chemical problem. Biochemistry **31:**6045-6056.

83. **Hondorp ER, Matthews RG.** 2004. Oxidative stress inactivates cobalamin-independent methionine synthase (MetE) in *Escherichia coli*. PLoS Biol **2:**e336.

84. **Bradbeer C.** 1982. Cobalamin transport in microorganisms., p 31-56. *In* Dolphin D (ed), B12, vol 2. John Wiley & Sons, New York.

85. **Warren MJ, Raux E, Schubert HL, Escalante-Semerena JC.** 2002. The biosynthesis of adenosylcobalamin (vitamin B12). Nat Prod Rep **19:**390-412.

86. **Scott AI, Roessner CA.** 2002. Biosynthesis of cobalamin (vitamin B12). Biochem Soc Trans **30:**613-620.

87. **Moore SJ, Lawrence AD, Biedendieck R, Deery E, Frank S, Howard MJ, Rigby SE, Warren MJ.** 2013. Elucidation of the anaerobic pathway for the corrin component of cobalamin (vitamin B12). Proc Natl Acad Sci U S A **110:**14906-14911.

88. **Blanche F, Debussche L, Thibaut D, Crouzet J, Cameron B.** 1989. Purification and characterization of S-adenosyl-L-methionine: uroporphyrinogen III methyltransferase from *Pseudomonas denitrificans*. J Bacteriol **171:**4222-4231.

89. **Cameron B, Briggs K, Pridmore S, Brefort G, Crouzet J.** 1989. Cloning and analysis of genes involved in coenzyme B12 biosynthesis in *Pseudomonas denitrificans*. J Bacteriol **171:**547-557.

90. **Thibaut D, Debussche L, Blanche F.** 1990. Biosynthesis of vitamin B12: isolation of precorrin-6x, a metal-free precursor of the corrin macrocycle retaining five S-adenosylmethionine-derived peripheral methyl groups. Proc Natl Acad Sci U S A **87:**8795-8799.

91. **Crouzet J, Cameron B, Cauchois L, Rigault S, Rouyez MC, Blanche F, Thibaut D, Debussche L.** 1990. Genetic and sequence analysis of an 8.7-kilobase *Pseudomonas denitrificans* fragment carrying eight genes involved in transformation of precorrin-2 to cobyrinic acid. J Bacteriol **172:**5980-5990.

92. **Debussche L, Couder M, Thibaut D, Cameron B, Crouzet J, Blanche F.** 1991. Purification and partial characterization of cob(I)alamin adenosyltransferase from *Pseudomonas denitrificans*. J Bacteriol **173:**6300-6302.

93. **Cameron B, Guilhot C, Blanche F, Cauchois L, Rouyez MC, Rigault S, Levy-Schil S, Crouzet J.** 1991. Genetic and sequence analyses of a *Pseudomonas denitrificans* DNA fragment containing two *cob* genes. J Bacteriol **173:**6058-6065.

94. **Blanche F, Couder M, Debussche L, Thibaut D, Cameron B, Crouzet J.** 1991. Biosynthesis of vitamin B12: stepwise amidation of carboxyl groups *b, d, e*, and *g* of cobyrinic acid *a,c*-diamide is catalyzed by one enzyme in *Pseudomonas denitrificans*. J Bacteriol **173:**6046-6051.

95. **Crouzet J, Levy-Schil S, Cameron B, Cauchois L, Rigault S, Rouyez MC, Blanche F, Debussche L, Thibaut D.** 1991. Nucleotide sequence and genetic analysis of a 13.1-kilobase-pair *Pseudomonas denitrificans* DNA fragment containing five *cob* genes and identification of structural genes encoding Cob(I)alamin adenosyltransferase, cobyric acid synthase, and bifunctional cobinamide kinase-cobinamide phosphate guanylyltransferase. J Bacteriol **173:**6074-6087.

96. **Blanche F, Debussche L, Famechon A, Thibaut D, Cameron B, Crouzet J.** 1991. A bifunctional protein from *Pseudomonas denitrificans* carries cobinamide kinase and cobinamide phosphate guanylyltransferase activities. J Bacteriol **173:**6052-6057.

97. **Cameron B, Blanche F, Rouyez MC, Bisch D, Famechon A, Couder M, Cauchois L, Thibaut D, Debussche L, Crouzet J.** 1991. Genetic analysis, nucleotide sequence, and products of two *Pseudomonas denitrificans cob* genes encoding nicotinate-nucleotide: dimethylbenzimidazole phosphoribosyltransferase and cobalamin (5'-phosphate) synthase. J Bacteriol **173:**6066-6073.

98. **Blanche F, Famechon A, Thibaut D, Debussche L, Cameron B, Crouzet J.** 1992. Biosynthesis of vitamin $B_{12}$ in *Pseudomonas denitrificans*: the biosynthetic sequence from precorrin-6y to precorrin-8x is catalyzed by the *cobL* gene product. J Bacteriol **174:**1050-1052.

99. **Blanche F, Maton L, Debussche L, Thibaut D.** 1992. Purification and characterization of Cob(II)yrinic acid *a,c*-diamide reductase from *Pseudomonas denitrificans*. J Bacteriol **174:**7452-7454.

100. **Blanche F, Thibaut D, Famechon A, Debussche L, Cameron B, Crouzet J.** 1992. Precorrin-6x reductase from *Pseudomonas denitrificans*: purification and characterization of the enzyme and identification of the structural gene. J Bacteriol **174:**1036-1042.

101. **Debussche L, Couder M, Thibaut D, Cameron B, Crouzet J, Blanche F.** 1992. Assay, purification, and characterization of cobaltochelatase, a unique complex enzyme catalyzing cobalt insertion in hydrogenobyrinic acid a,c-diamide during coenzyme $B_{12}$ biosynthesis in *Pseudomonas denitrificans*. J Bacteriol **174:**7445-7451.

102. **Pollich M, Klug G.** 1995. Identification and sequence analysis of genes involved in late steps of cobalamin (vitamin B12) synthesis in Rhodobacter capsulatus. J Bacteriol **177:**4481-4487.

103. **Pollich M, Wersig C, Klug G.** 1996. The *bluF* gene of *Rhodobacter capsulatus* is involved in conversion of cobinamide to cobalamin (vitamin $B_{12}$). J Bacteriol **178:**7308-7310.

104. **Vlcek C, Paces V, Maltsev N, Paces J, Haselkorn R, Fonstein M.** 1997. Sequence of a 189-kb segment of the chromosome of Rhodobacter capsulatus SB1003. Proc Natl Acad Sci U S A **94:**9384-9388.

105. **McGoldrick H, Deery E, Warren M, Heathcote P.** 2002. Cobalamin (vitamin B12) biosynthesis in *Rhodobacter capsulatus*. Biochem Soc Trans **30:**646-648.

106. **McGoldrick HM, Roessner CA, Raux E, Lawrence AD, McLean KJ, Munro AW, Santabarbara S, Rigby SE, Heathcote P, Scott AI, Warren MJ.** 2005. Identification and characterization of a novel vitamin B12 (cobalamin) biosynthetic enzyme (CobZ) from *Rhodobacter capsulatus*, containing flavin, heme, and Fe-S cofactors. J Biol Chem **280:**1086-1094.

107. **Heldt D, Lawrence AD, Lindenmeyer M, Deery E, Heathcote P, Rigby SE, Warren MJ.** 2005. Aerobic synthesis of vitamin B12: ring contraction and cobalt chelation. Biochem Soc Trans **33:**815-819.

108. **Campbell GR, Taga ME, Mistry K, Lloret J, Anderson PJ, Roth JR, Walker GC.** 2006. *Sinorhizobium meliloti bluB* is necessary for production of 5,6-dimethylbenzimidazole, the lower ligand of B12. Proc Natl Acad Sci U S A **103:**4634-4369.

109. **Raux E, Thermes C, Heathcote P, Rambach A, Warren MJ.** 1997. A role for *Salmonella typhimurium cbiK* in cobalamin (vitamin B12) and siroheme biosynthesis. J Bacteriol **179:**3202-3212.

110. **Beck R, Raux E, Thermes C, Rambach A, Warren M.** 1997. CbiX: a novel metal-binding protein involved in sirohaem biosynthesis in *Bacillus megaterium*. Biochem SocTrans **25:**77S.

111. **Fazzio TG, Roth JR.** 1996. Evidence that the CysG protein catalyzes the first reaction specific to B12 synthesis in *Salmonella typhimurium*, insertion of cobalt. J Bacteriol **178:**6952-6959.

112. **Moore SJ, Biedendieck R, Lawrence AD, Deery E, Howard MJ, Rigby SE, Warren MJ.** 2012. Characterization of the enzyme CbiH60 involved in anaerobic ring contraction of the cobalamin (vitamin B12) biosynthetic pathway. J Biol Chem **288:**297-305.

113. **Scott AI, Roessner CA, Stolowich NJ, Spencer JB, Min C, Ozaki SI.** 1993. Biosynthesis of vitamin $B_{12}$. Discovery of the enzymes for oxidative ring contraction and insertion of the fourth methyl group. FEBS Lett **331:**105-108.

114. **Chan CH, Escalante-Semerena JC.** 2011. ArsAB, a novel enzyme from Sporomusa ovata activates phenolic bases for adenosylcobamide biosynthesis. Mol Microbiol **81:**952-967.

115. **Newmister SA, Chan CH, Escalante-Semerena JC, Rayment I.** 2012. Structural insights into the function of the nicotinate mononucleotide:phenol/p-cresol phosphoribosyltransferase (ArsAB) enzyme from Sporomusa ovata. Biochemistry **51:**8571-8582.

116. **Seth EC, Taga ME.** 2014. Nutrient cross-feeding in the microbial world. Front Microbiol **5:**350.

117. **Gray MJ, Tavares NK, Escalante-Semerena JC.** 2008. The genome of *Rhodobacter sphaeroides* strain 2.4.1 encodes functional cobinamide salvaging systems of archaeal and bacterial origins. Mol Microbiol **70:**824-836.

118. **Gray MJ, Escalante-Semerena JC.** 2009. In vivo analysis of cobinamide salvaging in *Rhodobacter sphaeroides* strain 2.4.1. J Bacteriol **191:**3842-3851.

119. **Gray MJ, Escalante-Semerena JC.** 2009. The cobinamide amidohydrolase (cobyric acid-forming) CbiZ enzyme: a critical activity of the cobamide remodelling system of *Rhodobacter sphaeroides*. Mol Microbiol **74:**1198-1210.

120. **Dion HW, Calkins DG, Pfiffner JJ.** 1952. Hydrolysis products of pseudovitamin B12. J Amer Che Soc **74:**1108-1108.

121. **Friedmann HC, Fyfe JA.** 1969. Pseudovitamin B12 biosynthesis. Enzymatic formation of a new adenylic acid, 7-α-D-ribofuranosyladenine 5'-phosphate. J Biol Chem **244:**1667-1671.

122. **Erb TJ, Retey J, Fuchs G, Alber BE.** 2008. Ethylmalonyl-CoA mutase from *Rhodobacter sphaeroides* defines a new subclade of coenzyme B12-dependent acyl-CoA mutases. J Biol Chem **283:**32283-32293.

123. **Ford JE, Porter JW.** 1953. Vitamin B12 like compounds. II. Some properties of compounds isolated from bovine gut contents and faeces. Br J Nutr **7:**326-337.

124. **Ford JE, Holdsworth ES, Kon SK, Porter JW.** 1953. Differentiation of vitamin B12 active compounds by ionophoresis and microbiological assay; occurrence of the various vitamin B12 active compounds. Nature **171:**150-151.

125. **Dion HW, Calkins DG, Pfiffner JJ.** 1954. 2-Methyladenine, an hydrolysis product of pseudovitamin B12d. J Amer Che Soc **76:**948-949.

# CHAPTER 2

# ORF RSP_0788 IS THE L-THREONINE KINASE IN THE $B_{12}$ BIOSYNTHETIC PATHWAY OF *RHODOBACTER SPHAEROIDES* 2.4.1[1]

[1] Tavares N.K. and Escalante-Semerena J.C. To be submitted to *Journal of Bacteriology*.

**ABSTRACT**

The end of the late cobalt insertion pathway for cobamide biosynthesis has several steps with unidentified enzymes. Using bioinformatics we have identified the gene *bluE* (locus tag RSP_0788) in *Rhodobacter sphaeroides* 2.4.1 as the functional homolog of the L-threonine kinase, PduX (EC 2.7.1.177) in the early Co insertion pathway. Here we show that *Rs*BluE has ATPase activity *in vivo* and *in vitro*. Using $^{31}$P-NMR spectroscopy we demonstrate that *Rs*BluE transfers a phosphate from ATP to generate ADP and L-threonine-*O*-*3*-phosphate as products. BluE from *R. sphaeroides* and *Rhodobacter capsulatus* complements a *Salmonella enterica pduX* strain under conditions that demand cobamide synthesis for growth. Phylogenic analysis reveals that *bluE* is restricted to the genomes of a few Rhodobacterales that appear to have a strict requirement for the cobamide species of cobalamin, which has 5,6-dimethylbenzimidazole as the lower ligand base and 1-amino-2-propanol-*O*-2-phosphate at the nucleotide linker.

**INTRODUCTION**

There is generally regarded to be two major pathways for the biosynthesis of cobamides (Cba; e.g. adenosylcobalamin, AdoCbl) in bacteria and archaea. There is the early cobalt insertion (anaerobic) pathway and the late cobalt insertion (aerobic) pathway (Fig. 2.1). The 'anaerobic pathway' has an oxygen sensitive corrinoid intermediate (1) and has been well characterized in organisms such as *Salmonella enterica* sv Typhimurium LT2 *(2-4)*. The 'aerobic pathway' is functional under aerobic and anaerobic conditions, and has been primarily studied in *Pseudomonas denitrificans* (5-11) and *Rhodobacter capsulatus* (12-16). Both pathways share several homologous enzymes, however, while all the major enzymes within the anaerobic pathway of *S.* Typhimurium have been identified, the functional equivalents of some enzymes have not been identified in the aerobic pathway. Yet to be identified is the phosphatase that dephosphorylates cobalamin phosphate to produce the final product, AdoCbl. Also unidentified is the kinase that generates L-threonine-*O*-3-phosphate, which is converted to 1-amino-2-propanol-*O*-2-phosphate (AP-P) and incorporated as the linker between the lower ligand nucleotide and

the corrin ring (Fig. 2.1). In *S.* Typhimurium PduX is the enzyme that catalyzes this reaction (17, 18). The functional equivalent of PduX has not been identified in any organism that uses the aerobic pathway. The *bluE* gene was first identified in *R. capsulatus* BB1 by Pollich *et. al*. as the second gene in an operon named *bluFEDCB* because disruption of the operon resulted in a reduction of red pigmentation or a "blush" phenotype (15). Disruption of the *bluE* gene by a Tn5 insertion was correctable by the addition of AdoCbl and not cobinamide (Cbi) (15). BluD is homologous to the adenosylcobinamide-phosphate synthase (CbiB/CobD, EC 6.3.1.10), BluC is homologous to the L-threonine-*O*-3-phosphate decarboxylase (CobC/CobD, EC 4.1.1.81), and BluB was later identified as a 5,6-dimethylbenzimidazole (DMB) synthase (19, 20). The functions encoded by the *bluF* and *bluE* genes have remained unexplored to date. Here we report the identification and initial characterization of BluE (locus tag RSP_0788), the L-threonine kinase from *Rhodobacter sphaeroides* 2.4.1. This enzyme was found to be restricted to a subclass of Rhodobacterales with specific requirements for the cobalamin (Cbl) species of Cba only. That is to say they require and synthesize only a Cba with DMB as the lower ligand and AP-P as the nucleotide linker.

**METHODS AND MATERIAL**

**Bacterial strains.** Strains and plasmids are described in Table S2.1. All *S.* Typhimurium strains used in this work carred a null allele of the *metE* gene that encodes the AdoCbl-independent methionine synthase (MetE) enzyme (21). In the absence of MetE the cell to use the Cbl-dependent methionine synthase (MetH) enzyme (22-24). All *S.* Typhimurium strains used in this work also carry a mutation (allele *ara*-9) which prevents the utilization of arabinose as a carbon and energy source. Deletions of the *metE* and *pduX* genes in *S.* Typhimurium were constructed using the phage lambda Red recombinase system as described elsewhere (25).

**Plasmid construction.** *R. sphaeroides* strain 2.4.1 and *R. capsulatus* strain SB1003 were gifts from Tim Donohue (University of Wisconsin, Madison WI). Genomic DNA was isolated using Wizard SV Genomic DNA Purification kit (Promega). Oligonucleotide primers were purchased from Integrated DNA Technologies Inc. (IDT, Coralville, IA). Primers for cloning were designed using the Saccharomyces

Genome Database web based primer design tool http://www.yeastgenome.org/cgi-bin/web-primer. Genes were PCR amplified from the appropriate genomic DNA template with PCR Extender Polymerase (5 Prime). PCR products and vectors were treated with restriction endonucleases (Fermentas) and purified with the Wizard SV Gel and PCR Clean-Up kit (Promega). Vectors were treated with Fast AP alkaline phosphatase (Fermentas). PCR fragments and vectors were ligated together using T4 Ligase (New England BioLAB) and introduced into *Escherichia coli* DH5α (26, 27) via electroporation. Plasmid DNA was purified using the Wizard Plus SV Miniprep kit (Promega). Plasmid sequence was confirmed by using BigDye (ABI PRISM) sequencing protocols (University of Wisconsin-Madison Biotechnology Center). Table S2.1 lists the resulting plasmids. Primers for site directed mutagenesis were designed with the QuikChange Primer Design tool found at http://www.genomics.agilent.com/primerDesignProgram.jsp. DNA for site directed mutagenesis was amplified using PfuUltra II Fusion DNA polymerase (Stratagene), and site-directed mutagenesis was performed using the QuikChange protocol from Stratagene. All genes used in this study were cloned into pBAD24 or pBAD30 (28) for complementation and pTEV5 (29) for overexpression and protein production.

**Growth media and conditions.** No carbon essential (NCE) (30) was used as minimal growth medium with either glycerol (22 mM) or ethanolamine (90 mM) as the carbon and energy source. When added to the medium, the following supplements were at the indicated concentrations: trace minerals, 10 mL/liter (31); $MgSO_4$ (1 mM), 5,6-dimethylbenzimidazole (DMB) (0.15 mM), ampicillin (100 μg/mL), arabinose (500 μM). All corrinoids (dicyanide cobyric acid [$(CN)_2$Cby], cobinamide dicyanide [$(CN)_2$Cbi], and cobalamin dicyanide [$(CN)_2$Cbl] were added at (1 nM) final concentration when glycerol was used or 300 nM for ethanolamine. Fe-citrate (50 μM) was also added to the medium when ethanolamine was used. Cobyric acid was a gift from Paul Renz (Universität-Hohenheim, Stuttgart, Germany). All other chemicals were purchased from Sigma-Aldrich. *S.* Typhimurium strains were cultured in Nutrient Broth (NB, Difco Laboratories) (0.8% w/v) containing NaCl (85 mM). Lysogenic broth (LB) (32, 33) was used as rich medium to culture *E. coli* strains unless otherwise indicated. For complementation studies strains

were grown in minimal medium in triplicate in sterile 96-well tissue culture plates (Falcon). A 1% inoculum of an overnight cell culture grown in Nutrient Broth was used when glycerol was the carbon and energy source, and a 5% inoculum was used for ethanolamine minimal medium in 200 µL total volume. Cultures were monitored using Gen5 software (BioTek Instruments) while growing at 37˚C with continuous shaking (19 Hz) in an EL808 Ultra or PowerWave XS Microplate Reader (BioTek Instruments). Cell density measurements at 630 nm were acquired every 30 min for 24 or 72 h. Data were analyzed using the Prism v6 software package (GraphPad Software).

**Solubilization and purification of *Rs*BluE and PduX proteins.** *E. coli* C43 λDE3 (JE6664) (34) cells carrying plasmids pPDU23 or p*Rs*BLUE4 encoding *N*-terminally $His_6$ tagged *S.* Typhimurium PduX or *R. sphaeroides* BluE genes cloned into pTEV5 (29) respectively, were grown in 4 L of LB. Cells were centrifuged at 6,000 x g for 10 min. Cell pellets were placed at -20°C until used. Cell pellets were gently thawed in 30 mL bind buffer [4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES, 50 mM, pH 7.9), NaCl (500 mM), Imidazole (5 mM)], and cells were lysed at 27 kPsi using a 1.1 kW TS Series Bench Top cell disrupter (Constant Systems Ltd.), equipped with a cooling jacket on the disruptor head to maintain a 6°C temperature using a Neslab ThermoFlex 900 recirculating chiller (Thermo Scientific). Cell contents were separated by centrifugation at 39,000 x g for 20 min. The supernatant containing soluble proteins was discarded. The pellet containing insoluble *Rs*BluE or PduX was resuspended in 30 mL bind buffer [(HEPES (50 mM, pH 7.9), NaCl (500 mM), Imidazole (5 mM)] along with ~12% sarkosyl and incubated at room temp for 30 min with shaking until the majority of the pellet was solubilized. The remaining debris was removed by centrifugation at 20,000 x g for 10 min. Sarkosyl was diluted stepwise by dialysis into bind buffer containing 2%, then 1% sarkosyl, followed by three rounds of dialysis into bind buffer without sarkosyl. Each dialyzed was in 2 L at 4C for 3 h. The presence of solubilized protein was confirmed by SDS-PAGE. Precipitated protein was removed by centrifugation at 20,000 x g for 10 min. Ni-affinity chromatography was used with a 1.5-ml bed volume of HisPur Ni-NTA resin (Thermo Scientific). The column was equilibrated with bind buffer [HEPES (50 mM pH7.9 at 4°C), NaCl (500 mM), imidazole (5 mM)] before supernatant was applied to the column. Due to the

viscosity of the caused by sarkosyl, movement of the supernatant across the column was facilitated by the use of a GE Healthcare P1 peristaltic pump with a flow rate of 1 mL/min. After binding to the column, the column was washed with 4 column volumes of bind buffer. Bound proteins were eluted stepwise by increasing the imidazole concentration from 20 to 100 mM, in 20 mM steps. 5 mL fractions were collected for each step. The $His_6$ tagged proteins did not bind to the column due to interference by sarkosyl and were found in the flow through. $His_6$ tagged proteins were treated with rTEV protease (1:100 rTEV:$His_6$-protein ratio) for 3 h at 25°C in bind buffer containing 1 mM 1,4-dithiothreitol (DTT). It is unclear if cleavage of the tag was successful as sarkosyl solubilized protein will not bind a Ni column to separate the potentially cleaved peptide tag from the protein. The protein was dialyzed into bind buffer containing 1 mM EDTA at room temperature for 1 h, then dialyzed twice more in the same buffer without EDTA. Final dialysis was performed until the protein began to slowly precipitate. Precipitated protein was removed via centrifugation at 6,000 x g for 5 min. Proteins were concentrated using Ultracel 10,000 MWCO centrifugal filters (Amicon Ultra) and frozen drop-wise into liquid $N_2$, and stored at -80ºC until used. Further attempts to completely remove sarkosyl by treatment with BioBeads™ SM-2 Resin (BioRad) or Pierce™ Detergent Removal Spin (ThermoFisher) resulted in protein precipitation. Cells expressing the empty cloning vector, pTEV5, were subjected to the same purification process and sarkosyl solubilized pellet was used as a control for enzyme assays described below. Protein concentrations were determined using a NanoDrop 1000 spectrophotometer (Thermo Scientific), using molecular weights and $A_{280}$ molar extinction coefficients for each protein, which were obtained from the ExPASy Protparam database (35). Proteins were resolved by SDS-PAGE gel. Protein purity was estimated using band densitometry with a Fotodyne imaging system and Foto/Analyst v.5.00 software (Fotodyne Inc) for image acquisition and TotalLab v.2005 software for analysis (Nonlinear Dynamics). *Rs*BluE and PduX were purified to ~70% purity.

**Phylogenetic analysis and tree construction**. Sequences were obtained using Basic Local Alignment Search Tool (BLAST) (36) search for homologs of *Rs*BluE (RSP_0788), *Rcbl*uE (RCAP_rcc02055), *Se*PduX (STM2058), *Se*CbiK (STM2025), *Rs*BluB (RSP_3218), and *Rs*CobN (RSP_2827), available as

of 1/21/2015 in the Integrate Microbial Genome (IMG) database (37). Only finished genomes with bit scores > 50 or *e* values > 1.0e-7 were used in the analysis. Sequence header files were simplified with the find/replace and *grep* functions of TextWrangler (Bare Bones Software, North Chelmsford, MA). Outliers and sequences with extreme divergence were not included nor were alleles not associated with AdoCbl biosynthetic genes cluster in the chromosome. FASTA formatted sequences were aligned using the MUSCLE (38) plugin within Geneious R8 software (Biomatters Ltd., Auckland NZ) with 100 iterations and default settings. Phylogenetic trees were constructed with Bayesian estimation using MrBayes (39) plugin and maximum likelihood using PhyML (40) with 500 Bootstrap replicates with default settings in Geneious R8. Resulting trees were edited using Figtree (41) and Adobe Illustrator CS6 (Adobe). We used the packages *ape* (42), *geiger* (43) and *diversitree* (44) of the R program (45) to represent in the phylogeny the presence or absence of CbiK, CobN, or BluB within the genome of each species. We plotted the presence/absence data in the phylogeny using the function *trait.plot* of the *diversitree* (44) package. ESPript 3.x (46) was used to generate additional images of alignments. Sequences and alignment files are available in the supplementary Material.

**ATPase activity assay.** ATPase activity was assessed using the ADP-glo™ Kinase assay kit (Promega) (47) per manufacturer's instructions. Briefly, After the *Rs*BluE/PduX ATPase reaction is complete the assay uses a proprietary reagent to deplete any remaining ATP in the reaction mixture, then a secondary reagent converts ADP to ATP which is measured by a luciferase reaction. The resulting luminescence was measured with a SpectraMax Plus Gemini EM microplate spectrophotometer (Molecular Devices) equipped with SoftMax Pro v4 software. 25 µL reaction mixtures consisting of HEPES buffer (50 mM, pH 7.9), $MgCl_2$ (1 mM), ATP (100 µM or 10 µM), L-Thr (100 µM or 50 mM) and protein (100ng or 12 µM) were incubated at 25°C for 1 h. 5 µL of the reaction are used in the ADP-glo assay kit per the manufacturer's instructions. Nunc 96-well round bottom black polypropylene microtiter plates (Thermo Fisher) were used to minimize background. ATP to ADP conversion was quantified from the luminescence relative light units (RLU) after subtracting the background from the no enzyme control and comparing the value to a standard curve (Fig. S2.1). For ATPase inhibition assay the following inhibitors

were used adenosine triphosphate (ADP, 20 mM), sodium pyrophosphate (PPi, 10 mM), sodium phosphate monobasic (Pi, 10mM), adenosine 5′-[γ-thio]triphosphate (ADP-γ-S, 0.2 mM), α,β-Methyleneadenosine 5'-triphosphate (AMP-CPP, 10 mM), P1,P3-Di(adenosine-5') triphosphate ammonium salt (Ap3A, 10 mM).

**Phosphorous Nuclear Magnetic Resonance (NMR) analysis of L-threonine kinase reaction products.** Proton-decoupled $^{31}$P-NMR spectra were obtained using a Varian Unity Inova 500 MHz spectrometer (Chemical Sciences Magnetic Resonance Facility, University of Georgia) with the following parameters: pulse angle 45°, repetition delay 1 s, excitation pulse 3.88 μs, spectral width 12.11 kHz, acquisition time 0.810 s. Chemical shifts were referenced to a $H_3PO_4$ (85%) standard set to 0.0 ppm. Reaction mixtures (500 μL) consisted of HEPES (50 mM, pH 7.5) buffer, $MgCl_2$ (1 mM), ATP (0.3 mM), L-Thr (0.3 mM) and 3 μM of protein incubated at 25°C for 2 h. Reaction mixtures were brought up to a final volume of 600 μL in $D_2O$ (17%v/v). Spectra were processed with MestReNova software version v7.0.0-8333 (Mestrelab Research, Santiago de Compostela, Spain).

**Spectrophotometric kinase assay.** ATPase and kinase activities were measured using an NADH-consuming assay (48-51). All substrate stocks were made fresh. Reactions (100 μL) contained HEPES buffer (50 mM, pH 8.5), $MgCl_2$ (5 mM), PEP (3 mM), NADH (0.1 mM), pyruvate kinase (1 U), lactate dehydrogenase (1.5 U) incubated at 25°C with measurements taken every 11 sec over a 20 min period. For ATPase specific activity, L-Thr concentration was held at 50 mM while ATP concentration was varied (5 – 100 mM). To measure the effect of the co-substrate on ATPase activity, ATP concentration was held at 50 mM while the concentration L-Thr was varied (0 – 100 mM). Reactions were started by the addition of *Rs*BluE or PduX (3 μM). The absorbance at 340 nm was monitored in a 96-well plate using the Spectramax Plus UV-visible spectrophotometer (Molecular Devices) equipped with SoftMax Pro v6.2. Enzyme activities were calculated as described elsewhere (50). Specific activity data are presented with standard deviations from triplicate experiments.

## RESULTS

**BluE is the L-Thr kinase of Rhodobacterales.** Previous bioinformatics analysis of the genomes of AdoCba producing organisms (52), where unable to identify homologs of the L-Thr kinase (PduX in *S.* Typhimurium) in organisms that use the late Co insertion pathway for AdoCbl biosynthesis (Fig. 2.1A). Because of significant sequence differences, BLAST (36) searches using *S.* Typhimurium PduX do not retrieve BluE sequences and vice versa. The BluE (locus RCAP_rcc02055, accession number Z46611) gene from *R. capsulatus* was first identified by Pollich *et al.* (15) as part of a AdoCbl biosynthetic gene cluster but its function has not been elucidated. All *bluE* homologs identified thus far are clustered with other AdoCbl biosynthetic genes. BluE proteins are about 35 residues (~4 kDa) shorter than PduX from *S.* Typhimurium (Supplemental Fig. S2.2). The BluE proteins (27.1 kDa) from *R. capsulatus* SB1003 and *R. sphaeroides* (locus RSP_0788, 28.0 kDa) have a 52% end-to-end identity, but only a 19% identity to *S.* Typhimurium PduX (32.8 kDa). This falls below the threshold of most default BLAST searches, and is likely the reason it was overlooked in previous bioinformatics analysis of AdoCbl genes (52). Despite the low sequence identity, and notable gaps in the sequence of BluE relative to PduX, the sequence alignment revealed multiple conserved residues. Several of these residues were previously identified as important for PduX function (17, 18) (Fig. S2.2). A phylogenetic tree was constructed from *S.* Typhimurium PduX and BluE from *R. sphaeroides* and *R. capsulatus* protein sequences (Fig. 2.2). There are three distinct phylogenetic clusters. At the bottom of this tree are closely related homologs of *S.* Typhimurium PduX. This cluster is also made up of organisms that possess the cobalt chelatase CbiK (EC 4.99.1.3, red squares) and therefore uses the early Co insertion (anaerobic) pathway for AdoCbl biosynthesis (Fig. 2.1A). At the top of the tree is a cluster containing organisms that possess BluE homologs. This BluE cluster contains members only from the order Rhodobacterales, with one exception from the Rhizobiales. This cluster also contains only organisms that contain the cobalt chelatase, CobN (EC 6.6.1.2, green squares), which is a marker for the late cobalt insertion (aerobic) pathway. Rhodobacterales are known to use the late Co insertion pathway (reviewed in (4)). The one exception in the cluster being a *Roseobacter* species, which appears to lack all the gene of the early *de novo* biosynthetic pathway and must therefore

salvage intermediates of the late steps of the pathway such as cobyric acid (Cby) or cobinamide (Cbi) in order to assemble AdoCbl. The middle cluster contains Firmicutes, a Phylum that employs a broad array of strategies for AdoCbl biosynthesis and includes members that use either the early or late Co insertion pathway. However, all the Firmicutes possess homologs of *S.* Typhimurium PduX rather than the Rhodobacterales BluE. The Rhodobacterales that possess BluE also encode the 5,6-dimethylbenzimidazole (DMB) synthase, BluB (EC 1.14.99.40, purple squares) (19, 20). Implications of this genetic affiliation are explored in the Discussion section.

***R. capsulatus* and *R. sphaeroides* BluE restores AdoCbl-dependent growth of a *S.* Typhimurium *pduX* strain.** We used a *S.* Typhimurium *pduX* strain to verify BluE had L-Thr kinase activity *in vivo*. It has been observed by us and others (17), that when grown on minimal glycerol medium at high levels of Cby (> 10 nM) a *S.* Typhimurium *metE pduX* strain does not display a robust growth phenotype relative to wild type. One reason for this is very low levels of AdoCba are required for the Cba-dependent methionine synthesis (MetH) to satisfy the needs of the cells. It is estimated that as low as 10-25 molecules of Cba per cell are needed (53). It also suggests the presence of at least one other enzyme with L-Thr kinase activity within *S.* Typhimurium. This alternate kinase is only cable of recovering growth of a *pduX* strain when intracellular levels of Cby are high. Likewise when cells are grown under conditions that demand a high level of AdoCbl production, such as growth on ethanolamine as a carbon and energy source, a *pduX* strain is not able to recover (Fig. 2.3 & 2.4B) (17). It is estimated that as much as 500 molecules of AdoCba per cell (54) are required for growth on ethanolamine, which is an AdoCba-dependent process (55).

Figure 2.3 shows growth analysis of *pduX* strains carrying plasmids encoding *R. sphaeroides* locus RSP_0788 (p*RsbluE*, open square), *R. capsulatus* locus RCAP_rcc02055 (p*RcbluE*, closed squares), *S.* Typhimurium PduX (p*SepduX*, closed triangle) and the corresponding controls. The growth medium contains ethanolamine (90 mM) as the sole carbon and energy source supplemented with the cobamide precursors Cby (300 nM) and DMB (150 µM) (Fig. 2.1). *Rc*BluE was capable of complementing a *pduX* strain to the same level as wild-type *S.* Typhimurium (3.7 h). The strains expressing *Rs*BluE or *Se*PduX

have a much shorter lag time (12 and 16 h, respectively) than the wild-type empty vector only control (VOC) strain or the strain expressing *Rc*BluE (26 h). These results demonstrate that BluE from *R. sphaeroides* and *R. capsulatus* can substitute for PduX *in vivo* in *S.* Typhimurium. Using sequence alignments of both PduX and BluE homologs (Supplemental Fig. S2.2) we targeted conserved residues for substitution. We isolated a variant, *Rs*BluE$^{G99A}$, which was unable to complement a *pduX* strain under the low (Fig. 2.4A) or high (Fig. 2.4B) AdoCbl demanding conditions. This variant was purified and used to assess the enzymatic activity relative to wild-type *Rs*BluE *in vitro*.

**Purification of *Rs*BluE and PduX.** Overproduction of *Rs*BluE and PduX resulted in the production of substantial quantities (~12 mg/g cells) of insoluble protein for both. Several strains of *E. coli* overexpression strains (BL21, BL21/RIL, C41, C43, MDS 42, and Lemo21), at various temperatures (15, 25, 37ºC), with different lysis methods (chemical lysis, French press, and sonication) were tried with little success. The fusion of tags to aid is solubility such as maltose binding-protein (MBP) resulted in the inability to detect protein production either due to poor expression or proteolysis prior to cell harvesting. Several different detergents were screened in an attempt to solubilize these proteins; 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS), *n*-dodecyl-β-*D*-maltopyranoside (DDM), Fos-choline-16 (FOS16), Triton-X, sarkosyl, and sodium dodecyl sulfate (SDS). The proteins were only soluble in SDS or sarkosyl. Attempts to completely remove the detergent after solubilization resulted in protein precipitation. A concentration of ~1% sarkosyl or SDS had to be maintained to keep the proteins in solution. *N*-terminally $His_6$ tagged protein was overproduced and solubilized and purified to ~70% purity with the detergent sarkosyl as described in the Methods and Materials section. The sarkosyl concentration was reduced to ~1% by dialysis. Figure 2.5A shows a representative SDS-PAGE gel of purified *Rs*BluE and PduX. The proteins were found to be active in the presence of ~1% sarkosyl.

***R. sphaeroides* BluE has ATPase activity *in vitro*.** Figure 2.5B shows a representative results of an ATP activity assay using ADP-glo™ Kinase Assay Kit (Promega) with biological replicates in triplicate. *Rs*BluE and PduX exhibit comparable ATPase activity when provided ATP (10 mM) and L-threonine (L-Thr, 50 mM). ATP to ADP conversion was quantified by comparison to a standard curve (Supplemental

Fig. S2.1). PduX converted 100% of the ATP to ADP, *Rs*BluE$^{WT}$ converted 61% of the ATP to ADP and the *Rs*BluE$^{G99A}$ variant converted 21% after 1 h incubation at 25ºC. There was residual 8% ATPase activity detected in the empty vector only control (VOC). With L-serine (L-Ser) as a co-substrate PduX converted 90% of the ATP to ADP, *Rs*BluE$^{WT}$ converted 51%, and the *Rs*BluE$^{G99A}$ variant converted 30%. There was no activity detected in the vector only control. These data seem to imply that PduX from *S.* Typhimurium is a more efficient enzyme than *Rs*BluE *in vitro*. These data also indicate that both PduX and *Rs*BluE are able to use L-Ser as substrates to generate L-serine phosphate (L-Ser-P). This is the first time L-Ser has been shown to be a substrate for PduX or *Rs*BluE. L-Ser-P can be dephosphorylated by CobD (56, 57) to generate ethanolamine phosphate (EA-P) which can be incorporated as the linker in the corrin ring to generate norCbas (58).

***R. sphaeroides* BluE generates L-threonine-*O*-3-phosphate *in vitro*.** We used $^{31}$P-NMR to verify that *Rs*BluE generates ADP from ATP, and the phosphate from ATP is transferred to L-Thr to generate L-Thr-P. PduX was included as a positive control as this enzyme and been shown to be a genuine L-Thr kinase which generates ADP and L-Thr-P as products (17, 18). Complete conversion of ATP to ADP was observed in the reaction containing PduX, as evidenced by the disappearance of the peak corresponding to the β-phosphate of ATP with a chemical shift of -18.8 ppm (Fig. 2.6C). This peak is present in the reaction containing *Rs*BluE (Fig. 2.6D), indicating that not all of the ATP was converted to ADP. This is in agreement with other *in* vitro data (Fig. 2.5B) that *Rs*BluE is not as efficient as PduX. The appearance of a peak with a chemical shift at 2.7 ppm in both the PduX and *Rs*BluE reactions indicates the transfer of the β-phosphate from ATP to L-Thr to generate L-Thr-P. This chemical shift for L-Thr-P is consistent with what has been reported by Fan *et. al.* (17) (2.2 ppm). However, the chemical shift of this new peak at 2.7 ppm is inconsistent when compared to the chemical shift for the L-Thr-P standard at 3.7 ppm (Fig. 2.6B). We suspect the difference in the chemical shift between the L-Thr-P standard and the L-Thr-P product of the kinase reaction is due to continued binding of the product by the enzyme, or interaction of the product with sarkosyl which is present in the reactions but not in the L-Thr-P standard. A further explanation for this chemical shift is discussed below.

**ATP is the limiting substrate for PduX and *Rs*BluE.** We measured the specific activity of PduX and *Rs*BluE using a pyruvate kinase/lactate dehydrogenase coupled assay that measures the consumption of NADH, described in Methods and Materials (48) (Fig. 2.7). Shown are the specific activities of *Rs*BluE (Fig. 2.7A) and PduX (Fig. 2.7B) as a function of ATP, L-Thr (Fig. 2.7C, D), and L-Ser (Fig. 2.7E, F) concentration. L-Thr is held constant at 50 mM as the ATP concentration is increased or the ATP concentration is held at 50 mM while L-Thr or L-Ser concentrations are increased. The specific activities are reported in Table 2.1. PduX and *Rs*BluE have comparable specific activities for ATP at 1.0 and 0.93 μM ATP per minute per mg of protein. However complete substrate saturation was not reached. The concentration of L-Thr as the co-substrate did not change the rate of ATP consumption (Fig. 2.7 C & D). Even when no L-Thr was present both enzymes hydrolyzed ATP to ADP. This is also illustrated in supplemental Figure S2.3 G and F. The same is true for PduX when L-Ser is the co-substrate, however there does appear to be a slight decline in activity for *Rs*BluE when L-Ser is the co-substrate. Supplemental Figure S2.3-F shows a more dramatic decline in activity as the concentration of L-Ser is increased. The addition of L-Thr as a co-substrate does reduce the overall specific activity of both enzymes by 9% for PduX and 19% for *Rs*BluE. In light of the fact that these enzymes hydrolyze ATP in the absence of a co-substrate, the slight decline in specific activity is likely reflective of the enzymes slowing down in order to bind and transfer the phosphate from ATP to L-Thr to generate the desired product L-Thr-P. The effect of L-Ser seems quite different. PduX specific activity declines by 35% relative to the no co-substrate activity and 29% relative to the activity with L-Thr as the co-substrate. *Rs*BluE specific activity declines by 55% relative to the no co-substrate activity and 46% relative to the activity with L-Thr as the co-substrate. Both enzymes prefer L-Thr to L-Ser; however, *Rs*BluE appears to be specifically inhibited by L-Ser in a way that PduX is not.

**Optimization of *Rs*BluE activity.** The reaction conditions for the kinase reactions were optimized (Supplemental Fig. S2.3). The optimal pH was 8 in HEPES buffer (50 mM) (Fig. S2.3A). *Rs*BluE activity as a function of protein concentration shows 57% more activity at 2 μM of protein relative to 12 or 24 μM (Fig. S2.3B). Reasons for this are explored in the Discussion section. Salts were not required

for activity, and $CaCl_2$ (100 mM) decreases activity. Protein or Ca precipitation was observed (Fig. S2.3C). A divalent metal (1 mM) is required for *Rs*BluE ATPase activity. *Rs*BluE ATPase activity was optimal with $MnCl_2$ (1 mM). *Rs*BluE can use $MgCl_2$, $ZnCl_2$, $NiCl_2$ and $CaCl_2$ equally well, however, reduced activity was observed with $CoCl_2$. (Fig. S2.3D). The concentration of L-Thr does not affect the ATPase activity or *Rs*BluE. There does appear to be a slight drop in activity at 50 and 100 mM L-Thr concentration (Fig. 2.3E). Increasing the concentration of L-Ser inhibits *Rs*BluE ATPase activity (Fig. S2.3F). *Rs*BluE hydrolyzes ATP in the absence of a co-substrate (L-Thr). *Rs*BluE may be capable of using D-Ser or D-Thr as co-substrates (Fig. S2.3G), given the activity is ~28% higher with D-Ser and 10% higher with D-Thr as co-substrates relative to the ATP only sample, though 6% less than the L-Thr sample and 24% higher than the L-Ser sample. *Rs*BluE might simply perform better in the presences of either enantiomer of Thr or Ser or any amino acid without a bulky or hydrophobic side chain. The bulky amino acid L-tyrosine (L-Tyr) caused a decline in activity by 2% and the hydrophobic L-valine (L-Val) only increased activity by 3% relative to the ATP only control. Supplemental Figure S2.3H shows the activity of *Rs*BluE as a function of ATP concentration. As was expected and was shown in Figure 2.7, the relationship between ATP concentration and activity is roughly parabolic. *Rs*BluE was assayed for inhibition by known ATPase inhibitors. ADP was the most effective inhibitor followed by pyrophosphate (PPi), with a 51% and 43% decline in activity respectively. Surprisingly, monophosphate (Pi) and adenosine 5′-[γ-thio]triphosphate (ADP-γ-S) increased the activity of *Rs*BluE by 8% and 22%, respectively.

**DISCUSSION**

***bluE* from *R. capsulatus* and *R. sphaeroides* are L-Thr kinases.** BluE from *R. capsulatus* (*Rcbl*uE) and *R. sphaeroides* (*RsbluE*) both complement a *S.* Typhimurium *pduX* strain (Fig. 2.3). BluE from *R. capsulatus* complements to the same level as wild-type cells grown on ethanolamine. However, expression of *S.* Typhimurium PduX (*Se*PduX) or *R. sphaeroides* BluE from a plasmid in either a wild-type or a *pduX* strain results in a 10 h decreased lag time while maintaining a similar doubling time as the wild-type vector only control cells expressing *Rcbl*uE. Firstly, these results suggest that *Rcbl*uE is either

not as highly expressed, or is not as enzymatically efficient *in vivo* as *Rs*BluE or PduX. The decreased lag time which results from the expression of *Se*PduX and *Rs*BluE may simply be the effect of high gene dosage, or an unknown benefit resulting from increased intercellular phosphorylation of L-Thr. Likewise, the cause may be due to increased flux through the AdoCbl biosynthetic pathway resulting in a higher production of AdoCbl, resulting in more efficient catabolism of ethanolamine. This result is surprising in light of the *in vitro* data that shows *Se*PduX and *Rs*BluE will hydrolyze ATP to ADP in the absence of the amino acid co-substrate. It seems strange that the cell could benefit from the overexpression of an enzyme that is possibly cleaving ATP indiscriminately. The benefit of the overexpression of these L-Thr kinases is more complicated than can be teased apart here.

**The effect of sarkosyl solubilization on the results presented herein.** Sarkosyl solubilizes by unfolding the protein. Gentile removal of the detergent by dialysis, dilution, or other methods, allows the to protein to refold in to the proper conformation (59). High concentrations of sarkosyl can result in inactive protein. *Rs*BluE and PduX are inactive at the high concentrations of sarkosyl needed for solubilization and separation for cellular debris (~12%, see Methods and Material). The proteins begin to show activity at a concentration of ~1% but also begins to precipitate from solution at this concentration. There is a very fine line between acquiring both active and soluble protein. During purification the proteins were dialyzed into buffer without sarkosyl several times until the first signs of protein precipitation began to occur. At which time dialysis was stopped and the protein was frozen for later use. This stopping point was intended to maximize the population of refolded and active protein while maintaining the minimum concentration of sarkosyl to maintain solubility. There is likely a portion of the population of protein that is partially unfolded due to the presence of the sarkosyl. We believe the presence of the sarkosyl and unfolded protein may be the cause of some of the ambiguous results. When the protein is added to a reaction mixture it results in dilution of the sarkosyl and by extension the likely refolding of some portion of the protein population, and the precipitation of others depending on the volume of the reaction/dilution. The difference in activity between PduX and *Rs*BluE may be a result of slightly differing sarkosyl concentrations verses the concentration of protein in each preparation. It may not be a true reflection of

the enzyme efficiencies relative to one another. This is illustrated in supplemental Figure S2.3B. Counterintuitively, the specific activity of the enzyme increases as the concentration decreases (from 24 to 2 μM). However, the total percent conversion after 1 h incubation at 25°C is proportionally much higher at higher concentrations of protein; 81% at 24 μM, 54% at 12 μM, and 20% at 2 μM (data not shown). This contradiction might be accounted for in the following manner. At more dilute protein concentrations, the percentage of the protein population that is correctly folded is higher and more active, but very dilute. In other words, fewer but very efficient enzymes. At higher protein concentrations the concentration of sarkosyl is also high and so is the percentage of unfolded protein in the population, resulting in more inefficient enzymes but a greater number of them. The *in vivo* data support the idea that both enzymes are very active and efficient inside the cell, as they complement a *S.* Typhimurium *pduX* strain to better than wild-type levels (Fig. 2.3 & 2.4). The presence of sarkosyl may explain the anomalous $^{31}$P-NMR results (Fig. 2.6). We believe sarkosyl is the cause of the difference in the chemical shift in the peak that we believe to be L-Thr-P (2.7 ppm) in the PduX and *Rs*BluE reactions relative to the L-Thr-P standard (3.6 ppm). We believe the L-Thr-P product is interacting directly with the sarkosyl in the reaction mixtures, or remaining bound to the partially unfolded protein. There are very few other possibilities to account for the appearance of this peak at 2.7 ppm. $^{31}$P-NMR only detects phosphorylated compounds, and ATP is the only source of phosphate in the reactions. Only two substrates were provided, ATP and L-Thr, and one product was confirmed to be ADP, leaving L-Thr-P as the most likely other product. This peak is not AMP, which has a chemical of 4.4 ppm under these conditions.

**BluE is not the L-Thr kinase of the late Co insertion (aerobic) pathway.** At first glance it appears that BluE is primarily found in the organisms with CobN, the cobalt chelatase of the late Co insertion (aerobic) pathway (Fig 2.2). BLAST searches using *S.* Typhimurium PduX pulls up PduX homologs primarily in organisms that use the early Co insertion (anaerobic) pathway, with only a few exceptions within the Firmicutes Phylum. However, the Firmicutes represent a Phylum with many exceptions among AdoCbl producing organisms and they do not fit neatly within the two primary AdoCbl pathways. Some members of the Firmicutes such as *Clostridium difficile, Mahella australiensis, Listeria monocytogenes,*

*Eubacterium limosum, Desulfotomaculum carboxydivorans,* and *Veilonella parvula* appear to utilize the early Co insertion pathway, while others such as *Thermincola sp., Thremacetogeniuym phaeum,* and *Desulfitobacterium hafnienes* possess the genes for the late Co insertion pathway. BluE appears to be restricted only to the order Rhodobacterales, and is not found in any other group that uses the late or early Co insertion pathway.

**The BluE protein is the L-Thr kinase of a subgroup of Rhodobacterales but not of purple photosynthetic bacteria in general.** While BluE is present in *R. sphaeroides* and *R. capsulatus* it is not found in other AdoCba producing purple photosynthetic bacteria order such as *Rhodopseudomonas palustris* (Order Rhizobiales) or *Rhodospirillum ruburm* (Order Rhodospirillales). BluE is also absent in other Rhodobacterales such as *Silicibacter pomeroyi* and *Jannashia*. No homologs of PduX or BluE have been identified in these other purple photosynthetic bacteria or in archaea. This leaves open the question of how do these organisms produce L-Thr-P. It is not clear what evolutionary or metabolic distinction might be responsible for the restriction of BluE to this small group of Rhodobacterales.

**BluE is the L-Thr kinase of a subgroup of Rhodobacterales with a strict requirement for AdoCbas with AP-P as the nucleotide linker and DMB as the ribotide base; AdoCbl.** Rhodobacterales which possess BluE also encode the 5,6-dimethylbenzimidazole (DMB) synthase, BluB (EC:1.13.11.79) (19, 20) (Fig. 2.2). Organisms that encode BluB have a strict requirement for DMB as the lower ligand base and do not synthesize or use Cbas with other bases (19, 20, 60). Several of these Rhodobacterales also encode CbiZ, an adenosylcobinamide amidohydrolase (EC 3.5.1.90) (61-63), which cleaves any lower ligand that does not have DMB as the base, allowing the remodeling of salvaged Cbas with the preferred DMB base as the lower ligand. CbiZ hydrolysis occurs at the amide group in the nucleotide linker, removing the linker (AP-P, EA-P, AP, or EA) along with any attached lower ligand. Other organisms have more flexibility with the Cbas they synthesize and use. *S.* Typhimurium can substitute bases such as adenine or *2*-methyladenine for the lower ligand to produce the Cbas, pseudoCbl or Factor A, respectively. Other organisms substitute the 1-amino-2-propanol-*O*-2-phosphate (AP-P) moiety linker with ethanolamine phosphate (EA-P), which generates a linker that is shorter by one methyl group. The

resulting cobamide is norCbl (58). The presence of both BluE and BluB in these Rhodobacterales implies a strict requirement for the Cba with DMB as the base and AP-P as the nucleotide linker, known as adenosylcobalamin (AdoCbl). Animals also have this requirement for only AdoCbl.

**Conclusions.** PduX was the first enzyme reported to phosphorylate free L-Thr. PduX is annotated as a member of the GHMP kinase family (64). Here we add BluE from the Rhodobacterales as a new member of that family. Further investigation of BluE may reveal mechanistic difference with PduX. There are several gaps in the in late Co insertion AdoCba biosynthesis pathway, and within AdoCbl producers like *R. sphaeroides* which appear to have strict configuration requirements for the Cbas they synthesize and use. Here we have identified BluE as the L-Thr kinase for some of these Rhodobacterales. How other organisms that do not possess *bluE* or *pduX* generate L-Thr-P remains an open question for investigation.

**ACKNOWLEDGMENTS**

This work was supported by NIH grant R37 GM40313 to J.C.E.-S. N.K.T. was supported in part by NIH grant F31 GM095230 and Advanced Opportunity Fellowship awarded by the Graduate School of the University of Wisconsin, Madison. We thank Paul Renz (Universität Stuttgart) for his gift of CNCby, Tim Donohue (University of Wisconsin-Madison) for providing *R. sphaeroides* strain 2.4.1 and *R. capsulatus* strain SB1003, (University of Wisconsin-Madison) and Paul Merchant for technical assistance. Thanks also to D. Cui at the University of Georgia Chemical Sciences Magnetic Resonance facility for her assistance with NMR. We are grateful to Paula Pappalardo (University of Georgia) for assistance with R and phylogenetic analysis.

**FIGURES**

**Table 2.1. Specific activity.** Specific activity values for *Rs*BluE and *Se*PduX enzyme were assayed for ATPase activity in the presence and absence of L-Thr or L-Ser as a cosubstrates. Values are reported as mean ± standard deviation of three activity measurements of activity at 100 mM ATP and 50 mM L-Thr or L-Ser. Activity was measured with a NADH consumption assay (see *materials and methods*).

| **Protein** | **ATP** (μmol ATP $min^{-1}$ $mg^{-1}$) | **L-Thr** (μmol ATP $min^{-1}$ $mg^{-1}$) | **L-Ser** (μmol ATP $min^{-1}$ $mg^{-1}$) |
|---|---|---|---|
| *Rs*BluE | 0.93 ± 0.02 | 0.77 ± 0.02 | 0.42 ± 0.01 |
| *Se*PduX | 1.00 ± 0.03 | 0.91 ± 0.03 | 0.65 ± 0.02 |

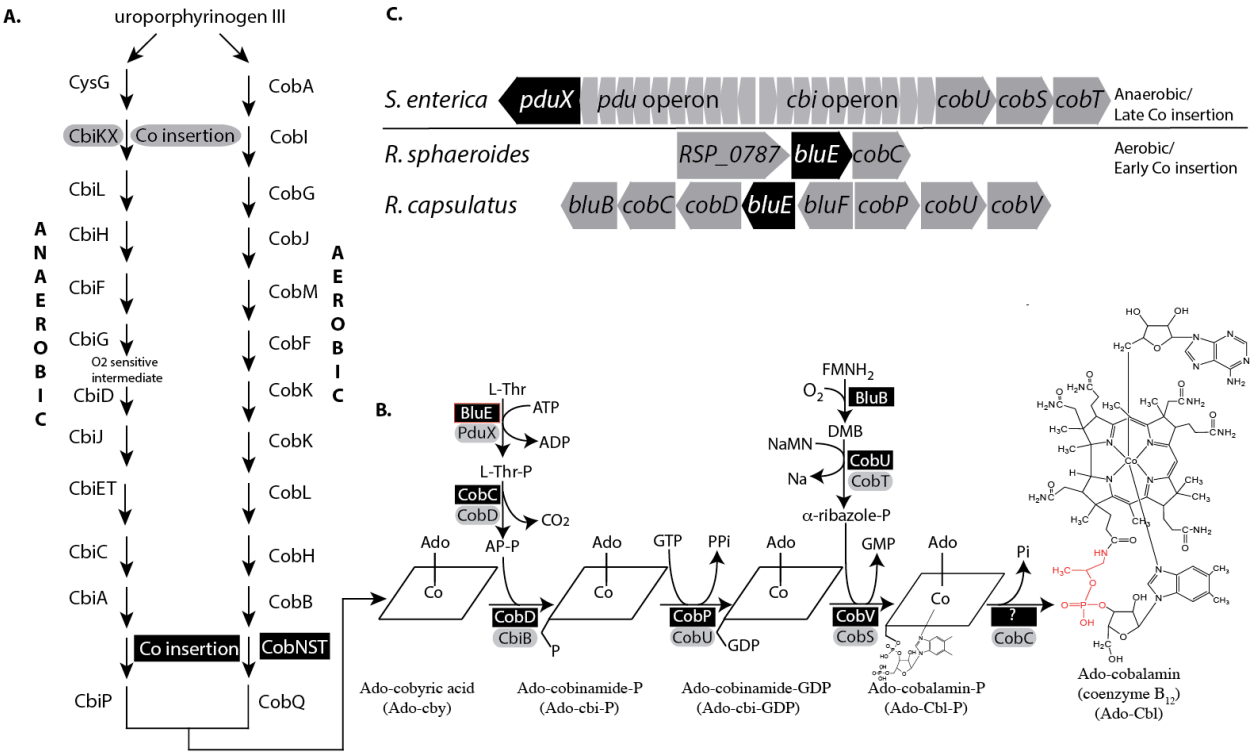



**Figure 2.1**. **Cbl biosynthetic pathway in bacteria and archaea. A.** Schematic of the early (anaerobic) and late (aerobic) cobalt insertion pathways of Cbl biosynthesis. **B.** Late steps of the $B_{12}$ biosynthetic pathway, with proteins in the aerobic/early (black) and anaerobic/late (gray) cobalt insertion pathways. BluE enzyme boxed in red. (*R*)-*1*-amino-*2*-propanol *O*-*2*-phosphate (AP-P) moiety highlighted in red. **C.** Genetic layout of the L-threonine kinase encoding genes *bluE* in the aerobic/late cobalt insertion $B_{12}$ synthesizing bacteria *R. sphaeroides* and *R. capsulatus*, and *pduX* in *S.* Typhimurium an anaerobic/early cobalt insertion Cbl synthesizing bacteria.

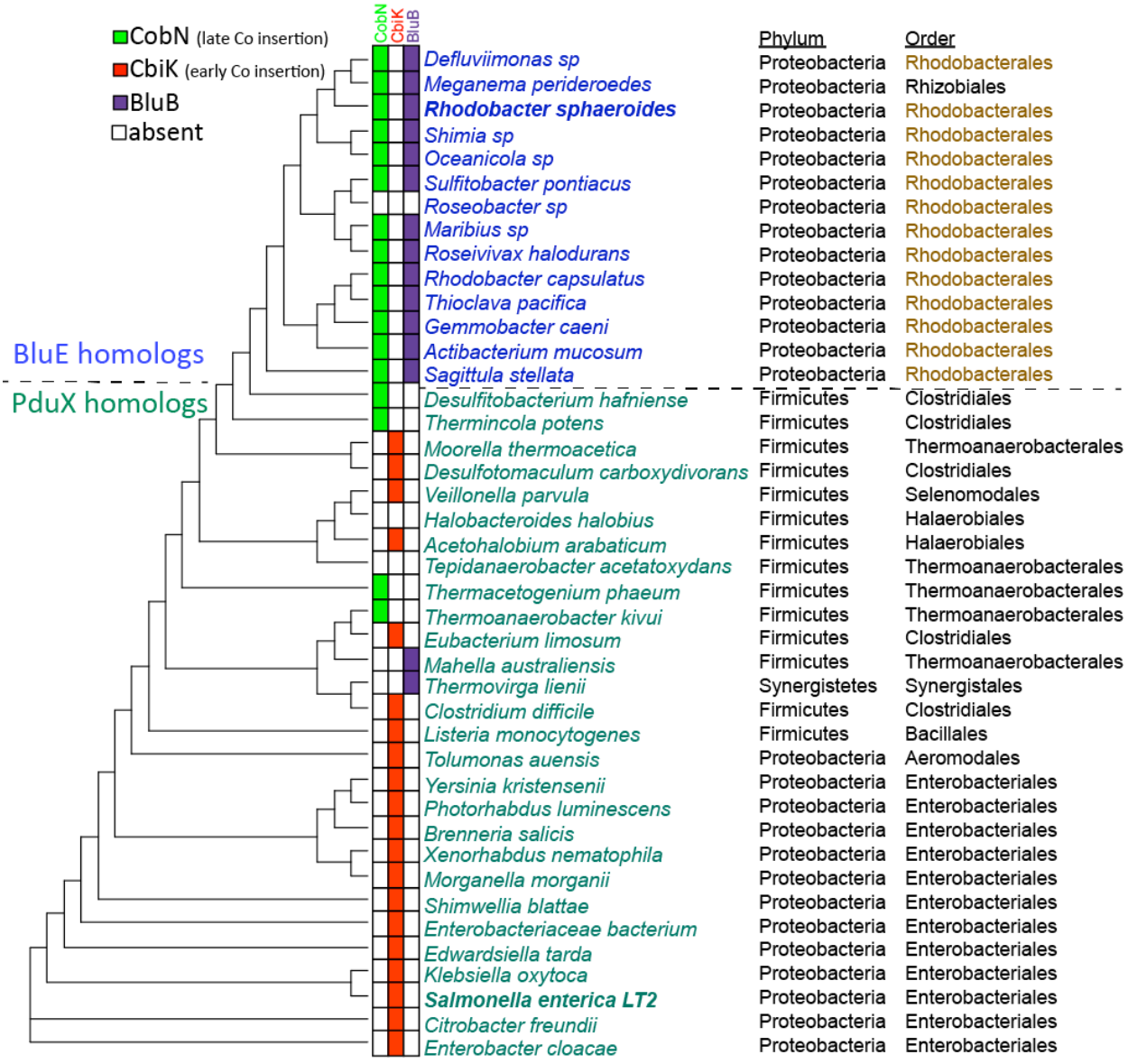



**Figure 2.2. Phylogenetic analysis of the distribution of PduX and BluE proteins.** Maximum likelihood phylogenetic tree of homologues base on amino acid sequence of PduX (green) and BluE (blue). Order Rhodobacterales is highlighted in brown for BluE containing organisms. Color-coded table of the presence or absence of the cobalt chelatase CbiK (red squares), which is indicative of the early Co insertion pathway, CobN (green squares) which is present in organisms that use the late Co insertion pathway, and BluB (blue squares) which is indicative of $O_2$-dependent DMB synthesis and physiological reliance on $B_{12}$ with DMB as the lower base.

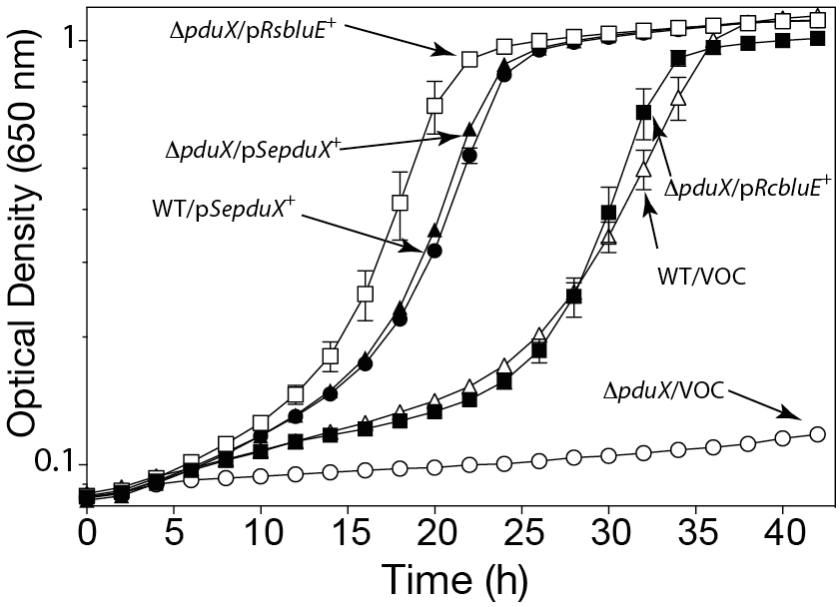



**Figure 2.3. Complementation of *S.* Typhimurium *pduX* strain.** Growth analysis of *S.* Typhimurium wild-type (WT) and *pduX* strains with plasmids expressing BluE from *R. sphaeroides* (*RsbluE*), *R. capsulatus* (*RcbluE*), PduX from *S.* Typhimurium (*SepduX*) or pBAD24 empty vector only control (VOC). Cells grown aerobically at 37ºC in NCE minimal medium with ethanolamine (90 mM) as the sole carbon and energy source, supplemented with Cby (300 nM), DMB (150 μM), arabinose (500 μM), ampicillin (100 μg/mL), $MgSO_4$ (1 mM), and Fe-citrate (50 μM).

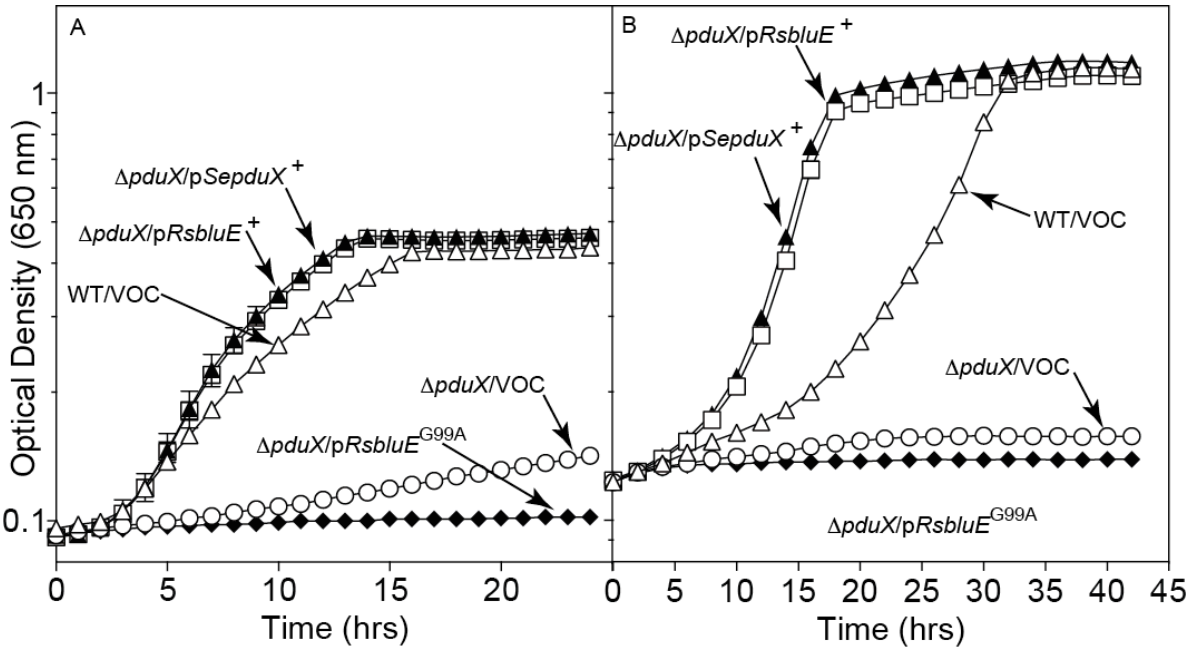



**Figure 2.4. Complementation of *S.* Typhimurium *pduX* strain.** Growth analysis of *S.* Typhimurium cells grown aerobically at 37ºC in NCE minimal medium with **A.** glycerol (22 mM) and Cby (1 nM) or **B.** ethanolamine (90 mM) and Cby (300 nM), supplemented with, DMB (150 μM), ampicillin (100 μg/mL), $MgSO_4$ (1mM).

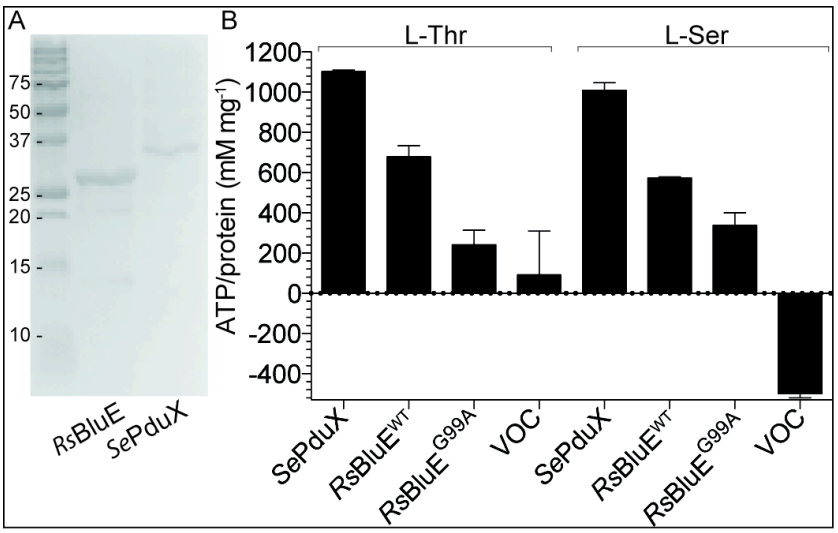



**Figure 2.5. *In vitro* assay *Rs*BluE of ATPase activity. A.** SDS-PAGE gel of purified BluE protein from *R. sphaeroides* (*Rs*BluE) and PduX from *S.* Typhimurium (*Se*PduX). **B.** ATPase activity assay performed with ADP-glo Kinase Assay Kit (Promega). Reaction mixture contained HEPES buffer (50 mM, pH 7), $MgCl_2$ (1 mM), ATP (10 mM), L-Thr or L-Ser (10 mM), protein (12 μM) incubated at 25°C for 1 h. Vector only control (VOC) contains protein extracts from sarkosyl solubilized cell pellet expressing the empty vector pTEV5.

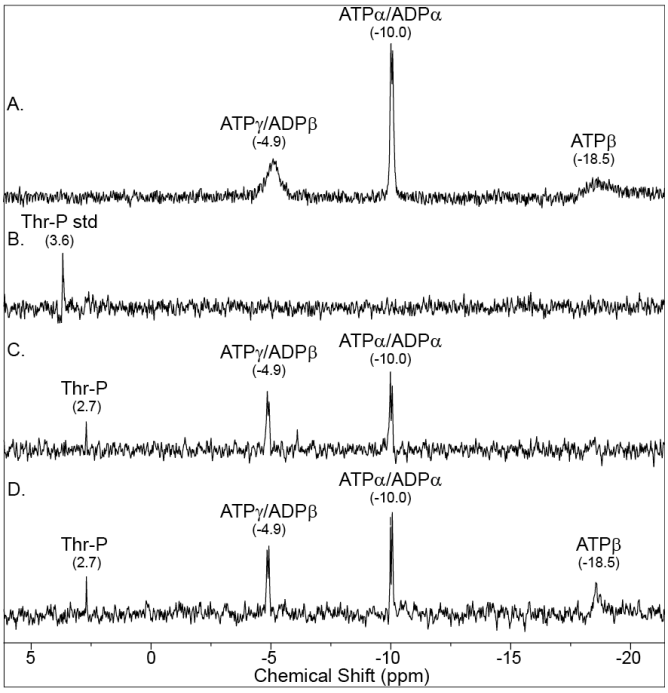



**Figure 2.6. $^{31}$P-NMR spectra of *Rs*BluE reaction.** Reactions mixtures containing $MgCl_2$ (1 mM), ATP (0.3 mM), L-Thr (0.3 mM) and purified sarkosyl solubilized protein (12 μM) were incubated for 1 hr at 25ºC. Each peak is labeled with the chemical shift value and the substrate that it represents based on the chemical shifts of the standards. **A.** ATP standard. **B.** L-threonine-*O*-3-phosphate-P (L-Thr-P) standard. **C.** Reaction containing ATP, L-Thr, and *Se*PduX. **D.** Reaction containing ATP, L-Thr, and *Rs*BluE.

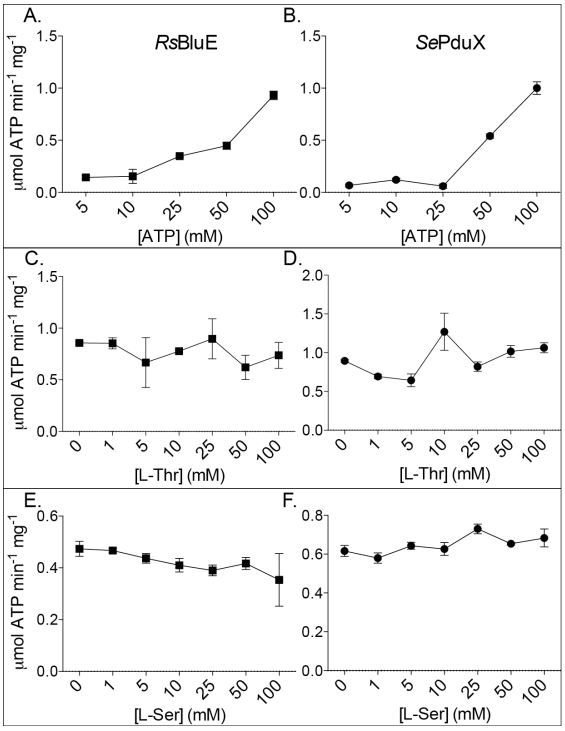



**Figure 2.7. Specific activity.** Specific activity of *R. sphaeroides* BluE (*Rs*BluE) and *S.* Typhimurium (*Se*PduX) proteins as a function of **A.** & **B.** ATP concentration **C.** & **D.** L-threonine (L-Thr) and **E.** & **F.** L-serine (L-Ser) concentration, expressed as µmol of ATP per min per mg of protein with the mean standard error of triplicate reactions represented by the error bars. Activity was measured in by an NADH consumption assay described in *materials and methods*. Assay was performed with purified and sarkosyl solubilized *Rs*BluE and *Se*PduX (12 µM) enzyme. For ATPase specific activity L-Thr concentration was held at 50 mM while ATP concentration was varied (5 – 100 mM). To measure the effect of the co-substrate for ATPase activity, ATP concentration was held at 50 mM while the concentration L-Thr or L-Ser was varied (0 – 100 mM).

**SUPPLEMENTAL FIGURES**

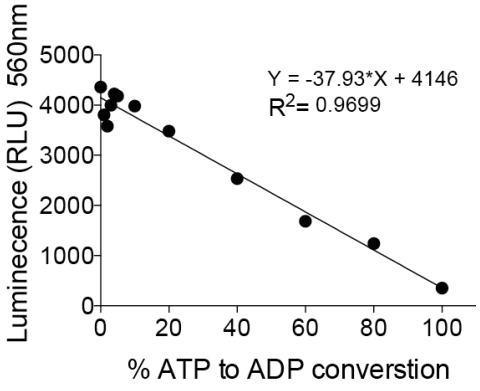



**Figure S2.1. Standard curve.** Standard curve used to calculate the percent conversion of ADP to ATP using the ADP-glo Kinase Assay Kit (Promega). RLU; relative light units. Luminescence measured at wavelength 560 nm.

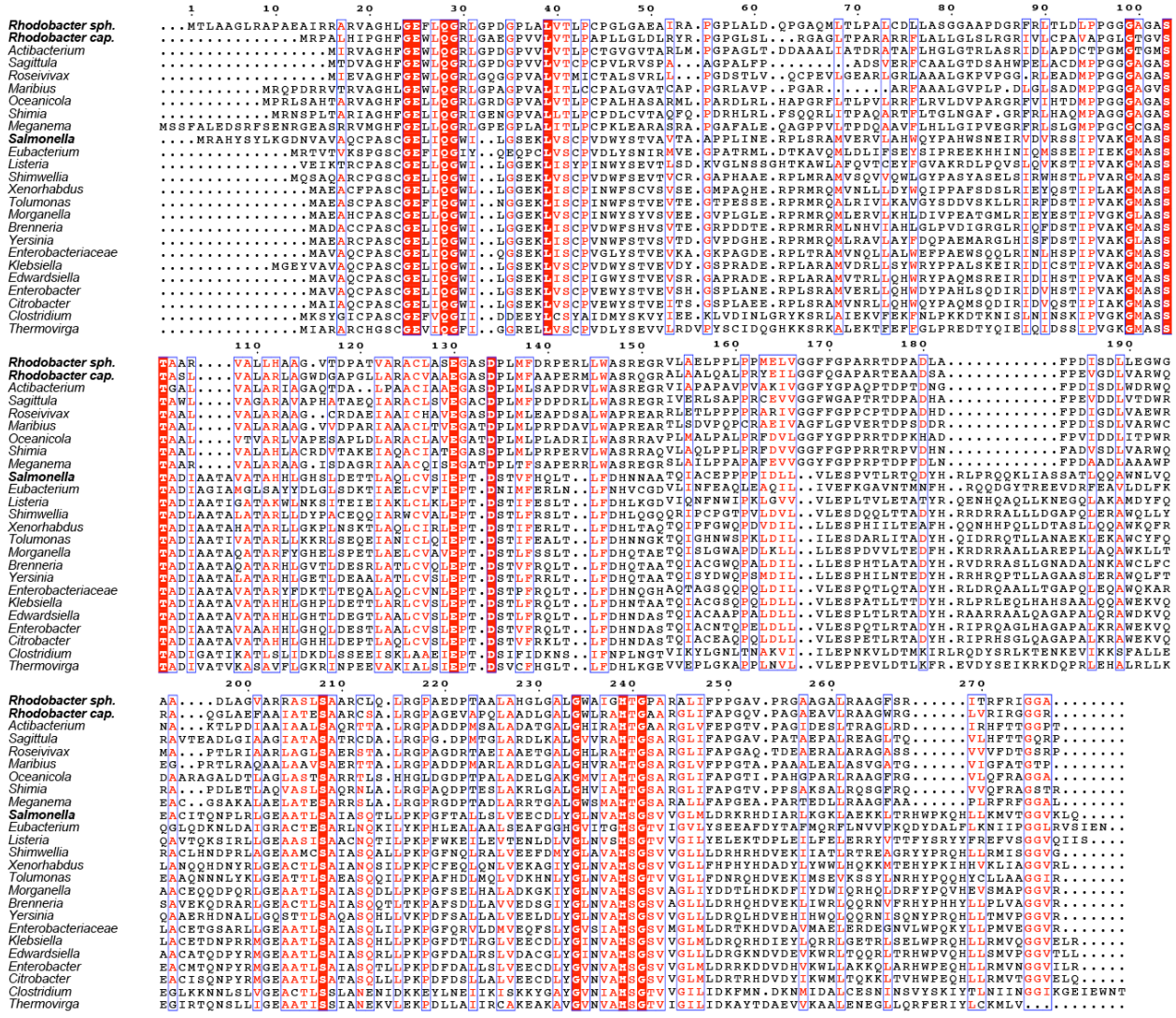



**Figure S2.2**. **Sequence alignment of representative PduX and BluE proteins.** Conserved residues are highlighted in red. Residues with similar properties are boxed in blue.

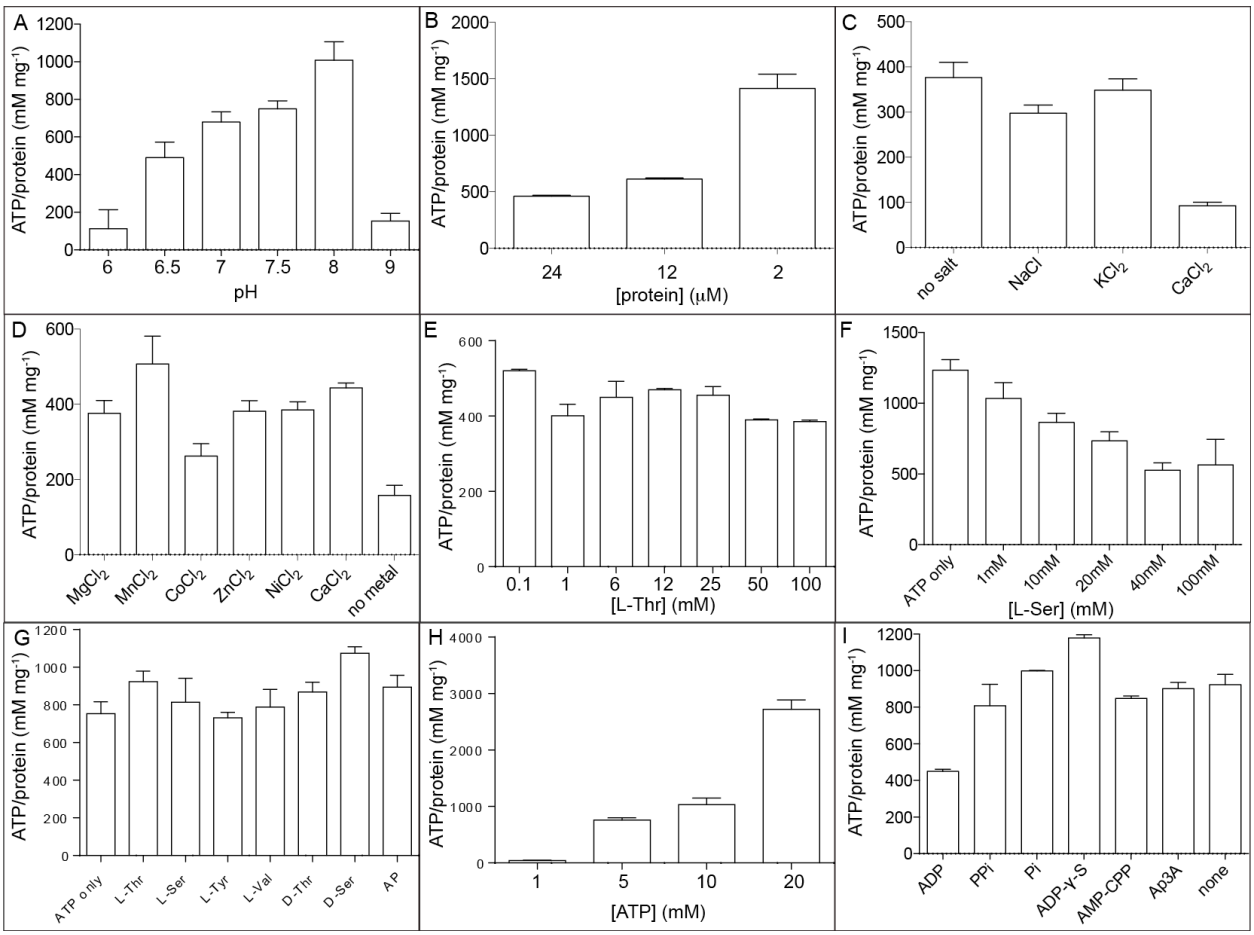



**Figure S2.3. Optimization of *Rs*BluE reaction conditions.** ATPase activity assay measured with ADP-glo Kinase Assay Kit (Promega), expressed as mM of ATP $mg^{-1}$ of protein, with the mean standard error of triplicate reactions represented by the error bars. Unless otherwise indicated, reaction mixture contained HEPES buffer (50 mM, pH 7), $MgCl_2$ (1 mM), ATP (10 mM), L-Thr (10 mM), protein (12 μM) in incubated at 25°C for 1 h. **A.** *Rs*BluE activity as a function of pH. **B**. *Rs*BluE activity as a function of protein concentration (2 – 24 mM). **C**. *Rs*BluE activity in the presence of added salts (100 mM). **D**. *Rs*BluE activity as a function of added divalent metals (1 mM). **E**. *Rs*BluE activity as a function of L-threonine (L-Thr) concentration (0.1 – 100 mM). **F**. *Rs*BluE activity as a function of L-serine (L-Ser) concentration (0 – 100 mM). **G**. *Rs*BluE activity as a function of added L- and D- amino acids (10 mM). **H**. *Rs*BluE activity as a function of ATP concentration (1 - 20 mM) with L-Thr (50 mM). **I**. *Rs*BluE activity in the presence of ATPase/kinase inhibitors ADP (20 mM), sodium pyrophosphate (PPi, 10 mM), sodium phosphate monobasic (Pi, 10mM), Adenosine 5′-[γ-thio]triphosphate (ADP-γ-S, 0.2 mM), α,β-methyleneadenosine 5'-triphosphate (AMP-CPP, 10 mM), P1,P3-Di(adenosine-5') triphosphate ammonium salt (Ap3A, 10 mM).

**Table S2.1. Strains and plasmids list**. *S. enterica* strains are derivatives of sv. Typhimurium strain LT2. Strains and plasmids were constructed during the course of this work unless stated otherwise.

| Strain/Plasmid | Relevant genotype | Reference Source |
|---|---|---|
| ***Salmonella enterica*** | | |
| JE7088 | *metE2702 ara-9* | Laboratory collection |
| **Derivatives of JE7088** | | |
| JE11685 | / pBAD24 *bla*$^+$ | |
| JE11686 | / pBAD24 *bla*$^+$ | |
| JE12914 | / pPDU15 *bla*$^+$ | |
| JE12686 | *ΔpduX516* / pPDU15 *bla*$^+$ | |
| JE13063 | *ΔpduX516* / pBAD30 *bla*$^+$ | |
| JE15873 | *ΔpduX516* / p*Rc*BLUE3*bla*$^+$ | |
| JE16285 | *ΔpduX516* / p*Rs*BLUE3 *bla*$^+$ | |
| JE21279 | *ΔpduX516* / pBAD24 *bla*$^+$ | |
| ***Escherichia coli*** | | |
| C43(DE3) | F$^-$ ompT gal hsd$S_B$ (rB$^-$mB) [dcm] [Ion] | (65) |
| DH5α | F$^-$/ endA1 hsdR17(rk$^-$, mk$^+$) glnV44 thi-1 recA1 gyrA96 (Nx$^R$) relA1 U169 deoR (Φ80-dlacZ M15 Δ(lacZYA-argF) phoAsupE44 relA1 | (27) |
| **Plasmids** | | |
| pPDU15 | *S.* Typhimurium *pduX*$^+$ in pBAD30 *bla*$^+$ | |
| p*Rc*BLUE3 | *R. capsulatus bluE*$^+$ in pBAD24 *bla*$^+$ | |
| p*Rs*BLUE3 | *R. sphaeroides bluE*$^+$ in pBAD24 *bla*$^+$ | |
| pPDU23 | *S.* Typhimurium *pduX*$^+$ in pTEV5 *bla*$^+$ | |
| p*Rs*BLUE4 | *R. sphaeroides bluE*$^+$ in pTEV5 *bla*$^+$ | |
| p*Rs*BLUE7 | *R. sphaeroides bluE* in (encodes *Rs*CobD$^{G99A}$) pTEV5 *bla*$^+$ | |
| pBAD24 | cloning/complementation vector *bla*$^+$ | (28) |
| pBAD30 | cloning/complementation vector *bla*$^+$ | (28) |
| pTEV5 | Cloning/overexpression vector, N-terminal rTEVcleavable $His_6$ tag *bla*$^+$ | (29) |

## REFERENCES


1. **Moore SJ, Lawrence AD, Biedendieck R, Deery E, Frank S, Howard MJ, Rigby SE, Warren MJ.** 2013. Elucidation of the anaerobic pathway for the corrin component of cobalamin (vitamin B12). Proc Natl Acad Sci U S A **110:**14906-14911.

2. **Escalante-Semerena JC.** 2007. Conversion of cobinamide into adenosylcobamide in bacteria and archaea. J Bacteriol **189:**4555-4560.

3. **Escalante-Semerena JC, Warren MJ.** 2008. Biosynthesis and Use of Cobalamin ($B_{12}$). *In* Böck A, Curtiss III R, Kaper JB, Karp PD, Neidhardt FC, Nyström T, Slauch JM, Squires CL (ed), *EcoSal - Escherichia coli* and *Salmonella*: cellular and molecular biology. ASM Press, Washington, D. C.

4. **Warren MJ, Raux E, Schubert HL, Escalante-Semerena JC.** 2002. The biosynthesis of adenosylcobalamin (vitamin B12). Nat Prod Rep **19:**390-412.

5. **Blanche F, Couder M, Debussche L, Thibaut D, Cameron B, Crouzet J.** 1991. Biosynthesis of vitamin B12: stepwise amidation of carboxyl groups *b, d, e*, and *g* of cobyrinic acid *a,c*-diamide is catalyzed by one enzyme in *Pseudomonas denitrificans*. J Bacteriol **173:**6046-6051.

6. **Cameron B, Guilhot C, Blanche F, Cauchois L, Rouyez MC, Rigault S, Levy-Schil S, Crouzet J.** 1991. Genetic and sequence analyses of a *Pseudomonas denitrificans* DNA fragment containing two *cob* genes. J Bacteriol **173:**6058-6065.

7. **Cameron B, Blanche F, Rouyez MC, Bisch D, Famechon A, Couder M, Cauchois L, Thibaut D, Debussche L, Crouzet J.** 1991. Genetic analysis, nucleotide sequence, and products of two *Pseudomonas denitrificans cob* genes encoding nicotinate-nucleotide: dimethylbenzimidazole phosphoribosyltransferase and cobalamin (5'-phosphate) synthase. J Bacteriol **173:**6066-6073.

8. **Crouzet J, Cameron B, Cauchois L, Rigault S, Rouyez MC, Blanche F, Thibaut D, Debussche L.** 1990. Genetic and sequence analysis of an 8.7-kilobase *Pseudomonas denitrificans* fragment carrying eight genes involved in transformation of precorrin-2 to cobyrinic acid. J Bacteriol **172:**5980-5990.

9. **Debussche L, Couder M, Thibaut D, Cameron B, Crouzet J, Blanche F.** 1991. Purification and partial characterization of cob(I)alamin adenosyltransferase from *Pseudomonas denitrificans*. J Bacteriol **173:**6300-6302.

10. **Thibaut D, Couder M, Crouzet J, Debussche L, Cameron B, Blanche F.** 1990. Assay and purification of S-adenosyl-L-methionine:precorrin-2 methyltransferase from *Pseudomonas denitrificans*. J Bacteriol **172:**6245-6251.

11. **Zumft WG.** 1997. Cell biology and molecular basis of denitrification. Microbiol Mol Biol Rev **61:**533-616.

12. **Biel AJ.** 1992. Oxygen-regulated steps in Rhodobacter capsulatus tetrapyrrole biosynthetic pathway. J BActeriol **174:**5272-5274.

13. **Heldt D, Lawrence AD, Lindenmeyer M, Deery E, Heathcote P, Rigby SE, Warren MJ.** 2005. Aerobic synthesis of vitamin B12: ring contraction and cobalt chelation. Biochem Soc Trans **33:**815-819.

14. **McGoldrick HM, Roessner CA, Raux E, Lawrence AD, McLean KJ, Munro AW, Santabarbara S, Rigby SE, Heathcote P, Scott AI, Warren MJ.** 2005. Identification and characterization of a novel vitamin B12 (cobalamin) biosynthetic enzyme (CobZ) from *Rhodobacter capsulatus*, containing flavin, heme, and Fe-S cofactors. J Biol Chem **280:**1086-1094.

15. **Pollich M, Klug G.** 1995. Identification and sequence analysis of genes involved in late steps of cobalamin (vitamin B12) synthesis in Rhodobacter capsulatus. J Bacteriol **177:**4481-4487.

16. **Vlcek C, Paces V, Maltsev N, Paces J, Haselkorn R, Fonstein M.** 1997. Sequence of a 189-kb segment of the chromosome of Rhodobacter capsulatus SB1003. Proc Natl Acad Sci U S A **94:**9384-9388.

17. **Fan C, Bobik TA.** 2008. The PDUX enzyme of *Salmonella enterica* is an L-threonine kinase used for coenzyme $B_{12}$ synthesis. J Biol Chem **283:**11322-11329.

18. **Fan C, Fromm HJ, Bobik TA.** 2009. Kinetic and functional analysis of L-threonine kinase, the PduX enzyme of *Salmonella enterica*. J Biol Chem **284:**20240-20248.

19. **Gray MJ, Escalante-Semerena JC.** 2007. Single-enzyme conversion of $FMNH_2$ to 5,6-dimethylbenzimidazole, the lower ligand of $B_{12}$. Proc Natl Acad Sci U S A **104:**2921-2926.

20. **Taga ME, Larsen NA, Howard-Jones AR, Walsh CT, Walker GC.** 2007. BluB cannibalizes flavin to form the lower ligand of vitamin B12. Nature **446:**449-453.

21. **Peariso K, Zhou ZS, Smith AE, Matthews RG, Penner-Hahn JE.** 2001. Characterization of the zinc sites in cobalamin-independent and cobalamin-dependent methionine synthase using zinc and selenium X-ray absorption spectroscopy. Biochemistry **40:**987-993.

22. **Drennan CL, Huang S, Drummond JT, Matthews RG, Ludwig ML.** 1994. How a protein binds $B_{12}$: A 3.0Å X-ray structure of $B_{12}$-binding domains of methionine synthase. Science **266:**1669-1674.

23. **Hall DA, Vander Kooi CW, Stasik CN, Stevens SY, Zuiderweg ER, Matthews RG.** 2001. Mapping the interactions between flavodoxin and its physiological partners flavodoxin reductase and cobalamin-dependent methionine synthase. Proc Natl Acad Sci USA **98:**9521-9526.

24. **Taylor RT, Weissbach H.** 1973. $N^5$-methylenetetrahydrofolate-homocysteine methyltransferases, p 121-165. *In* Boyer PD (ed), The Enzymes, vol 9. Academic Press, Inc., New York.

25. **Datsenko KA, Wanner BL.** 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci USA **97:**6640-6645.

26. **Raleigh EA, Lech K, Brent R.** 1989. Selected topics from classical bacterial genetics, p 1.4. *In* Ausubel FA, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (ed), Current Protocols in Molecular Biology, vol 1. Wiley Interscience, New York.

27. **Woodcock DM, Crowther PJ, Doherty J, Jefferson S, De Cruz E, Noyer-Weidner M, Smith SS, Michael MZ, Graham MW.** 1989. Quantitative evaluation of *Escherichia coli* host strains for tolerance to cytosine methylation in plasmid and phage recombinants. Nucl Acids Res **17:**3469-3478.

28. **Guzman LM, Belin D, Carson MJ, Beckwith J.** 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J Bacteriol **177:**4121-4130.

29. **Rocco CJ, Dennison KL, Klenchin VA, Rayment I, Escalante-Semerena JC.** 2008. Construction and use of new cloning vectors for the rapid isolation of recombinant proteins from *Escherichia coli*. Plasmid **59:**231-237.

30. **Berkowitz D, Hushon JM, Whitfield HJ, Jr., Roth J, Ames BN.** 1968. Procedure for identifying nonsense mutations. J Bacteriol **96:**215-220.

31. **Balch WE, Wolfe RS.** 1976. New approach to the cultivation of methanogenic bacteria: 2-mercaptoethanesulfonic acid (HS-CoM)-dependent growth of *Methanobacterium ruminantium* in a pressurized atmosphere. Appl Environ Microbiol **32:**781-791.

32. **Bertani G.** 1951. Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. J Bacteriol **62:**293-300.

33. **Bertani G.** 2004. Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. J Bacteriol **186:**595-600.

34. **Miroux B, Walker JE.** 1996. Over-production of proteins in Escherichia coli: Mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. Journal of Molecular Biology **260:**289-298.

35. **Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A.** 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res **31:**3784-3788.

36. **Altschul SF, Madden TL, Schaffer AA, Zhang J, Miller W, Lipmann DJ.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl Acids Res **25:**3389-3402.

37. **Markowitz VM, Korzeniewski F, Palaniappan K, Szeto E, Werner G, Padki A, Zhao X, Dubchak I, Hugenholtz P, Anderson I, Lykidis A, Mavromatis K, Ivanova N, Kyrpides NC.** 2006. The integrated microbial genomes (IMG) system. Nucleic Acids Res **34:**D344-348.

38. **Edgar RC.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32:**1792-1797.

39. **Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP.** 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol **61:**539-542.

40. **Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O.** 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol **59:**307-321.

41. **Rambaut A.** 2007. FigTree, vv1.4.2. Institute of Evolutionary Biology, University of Edinburgh, Ashworth Laboratories University of Edinburgh, Edinburgh UK. http://tree.bio.ed.ac.uk/software/figtree/.

42. **Paradis E, Claude J, Strimmer K.** 2004. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics **20:**289-290.

43. **Harmon LJ, Weir JT, Brock CD, Glor RE, Challenger W.** 2008. GEIGER: investigating evolutionary radiations. Bioinformatics **24:**129-131.

44. **FitzJohn RG.** 2012. Diversitree: comparative phylogenetic analyses of diversification in R. Meth Ecol Evol **3:**1084-1092.

45. **R Core Development Team.** 2015. R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

46. **Gouet P, Courcelle E, Stuart DI, Metoz F.** 1999. ESPript: multiple sequence alignments in PostScript. Bioinformatics **15:**305-308.

47. **Zegzouti H, Zdanovskaia M, Hsiao K, Goueli SA.** 2009. ADP-Glo: A Bioluminescent and homogeneous ADP monitoring assay for kinases. Assay Drug Dev Technol **7:**560-572.

48. **Bergmeyer HU, Bergmeyer J, Grassl M, Berger R.** 1985. Methods of enzymatic analysis. vol. iv. "enzymes 2: Esterases, glycosidases, lyases, ligases". 3rd Edition. Weinheim; Deerfield Beach, Florida; Basel: Verlag Chemie, 1984. 426 S., 258 DM. Acta Biotechnologica **5:**114-114.

49. **Horswill AR, Escalante-Semerena JC.** 2002. Characterization of the propionyl-CoA synthetase (PrpE) enzyme of *Salmonella enterica*: Residue Lys592 is required for propionyl-AMP synthesis. Biochemistry **41:**2379-2387.

50. **Wilson ACaABP.** 1962. Regulation of flavin synthesis by *Escherichia coli*. J Gen Microbiol **28:**283-303.

51. **Havemann GD, Bobik TA.** 2003. Protein content of polyhedral organelles involved in coenzyme B12-dependent degradation of 1,2-propanediol in *Salmonella enterica* serovar Typhimurium LT2. J Bacteriol **185:**5086-5095.

52. **Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS.** 2003. Comparative genomics of the vitamin $B_{12}$ metabolism and regulation in prokaryotes. J Biol Chem **278:**41148-41159.

53. **Bradbeer C.** 1982. Cobalamin transport in microorganisms., p 31-56. *In* Dolphin D (ed), B12, vol 2. John Wiley & Sons, New York.

54. **Bradbeer C.** 1965. The clostridial fermentation of choline and ethanolamine. II. Requirement for a cobamide coenzyme by an ethanolamine deaminase. J Biol Chem **240:**4675-4681.

55. **Babior BM.** 1982. Ethanolamine ammonia-lyase, p 263-288. *In* Dolphin D (ed), B12, vol 2. John Wiley & Sons, New York.

56. **Zayas CL, Claas K, Escalante-Semerena JC.** 2007. The CbiB protein of *Salmonella enterica* is an integral membrane protein involved in the last step of the de novo corrin ring biosynthetic pathway. J Bacteriol **189:**7697-7708.

57. **Keller S, Treder A, SH VR, Escalante-Semerena JC, Schubert T.** 2016. The SMUL_1544 gene product governs norcobamide biosynthesis in the tetrachloroethene-respiring bacterium Sulfurospirillum multivorans. J Bacteriol doi:10.1128/JB.00289-16.

58. **Kräutler B, Fieber W, Osterman S, Fasching M, Ongania K-H, Gruber K, Kratky C, Mikl C, Siebert A, Diekert G.** 2003. The cofactor of tetrachloroethene reductive dehalogenase of

*Dehalospirillum multivorans* is Norpseudo-B12, a new type of natural corrinoid. Helvetica Chimica Acta **86:**3698-3716.

59. **Burgess RR.** 1996. Purification of overproduced Escherichia coli RNA polymerase sigma factors by solubilizing inclusion bodies and refolding from Sarkosyl. Methods Enzymol **273:**145-149.

60. **Campbell GR, Taga ME, Mistry K, Lloret J, Anderson PJ, Roth JR, Walker GC.** 2006. *Sinorhizobium meliloti bluB* is necessary for production of 5,6-dimethylbenzimidazole, the lower ligand of B12. Proc Natl Acad Sci U S A **103:**4634-4369.

61. **Woodson JD, Escalante-Semerena JC.** 2004. CbiZ, an amidohydrolase enzyme required for salvaging the coenzyme $B_{12}$ precursor cobinamide in archaea. Proc Natl Acad Sci USA **101:**3591-3596.

62. **Gray MJ, Escalante-Semerena JC.** 2009. In vivo analysis of cobinamide salvaging in *Rhodobacter sphaeroides* strain 2.4.1. J Bacteriol **191:**3842-3851.

63. **Gray MJ, Escalante-Semerena JC.** 2009. The cobinamide amidohydrolase (cobyric acid-forming) CbiZ enzyme: a critical activity of the cobamide remodelling system of *Rhodobacter sphaeroides*. Mol Microbiol **74:**1198-1210.

64. **Bork P, Sander C, Valencia A.** 1993. Convergent evolution of similar enzymatic function on different protein folds: the hexokinase, ribokinase, and galactokinase families of sugar kinases. Protein Sci **2:**31-40.

65. **Miroux B, Walker JE.** 1996. Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. J Mol Biol **260:**289-298.

# CHAPTER 3

# THE *METHANOSARCINA MAZEI COBD* (MM2060) GENE ENCODES A PROTEIN WITH L-THREONINE-*O*-3-PHOSPHATE DECARBOXYLASE AND L-THREONINE KINASE ACTIVITIES FOR THE SYNTHESIS OF THE AMINOPROPANOL LINKER OF COBAMIDES[2]

[2] Tavares N.K., Zayas C.L. and Escalante-Semerena J.C. To be submitted to *Molecular Microbiology*.

**ABSTRACT**

It is known that methanogenic archaea utilize cobamides, but several of the enzymes required for assembly of these cofactors have not been studied in archaea, while others are yet to be identified. To expand our knowledge of this biosynthetic pathway in archaea, we focused on the last steps of the *de novo* corrin ring biosynthetic branch in the methanogenic archaeum *Methanosarcina mazei* strain Gö1. Here, we report *in vivo* and *in vitro* evidence that the open reading frame (ORF) MM2060 of the *M. mazei* Gö1 genome encodes an orthologous replacement of the bacterial L-threonine-*O*-3-phosphate (L-Thr-P) decarboxylase enzyme (CobD; EC 4.1.1.81). Heterologous expression of ORF MM2060 supported growth of a *S. enterica cobD* strain. L-Thr-P decarboxylase activity was confirmed in *vitro* with purified MM2060 protein. Surprisingly, MM2060 also complemented a strain lacking L-Thr kinase (EC 2.7.1.177) activity encoded by *pduX* in *S. enterica.* L-Thr kinase activity was confirmed *in vitro* with ATP and L-Thr as substrates. The products of the reaction were ADP, L-Thr-P, and 1-amino-2-propanol (AP-P). Notably, ORF MM2060 encodes a protein (*Mm*CobD) with a longer carboxy-terminal (*C*-terminal) domain, which contains a cysteine-rich region with a zinc finger-like metal-binding motif. The *C*-terminal domain alone did not complement a *cobD* or *pduX* strain, nor did it have decarboxylase or kinase activities *in vitro*. Complementation and *in vitro* studies with *C*-terminally-truncated *Mm*CobD protein demonstrated that this domain was not necessary for L-Thr-P decarboxylase or L-Thr kinase activities, but it did affect the activities. *Mm*CobD hydrolyzed ATP in the absence of the L-Thr co-substrate and had a lower specific activity when the *C*-terminus was absent. The specific activity for the kinase reaction of the wild-type enzyme decreased with the addition of L-Thr, but activity increased for the version lacking the *C*-terminus, suggesting a regulatory or substrate-gating role for the *C*-terminal domain. This work demonstrates that the MM2060 protein is a bifunctional enzyme encoding both L-Thr-P decarboxylase and L-Thr kinase activities, the latter being a new activity ascribed to this class of enzyme.

## INTRODUCTION

Cobamides (Cbas, *e.g.*, cobalamin) are required and synthesized by some archaea. Coenzyme $B_{12}$ ($CoB_{12}$, a.k.a. adenosylcobalamin, AdoCbl, $AdoB_{12}$) is the most structurally complex coenzyme known. In methanogenic archaea cobamides (Cbas) play a central role in methanogenesis (1, 2). In this process Cba-dependent enzymes serve as methyl-group carriers transferring the C1 unit from $N^5$-methyl-tetrahydropterin onto the sulfhydryl moiety of coenzyme M (3). Similarly, methionine synthase, MetH, also uses Cbas to transfer a methyl from $N^5$-methyltetrahydrofolate to the thiol moiety of homocysteine to yield methionine (4). The AdoCbl biosynthetic pathway has been extensively studied in bacteria such as *Salmonella enterica* sv Typhimurium LT2 (hereafter *S. enterica*), however there are still gaps in our knowledge of how archaea synthesize Cbas. In archaea, only about half of the genes have assigned functions, and a fraction of these do not have orthologues in bacteria (5, 6). Our work, and that of others, has employed comparative genomics as a tool for the identification of putative archaeal orthologues of bacterial cobamide biosynthetic genes (7-13). One of the missing enzymes in archaea is the L-threonine (L-Thr) kinase, which is encoded by *pduX* in *S. enterica* (14, 15). Methanogenic archaea share many AdoCba biosynthetic enzyme homologues with bacteria, including the L-Thr-P decarboxylase CobD protein. Notably, the CobD proteins encoded in the genomes of many methanogenic archaea have an extended *C*-terminal domain of about 108 amino acids, a domain that is not present in CobD from *S. enterica* and other bacteria. The function of this domain is unclear.

Comparative genomics studies done by others identified putative archaeal genes encoding orthologous proteins required for AdoCba biosynthesis in bacteria. Such studies revealed the absence of orthologues of several genes required by bacteria to synthesize AdoCba. While bioinformatics analysis of AdoCba biosynthesis gene clusters has been previously performed (10, 16), orthologues of the MM2060 protein have been overlooked despite its frequent association with *cob* biosynthetic genes.

The studies presented herein focused on the identification of the archaeal enzyme responsible for the pyridoxal-5′-phosphate (PLP)-dependent decarboxylation of L-threonine-*O*-3-phosphate (L-Thr-P) that produces 1-amino-2-propanol phosphate (AP-P), which in turn is used as co-substrate in the last step of

*de novo* corrin ring biosynthesis (Fig. 3.1). In *S. enterica*, the reaction alluded to above is performed by the CobD enzyme (EC 4.1.1.81; hereafter *Se*CobD) (17). Here, we report genetic and biochemical data that support the conclusion that the MM2060 protein from *M. mazei* possesses the expected L-Thr–P decarboxylase activity and, surprisingly, it also has L-Thr kinase activity. We also show that although the putative metal-binding *C*-terminus of MM2060 is not required for L-Thr-P decarboxylase or L-Thr kinase activities, disruptions to this domain affects the activity of both functions.

**MATERIAL AND METHODS**

**Bacterial strains.** Strains and plasmids used in this work are described in Table S3.1. *S. enterica* strains carried a null allele of the *metE* gene that encodes the Cba-independent methionine synthase (MetE) enzyme (18). In the absence of MetE the cell uses the Cba-dependent methionine synthase (MetH) enzyme (4, 19-21). All *S. enterica* strains also carry an undefined mutation, (allele *ara*-9), which prevents the utilization of arabinose as a carbon and energy source. Gene deletions in *S. enterica* were constructed using the phage lambda Red site specific recombinase system as described (22).

**Culture media and growth conditions.** No carbon essential (NCE) (23) with glycerol (22 mM) as the carbon and energy source was used as minimal growth medium. When added to the medium, the following supplements were at the indicated concentrations: trace minerals, 10 mL/liter (24); $MgSO_4$ (1 mM), 5,6-dimethylbenzimidazole (DMB) (0.15 mM), ampicillin (100 μg/mL), arabinose (500 μM). All corrinoids (cobyric acid dicyanide [$(CN)_2$Cby], cobinamide dicyanide [$(CN)_2$Cbi], and cyanocobalamin (CNCbl) were added at (1 or 10 nM) final concentrations. When ethanolamine (90 mM) was used as a carbon and energy source Fe-citrate (50 μM) was also added to the medium with corrinoids (300 nM). Cobyric acid was a gift from Paul Renz (Universität-Hohenheim, Stuttgart, Germany). All other chemicals were purchased from Sigma-Aldrich. *S. enterica* strains were cultured in Nutrient Broth (NB, Difco Laboratories) (0.8% w/v) containing NaCl (85 mM). Lysogenic broth (LB) (25, 26) was used as rich medium to culture *Escherichia coli* strains unless otherwise indicated.

**Plasmid construction.** *M. mazei* strain Gö1 gDNA for PCR-gene amplification was a gift from Gerhard Gottschalk (Göttingen, Germany). Genomic DNA from *S. enterica* strain JE7088 (*metE2702 ara*-9) were

extracted by heating cells at 90°C suspended in double distilled $H_2O$ for 5 min to release DNA. Cell debris was separated from DNA in the supernatant by centrifugation; this was the source of DNA used as a template for PCR amplification. Oligonucleotide primers were purchased from Integrated DNA Technologies Inc. (IDT, Coralville, IA). A list of primer sequences can be found in Table S3.2. Primers for cloning were designed using the Saccharomyces Genome Database web based primer design tool available at http://www.yeastgenome.org/cgi-bin/web-primer. Genes were PCR amplified from the appropriate gDNA template with PCR Extender Polymerase (5 Prime) and the primer pairs listed in Table S3.2. PCR products and vectors were treated with restriction endonucleases indicated in the primer name in Table S3.2 and purified with the Wizard SV Gel and PCR Clean-Up kit (Promega). Cloning vectors were treated with Fast AP alkaline phosphatase (Fermentas). PCR fragments and vectors were ligated together using Fastlink Ligase (Fermentas) and introduced into *E. coli* DH5α (27, 28) via electroporation. Plasmid DNA was purified using the Wizard Plus SV Miniprep kit (Promega). Plasmid sequence was confirmed by using BigDye® (ABI PRISM) protocols (University of Wisconsin-Madison Biotechnology Center & University of Georgia Genomics Facility). Table S3.1 lists the resulting plasmids. The start codon for wild-type *M.m. cobD* (ORF MM2060) was changed from GTG to ATG. The *N*-terminus (*Mm*CobD$^{1\text{-}385}$) was cloned from 1-385 with two stop codons TAA TAA added after the last residue. The *C*-terminal domain (*Mm*CobD$^{386\text{-}497}$) was cloned separately from codons encoding 386-497 with the addition of a methionine as the first residue. pBAD24 vectors were used for complementation (29) and pTEV5 overexpression vectors (30).

**Complementation of function.** To determine whether or not a protein of interest was functional *in vivo*, plasmids were introduced into *S. enterica* by electroporation as described elsewhere (31). *S. enterica* strains were grown to full density (~$2 \times 10^9$ cfu/mL) in nutrient broth (NB, Difco) supplemented with ampicillin (Amp, 100 μg/mL) for plasmid maintenance. Strains were grown in triplicate in sterile 96-well tissue culture plates (Falcon) where 2 μL of an overnight cell culture was used to inoculate 198 μL of fresh minimal (NCE) medium supplemented with glycerol, $MgSO_4$, and trace minerals. Corrinoids and 1-

amino-2-propanol (AP) were added as indicated under above. Growth behavior was monitored using Gen5 software (BioTek Instruments) during growth at 37˚C with continuous shaking (19 Hz) in an EL808 Ultra Microplate Reader (BioTek Instruments). Cell density measurements at 630 nm were acquired every 15 or 30 min for 24 or 60 h. Data were analyzed using the Prism v6 software package (GraphPad Software).

**Overproduction of *S. enterica* CobD$^{WT}$ (*Se*CobD$^{WT}$).** *Se*CobD$^{WT}$ was overproduced and purified as described elsewhere (17, 32).

**Overproduction of *M. mazei* CobD (*Mm*CobD$^{WT}$).** *N*-terminal, TEV-cleavable $His_6$ tagged *Mm*CobD and truncated (*Mm*CobD$^{1-385}$ and *Mm*CobD$^{386-497}$) proteins were overproduced in *E. coli* C43 (λDE3) (33) cells (Lucigen) from plasmids pMmCOBD18, pMmCOBD9, pMmCOBD19, respectively; cells carrying pTEV5 plasmid (30) were used as controls. The following cell culture method was used to increase the amount of heme and PLP available inside the cells. A sample (10 mL) of an overnight culture carrying the above plasmids was used to inoculate two liters of LB containing ampicillin (100 μg/mL) supplemented with pyridoxine HCl (1 mM), iron(III)citrate (50 μM), and δ-aminolevulinic acid (δ-ALA, 0.5 mM) to increase the intracellular concentration of heme (34). Cultures were grown at 37°C with shaking (180 rpm) to an $OD_{600}$ ~0.8, followed by a temperature down shift to 25°C and induction of gene expression by the addition of isopropyl-*β*-D-thiogalactopyranoside (IPTG, 0.3 mM) to the medium. Cultures were incubated overnight at 25°C with shaking. Cells were harvested by centrifugation at 6,000 × *g* at 4°C for 15 min in an Avanti J-20 XPI Beckman/Coulter refrigerated centrifuge equipped with a JLA 8.1000 rotor; cell paste was stored at -20°C until used. Cells were resuspended in 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES-NaOH) buffer (50 mM, pH 7.9 at 4°C), with NaCl (500 mM), imidazole (5 mM), lysozyme (1 mg/mL), DNaseI (1 mg/mL), and protease inhibitor phenylmethanesulfonyl fluoride (PMSF, 100

μM). Cells were lysed at 27 kPa using a 1.1 kW TS Series Bench Top cell disrupter (Constant Systems Ltd.), equipped with a cooling jacket on the disruptor head to maintain a 6°C temperature using a Neslab ThermoFlex 900 recirculating chiller (Thermo Scientific). Debris was removed by centrifugation at 39,000 x *g* for 20 min. Proteins were purified at 4°C by Ni-affinity chromatography using a 1.5-mL bed volume of HisPur Ni-NTA resin (Thermo Scientific). The resin was equilibrated with bind buffer [HEPES (50 mM, pH 7.9 at 4°C), NaCl (500 mM), imidazole (5 mM)] before clarified supernatant was applied to the column. After binding to the column, the column was washed with four column volumes of bind buffer, before proteins were eluted stepwise by increasing the imidazole concentration from 20 to 100 mM, 20 mM per step. Fractions (5 mL each) were collected for each step. A final wash step was performed with 300 mM imidazole. $His_6$-*Mm*CobD$^{1-385}$ eluted with 80-100 mM imidazole and $His_6$-*Mm*CobD and *Mm*CobD$^{386-497}$ eluted with 100-300 mM imidazole. Fractions containing $His_6$-tagged proteins were pooled and the tag was cleaved with rTEV protease (1:100 rTEV:$His_6$-protein ratio) for 3 h at 25°C in bind buffer containing 1,4-dithiothreitol (DTT, 1 mM). The cleaved protein was dialyzed into bind buffer containing ehtylenediaminetetracetic acid (EDTA, 1 mM) at 25°C for 20 min, then dialyzed twice more in the same buffer without EDTA. The protein was passed over a Ni(NTA) column again to remove the cleaved $His_6$ tag and $His_7$-tagged rTEV protease, using the buffers employed in the first purification step. Untagged proteins eluted in the flow through and were pooled and dialyzed at 25°C for 20 min into desalting buffer 1 [HEPES (50 mM, pH 7.9 at 4 ºC), NaCl (300 mM), EDTA (1 mM)], followed by desalting buffer 2 [(HEPES (50 mM, pH 7.9 at 4 ºC), NaCl (200 mM), and storage buffer 3 [HEPES (50 mM, pH 7.9 4 ºC), NaCl (100 mM), glycerol (10%, v/v)]. Proteins were concentrated using Ultracel 10,000 MWCO centrifugal filters (Amicon Ultra) and frozen drop-wise into liquid $N_2$,

and stored at -80ºC until used. Protein concentrations were determined using a NanoDrop 1000 spectrophotometer (Thermo Scientific), using theoretical molecular weights and $A_{280}$ molar extinction coefficients for each protein, which were obtained from the ExPASy Protparam database (35, 36). Purified proteins were run resolved using a 15% SDS-PAGE gel. Bands were cut out and protein identity verified by in-gel trypsin digestion, followed by MALDI mass spectrometry and peptide mass fingerprinting and protein identification via Mascot (Matrix Science) database performed by the Proteomics and Mass Spectrometry Core Facility at the University of Georgia (Athens, GA).

***In vitro* L-Thr-P decarboxylase activity assay.** Reaction mixtures contained HEPES buffer (50 mM, pH 8.5); pyridoxal 5′-phosphate (PLP; 2 nmol); L-Thr-P (5 nmol) and purified protein (0.074 μM). When required for radiolabeled assays, a mixture of [$^{14}$C-U]-L-Thr-P and L-Thr-P in a 1:10 ratio was used as substrate; the final volume of the reaction was 25 μL. Reactions were incubated at 37°C for one h. Reactions (5 μL) were spotted onto cellulose thin layer chromatography (TLC) plates for product separation and analysis (see below for details).

**Thin-layer chromatography (TLC) analysis.** [$^{14}$C-U]L-Thr-P decarboxylation products were detected by TLC on cellulose plates developed with ammonium acetate (2.5 M):ethanol (95%; v/v) (30:70 ratio) mobile phase. Plates were pre-developed with distilled water and allowed to air-dry prior to use. Use of pre-developed plates provided the best resolution. A Typhoon Trio Variable Mode Imager (GE Healthcare) with ImageQuant v5.2 software was used to visualize the results.

**ATPase activity assay.**

ATPase activity was assessed using the ADP-glo™ Kinase assay kit (Promega) as per manufacturer's instructions. Briefly, the assay uses a proprietary reagent to deplete any remaining ATP in the reaction mixtures, then a secondary reagent converts ADP to ATP which is measured by a luciferase/luciferin reaction which is measured with a SpectraMax Plus Gemini EM microplate spectrophotometer (Molecular Devices) equipped with SoftMax Pro v4 software. Reaction mixture consisting of HEPES (50

mM, pH 6.8 or 8.5), *tris*(2-carboxyethyl)phosphine (TCEP) (1-2 mM), $MgCl_2$ (1 mM), ATP (0.1-10 mM), L-Thr (0.3-50 mM) and protein (100 ng or 100 nM) were incubated at 37°C for 1 h. Nunc 96-well round bottom black polypropylene microtiter plates (Thermo Fisher) were used to minimize background. Standard curves for ATP were used for quantification. For ATPase inhibition assay the following inhibitors were used ADP (200 mM), AMP (200 mM), sodium pyrophosphate (PPi, 10 mM), sodium triphosphate (PPPi, 10 mM), ADP-γ-S, (100 μM), sodium *ortho*-vanadate ($Na_3VO_4$, 1 mM), and sodium beryllium fluoride ($BeF_2$, 2 mM) in HEPES buffer (50 mM, pH 7.5) with *Mm*CobD (6 μM).

**$^{31}$Phosphorous nuclear magnetic resonance ($^{31}$P-NMR) analysis of L-Thr kinase reaction products.** Proton-decoupled $^{31}$P-NMR spectra were obtained using a Varian Unity Inova 500 MHz spectrometer (Chemical Sciences Magnetic Resonance Facility, University of Georgia) with the following parameters: pulse angle 45°, repetition delay 1 s, excitation pulse 3.88 μs, spectral width 12.11 kHz, acquisition time 0.810 s. 500 μL reaction mixture consisted of HEPES (50 mM, pH 8.5), $MgCl_2$ (1 mM), ATP (3 mM), L-Thr (3 mM) and enzyme (3 μM) incubated at 37°C for 1 h. Protein was removed from reaction mixtures by filtration using Amicon Untracel filters (Millipore) with 10 kDa molecular mass size exclusion. Reaction mixtures (500 μL) were brought up to a final volume of 600 μL in $D_2O$ (17%v/v). Spectra were processed with MestReNova software version v7.0.0-8333 (Mestrelab Research, Santiago de Compostela, Spain).

**Bioinformatics and phylogenetic analysis.** Sequences were obtained using BLAST (37) search for homology among sequences available as of 1/21/2015 in the IMG database (38). The protein sequence for ORFs MM2060 was used as the query sequences. Only finished genomes with bit scores >50 or e values >1.0e-7 were used in the analysis. Sequence header files were simplified with the find/replace and grep functions of TextWrangler (Bare Bones Software, North Chelmsford, MA). Outliers with extreme sequence divergence were not included nor were alleles not associated with Cbl biosynthetic genes on the chromosome. FASTA formatted sequences were aligned using the MUSCLE (39) plugin within Geneious R8 software (Biomatters Ltd., Auckland NZ) with default setting. ESPript 3.x (40) was used to generate images of alignments.

**Oligomeric state analysis of *Mm*CobD$^{WT}$ and truncations**. Gel filtration was performed using a HiPrep 26/60 Sephacryl S-100 High Resolution Column (GE Healthcare) connected to a computer-controlled ÄKTA FPLC system. The column was equilibrated with HEPES buffer (50 mM, pH 7.5 at 4ºC) containing NaCl (150 mM). *Mm*CobD and truncated proteins (2 mg) was applied to the column, which was developed isocratically at a rate of 2 ml min$^{-1}$. Molecular mass calibrations were performed using ovalbumin (44 kDa), myoglobin (17 kDa), and vitamin $B_{12}$ (1.35 kDa) components of the Bio-Rad gel filtration standards along with bovine serum albumin (66.4 kDa, Promega) and DNaseI (31 kDa, Sigma).

**Spectrophotometric kinase assay.** *Mm*CobD ATPase and kinase activities were measured using an NADH-consuming assay (41-44). All substrate stocks were made fresh. Reactions (100 μL) contained HEPES buffer (50 mM, pH 8.5), $MgCl_2$ (5 mM), PEP (3 mM), NADH (0.1 mM), pyruvate kinase (1 U), lactate dehydrogenase (1.5 U) incubated at 25°C with measurements taken every 11 sec over a 20 min period. For ATPase specific activity L-Thr concentration was held at 50 mM while ATP concentration was varied (5 – 100 mM). To measure the effect of the co-substrate on ATPase activity, ATP concentration was held at 50 mM while the concentration L-Thr was varied (0 – 100 mM). Reactions were started by the addition of *Mm*CobD (3 μM). The absorbance at 340 nm was monitored in a 96-well plate using the Spectramax Plus UV-visible spectrophotometer (Molecular Devices) equipped with SoftMax Pro v6.2. Enzyme activities were calculated as described elsewhere (43). Specific activity data are presented with standard deviations from triplicate experiments.

**RESULTS**

**Bioinformatics analysis of *M. mazei* ORF MM2060 and its homologues in other Methanosarcinales.** Figure 3.2 shows a protein sequence alignment of the *M. mazei* MM2060 (*Mm*CobD) and the *S. enterica* CobD (*Se*CobD), and supplemental Figure S3.1 shows a protein sequence alignment of CobD from several methanogenic archaea and *S. enterica*. As seen in Figure 3.2, the sequence of these proteins is ~37% similar and ~23% identical from end to end. The *M. mazei* protein has an *N*-terminus that is 13 amino acids longer than the *S. enterica* protein, which also has two small deletions spanning position 56-62, and 262-266. The most striking difference is the 108-residue extended *C*-terminus of the *M. mazei*

protein. Information currently available in databases shows that genome sequences of all Methanosarcinales encode similar CobD proteins. Notably, the *C*-terminus of *Mm*CobD contains a cysteine-rich putative metal-binding/zinc finger-like domain, starting at residue 390. This domain contains five histidines and nine cysteines arranged in four clusters ($CX_4CH$, $CX_2CXCX_4C$, $CX_2CHX_2H$, $CX_2H$) including what appears to be a heme-binding motif (CXXCH). *Methanosarcina acetivorans* and *Methanosarcina barkeri* possess this cytochrome *c*-like heme-binding motif, whereas *M. burtonii*, *M. psychrophilus*, and *M. hallandica* appear to have two slightly different copies of the *cobD* gene adjacent to each other on the chromosome; the alluded *cobD* genes appear to encode proteins with variant CXXCV or CXXCN motifs, which are unlikely to bind heme. CobD from *M. psychrophilus*, also has a particularly long *N*-terminus which is 37 amino acids longer than *Se*CobD.

**A gene encoding a protein of unknown function with a cysteine-rich putative metal-binding motif is associated with *cob* genes across many genera of bacteria and archaea.** Further bioinformatics analysis of the putative metal-binding/zinc finger domain of *Mm*CobD$^{WT}$ revealed that this domain was found fused to other AdoCbl biosynthetic proteins or encoded separately by open reading frames associated with *cob* genes. Most methanogenic archaea genomes encode proteins with the zinc finger motif fused to the *C*-termini of CobDs, however in *Methanocorpusculum labreanm* the 'zinc finger' is fused to the corrinoid amidohydrolase CbiZ enzyme, in *Methanothermus fervidus* it is fused to the methyltransferase CbiH enzyme, and in the extremophile *Methanopyrus kandleri* it is not fused to any other protein. This 'zinc finger' can also be found fused to the corrinoid transporter protein BtuC in *Butyrivibrio,* to cobyrinic acid *a,c*-diamide synthetase CbiA in *Treponema* sp, and to the *N*-terminus of the methyltransferase CbiD enzyme in *Slackia heliotrinireducens*. Free standing 'zinc fingers' are found in a wide array of bacteria including representatives of the genera *Clostridum*, *Bacillus*, *Spirochaetes*, as well as Rhodobacteriales, Rhizobiales, Alpha-, Beta- and Gamma-proteobacteria. The gene encoding this putative zinc finger is found in organisms that utilize both the anaerobic/early cobalt-insertion or aerobic/late cobalt-insertion pathways (reviewed by (45)). It is notable that this 'zinc finger' protein is found only in AdoCba producers that do not encode the L-Thr kinase PduX enzyme. To date, non-

orthologous replacements of PduX or alternative pathways for the production of L-Thr-P have not been found in any AdoCba producer that lacks a homologue of PduX. The function of this putative 'zinc finger' is unknown.

***Mm*CobD restores AdoCba-dependent growth of a *S. enterica cobD* strain.** To verify that *Mm*CobD had L-Thr-P decarboxylase activity *in vivo* we used a *S. enterica cobD* strain to block the synthesis of 1-amino-2-propanol phosphate (AP-P). All *S. enterica* strains used in this study lacked the Cba-independent methionine synthase enzyme, MetE, to force the methylation of homocysteine to occur via the Cba-dependent MetH enzyme. Cultures were grown under oxic conditions to block *de novo* synthesis of the corrin ring (46, 47). The medium was supplemented with dicyanocobyric acid [$(CN)_2$Cby], whose conversion to AdoCba requires the synthesis and attachment of AP-P from L-Thr-P for the synthesis of adenosylcobinamide-phosphate (AdoCbi-P), the last intermediate in the corrin ring biosynthetic pathway (17) (Fig. 3.1).

Plasmids carrying genes encoding the full-length *Mm*CobD$^{WT}$ protein (pMmCOBD17), the *N*-terminal domain only *Mm*CobD$^{1-385}$ (pMmCOBD7), or the *C*-terminal, putative metal-binding domain only *Mm*CobD$^{386-497}$ (pMmCOBD13) (Table S3.1) were individually introduced into a *S. enterica cobD*. A *cobD* strain carrying a plasmid encoding *Se*CobD$^{WT}$ (pCOBD6) was used as positive control. The growth behavior of the strain expressing *Mm*CobD$^{WT}$ was comparable to that of the wild-type strain, and to that of the strain synthesizing *Se*CobD$^{WT}$ from a plasmid (4 h lag, 1.5 h doubling time) (Fig. 3.3A). We observed a lag before the onset of exponential growth of cells expressing only the *N*-terminal region, (13 h lag, 4.1 h doubling time), suggesting that although the cysteine-rich, putative metal-binding domain was not required for L-Thr-P decarboxylase activity, it appeared to be important for efficient CobD activity *in vivo*. Not surprisingly, the *C*-terminal cysteine-rich domain failed to complement the *cobD* strain.

**The *Mm*CobD protein decarboxylates L-Thr-P yielding AP-P.** Figure 3.3B shows a representative set of results from experiments aimed at detecting L-Thr-P decarboxylase enzymatic activity. Lanes 1 and 2 of the phosphorimage show that full-length *Mm*CobD$^{WT}$ and the *N*-terminal domain containing only the CobD domain (*Mm*CobD$^{1-385}$) converted [$^{14}$C-U]-L-Thr-P to [$^{14}$C-U]-AP-P, as did the *Se*CobD positive

control (lane 4). Small amounts of [$^{14}$C-U]-L-Thr can be seen in each lane as a decomposition product. The cysteine rich *C*-terminal domain alone (*Mm*CobD$^{386\text{-}497}$) (lane 3) failed to generate [$^{14}$C-U]-AP-P from [$^{14}$C-U]-L-Thr-P. These data support the *in vivo* result that indicated the *N*-terminal domain of MmCobD$^{WT}$ possesses the L-Thr-P decarboxylase activity, and that the removal of the *C*-terminal domain reduces but does not abolish activity under the conditions tested.

**The intracellular concentration of Cby is the limiting factor for the complementation of *S. enterica* by *Mm*CobD.** We performed complementation studies with varying concentrations of Cby and inducer. We assessed the ability of the full-length or truncated *Mm cobD* genes to restore AdoCba biosynthesis in a *S. enterica cobD* strain. For this purpose, we placed the appropriate *Mm cobD* allele under the control of the arabinose inducible promoter ($P_{araBAD}$) of cloning vector pBAD24 (29). The concentration of inducer had no effect on the level of complementation (data not shown), suggesting that substrate concentration, not protein level was the limiting factor for complementation by the *Mm*CobD$^{1\text{-}385}$ protein. Supplemental Figure S3.2 shows the effect of increasing levels of Cby on the ability of the *Mm*CobD$^{1\text{-}385}$ protein (Panel A) and *Mm*CobD$^{WT}$ protein (Panel B) to restore AdoCbl biosynthesis in a *S. enterica cobD* strain; all comparisons were relative to the growth behavior of a *S.e. cobD* strain expressing *S.e. cobD*$^{+}$ (black squares). For *Mm*CobD$^{WT}$ at low levels of Cby (1 nM, light gray) we observed a 16 h lag and a doubling time of 9.2 h relative to a *Se*CobD$^{WT}$ strain (3 h lag, 2.2 h) under the same conditions (Panel C). The *Se*CobD$^{WT}$ strain showed no difference in lag time at any of the Cby concentrations tested. The doubling time remained relatively the same (2.2 h) above 10 nM Cby. Growth rates, lag times, and doubling times are reported in supplemental Figure S3.1-C. By altering the concentration of Cby we created a set of conditions that allowed us to assess the effects of growth under high and low AdoCbl demanding growth conditions to reveal phenotypes and changes in phenotypes that may otherwise have been masked or overlooked.

***Mm*CobD restores AdoCba biosynthesis in a *S. enterica pduX* strain.** *M. mazei* and other AdoCba producers that encode a version of the cysteine-rich, putative metal-binding domain also lack genes encoding homologues of the L-Thr kinase, PduX. This observation led us to investigate the possibility

that the *C*-terminal domain of *Mm*CobD$^{WT}$ may generate L-Thr-P. Figure 3.4 shows the growth analysis of *S. enterica pduX* strains carrying plasmids encoding *Mm*CobD$^{WT}$ protein, or the *N*- (*Mm*CobD$^{1-385}$) and *C*-terminal (*Mm*CobD$^{386-497}$) domains. Panel A shows growth on glycerol supplemented with Cby (1 nM), a condition that requires only low levels of AdoCbl biosynthesis for growth. In contrast. panel B shows growth on ethanolamine as the sole source of carbon and energy supplemented with Cby (300 nM), a condition that demands a high level of AdoCbl production for growth (48-50). Both the *Mm*CobD$^{WT}$ and *Mm*CobD$^{1-385}$ proteins supported growth of a *pduX* strain under the less demanding condition (Fig. 3.4A) although not as well as the *Se*PduX$^{WT}$ did. Under growth conditions where higher levels of AdoCba was required (*e.g.,* ethanolamine catabolism) only *Mm*CobD$^{WT}$ supported growth (Fig. 3.4B). The strain that synthesized *Mm*CobD$^{1-385}$ grew poorly on ethanolamine but discernably better than the strain synthesizing *Mm*CobD$^{386-497}$ (Fig. 3.4B). On glycerol the strain that synthesized *Mm*CobD$^{WT}$ had a 12-h lag time and a doubling time of 6.2-h, while the strain that synthesized *Mm*CobD$^{1-385}$ strain had a 14-h lag time and 7.7-h doubling time, compared to the 3-h lag time and 3.3-h and 2.1-h doubling time of the wild-type strain and strains expressing *Se*PduX$^{WT}$, respectively. Cells that synthesized *Mm*CobD$^{386-497}$ on glycerol had a lag time of 13-h and a 16.2-h doubling time, relative to the vector only control, which had a similar a growth rate. Likewise, when we demanded growth on ethanolamine, strains that synthesized *Mm*CobD$^{WT}$ had an 11-h lag time and 10.7-h doubling time compared to the 4-h lag time and 3.4-h and 2.5-h doubling times for the wild-type strain and for a strain that synthesized *Se*PduX$^{WT}$, respectively. Collectively, these results support the hypothesis that the *C*-terminal domain of *Mm*CobD$^{WT}$ contributes to the production of L-Thr-P, but is not required for function, at least under growth conditions that demand low levels of AdoCba. These results strongly suggested that the *N*-terminus of *Mm*CobD$^{WT}$ has L-Thr kinase activity.

***Mm*CobD has ATPase activity.** To determine if *Mm*CobD$^{WT}$ had L-Thr kinase activity *in vitro,* we assayed the purified protein for ATPase activity as described under *Materials and Methods*. Data presented in Figure 3.4C support the conclusion that the *N*-terminal domain (*Mm*CobD$^{1-385}$) has ATPase activity comparable to that of *Se*PduX$^{WT}$. The *Mm*CobD$^{WT}$ protein also had ATPase activity, but only ~57% relative to the activity of the *N*-terminus alone or to *Se*PduX$^{WT}$. Surprisingly, the *Mm*CobD$^{386-497}$ *C*-

terminal domain also had about 20% of the ATPase activity associated with *Mm*CobD$^{1\text{-}385}$ and 37% of the activity of *Mm*CobD$^{WT}$.

The apparent activity for *Mm*CobD$^{386\text{-}497}$ was considered to be background when compared to the conditions used in Figure 3.6A. The apparent activity decreased substantially when less protein (100 ng) and higher concentration of ATP (10 mM) were present in the reaction mixture (Fig. 3.6A). *Se*CobD$^{WT}$ had no detectable ATPase activity under the same conditions. Figure 3.4D shows that *Mm*CobD$^{WT}$ had ATPase activity in the absence of L-Thr. However, the ATPase activity of *Mm*CobD$^{WT}$ was enhanced by 33% when either L-Thr, or the product L-Thr-P was present in the reaction mixture. L-Ser also slightly enhanced the ATPase activity of *Mm*CobD$^{WT}$ by 15%, indicating the possibility that *Mm*CobD may be capable of generating ethanolamine phosphate (EA-P) from L-Ser and ATP. EA-P is the nucleotide loop linker found in norCbl (51). However, it is unlikely that L-Ser is the natural substrate of *Mm*CobD as *M. mazei* and other methanogens do not produce norCbl (52). *Mm*CobD did not respond to D-threonine, D-serine, L- or D-tyrosine, L-homoserine, L-valine, L-alanine, L- homoalanine, 1-amino-2-propanol, Cby, or Cbi as co-substrates with ATP (data not shown).

***M. mazei* CobD produces L-Thr-P *in vitro*.** We used $^{31}$P-NMR to verify the *Mm*CobD$^{WT}$-dependent production of L-Thr-P from L-Thr and ATP, and the subsequent conversion of L-Thr-P to AP-P. Figure 3.5 shows representative NMR spectra of no enzyme controls for ATP (panel A) and Thr-P (panel B). To generate a standard for AP-P, we obtained the $^{31}$P-NMR spectrum for the *Se*CobD (panel E) and *Mm*CobD (panel D) decarboxylation reaction with L-Thr-P as the substrate. In these decarboxylation reactions, enzymatically produced L-Thr-P had a chemical shift of 3.6 ppm, consistent with that of commercially available standard (panel B). AP-P had a chemical shift of 4.0-4.2 ppm and was detected in both the *Mm*CobD$^{WT}$ and *Se*CobD$^{WT}$ reactions with L-Thr-P. The *Se*CobD$^{WT}$ reaction has a third peak with a chemical shift of 2.9 ppm that was not present in the *Mm*CobD$^{WT}$ reaction. This peak did not correspond to the chemical shift of free PLP in solution under these conditions. Enzyme associated PLP was not detectable at the concentration of protein (3 μM) used in these reactions. The identity of this peak remains unknown. Results shown in panel C confirmed that *Mm*CobD$^{WT}$ had ATPase activity. The

absence of the β phosphate peak of ATP, represented by the peak with the chemical shift at 18.5 ppm, suggested that the enzyme converted all ATP (10 mM) to ADP (peaks 5.8, 9.6, and 10.0 ppm). *Mm*CobD$^{1-385}$ generated a $^{31}$P-NMR spectrum identical to that of *Mm*CobD$^{WT}$, while the reaction mixture containing the *Mm*CobD$^{386-497}$ protein did not generate ADP or L-Thr-P from ATP and L-Thr (data not shown). These data confirmed that *Mm*CobD$^{WT}$ has L-Thr kinase activity and that the enzyme generates L-Thr-P, ADP, and AP-P at the expense of ATP and L-Thr.

**Removal of the *C*-terminal domain (*Mm*CobD$^{386-497}$) alters ATPase activity in the presence of L-Thr.** Data presented in Figure 3.4D show that the presence L-Thr or L-Ser or the product L-Thr-P enhanced the ATPase activity of the full-length *Mm*CobD$^{WT}$ protein. Figure 3.6A confirms this point with L-Thr as the co-substrate, however, the addition of L-Thr did not have an effect on the ATPase activity of the *N*-terminal (*Mm*CobD$^{1-385}$) domain. This result suggested that removal of the *C*-terminal domain had a negative effect on affinity of the protein for ATP. At higher substrate concentrations (10 mM ATP, 50 mM L-Thr) the ATPase activity of the *C*-terminus *Mm*CobD$^{386-497}$ domain was negligible and was considered background level.

The presence of a putative metal-binding domain from an anaerobe prompted us to consider assaying the enzyme under anoxic conditions in the presence of a reducing agent. As purified, the protein was brown which lead us to speculate that the protein might contain iron and be redox sensitive. Figure 3.6B shows the ATPase activity of *Mm*CobD$^{WT}$ under oxic (black bar) and anoxic conditions. Consistent with previous results, the addition of L- enhanced ATPase activity under oxic and anoxic conditions. However, there was no discernable difference in activity in the presence or absence of air.

**A *Mm*CobD *N*-terminal truncation has lower specific activity relative to wild type**. Figures 3.6C and D show the specific activity of full-length *Mm*CobD$^{WT}$ and the *N*-terminal domain (*Mm*CobD$^{1-385}$) as a function of ATP concentration. The activity was measured indirectly via coupled assay that measures the consumption of NADH (see *Materials and Methods*) (41). Both the *Mm*CobD$^{WT}$ and *Mm*CobD$^{1-385}$ enzymes responded to increasing ATP concentrations. However, the activity of *Mm*CobD$^{WT}$ was ~8 fold higher than that of the truncated protein (Table 3.1). The full-length enzyme reached saturation at 50 mM

ATP while the truncated protein did not reach saturation at twice that concentration (Figs. 3.6C, D). *Mm*CobD$^{WT}$ activity as a function of ATP concentrations was also shown in supplemental figure S3.3I. These results suggested that the full-length protein likely had a significantly higher affinity for ATP than the truncated protein. Figures 3.6E and F show the specific activity as a function of L-Thr concentration, where the concentration of ATP is held constant at 50 mM while the concentration of L-Thr was varied. In contrast with the previous *in vitro* data, the addition of L-Thr decreased the specific activity from 10.3 μmol min$^{-1}$ mg$^{-1}$ to 7.6 μmol min$^{-1}$ mg$^{-1}$ for *Mm*CobD$^{WT}$, and conversely increased the specific activity for *Mm*CobD$^{1-385}$ (1.3 to 3.7 μmol min$^{-1}$ mg$^{-1}$). However, there was no significant change in the activity as the concentration of L-Thr was increased (Fig. 3.6E and supplemental Fig. S3.3F), suggesting L-Thr is the limiting component as the enzyme is saturated by it at very low levels. The same observation was made with truncated enzyme, however the activity did decrease at L-Thr concentrations of 50 mM and above. The implications of the influence of the *C*-terminal domain on substrate affinity are discussed below.

**Factors influencing L-Thr kinase activity of *Mm*CobD$^{WT}$.** We optimized the reaction conditions for the kinase reactions (Supplemental Fig. S3.3). The optimal pH was 8.5 in HEPES buffer (50 mM) (Fig. S3.3A). Salts were not required for function, however, the presence of KCl (100 mM) increased activity. Reducing agents such as TCEP (1 mM) or sodium dithionite (1 mM) provided minor increases in activity (Fig. S3.3B). *Mm*CobD$^{WT}$ activity was optimal with $MnCl_2$ (1 mM); reduced activity was observed with $CoCl_2$ (1 mM) and $MgCl_2$ (1 mM); activity with $MgCl_2$ increased at higher concentrations (10 mM) (Fig. S3.3C). We also measured the ATPase activity of *Mm*CobD$^{WT}$ and *Mm*CobD$^{1-385}$ in the presence of L-Ser or L-Thr (Figs. S3.3D, E) using *Se*PduX$^{WT}$ and *Se*CobD$^{WT}$ as positive and negative controls, respectively. The behavior was the same as previously observed (Figs. 3.4D, 3.6A), with the exception that *Mm*CobD$^{WT}$ was slightly more active (10%) with L-Ser as co-substrate relative to *Se*PduX$^{WT}$ or *Mm*CobD$^{1-385}$. We tested ATPase activity as a function of L-Thr concentration and found no change in activity with increasing concentrations of L-Thr (Fig. S3.3F).

ATPase activity in the presence of known inhibitors ADP, AMP, PPPi, PPi, ADP-γ-S, $Na_3VO_4$, and $BeF_2$ was tested. AMP and PPi did not inhibit *Mm*CobD$^{WT}$. *Mm*CobD was inhibited in the following order from most inhibitory to least: $BeF_2$>ADP>ADP-γ-S>PPPi>$Na_3VO_4$. Additional testing was conducted with ADP-γ-S on *Mm*CobD$^{WT}$ and *Mm*CobD$^{1-385}$ with *Se*EutP$^{WT}$ acetate kinase (53) as a positive control. Both *Mm*CobD$^{WT}$ and *Se*EutP$^{WT}$ were inhibited by ADP-γ-S but surprisingly, *Mm*CobD$^{1-385}$ was not (Fig. S3.3H). ATPase activity was assayed as a function of ATP concentration displayed direct proportionality (Fig. S3.3I).

**The *C*-terminal domain of *Mm*CobD$^{WT}$ does not affect the multimeric state of the protein.** We investigated whether the *C*-terminus affected the multimeric state of the protein. Supplemental Figure S3.5 shows the results of gel filtration analysis. *Mm*CobD$^{WT}$ (theoretical MW = 55.5 kDa) formed dimers in solution, as did the *Mm*CobD$^{1-385}$ protein (theoretical MW = 45.8 kDa). The *C*-terminal domain (*Mm*CobD$^{386-497}$, theoretical MW =12.9 kDa) was detected as a monomer. These data demonstrated that the *C*-terminal domain was not required for dimerization. Given that it has been reported that *Se*CobD$^{WT}$ is a dimer (32, 54), it was not surprising that *Mm*CobD$^{WT}$ was also a dimer, making it unlikely that the *C*-terminal domain of each subunit of the *Mm*CobD dimer would be able to interact if *Mm*CobD$^{WT}$ was in an inverted orientation as *Se*CobD$^{WT}$ is.

**DISCUSSION**

In this work we show that CobD from *M. mazei* is a bifunctional metalloprotein with L-Thr-P decarboxylase and L-Thr kinase activities. The L-Thr-P decarboxylase activity was expected, as the protein is homologous to the well-characterized CobD from *S. enterica* (17, 32, 54). Here we uncover a new activity for *Mm*CobD as a L-Thr kinase, an activity that is not present in *Se*CobD (Fig. 3.4C). The L-Thr kinase and L-Thr-P decarboxylase activities are both associated with the *N*-terminal domain (residues 1-385). This is the first report of these two enzymatic activities being fused into a single polypeptide. The function of the putative 'zinc finger' metal-binding *C*-terminal domain remains unclear. On the basis of the data reported herein, we propose that *M. mazei*, and other bacteria and archaea that lack homologues of the known L-Thr kinase (PduX) have evolved a new class of L-Thr-*O*-3-P decarboxylase, one that also

has L-Thr kinase activity. This work fills an important gap in our understanding of AdoCba biosynthesis in methanogenic archaea and other AdoCba producing bacteria.

Bioinformatics analyses suggest that the metal-binding domain is fused to the *C*-terminus of CobD exclusively in methanogens (Fig. 3.2), but it is also found encoded as a *N*- or *C*-terminal fusion in other AdoCba biosynthetic or transport genes such as *cbiA, cbiD, cbiH, cbiZ,* or *btuD* as well as a stand-alone ORF in a variety of bacteria. To date these CobD orthologues have only been found in organism that lack *pduX* homologues. The function of the putative metal-binding domain is unclear, as CobD from *M. mazei* retains L-Thr decarboxylase activity when it is removed. Its fusion to CobD appears to be ideal placement for the shuttling of a product to serve as substrate for CobD. We hypothesize that the metal-binding domain is involved in L-Thr kinase activity and or regulation.

**Is the *C*-terminus a 'zinc finger'?**

Supplemental Figure S3.4 shows an SDS-PAGE gel of the purified full-length *Mm*CobD, *N*- and *C*-terminal truncations and *Se*CobD. The *Mm*CobD$^{WT}$ protein produces two bands, both of which were verified by mass spectrometry to be *Mm*CobD. It is possible that the lower molecular weight band was produced as a result of the loss of the PLP cofactor or the putative metal or heme cofactor from the *C*-terminus, though this seems not to have affected the activities of the enzyme. We were unable to separate the two species of protein. We have begun to investigate the type and form of the metal bound to *Mm*CobD including the possibility of Zn, Fe, or heme. The deep brown color of the purified wild-type and *C*-terminal proteins leads us to believe the protein contains iron and not zinc.

**Intercellular substrate concentration is the limiting factor for *in vivo* heterologous expression of *Mm*CobD$^{WT}$ in *S. enterica cobD* or *pduX* strains.** A *S. enterica metE pduX* strain displays a weak growth phenotype when grown in minimal medium supplemented with glycerol as the sole source of carbon and energy, even when the medium contains high levels of Cby (>1 nM) (data not shown) (14). This is likely because very low levels of AdoCba are required to meet the methionine synthesis needs of the cells. It also suggests the existence of one or more enzymes with L-Thr kinase activity in *S. enterica*. The presumed redundant activity of this kinase is not sufficient to fully recover growth of a *pduX* strain at

levels of Cby below 1 nM (Fig. 3.4A), or when cells are grown under conditions that require a high level of AdoCba production, such as growth on ethanolamine (Fig. 3.4B) or 1,2-propanediol as a carbon source (14).

Increasing protein level of *Mm*CobD$^{WT}$ did not restore AdoCba biosynthesis (hence growth) of a *cobD* or *pduX* strain (data not shown). However, increasing the concentration of Cby in the medium did improve the growth of cells carrying plasmids encoding *Mm*CobD$^{WT}$ or *Mm*CobD$^{1-385}$ (Fig. S3.2). It is likely that increasing the level of Cby augments flux through the late steps of the AdoCba biosynthetic pathway (Fig. 3.1). There is precedence for this kind of metabolic flux response within the AdoCba biosynthetic pathway. Increasing the level of exogenously provided precursors such as AP, DMB, α-ribazole, Cby, or Cbi improves the growth of a *S. enterica cobC* strain due to increased flux through the pathway along with the presence of a phosphatase with redundant CobC activity (55).

**The *C*-terminus of *Mm*CobD$^{WT}$ is not required for, but affects the kinase activity.** The *N*-terminal domain of *Mm*CobD$^{WT}$ complements a *S. enterica cobD* or *pduX* strain. However, the *Mm*CobD$^{1-385}$ protein appears to be a less efficient enzyme *in vivo,* as it does not fully complement a *cobD* strain grown in low AdoCba demanding conditions (Figs. 3.4A & S3.2), or a *pduX* strain grown on ethanolamine, a high AdoCbal demanding condition (Fig. 3.4B). *In vitro*, the *Mm*CobD$^{WT}$ enzyme has 8-fold higher specific activity for ATP than does the *N*-terminal domain alone. While the *C*-terminus is not required for activity, its presence greatly enhances the activity (Table 3.1). The addition of L-Thr lowers the specific activity of the *Mm*CobD$^{WT}$ enzyme (10.25 to 7.61 μmol min$^{-1}$ mg$^{-1}$), but increases the activity of the *N*-terminal domain (1.33 to 3.65 μmol min$^{-1}$ mg$^{-1}$) (Figs. 3.6D, F). This seems to contradict the finding that the addition of L-Thr increases the activity of *Mm*CobD$^{WT}$ but has little effect on *Mm*CobD$^{1-385}$ (Fig 3.6A). The reactions in figures 3.4, 3.6C, 3.6D and figure S3.3 are endpoint assays measured in 1-h incubations to allow completion of the reaction. We have shown that both enzymes hydrolyzed ATP to ADP in the absence of the L-Thr co-substrate. When the reactions are allowed to go to completion, the *Mm*CobD$^{1-385}$ enzyme simply converts more ATP to ADP, regardless of the L-Thr concentration, while the *Mm*CobD$^{WT}$ protein likely generates more of the desired product of L-Thr-P when both ATP and L-

Thr are present. It is noteworthy that the addition of either L-Thr or L-Thr-P enhances the activity of *Mm*CobD$^{WT}$ (Fig. 3.4D). Taken together, the endpoint and specific activity determinations suggest that the *N*-terminal domain appears to have a higher turnover rate than does the full-length protein (Figs. 3.4C, Figs. S3.3D, S3.3H) suggesting that while *Mm*CobD$^{WT}$ is a faster enzyme, *Mm*CobD$^{1\text{-}385}$ can convert more of the substrate to product if the reaction is allowed to go to completion. This suggests that removal of the *C*-terminal domain increases the affinity of the enzyme for its substrates allowing more product to be generated. That is to say, the presence of the *C*-terminal domain somehow increases the affinity of the enzyme for the substrates. The *C*-terminal domain may prevent rapid hydrolysis of ATP, enhance or inhibit binding of L-Thr, or the release of the L-Thr-P, ADP, or AP-P products. The bottleneck might also occur with the shifting of the L-Thr-P product within the active site or to a second site where the decarboxylation reaction occurs. That the *N*-terminal truncation protein reaches saturation between 25 and 50 mM L-Thr concentration and the full-length protein does not up to 100 mM L-Thr indicates that the binding of L-Thr is unhindered in the *Mm*CobD$^{1\text{-}385}$ protein relative to the *Mm*CobD$^{WT}$ protein. The *C*-terminal domain may regulate the entry of L-Thr into the active site and the release of the final products, ADP and AP-P, hence functioning as a 'gate' to the active site.

*Mm*CobD$^{WT}$ displayed sensitivity to several ATPase inhibitors including $BeF_2$, ADP, ADP-γ-S, PPPi, and $Na_3VO_4$ (Fig. S3.3G). In contrast, the *Mm*CobD$^{1\text{-}385}$ protein did not display any loss of activity in the presence of the non-hydrolyzable competitive inhibitor ADP-γ-S (Fig. S3.3H). This surprising result supports the idea that the absence of the *C*-terminal domain allows unhindered access to the active site by substrates. We hypothesize that the binding of ATP to the *Mm*CobD$^{1\text{-}385}$ protein is not very strong and the on/off rate is rapid, allowing any bound ADP-γ-S to be quickly displaced by ATP without having a detrimental effects on activity. It is also possible that ADP-γ-S is simply not able to bind or access the active site of the truncated protein.

**The metal-binding domain of *Mm*CobD may facilitate protein-protein interactions.** There are a few examples of zinc fingers fused to kinases (56, 57) or NTPases (58) in which the zinc finger facilitates the enzymatic reaction or binds DNA and modulates transcription (58). The 'zinc finger' domains of other

enzymes have been found to be necessary for protein-protein interactions. The eukaryotic protein Ikaros forms homodimers via *C*-terminal zinc finger, which in turn allows it to form heterodimers with its partner protein, Aiolos (59, 60). A similar zinc finger mediated interaction occurs between the erythroid DNA-binding protein, GATA-1 and FOG-1 (58). The role of the putative 'zinc finger'/metal-binding domain in *Mm*CobD$^{WT}$ is unclear but it may be facilitating protein-protein interactions between CobD and the next enzyme in the pathway, CbiB (Fig. 3.1).

***Mm*CobD$^{WT}$ may form a phosphoenzyme intermediate.** The $^{31}$P-NMR results presented in Figure 3.5 consists of reactions in which the proteins were removed by filtration prior to the acquisition of $^{31}$P-NMR spectra. In *Mm*CobD$^{WT}$ reactions with ATP and L-Thr in which the protein was not removed, an additional peak appeared with a chemical shift of 1.95 ppm (data not shown). The peak was associated with protein, as it is absent in reactions where the protein had been removed. We suspect that it may represent a phosphoenzyme intermediate. It is unlikely to be PLP, as the chemical shift did not correspond to the chemical shift of free PLP in solution (4.2/3.9 ppm). PLP bound to *Mm*CobD$^{WT}$ or *Se*CobD$^{WT}$ was not detectable at the concentration of protein (3 μM) used in the reactions.

**Is L-thr kinase activity a feature of CobD that is widespread in other Cba producers?**

The presence of the ATPase activity in the *N*-terminus of *Mm*CobD$^{WT}$ in the absence of the metal-binding domain poses the question of whether CobD from other organisms also have L-Thr kinase function? If so how can these bifunctional CobDs be identified and differentiated from CobDs like the one found in *S. enterica*, which does not have L-thr kinase activity? On average, CobD homologues are about 28% identical, but CobDs are commonly mis-annotated as histidinol-phosphate aminotransferases (*hisC*) due to structural and amino acid sequence similarity between CobD and HisC (32, 54). Given this hurdle, the identification of a motif(s) responsible for L-Thr and ATP binding may be challenging. The protein sequence of *Mm*CobD$^{WT}$ does not encode any canonical ATP binding motifs. Potential novel ATP or L-Thr binding motifs are the two small insertions spanning position 68-74, and 278-283 (Fig. 3.2) and the 13-amino acid extension at the *N*-terminus. The role that these extensions may play in the L-Thr kinase activity is currently under investigation.

**Conclusions.** CobD from *M. mazei* is a novel, bifunctional enzyme that encodes L-Thr decarboxylase activity and L-Thr kinase activity. Also unique to this enzyme is the fusion of a 108-residue *C*-terminal metal binding domain that is not required for either activity but does influence both activities. The function of the *C*-terminal domain of *Mm*CobD$^{WT}$ remains to be determined.

**ACKNOWLEDGMENTS**

This work was supported by NIH grant R37 GM40313 to J.C.E.-S. N.K.T. was supported in part by NIH grant F31 GM095230 and by an Advanced Opportunity Fellowship awarded by the Graduate School of the University of Wisconsin, Madison. C.L.Z. was supported in part by the NIH grant F31 GM64009. We thank P. Renz for his gift of CNCby, G. Gottschalk for his gift of *M. mazei* genomic DNA, and A. I. Siesarev for the gift of *M. kandleri* genomic DNA. Thanks also to D. Cui at the University of Georgia Chemical Sciences Magnetic Resonance facility for her assistance with NMR. We are grateful to M. P. Pappalardo for her assistance with R and phylogenetic analysis.

**FIGURES**

**Table 3.1. Specific activity of *M. mazei* CobD.** Specific activity of truncated and full-length wild-type *M. mazei* CobD enzyme was test for ATPase activity in the presence and absence of L-Thr as a cosubstrate. Values are reported as mean ± standard deviation of three activity measurements. Activity was measured with an NADH consumption assay (see *material and methods*).

| **Protein** | **ATP only** (μmol ATP $min^{-1}$ $mg^{-1}$) | **ATP + L-Thr** (μmol ATP $min^{-1}$ $mg^{-1}$) |
|---|---|---|
| *Mm*CobD$^{WT}$ | 10.25 ± 0.27 | 7.64 ± 0.02 |
| *Mm*CobD$^{1-385}$ | 1.33 ± 0.04 | 3.65 ± 0.11 |

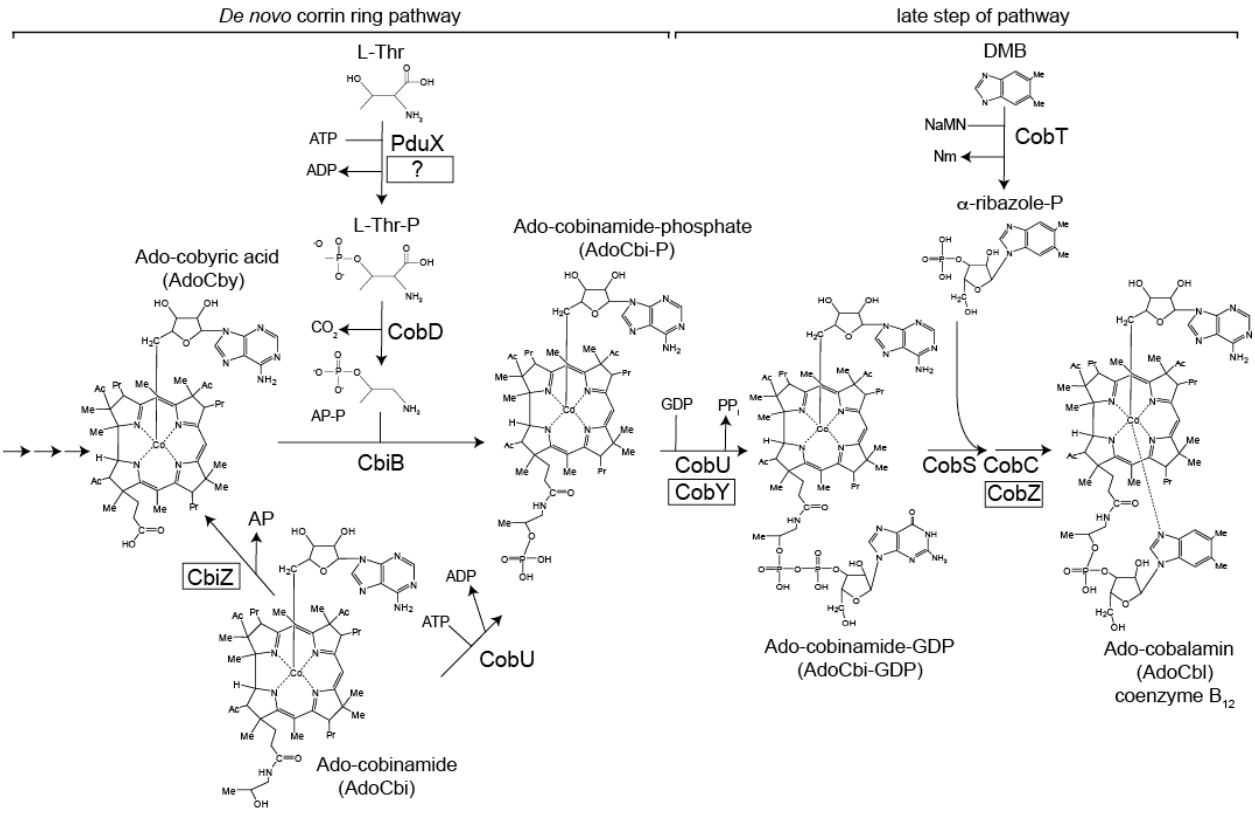



**Figure 3.1. Late steps of the anaerobic pathway of cobalamin biosynthesis in bacteria and archaea.** Names of intermediates are located below the structure. Non-homologous archaeal enzyme names are boxed in black. AdoCby, adenosylcobyric acid; AdoCbi, adenosylcobinamide; AdoCbi-P, adenosylcobinamide phosphate; AdoCbi-GDP, adenosylcobinamide-GDP; AP-P, *1*-amino-*2*-propanol phosphate; AP, *1*-amino-*2*-propanol; L-Thr-P, L-threonine-*O*-*3*-phosphate; L-Thr, L-threonine. CbiZ, adenosylcobinamide amidohydrolase; CbiB, adenosylcobinamide synthase; CobY, adenosylcobinamide-phosphate guanylyltransferase, CobS; adenosylcobalamin-5′-phosphate synthase; CobD, L-threonine-*O*-*3*-phosphate decarboxylase; CobT, NaMN:5,6-dimethylbenzimidazole phosphoribosyltransferase; CobU, adenosylcobinamide kinase / adenosylcobinamide-phosphate guanylyltransferase; PduX, L-threonine kinase; CobC, adenosylcobalamin-5′-phosphate phosphatase; CobZ; adenosylcobalamin-5′-phosphate phosphatase.

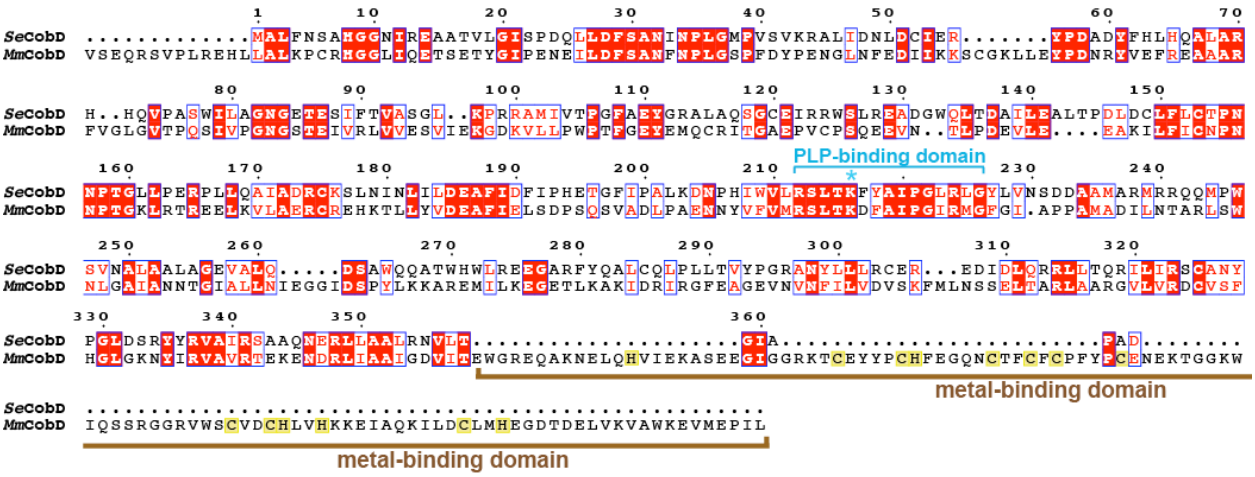



**Figure 3.2. Protein sequence alignment.** Protein sequence alignment of CobD from *S. enterica* and *M. mazei*. Conserved residues are highlighted in red. Residues with similar properties are boxed in blue. The pyridoxal phosphate (PLP) binding domain is bracketed in light blue and the active site lysine is marked with an asterisk. The cysteine rich/metal binding domain indicated with brown brackets and the cysteine and histidine residues in this region are highlighted yellow.

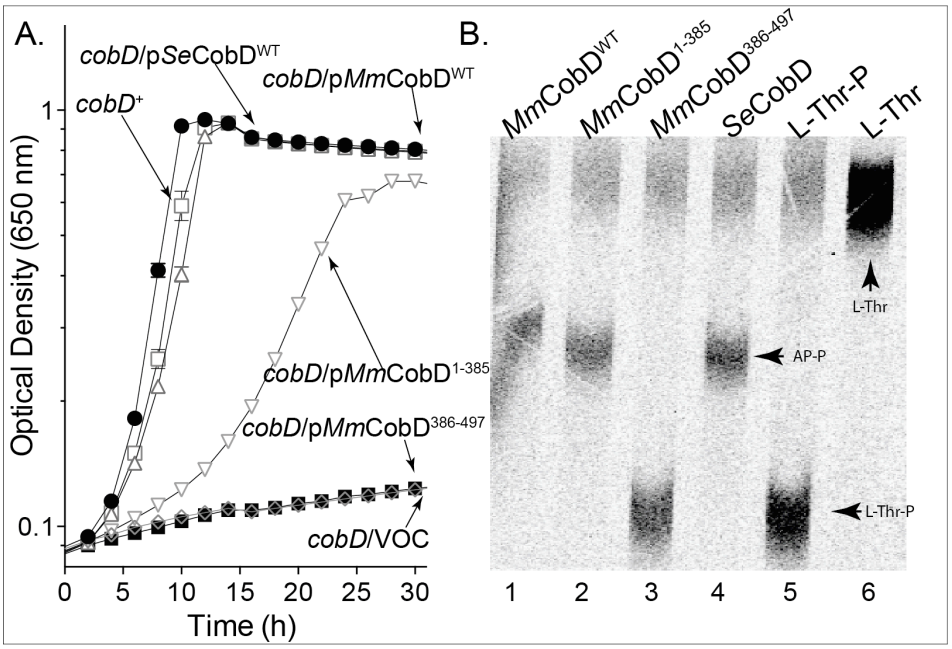



**Figure 3.3. *M. mazei cobD* has L-threonine decarboxylase activity *in vivo* and *in vitro*. A.** Cobalamin-dependent growth assessment of *S. enterica* wild-type (*cobD*$^+$) and Δ*cobD* strains with plasmids expressing wild-type (WT) or truncated *M. mazei* CobD (*Mm*CobD) proteins, CobD from *S. enterica* (*Se*CobD), or pBAD24 empty vector only control (VOC). Cells were grown aerobically at 37ºC in NCE minimal medium with glycerol (22 mM) as the sole carbon and energy source, supplemented with Cby (1 nM), arabinose (250 μM), ampicillin (100 μg/mL), and $MgSO_4$ (1 mM). **B.** Phosphorimage of the resolution of products and reactants by TLC of the *M. mazei* and *S. enterica* CobD L-Thr-P decarboxylation reaction. AP-P, *1*-amino-*2*-propanol phosphate; L-Thr-P, L-Threonine-*O*-*3*-phosphate; L-Thr, L-threonine.

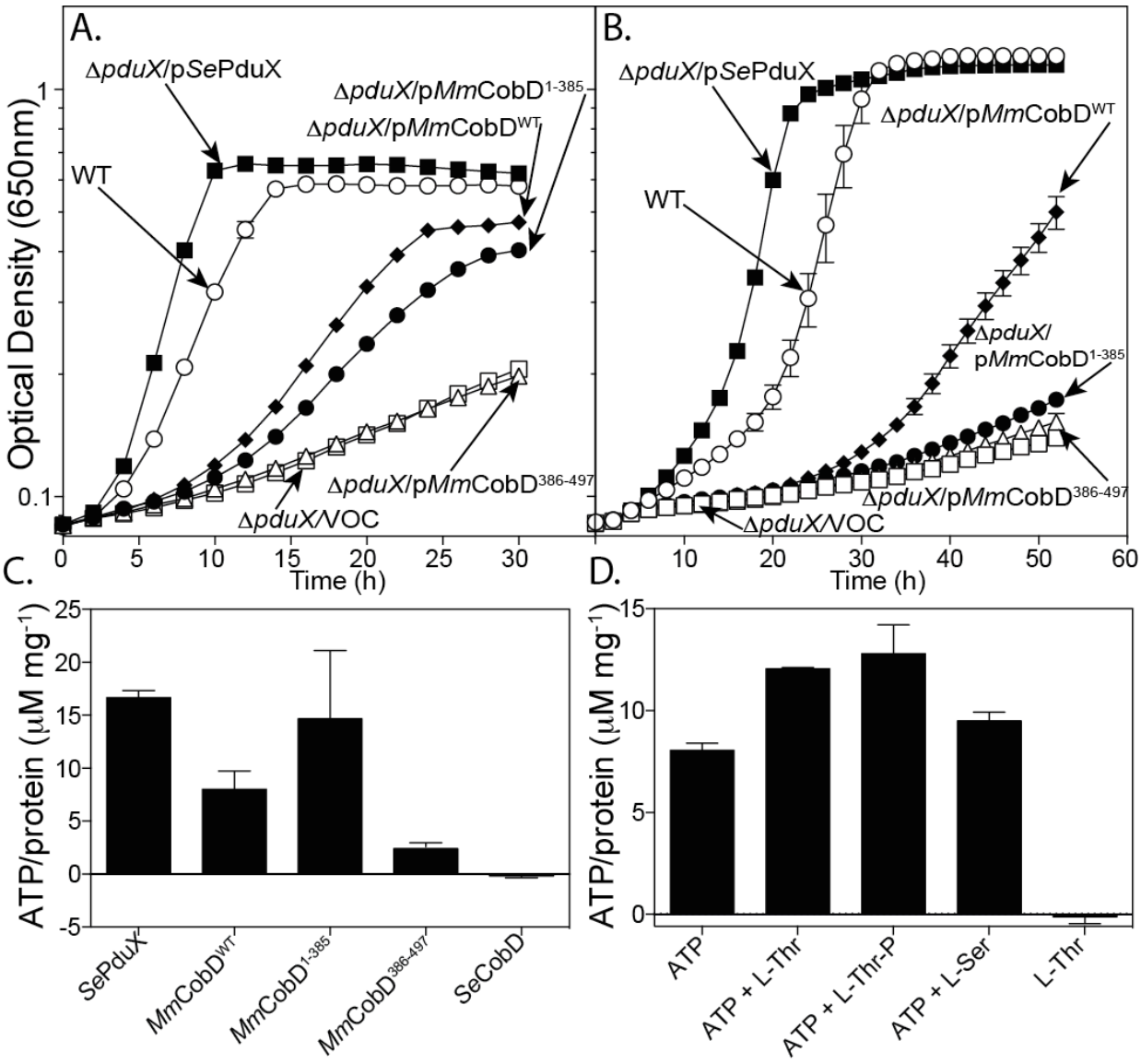



**Figure 3.4.** ***Mm*CobD complements a *S. enterica pduX* strain *in vivo* and has ATPase activity *in vitro*.** Cobalamin-dependent growth was assessed with *S. enterica* cells grown aerobically at 37ºC in NCE minimal medium with **A.** glycerol (22 mM) and Cby (1 nM) or **B.** ethanolamine (90 mM) and Cby (300 nM), supplemented with, DMB (150 μM), ampicillin (100 μg/mL), $MgSO_4$ (1 mM). *S. enterica* wild-type (WT) and *pduX* strains with plasmids expressing wild-type or truncations of *M. mazei* CobD (*Mm*CobD), PduX from *S. enterica* (*Se*PduX), or pBAD24 empty vector only control (VOC). **C**. ATPase activity assay measured with ADP-glo Kinase Assay Kit (Promega). Reaction mixture contained HEPES buffer (50 mM, pH 6.8), NaCl (20 mM), TCEP (1mM) $MgCl_2$ (1 mM), ATP (100 μM), L-Thr (300 μM), aerobically purified protein (100 nM) incubated at 25°C for 1 h. **D**. Reaction mixture contained HEPES buffer (50 mM, pH 7), TCEP (2 μM) $MgCl_2$ (1 mM), ATP (10 mM), L-Thr or L-Ser (10 mM), aerobically purified *Mm*CobD (72 nM) incubated at 25°C for 1 h. No enzyme controls were subtracted to reduce background. Values are compared to a standard curve for ATP conversion vs. luminescence and converted into units of mM ATP per μM protein, with the standard error mean of triplicate reactions represented by the error bars.

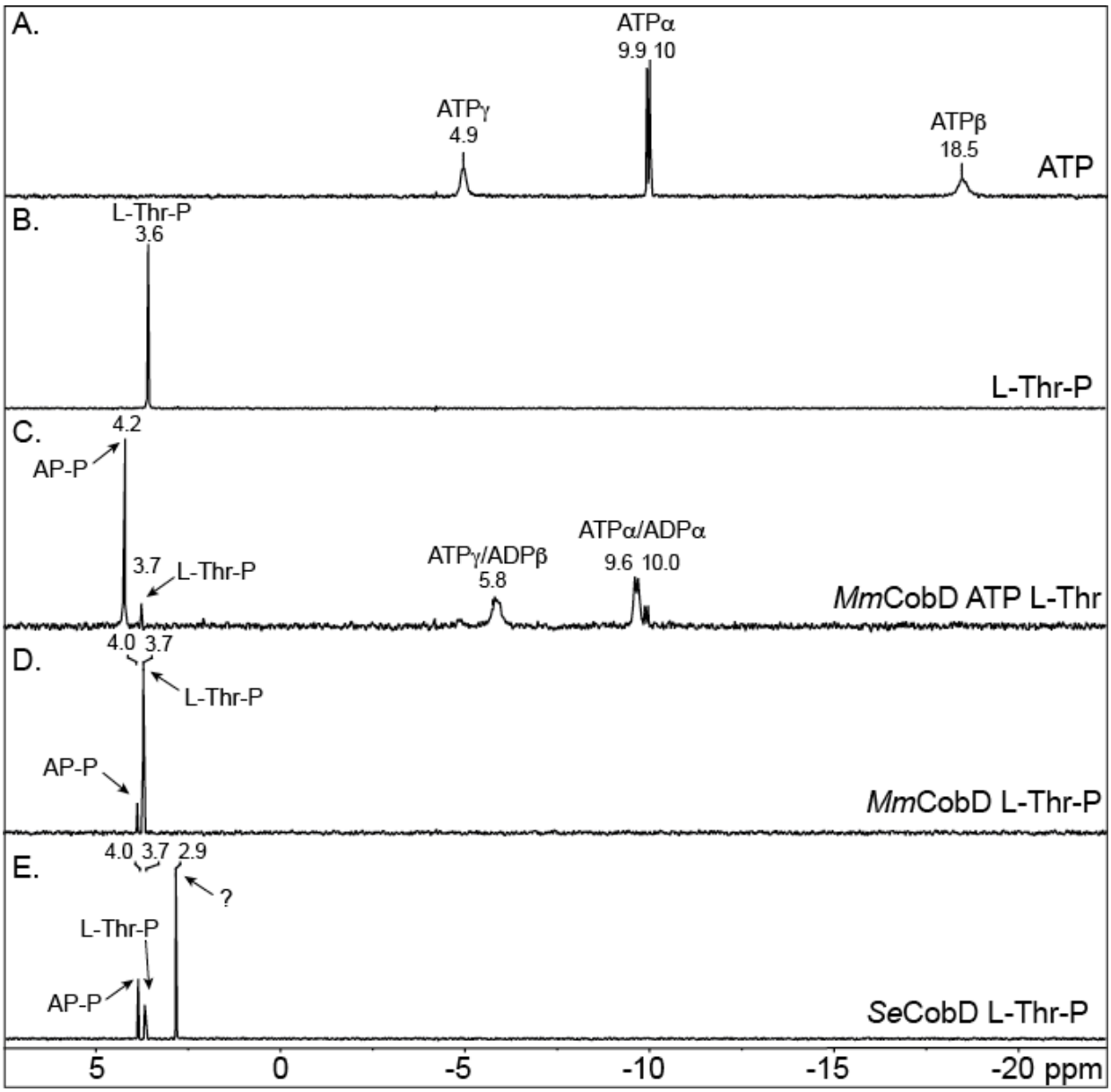



**Figure 3.5. $^{31}$P-NMR spectra of *Mm*CobD kinase reaction.** Reactions mixtures containing $MgCl_2$ (1 mM), ATP (0.3 mM), L-Thr (0.3 mM) and protein (3 μM) were incubated for 1 h at 25ºC. Each panel is labeled with the substrate and or protein reaction mixture. Each peak is labeled with the chemical shift value and the substrate that it represents based on the corresponding chemical shifts of the standards. **A.** ATP standard. **B.** L-threonine-*O*-3-phosphate-P (Thr-P) standard. **C.** Reaction containing ATP, L-Thr, and *Mm*CobD. **D.** Reaction containing L-Thr-P, and *Mm*CobD. **E.** Reaction containing L-Thr-P, and *Se*CobD.

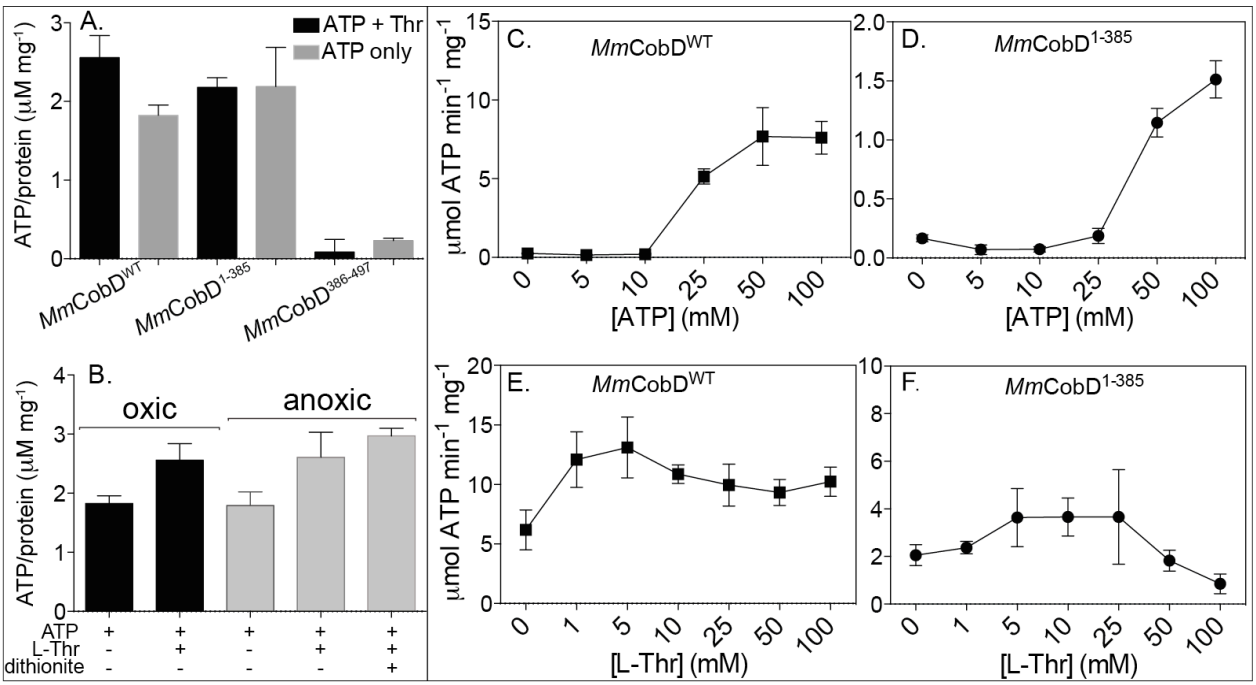



**Figure 3.6. Specific activity. A.** & **B.** ATPase activity measured with ADP-glo Kinase Assay Kit (Promega). Reaction mixture contained HEPES buffer (50 mM, pH 6.8), TCEP (2 μM) $MgCl_2$ (1 mM), ATP (10 mM), L-Thr (50 mM), protein (100 ng) and dithionite (2 mM) where indicated. Reactions incubated at 25°C for 1 h. Error bars represent standard error mean of triplicate reactions. Indirect measurement of specific activity of *M. mazei* full-length *Mm*CobD$^{WT}$ and truncated *Mm*CobD$^{M1\text{-}W385}$ proteins as a function of **C.** & **D.** ATP concentration and **E.** & **F.** L-threonine (L-Thr) concentration expressed as μmol of ATP per min per mg of protein with the standard error mean of triplicate reactions represented by the error bars. Activity was measured by a NADH consumption assay described in *materials and methods*. Assay was performed with purified protein (3 μM), HEPES buffer (50 mM pH 7.5), $MgCl_2$ (5 mM), PEP (3 mM), NADH (0.1 mM), pyruvate kinase (1 U), lactate dehydrogenase (1.5 U) incubated at 25°C with measurements taken every 11 sec for 20 min. For ATPase specific activity L-Thr concentration was held at 50 mM while ATP concentration was varied (5 – 100 mM). To measure the effect of the cosubstrate on ATPase activity, ATP concentration was held at 50 mM while the concentration L-Thr was varied (0 – 100 mM).

## SUPPLEMENTAL FIGURES

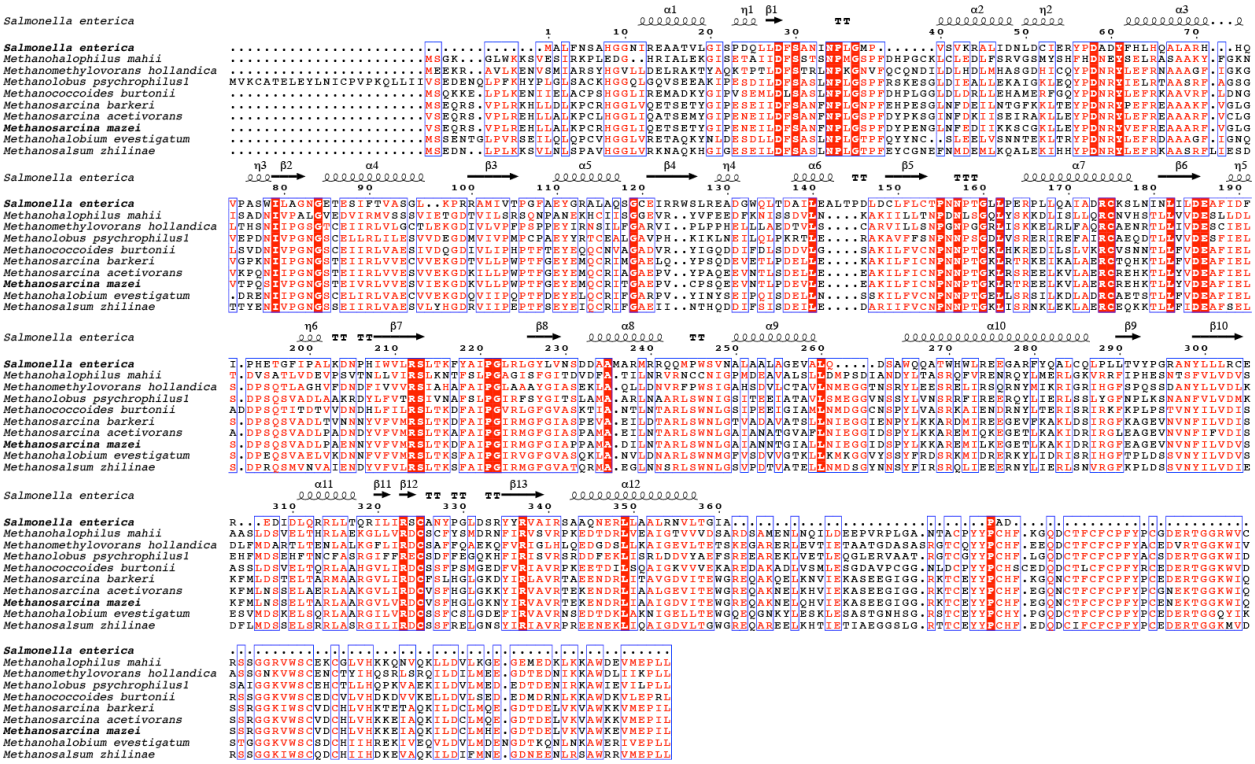
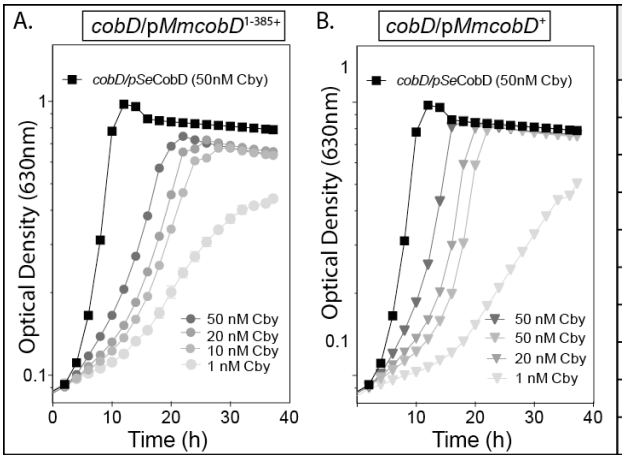


**Figure S3.1. Multiple sequence alignment of CobD from methanogenic archaea with CobD from *S. enterica*.** Conserved residues are highlighted in red. Residues with similar properties are boxed in blue.




| C. Strain | Cby (nM) | growth rate (h) | lag time (h) |
|---|---|---|---|
| *cobD/pMmcobD*$^{1\text{-}385}$ | 1 | 8.2 | 12 |
| | 10 | 4.5 | 10 |
| | 20 | 4.3 | 11 |
| | 50 | 3.7 | 4 |
| *cobD/pMmcobD* | 1 | 9.2 | 16 |
| | 10 | 2.2 | 12 |
| | 20 | 2.2 | 8 |
| | 50 | 2.3 | 4 |
| *cobD/pSecobD* | 1 | 2.2 | 3 |
| | 50 | 1.5 | 3 |

**Figure S3.2. The intracellular concentration of Cby is the limiting factor for the complementation of *S. enterica* by *Mm*CobD.** Cobalamin-dependent growth was assessed in minimal medium supplemented with glycerol and Cby. Shown is the effect increasing levels of Cby (1, 10, 20, and 50 nM) on the complementation of a *S. enterica cobD* strain with plasmids expressing **A.** *Mm*CobD$^{M1\text{-}W385}$ (p*Mm*COBD7) or **B**. *Mm*CobD$^{WT}$ (pMmCOBD18), relative to wild-type vector only control (VOC) (black squares). **C.** Table of growth rates, lag time and generation time as a function of Cby concentration.

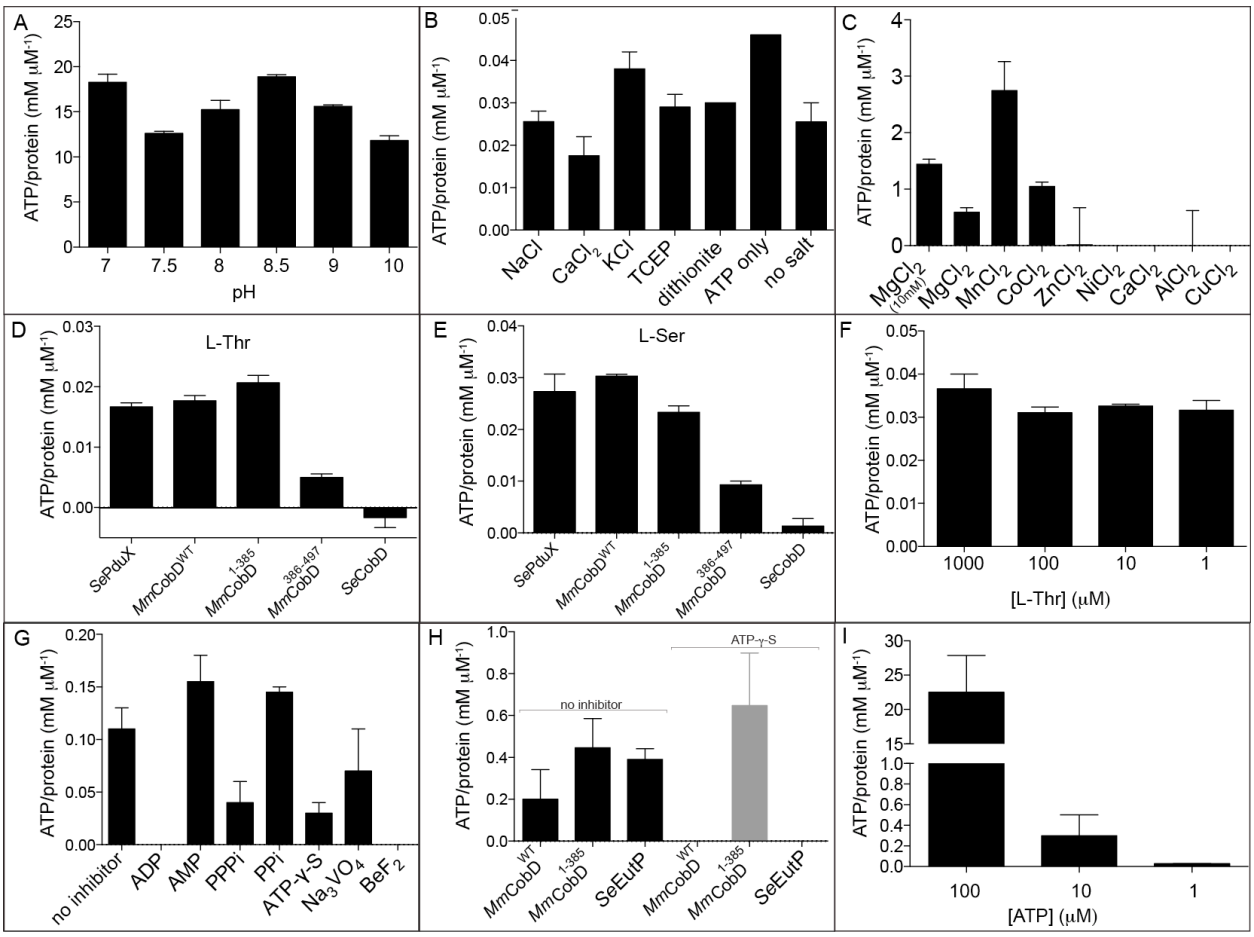



**Figure S3.3 Optimization of *Mm*CobD reaction conditions.** ATPase activity assay measured with ADP-glo Kinase Assay Kit (Promega), expressed as mM of ATP per μM of protein, with the standard error mean of triplicate reactions represented by the error bars. Unless otherwise indicated, reaction mixture contained HEPES buffer (50 mM, pH 8.5), $MgCl_2$ (1 mM), ATP (100 μM), L-Thr (100 μM), protein (100 nM), incubated at 25°C for 1 h. **A.** *Mm*CobD activity as a function of pH. **B**. *MmCobD* activity in the presences of salts (100 mM) or other additions TCEP (1 mM), sodium dithionite (1 mM). **C**. *MmCobD* activity as a function of added divalent metals (1 mM). **D**. Enzymatic activity of truncations and wild-type *MmCobD* in the presence of ATP (10 mM) and L-threonine (L-Thr, 50 mM) or **E**. L-serine (L-Ser, 50 mM). *Se*CobD and *Se*PduX proteins were used as negative and positive controls, respectively. **F**. *MmCobD* activity as a function of L-threonine (L-Thr) concentration (1 - 100 μM). **G**. *Mm*CobD (6 μM) activity in the presence of ATPase/kinase inhibitors ADP (200 mM), AMP (200 mM), sodium pyrophosphate (PPi, 10 mM), sodium triphosphate (PPPi, 10 mM), ADP-γ-S, (100 μM), sodium *ortho*-vanadate ($Na_3VO_4$, 1 mM), and sodium beryllium fluoride ($BeF_2$, 2 mM) in HEPES buffer (50 mM, pH 7.5). **H**. $MmCobD^{WT}$ and $MmCobD^{1\text{-}385}$ activity in the presence ADP-γ-S, (100 μM). *Se*EtuP enzyme was added as a positive control. Protein concentrations (100 μM). **I**. *MmCobD* activity as a function of ATP concentration (1 - 100 μM) with L-Thr (50 mM).

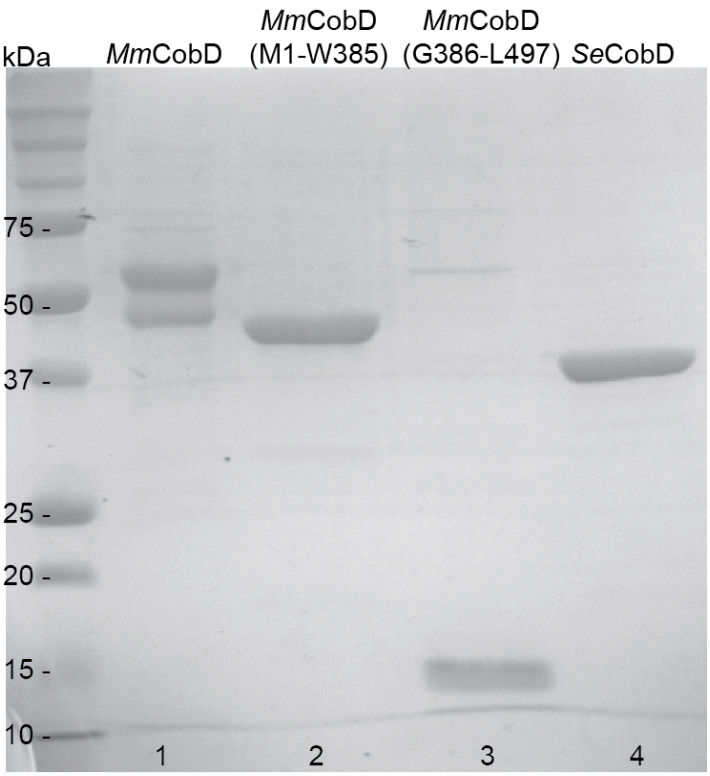



**Figure S3.4. Purification of *Mm*CobD proteins.** SDS-PAGE gel of aerobically purified proteins. Lane 1; full length *Mm*CobD, 2; *Mm*CobD$^{M1\text{-}W385}$, 3; *Mm*CobD$^{G386\text{-}L497}$, 4; *Se*CobD.

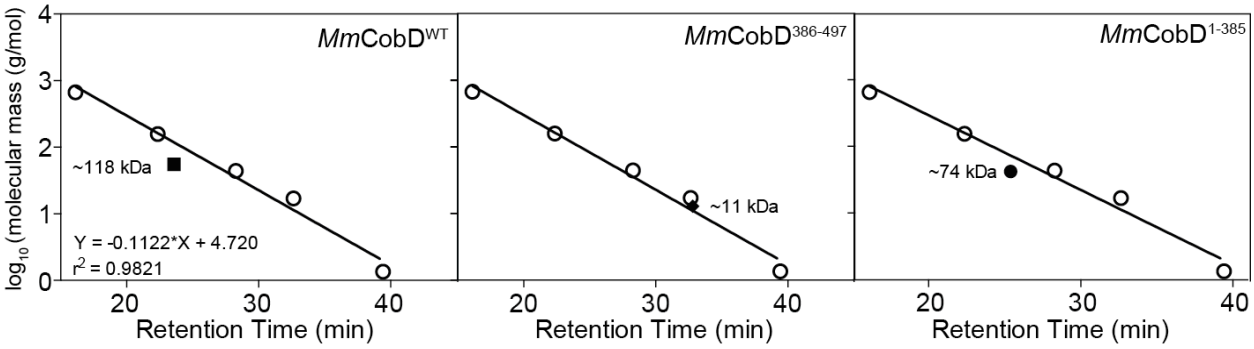



**Figure S3.5. Oligomeric state of *Mm*CobD protein.** Gel filtration analysis of *Mm*CobD and truncations. Samples were applied to a HiPrep 26/60 Sephacryl S-100 High Resolution gel filtration column using isocratic elution with sodium phosphate (50 mM, pH 7.4) containing 150 mM NaCl. Calibration was performed with Bio-Rad gel filtration standards supplemented with BSA and DNaseI with linear regression to generate the standard curve. *Mm*CobD$^{WT}$ and *Mm*CobD$^{1\text{-}385}$ correspond to the approximate molecular weight of a dimer and *Mm*CobD$^{386\text{-}497}$ corresponds to the molecular weight of a monomer.

**Table S3.1. Strains and plasmids list**. *S. enterica* strains are derivatives of sv. Typhimurium strain LT2. Strains and plasmids were constructed during the course of this work unless stated otherwise.

| Strains | Relevant genotype | Reference / Source |
|---|---|---|
| ***Salmonella enterica*** | | |
| JE7088 | *metE2702 ara-9* | Laboratory collection |
| **Derivatives of JE7088** | | |
| JE21644 | / pBAD24 *bla*$^+$ | |
| JE21557 | / pTEV5 *bla*$^+$ | |
| JE12656 | Δ*pduX*516 | |
| JE2216 | *cobD*1302::Tn*10d*(*cat*$^+$) | Laboratory collection |
| JE21654 | *cobD*1302::Tn*10d*(*cat*$^+$) / pBAD24 *bla*$^+$ | Laboratory collection |
| JE21653 | *cobD*1302::Tn*10d*(*cat*$^+$) / pTEV5 *bla*$^+$ | |
| JE6158 | *cobD*1302::Tn*10d*(*cat*$^+$) / pCOBD6 *bla*$^+$ | Laboratory collection |
| JE21656 | *cobD*1302::Tn*10d*(*cat*$^+$) / p*Mm*COBD7 *bla*$^+$ | |
| JE21781 | *cobD*1302::Tn*10d*(*cat*$^+$) / p*Mm*COBD9 *bla*$^+$ | |
| JE21657 | *cobD*1302::Tn*10d*(*cat*$^+$) / p*Mm*COBD13 *bla*$^+$ | |
| JE21646 | *cobD*1302::Tn*10d*(*cat*$^+$) / p*Mm*COBD17 *bla*$^+$ | |
| JE21659 | *cobD*1302::Tn*10d*(*cat*$^+$) / p*Mm*COBD18 *bla*$^+$ | |
| JE21782 | *cobD*1302::Tn*10d*(*cat*$^+$) / p*Mm*COBD19 *bla*$^+$ | |
| JE18765 | *pduX*516 / p*Mm*COBD7 *bla*$^+$ | |
| JE21619 | *pduX*516 / p*Mm*COBD9 *bla*$^+$ | |
| JE18970 | *pduX*516 / p*Mm*COBD13 *bla*$^+$ | |
| JE19204 | *pduX*516 / p*Mm*COBD17 *bla*$^+$ | |
| JE21336 | *pduX*516 / p*Mm*COBD18 *bla*$^+$ | |
| JE19205 | *pduX*516 / p*Mm*COBD19 *bla*$^+$ | |
| ***Escherichia coli*** | | |
| C43 (λDE3) | F$^-$ ompT gal hsd$S_B$ (rB$^-$mB) [dcm] [Ion] | (33) |
| DH5α | F$^-$/ endA1 hsdR17(rk$^-$, mk$^+$) glnV44 thi-1 recA1 gyrA96 (Nx$^R$) relA1 U169 deoR (Φ80-dlacZ M15 Δ(lacZYA-argF) phoAsupE44 relA1 | (28) |
| **Plasmids** | | |
| pCOBD6 | *S. enterica cobD*$^+$ in pT7-7 *bla*$^+$ | Laboratory collection |
| p*Mm*COBD7 | *M. mazei cobD*$^{1\text{-}385}$ in pBAD24 *bla*$^+$ | |
| p*Mm*COBD9 | *M. mazei cobD*$^{1\text{-}385}$ in pTEV5 *bla*$^+$ | |
| p*Mm*COBD13 | *M. mazei cobD*$^{386\text{-}497}$ in pBAD24 *bla*$^+$ | |
| p*Mm*COBD17 | *M. mazei cobD*$^+$ in pBAD24 *bla*$^+$ | |
| p*Mm*COBD18 | *M. mazei cobD*$^+$ pTEV5 *bla*$^+$ | |
| p*Mm*COBD19 | *M. mazei cobD*$^{386\text{-}497}$ pTEV5 *bla*$^+$ | |
| pPDU15 | *S. enterica pduX*$^+$ in pBAD30 *bla*$^+$ | Laboratory collection |
| pPDU23 | *S. enterica pduX*$^+$ in pTEV5 *bla*$^+$ | Laboratory collection |
| pBAD24 | cloning/complementation vector *bla*$^+$ | (29) |
| pBAD30 | cloning/complementation vector *bla*$^+$ | (29) |
| pTEV5 | cloning/overexpression vector, *N*-terminal rTEV cleavable $His_6$ tag *bla*$^+$ | (30) |

**REFERENCES**


1. **Banerjee R, Ragsdale SW.** 2003. The many faces of vitamin B12: catalysis by cobalamin-dependent enzymes. Annu Rev Biochem **72:**209-247.

2. **Yeliseev A, Gartner P, Harms U, Linder D, Thauer RK.** 1993. Function of methylcobalamin: coenzyme M methyltransferase isoenzyme II in *Methanosarcina barkeri*. Arch Microbiol **159:**530-536.

3. **DiMarco AA, Bobik TA, Wolfe RS.** 1990. Unusual coenzymes of methanogenesis. Annu Rev Biochem **59:**355-394.

4. **Taylor RT, Weissbach H.** 1973. $N^5$-methylenetetrahydrofolate-homocysteine methyltransferases, p 121-165. *In* Boyer PD (ed), The Enzymes, vol 9. Academic Press, Inc., New York.

5. **Allers T, Mevarech M.** 2005. Archaeal genetics - the third way. Nat Rev Genet **6:**58-73.

6. **Graham DE, Overbeek R, Olsen GJ, Woese CR.** 2000. An archaeal genomic signature. Proc Natl Acad Sci U S A **97:**3304-3308.

7. **Woodson JD, Zayas CL, Escalante-Semerena JC.** 2003. A new pathway for salvaging the coenzyme $B_{12}$ precursor cobinamide in archaea requires cobinamide-phosphate synthase (CbiB) enzyme activity. J Bacteriol **185:**7193-7201.

8. **Zayas CL, Woodson JD, Escalante-Semerena JC.** 2006. The *cobZ* gene of *Methanosarcina mazei* Gö1 encodes the nonorthologous replacement of the α-ribazole-5'-phosphate phosphatase (CobC) enzyme of *Salmonella enterica*. J Bacteriol **188:**2740-2743.

9. **Buan NR, Rehfeld K, Escalante-Semerena JC.** 2006. Studies of the CobA-type ATP:Co(I)rrinoid adenosyltransferase enzyme of *Methanosarcina mazei* strain Gö1. J Bacteriol **188:**3543-3550.

10. **Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS.** 2003. Comparative genomics of the vitamin $B_{12}$ metabolism and regulation in prokaryotes. J Biol Chem **278:**41148-41159.

11. **Woodson JD, Peck RF, Krebs MP, Escalante-Semerena JC.** 2003. The *cobY* gene of the archaeon *Halobacterium* sp. strain NRC-1 is required for de novo cobamide synthesis. J Bacteriol **185:**311-316.

12. **Woodson JD, Reynolds AA, Escalante-Semerena JC.** 2005. ABC transporter for corrinoids in *Halobacterium* sp. strain NRC-1. J Bacteriol **187:**5901-5909.

13. **Woodson JD, Escalante-Semerena JC.** 2006. The *cbiS* gene of the archaeon *Methanopyrus kandleri* AV19 encodes a bifunctional enzyme with adenosylcobinamide amidohydrolase and alpha-ribazole-phosphate phosphatase activities. J Bacteriol **188:**4227-4235.

14. **Fan C, Bobik TA.** 2008. The PDUX enzyme of *Salmonella enterica* is an L-threonine kinase used for coenzyme $B_{12}$ synthesis. J Biol Chem **283:**11322-11329.

15. **Fan C, Fromm HJ, Bobik TA.** 2009. Kinetic and functional analysis of L-threonine kinase, the PduX enzyme of *Salmonella enterica*. J Biol Chem **284:**20240-20248.

16. **Zhang Y, Rodionov DA, Gelfand MS, Gladyshev VN.** 2009. Comparative genomic analyses of nickel, cobalt and vitamin B12 utilization. BMC Genomics **10:**78.

17. **Brushaber KR, O'Toole GA, Escalante-Semerena JC.** 1998. CobD, a novel enzyme with L-threonine-*O*-3-phosphate decarboxylase activity, is responsible for the synthesis of (*R*)-1-amino-2-propanol *O*-2-phosphate, a proposed new intermediate in cobalamin biosynthesis in *Salmonella typhimurium* LT2. J Biol Chem **273:**2684-2691.

18. **Peariso K, Zhou ZS, Smith AE, Matthews RG, Penner-Hahn JE.** 2001. Characterization of the zinc sites in cobalamin-independent and cobalamin-dependent methionine synthase using zinc and selenium X-ray absorption spectroscopy. Biochemistry **40:**987-993.

19. **Drennan CL, Huang S, Drummond JT, Matthews RG, Ludwig ML.** 1994. How a protein binds $B_{12}$: A 3.0Å X-ray structure of $B_{12}$-binding domains of methionine synthase. Science **266:**1669-1674.

20. **Hall DA, Vander Kooi CW, Stasik CN, Stevens SY, Zuiderweg ER, Matthews RG.** 2001. Mapping the interactions between flavodoxin and its physiological partners flavodoxin reductase and cobalamin-dependent methionine synthase. Proc Natl Acad Sci USA **98:**9521-9526.

21. **Banerjee RV, Matthews RG.** 1990. Cobalamin-dependent methionine synthase. FASEB J **4:**1450-1459.

22. **Datsenko KA, Wanner BL.** 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci USA **97:**6640-6645.

23. **Berkowitz D, Hushon JM, Whitfield HJ, Jr., Roth J, Ames BN.** 1968. Procedure for identifying nonsense mutations. J Bacteriol **96:**215-220.

24. **Balch WE, Wolfe RS.** 1976. New approach to the cultivation of methanogenic bacteria: 2-mercaptoethanesulfonic acid (HS-CoM)-dependent growth of *Methanobacterium ruminantium* in a pressurized atmosphere. Appl Environ Microbiol **32:**781-791.

25. **Bertani G.** 1951. Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. J Bacteriol **62:**293-300.

26. **Bertani G.** 2004. Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. J Bacteriol **186:**595-600.

27. **Raleigh EA, Lech K, Brent R.** 1989. Selected topics from classical bacterial genetics, p 1.4. *In* Ausubel FA, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (ed), Current Protocols in Molecular Biology, vol 1. Wiley Interscience, New York.

28. **Woodcock DM, Crowther PJ, Doherty J, Jefferson S, De Cruz E, Noyer-Weidner M, Smith SS, Michael MZ, Graham MW.** 1989. Quantitative evaluation of *Escherichia coli* host strains for tolerance to cytosine methylation in plasmid and phage recombinants. Nucl Acids Res **17:**3469-3478.

29. **Guzman LM, Belin D, Carson MJ, Beckwith J.** 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J Bacteriol **177:**4121-4130.
30. **Rocco CJ, Dennison KL, Klenchin VA, Rayment I, Escalante-Semerena JC.** 2008. Construction and use of new cloning vectors for the rapid isolation of recombinant proteins from *Escherichia coli*. Plasmid **59:**231-237.

31. **O'Toole GA, Rondon MR, Escalante-Semerena JC.** 1993. Analysis of mutants of defective in the synthesis of the nucleotide loop of cobalamin. J Bacteriol **175:**3317-3326.

32. **Cheong CG, Bauer CB, Brushaber KR, Escalante-Semerena JC, Rayment I.** 2002. Three-dimensional structure of the L-threonine-*O*-3-phosphate decarboxylase (CobD) enzyme from *Salmonella enterica*. Biochemistry **41:**4798-4808.

33. **Miroux B, Walker JE.** 1996. Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. J Mol Biol **260:**289-298.

34. **Varnado CL, Goodwin DC.** 2004. System for the expression of recombinant hemoproteins in *Escherichia coli*. Protein Expr Purif **35:**76-83.

35. **Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A.** 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res **31:**3784-3788.
36. **Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF.** 1999. Protein identification and analysis tools in the ExPASy server. Methods Mol Biol **112:**531-552.

37. **Altschul SF, Madden TL, Schaffer AA, Zhang J, Miller W, Lipmann DJ.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl Acids Res **25:**3389-3402.

38. **Markowitz VM, Korzeniewski F, Palaniappan K, Szeto E, Werner G, Padki A, Zhao X, Dubchak I, Hugenholtz P, Anderson I, Lykidis A, Mavromatis K, Ivanova N, Kyrpides NC.** 2006. The integrated microbial genomes (IMG) system. Nucleic Acids Res **34:**D344-348.

39. **Edgar RC.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32:**1792-1797.

40. **Gouet P, Courcelle E, Stuart DI, Metoz F.** 1999. ESPript: multiple sequence alignments in PostScript. Bioinformatics **15:**305-308.

41. **Bergmeyer HU, Bergmeyer J, Grassl M, Berger R.** 1985. Methods of enzymatic analysis. vol. iv. "enzymes 2: Esterases, glycosidases, lyases, ligases". 3rd Edition. Weinheim; Deerfield Beach, Florida; Basel: Verlag Chemie, 1984. 426 S., 258 DM. Acta Biotechnologica **5:**114-114.

42. **Horswill AR, Escalante-Semerena JC.** 2002. Characterization of the propionyl-CoA synthetase (PrpE) enzyme of *Salmonella enterica*: Residue Lys592 is required for propionyl-AMP synthesis. Biochemistry **41:**2379-2387.

43. **Wilson ACaABP.** 1962. Regulation of flavin synthesis by *Escherichia coli*. J Gen Microbiol **28:**283-303.

44. **Havemann GD, Bobik TA.** 2003. Protein content of polyhedral organelles involved in coenzyme B12-dependent degradation of 1,2-propanediol in *Salmonella enterica* serovar Typhimurium LT2. J Bacteriol **185:**5086-5095.

45. **Escalante-Semerena JC, Warren MJ.** 2008. Biosynthesis and Use of Cobalamin ($B_{12}$). *In* Böck A, Curtiss III R, Kaper JB, Karp PD, Neidhardt FC, Nyström T, Slauch JM, Squires CL (ed), *EcoSal - Escherichia coli* and *Salmonella*: cellular and molecular biology. ASM Press, Washington, D. C.

46. **Escalante-Semerena JC, Roth JR.** 1987. Regulation of cobalamin biosynthetic operons in *Salmonella typhimurium*. J Bacteriol **169:**2251-2258.

47. **Jeter RM, Olivera BM, Roth JR.** 1984. *Salmonella typhimurium* synthesizes cobalamin (vitamin $B_{12}$) de novo under anaerobic growth conditions. J Bacteriol **159:**206-213.

48. **Roof DM, Roth JR.** 1988. Ethanolamine utilization in *Salmonella typhimurium*. J Bacteriol **170:**3855-3863.

49. **Stojiljkovic I, Baumler AJ, Heffron F.** 1995. Ethanolamine utilization in *Salmonella typhimurium*: nucleotide sequence, protein expression, and mutational analysis of the *cchA cchB eutE eutJ eutG eutH* gene cluster. J Bacteriol **177:**1357-1366.

50. **Kofoid E, Rappleye C, Stojiljkovic I, Roth J.** 1999. The 17-gene ethanolamine (*eut*) operon of *Salmonella typhimurium* encodes five homologues of carboxysome shell proteins. J Bacteriol **181:**5317-5329.

51. **Kräutler B, Fieber W, Osterman S, Fasching M, Ongania K-H, Gruber K, Kratky C, Mikl C, Siebert A, Diekert G.** 2003. The cofactor of tetrachloroethene reductive dehalogenase of *Dehalospirillum multivorans* is Norpseudo-B12, a new type of natural corrinoid. Helvetica Chimica Acta **86:**3698-3716.

52. **Stupperich E, Kräutler B.** 1988. Pseudo vitamin B12 or 5-hydroxybenzimidazolyl-cobamide are the corrinoids found in methanogenic bacteria. Arch Microbiol **149:**213-217.

53. **Moore TC, Escalante-Semerena JC.** 2016. The EutQ and EutP proteins are novel acetate kinases involved in ethanolamine catabolism: physiological implications for the function of the ethanolamine metabolosome in Salmonella enterica. Mol Microbiol **99:**497-511.

54. **Cheong CG, Escalante-Semerena JC, Rayment I.** 2002. Structural studies of the L-threonine-*O*-3-phosphate decarboxylase (CobD) enzyme from *Salmonella enterica*: the apo, substrate, and product-aldimine complexes. Biochemistry **41:**9079-9089.

55. **Zayas CL, Escalante-Semerena JC.** 2007. Reassessment of the late steps of coenzyme $B_{12}$ synthesis in *Salmonella enterica*: Evidence that dephosphorylation of adenosylcobalamin-5'-phosphate by the CobC phosphatase is the last step of the pathway. J Bacteriol **189:**2210-2218.

56. **Goldberg J.** 1999. Structural and functional analysis of the ARF1-ARFGAP complex reveals a role for coatomer in GTP hydrolysis. Cell **96:**893-902.

57. **Mandiyan V, Andreev J, Schlessinger J, Hubbard SR.** 1999. Crystal structure of the ARF-GAP domain and ankyrin repeats of PYK2-associated protein beta. EMBO J **18:**6890-6898.

58. **Fox AH, Liew C, Holmes M, Kowalski K, Mackay J, Crossley M.** 1999. Transcriptional cofactors of the FOG family interact with GATA proteins by means of multiple zinc fingers. EMBO J **18:**2812-2822.

59. **Morgan B, Sun L, Avitahl N, Andrikopoulos K, Ikeda T, Gonzales E, Wu P, Neben S, Georgopoulos K.** 1997. Aiolos, a lymphoid restricted transcription factor that interacts with Ikaros to regulate lymphocyte differentiation. EMBO J **16:**2004-2013.

60. **Sun L, Liu A, Georgopoulos K.** 1996. Zinc finger-mediated protein interactions modulate Ikaros activity, a molecular control of lymphocyte development. EMBO J **15:**5358-5369.

# CHAPTER 4

# THE *C*-TERMINUS OF COBD FROM *METHANOSARCINA MAZEI* GÖ1 INFLUENCES THE *N*-TERMINAL KINASE AND DECARBOXYLASE ACTIVITIES THROUGH A [4FE-4S] CLUSTER[3]

[3] Tavares N.K., Stracey N., Brunold T.C. and Escalante-Semerena J.C. To be submitted to *Journal of Biological Chemistry.*

**ABSTRACT**

We previously demonstrated that the open reading frame MM2060 from *Methanosarcina mazei* strain Gö1 encodes an unusual version of CobD (*Mm*CobD). *MmcobD* encodes both L-threonine-*O*-3-phosphate (L-Thr-P) decarboxylase (CobD; EC 4.1.1.81), and L-threonine (L-Thr) kinase (PduX, EC 2.7.1.177) activities. In addition to the unexpected L-Thr kinase activity, *Mm*CobD has an extended *C*-terminus annotated as a putative metal binding zinc finger-like domain. Here we demonstrate the *C*-terminal domain of *Mm*CobD is a ferroprotein that contains ~24 non-heme iron per monomer of protein. The *C*-terminus can be removed from *Mm*CobD without complete loss of activity, but its absence alters both activities. Here we show that the L-Thr kinase activity is 90% higher under anoxic conditions with anoxically purified protein. Substitutions of residues in the *C*-terminus demonstrate that while the domain is not required for either activity it does influence the activity. Single residue substitutions of cysteines in the *C*-terminus results in loss of Fe and alterations in the activities of the enzyme. Some substitutions resulted in enhanced activity while most resulted in a decline in activity *in vivo* and *in vitro*. These results point to a possible interaction between the Fe containing *C*-terminus and the active site as a possible regulatory domain. The domain no fully reduced by dithionite and has a UV-Vis spectrum similar to that of a [4Fe-4S] cluster. EPR and MCD analysis were inconclusive, but *Mm*CobD appears to contain one or more polynuclear low-potential redox inactive $[4Fe\text{-}4S]^{2+}$ center(s), which may play a structural role.

**INTRODUCTION**

Cobamide (Cba, *e.g.*, adenosylcobalamin, AdoCbl, vitamin $B_{12}$) biosynthesis in archaea is not well studied. AdoCba biosynthesis has been well studied in bacteria such as *Salmonella enterica* sv Typhimurium LT2. While archaea encode several enzymes that are homologous to those in bacteria, the enzymes that catalyze some steps within the Cba biosynthetic pathway are unknown. While some of the reactions appear to be conserved, several of the enzymes that catalyze these steps are non-orthologous to those in the bacterial pathways (reviewed here (1)). Other apparent homologues contain significantly different features. The L-threonine-*O*-3-phosphate (L-Thr-P) decarboxylase (CobD; EC 4.1.1.81) from the methanogenic archaeum *Methanosarcina mazei* strain Gö1 is one such enzyme (*Mm*CobD). The *N*-

terminus shares ~23% sequence identity with the CobD enzyme from *S. enterica*, but it has an extended *C*-terminus (108 amino acids). This *C*-terminus domain has been annotated as a cysteine rich putative metal binding/zinc finger protein. It appears to be a protein fusion, as homologues of this domain can be found fused to other Cba biosynthesis proteins or as independent open reading frames. The role of this putative metal binding domain is unknown.

We previously showed the CobD from *M. mazei* is a bifunctional enzyme with a new L-threonine (L-Thr) kinase activity in addition to the expected L-Thr-P decarboxylase activity needed to synthesize the 1-amino-*2*-propanol phosphate (AP-P) moiety of cobalamin (2). This former activity is not present in the well-characterized *S. enterica* CobD enzyme (3-5). The L-Thr kinase function is encoded separately in the PduX enzyme in *S. enterica* (6, 7) (Fig. 4.1).

Here we show that *Mm*CobD is a [4Fe-4S] cluster protein. Data presented herein suggests the *C*-terminus contains one or more low-potential $[4Fe\text{-}4S]^{2+}$ cluster(s) that appears to be redox in active. This suggests the role of the cluster may be structural. We confirm that although the [4Fe-4S] cluster contained in the *C*-terminus is not required for activity, perturbations in the domain results in loss of Fe and alterations in the enzyme activity of the *N*-terminus. We hypothesize that the structural role of the *C*-terminus may be as a regulator of the enzymatic function by gating the active site, or facilitating the transfer of the product of the kinase reaction (L-Thr-P) to function as the substrate for the decarboxylation reaction.

**EXPERIMENTAL PROCEDURES**

**Bacterial strains.** Strains and plasmids used in this work are described in Table S4.1. *S. enterica* strains carried a null allele of the *metE* gene that encodes the Cba-independent methionine synthase (MetE) enzyme (8). In the absence of MetE the cell uses the Cba-dependent methionine synthase (MetH) enzyme (9-11). All *S. enterica* strains also carry an undefined mutation, (allele *ara*-9) which prevents the utilization of arabinose as a carbon and energy source. Gene deletions in *S. enterica* were constructed using the phage lambda Red recombinase system as described (12).

**Culture media and growth conditions.** No carbon essential (NCE) (13) with glycerol (22 mM) as the carbon and energy source was used as minimal growth medium. When added to the medium, the following supplements were at the indicated concentrations: trace minerals, 10 mL/liter (14); $MgSO_4$ (1 mM), 5,6-dimethylbenzimidazole (DMB) (0.15 mM), ampicillin (100μg/mL), arabinose (500 μM). All corrinoids (cobyric acid dicyanide [$(CN)_2$Cby], cobinamide dicyanide [$(CN)_2$Cbi], and cobalamin dicyanide [$(CN)_2$Cbl] were added at (1 or 10 nM) final concentrations. When ethanolamine (90 mM) was used as a carbon and energy source Fe-citrate (50 μM) was also added to the medium with corrinoids (300 nM). Cobyric acid was a gift from Paul Renz (Universität-Hohenheim, Stuttgart, Germany). All other chemicals were purchased from Sigma-Aldrich. *S. enterica* strains were cultured in Nutrient Broth (NB, Difco Laboratories) (0.8% w/v) containing NaCl (85 mM). Lysogenic broth (LB) (15, 16) was used as rich medium to culture *Escherichia coli* strains unless otherwise indicated.

**Plasmid construction.** *M. mazei* strain Gö1 gDNA for PCR-gene amplification was a gift from Gerhard Gottschalk (Göttingen, Germany). Genomic DNA from *S. enterica* strain JE7088 (*metE2702*, *ara*-9) were extracted by heating cells at 90°C suspended in double distilled $H_2O$ for 5 min to release DNA. Cell debris was separated from DNA in the supernatant by centrifugation; this was the source of DNA used for the template for PCR amplification. Oligonucleotide primers were purchased from Integrated DNA Technologies Inc. (IDT, Coralville, IA). Primers for cloning were designed using the Saccharomyces Genome Database web based primer design tool available at http://www.yeastgenome.org/cgi-bin/web-primer. Genes were PCR amplified from the appropriate gDNA template with PCR Extender Polymerase (5 Prime) and the primer pairs listed in Table S4.2. PCR products and vectors were treated with restriction endonucleases indicated in the primer name in Table S4.2 and purified with the Wizard SV Gel and PCR Clean-Up kit (Promega). Vectors were treated with Fast AP alkaline phosphatase (Fermentas). PCR fragments and vectors were ligated together using Fastlink Ligase (Fermentas) and introduced into *E. coli* DH5α (17, 18) via electroporation. Plasmid DNA was purified using the Wizard Plus SV Miniprep kit (Promega). Plasmid sequence was confirmed by using BigDye (ABI PRISM) sequencing protocols (University of Georgia Genomics Facility). Primers for site directed mutagenesis were designed

with PrimerX (http://www.bioinformatics.org/primerx/. DNA for site directed mutagenesis was amplified using PfuUltra II Fusion DNA polymerase (Stratagene), and site-directed mutagenesis was performed using the QuikChange protocol from Stratagene. Table S4.1 lists the resulting plasmids. The start codon for wild-type *MmcobD* (ORF MM2060) was changed from GTG to ATG. The *N*-terminus (*Mm*CobD$^{1\text{-}385}$) was cloned from 1-385 with a hard stop codon TAA TAA added after the last residue. The *C*-terminal region (*Mm*CobD$^{386\text{-}497}$) was cloned separately from codons encoding 386-497 with the addition of a methionine as the first residue. pBAD24 vectors were used for complementation (19) and pTEV5 overexpression vectors (20).

**Complementation studies.** Plasmids were introduced into *S. enterica* by electroporation as described elsewhere (21). *S. enterica* strains were grown to full density (~2 × $10^9$ cfu/mL) in NB broth supplemented with Ampicillin (Amp, 100 μg/mL) to maintain plasmids. Strains were grown in triplicate in sterile 96-well tissue culture plates (Falcon) where 2 μL of an overnight cell culture were used to inoculate 198 μL fresh minimal (NCE) medium supplemented with glycerol, $MgSO_4$, and trace minerals. Corrinoids were added as indicated under Culture Media and Growth Conditions section. Ampicillin (100 μg/mL) was used to ensure maintenance of plasmids. Cultures were monitored using Gen5 software (BioTek Instruments) while grown at 37˚C with continuous shaking (19 Hz) in an EL808 Ultra Microplate Reader (BioTek Instruments). Cell density measurements at 630 nm were acquired every 15 or 30 min for 24 or 60 h. Data were analyzed using the Prism v6 software package (GraphPad Software).

**Overproduction of *S. enterica* CobD.** CobD from *S. enterica* was overproduced and purified as described elsewhere (4, 5).

**Overproduction of *M. mazei* CobD.** Aerobic overproduction and purification of *Mm*CobD is described elsewhere (2). Anoxic purification of *Mm*CobD and variants was performed as previously described for the oxic purification with the following exceptions. Cells were broken in an anaerobic chamber by sonication (5 min, continuous pulse with stirring, amplitude 80 m) with a Qsonica Q55 sonicator equipped with microtip. All handling of cell extracts and resulting proteins were conducted anoxically in the anaerobic chamber at 24°C. Purified proteins were desalted by dialysis into HEPES buffer (10 mM,

pH 8.5) and glycerol (50%). Protein samples were aliquoted (250 µL) and sealed in glass anoxic vials and stored at -20°C until used. Proteins were resolved using a 15% SDS-PAGE gel. Protein purity was estimated using band densitometry with a Fotodyne imaging system and Foto/Analyst v.5.00 software (Fotodyne Inc) for image acquisition and TotalLab v.2005 software for analysis (Nonlinear Dynamics). Proteins were purified to the following estimated purities *Mm*CobD$^{WT}$ and variants (85%), *Mm*CobD$^{1\text{-}385}$ (97%), *Mm*CobD$^{386\text{-}497}$ (65%), and *Se*CobD (84%).

**Reconstitution of Fe-S cluster in *M. mazei* CobD.** Oxically purified *Mm*CobD was reconstituted using previously published methods (22).

**Detection of hemoproteins with in-gel heme-dependent peroxides assay.** The detection of heme in purified protein was performed using a modified in-gel double-staining heme peroxidase assay (23, Thomas, 1976 #24730, 24). Purified *Mm*CobD, *Mm*CobD$^{M1\text{-}W385}$ and *Mm*CobD$^{G386\text{-}L497}$, RPA0994, *Se*CobD, horse heart cytochrome *c* (CtyC, positive control), and bovine serum albumen (BSA, negative control) were loaded onto a 12% (w/v) polyacrylamide gel with non-reducing loading buffer consisting of Tris-HCl (200 mM, pH 6.8.), lithium dodecyl sulfate (LDS) (1%), glycerol (10%), EDTA (0.50 mM), bromophenol blue (0.22 mM), Phenol Red (0.175 mM). LDS-PAGE gel was run for 50 min at 200V. The gel was rinsed in water for 5 min before staining in the dark for 2 h in a 0.14% *O*-dianisidine solution in glacial acetic acid. 495 µL of 30% $H_2O_2$ (30 mM final concentration) was added and incubated for 5 min to visualize green colored hemin. The gel was then stained with Coomassie Brilliant Blue R-250 (25) and destained with glacial acetic acid (10%) to visualize all proteins.

***In vitro* L-Thr-P decarboxylase activity assay.** Reaction mixtures contained HEPES buffer (50 mM, pH 8.5); L-Thr-P (5 nmol) and purified protein (0.1 µg). For the metal sensitivity reactions 1 mM of the following metals were added to the reaction; $MgCl_2$, $CoCl_2$, $ZnCl_2$, $NiCl_2$, Fe(III)citrate, and $Fe_2SO_4$.When required for radiolabeled assays, a mixture of [$^{14}$C-U]-L-Thr-P and L-Thr-P in a 1:10 ratio was used as substrate. The final volume of the reaction was 25 µL. Reactions were incubated at 37°C for 1 h. Reactions (5 µL) were spotted onto cellulose thin layer chromatography (TLC) plates for product separation and analysis.

**Thin-layer chromatography (TLC) analysis.** $C^{14}$ radiolabeled L-Thr-P decarboxylation products were detected by TLC on cellulose plates developed with ammonium acetate (2.5 M):ethanol (95%; v/v) (30:70 ratio) mobile phase. Plates were pre-developed with distilled water and allowed to air-dry prior to use. Use of pre-developed plates provided the best resolution. A Typhoon Trio Variable Mode Imager (GE Healthcare) with ImageQuant v5.2 software was used to visualize the results.

**ATPase activity assay.** ATPase activity was assessed using the ADP-glo™ Kinase assay kit (Promega) per manufacturer's instructions. Briefly, the assay uses a proprietary reagent to deplete any remaining ATP in the reaction mixtures, then a secondary reagent converts ADP to ATP which is measured by a luciferase/luciferin reaction which is measured with a SpectraMax Plus Gemini EM microplate spectrophotometer (Molecular Devices) equipped with SoftMax Pro v4 software. Reaction mixture consisting of HEPES (50 mM, pH 6.8 or 8.5), tris(2-carboxyethyl)phosphine (TCEP) (1-2 mM), $MgCl_2$ (1 mM), ATP (0.1-10 mM), L-Thr (0.3-50 mM) and protein (100 ng or 100 nM) were incubated at 37°C for 1 h. Nunc 96-well round bottom black polypropylene microtiter plates (Thermo Fisher) were used to minimize background. Standard curves for ATP were used for quantification. For ATPase inhibition assay the following inhibitors were used ADP (200 mM), AMP (200 mM), sodium pyrophosphate (PPi, 10 mM), sodium triphosphate (PPPi, 10 mM), ADP-γ-S, (100 μM), sodium orthovanadate ($Na_3VO_4$, 1 mM), and sodium beryllium fluoride ($BeF_2$, 2 mM) in HEPES buffer (50 mM, pH 7.5) with *Mm*CobD (6 μM).

**Phosphorous nuclear magnetic resonance (NMR) analysis of L-Thr kinase reaction products.** Proton-decoupled $^{31}P$-NMR spectra were obtained using a Varian Unity Inova 500 MHz spectrometer (Chemical Sciences Magnetic Resonance Facility, University of Georgia) with the following parameters: pulse angle 45°, repetition delay 1 s, excitation pulse 3.88 μs, spectral width 12.11 kHz, acquisition time 0.810 s. 500 μL reaction mixture consisted of HEPES (50 mM, pH 8.5), $MgCl_2$ (1 mM), ATP (3 mM), L-Thr (3 mM) and enzyme (3 μM) incubated at 37°C for 1 h. Protein was removed from reaction mixtures by filtration using Amicon Untracel filters (Millipore) with 10 kDa molecular mass size exclusion. Reaction mixtures (500 μL) were brought up to a final volume of 600 μL in $D_2O$ (17%v/v). Spectra were

processed with MestReNova software version v7.0.0-8333 (Mestrelab Research, Santiago de Compostela, Spain).

**Metal analysis of *Mm*CobD protein.** Inductively coupled plasma optical emission spectroscopy (ICP-OES) was performed on *Mm*CobD samples at the University of Center for Applied Isotope Studies Plasma Chemistry Laboratory (Athens, GA). To remove any non-protein associated contaminating metals from the protein, samples were subjected to buffer exchange by filtration using Amicon Untracel filters (Millipore) with 10 kDa molecular mass size exclusion and washed three times with Chelex 100 resin (Bio-Rad) treated buffer. All other proteins were dialyzed three times in HEPES buffer (50 mM, pH 7.5) treated with Chelex 100. Glassware used were acid washed to remove any residual metal contamination. Inductively coupled plasma mass spectroscopy (ICP-MS) was used to detect Zn. Values represent the concentration of protein or metal in nmol ± standard deviation of the mean of quadruplicate samples.

**X-band electron paramagnetic resonance (EPR) analysis of *Mm*CobD and variants**. Protein (67 μM) samples in HEPES buffer (10 mM, pH 7.5) with 55% glycerol were prepared for EPR under anoxic conditions in an anaerobic chamber and frozen in liquid $N_2$ immediately. EPR spectra were obtained using a Bruker ESP 300E spectrometer fitted with a Vairan EIP model 625A continuous wave counter. Samples were kept at 15 K using an Oxford ESR 900 continuous flow liquid He cryostat that was regulated by an Oxford ITC4 temperature controller. The microwave frequency was measured using a Varian EIP model 625A CW frequency counter. EPR data were fitted and simulate using the Easy Spin (26) software (Stoll and Schweiger). Double integration of simulated data was performed using IGOR version 6.3 (WaveMetrics).

**Magnetic Circular Dichroism (MCD) analysis of *Mm*CobD and variants**. Protein samples were suspending in HEPES buffer (10 mM, pH 7.5) with 55% glycerol. Protein (31 μM) solutions were injected into MCD sample cells under anoxic conditions. When indicated sodium dithionite was used at 7 mM. Room temperature absorption (RT Abs) data were collected under a $N_2$ atmosphere, after which the samples were immediately frozen and stored in liquid $N_2$. MCD and low-temperature (LT) absorption spectra were collected using a Jasco J-715 spectropolarimeter with an Oxford instruments SM4000-8T

agnetocryostat. The MCD spectra were obtained via subtraction of the spectra obtained with the magnetic field oriented antiparallel and parallel to light-axis of propagation, to reduce effects from the natural CD and glass strain.

**RESULTS**

**CobD from *M. mazei* does not incorporate Zn.** The CobD protein from *M. mazei* is unusual relative to other CobD homologues because the *C*-terminus is a cysteine rich domain annotated as a putative metal-binding domain or 'zinc finger' protein. This *C*-terminal domain appears to be a protein fusion, as homologues of the 'zinc finger' are found alone as open reading frames. Genes encoding these proteins are often chromosomally adjacent to other cobalamin biosynthetic genes and are therefore assumed to have some function in AdoCba biosynthesis. The function is currently unknown. When purified the full-length and *C*-terminal only truncation (*Mm*CobD$^{G386\text{-}L497}$) proteins are brown in color. Which lead us to believe the protein binds iron rather than zinc. Figure 4.2 shows a SDS-PAGE gel of wild-type *M. mazei* CobD purified from *E. coli* C43 (27) cells grown in either LB rich medium (lane 4) or NCE minimal glycerol (22 mM) medium supplemented with pyridoxine (1 mM) to stimulate PLP production. Fe(III)citrate (50 mM) (lane 2) or $ZnSO_4$ (50 mM) (lane 3) was added to the medium to encourage incorporation of either metal into the protein. All cultures grew to full density (~$2x10^9$ cfu/mL). The addition of $ZnSO_4$ resulted in no protein production. Expression of *Mm*CobD in minimal medium without any metal supplementation results in two bands (lane 1). Both band were verified to be *Mm*CobD by mass spectrometry peptide finger printing. The quantity of the lower molecular weight band is reduced with overexpression in rich medium or (lane 4) or with Fe supplementation (lane 2). This suggests the lower molecular weight band is *Mm*CobD that was unable to incorporate a full complete complement of Fe. Co-purified with *Mm*CobD was D-tagatose-1,6-bisphosphate aldolase (GatY), and the reason for this unclear. GatY is a metalloenzyme that requires a divalent metal for function (28, 29). It is possible that the addition of metals to the medium stimulated GatY production.

***M. mazei* CobD is not a heme protein.** The presence of the CXXCH motif, which is a typical heme-binding motif, led us to investigate the possibility that *Mm*CobD might be a hemeprotein. We performed

an in-gel double-staining heme-dependent peroxidase assay for the identification of heme (23, Thomas, 1976 #24730, 30) (see *materials and methods*). Supplemental Figure S4.1 shows a lithium dodecyl sulfate (LDS)-PAGE non-reducing gel that has been treated and stained to reveal the presence of heme. Heme is converted to hemin and stains green when acidified and exposed to peroxide. Only the positive control horse heart cytochrome *c* (CytC, lane 5) was stained on the gel. Panel B shows a Coomassie Brilliant Blue (25) stained gel to reveal all proteins. These results demonstrate that heme was not present in *Mm*CobD$^{WT}$ or RPA0994, a representative 'zinc finger' homologue protein from *Rhodopseudomonas palustris,* that exists as an independent open reading frame associated with AdoCbl genes.

***M. mazei* CobD is more active under anoxic conditions.** We previously showed that CobD from *M. mazei* has L-Thr kinase activity when purified and assayed under oxic conditions (2). The oxically purified enzyme exhibited no change in activity when assayed anoxically but did display a very slight increase in activity in the presence of the reducing agent sodium dithionite. The possibility of the presences of a potentially redox sensitive metal such as Fe caused us to rethink our purification and assay conditions. We purified *M. mazei* CobD anoxically as described in experimental procedures and assayed the ATPase activity anoxically comparing it to protein purified under oxic conditions, and anoxically purified protein after 15 min of exposure to air (Fig. 4.3). The anoxically purified protein had ~90% higher activity, and there was a loss 75% of the enzymatic activity following exposure to air. These results demonstrate that while the ATPase activity is not abolished under oxic conditions, *Mm*CobD is much more active anoxically.

**Disruptions in the *C*-terminal metal binding domain results in the production of a lower molecular weight protein.** We previously demonstrated that the putative metal binding *C*-terminus of *Mm*CobD was not required *in vivo* or *in vitro* for L-Thr kinase activity, but removal of this domain resulted in altered activity (2), suggesting an influence of this domain on L-Thr kinase activity. In an effort to uncover the role of the putative metal-binding domain several conserved residues were mutated to alanine, with particular focus on the 9 cysteines and 5 histidines that might coordinate metal ions, or Fe-S clusters. The locations of these amino acid substitutions are marked on the sequence alignment of the *C*-terminal

domain of *Mm*CobD along with other annotated 'zinc finger' proteins from other cobamide producers (Supplemental Figure S4.3). These protein variants were purified under anoxic condition along with the wild-type and *C*-terminus proteins. Figure 4.4A shows a SDS-PAGE gel of the wild-type, truncated proteins and select variants. When cells are grown in medium with insufficient Fe a second smaller molecular weight version of *Mm*CobD co-purifies with the full-length protein (lane 1), but not with *N*- (lane 2) or *C*-terminal (lane 3) proteins. Several variants of *Mm*CobD also displayed this second band, and for some the majority of the protein was shifted to this lower molecular weight band (lane 6). Both bands were verified to be *Mm*CobD protein via mass spectroscopy peptide finger printing. We speculate that the lower molecular weight band is a species of *Mm*CobD that has less Fe due to insufficient Fe in the medium, or in the case of the variant *Mm*CobD$^{C458A}$ (lane 6), the substitution of a cysteine residue for an alanine has sufficiently disrupted this protein's ability to bind Fe molecules. The resulting shift in molecular weight may be due to 1) the absence of the iron molecules themselves, disruption of Fe-S centers, misfolding, or proteolysis. Proteolysis is unlikely as mass spectroscopy in-gel Trypsin digest results gives coverage of most of the *N*- and *C*-termini (data not shown). We know these shifts in molecular weight are not due to poor incorporation of PLP since variants unable to bind PLP, (*Mm*CobD$^{K234A}$, lane 7) behave identical the wild-type protein.

**The *C*-terminus of *M. mazei* CobD is not required for L-Thr kinase activity but influences it.** We previously showed that removal of the *C*-terminus does not greatly affect the ability of the *Mm*CobD protein to complement a *S. enterica cobD* or *pduX* strain when grown under conditions that require low levels of AdoCbl synthesis. We tested the growth behavior of a *S. enterica pduX* strains carrying plasmids expressing single-residue substitutions of *Mm*CobD variants. Variants H396A, Y414A, Y414S/Y415A, C417C H418A, C427A, C429A, C434A, and S454A had no effect on growth (data not shown). Variants H476A, S446A (Supplemental Fig. S4.4C), C458A, and C424A (Fig. 4.4B) all had impaired growth in a *S. enterica pduX* strain on glycerol supplemented with Cby. Variant H459A had the surprising effect of improving the growth (3 h) of a *pduX* strain to the level of wild type or the *pduX*/p*Se*CobD positive control (3.7 and 3.4 h, respectively) (Supplemental Fig. S4.4C). This suggests the H459A substitution

resulted in a more effective enzyme with respect to the L-Thr kinase activity. We next tested the ATPase activity of these variant. The variants with growth defects *in vivo* also had decreased ATPase activity *in vitro*. Variant C424A had the most severe growth defect similar to that of the *pduX* empty vector only control strain (Fig. 4.4C). For simplification we will focus on variants C424A, C434A, and C458A throughout this paper.

**Substitutions within the *C*-terminal, putative metal-binding domain of *M. mazei* CobD affects the decarboxylation activity associated with the *N*-terminal domain of the protein.** Plasmids expressing single residue substitutions variants were expressed in a *S. enterica cobD* strain and assayed for their ability to restore AdoCbl synthesis when the strain was grown in minimal medium supplemented with Cby. The following variants had no effect under any conditions tested in a *S. enterica cobD* strain: H396A, Y414A, C417A, H418A, C427A, S454A, and H459A (data not shown). Substitutions Y414S/Y415A, C412A, (data not shown) and S446A had a very slight growth defects relative to wild type (Supplemental Fig. S4.3C). We will focus on the substitutions with the most severe effects relative to wild type, H476A, C424A, C434A, C429A, and H476A. Although all variant proteins were overproduced and visualized on a SDS-PAGE gel (Figure 4.4A), we cannot rule out the possibly of misfolded proteins. *S. enterica cobD* strains synthesizing *Mm*CobD$^{C429A}$ and *Mm*CobD$^{C434A}$ variants displayed growth defects similar to the those observed when the empty cloning vector was used as a negative control in cells grown on glycerol in the presence of a low concentration of Cby (1 nM, Fig. 4.4B). These growth defects were present when the cells were grown on a higher concentration of Cby (10 nM), but the effects were less severe. However, on 10 nM Cby the strain synthesizing the *Mm*CobD$^{C429A}$ variant eventually reached full density while the strain synthesizing the *Mm*CobD$^{C434A}$ variant grew poorly (data not shown). Unexpectedly, the effect of variant the *Mm*CobD$^{C458A}$ was to enhance growth relative to *Mm*CobD$^{WT}$ on low or high Cby. The *cobD* strain expressing the *Mm*CobD$^{C458A}$ variant maintained a growth rate (3.4 h) faster than that of the wild-type *Mm*CobD strain (9.2 h), and similar to that of the *S. enterica* wild-type (3.7 h) and the *Se*CobD complemented strain (2.4 h). These results suggest that *Mm*CobD$^{C458A}$ variant is a more efficient enzyme than wild type. This

behavior is similar to the effect observed with the substitution of the adjacent residue H459A, with the exception that the substitution resulted in improved growth in a *pduX* strain background (Fig. 4.4B) whereas the same mutation had no effect in a *cobD* strain. *In vitro* assessment of the activity of the variant proteins showed the L-Thr-P decarboxylase activity was not completely abolished, which is in agreement with the *in vivo* data. However, while the assay is very sensitive we are unable to easily quantify the subtle difference in activity of the variant proteins. The *in vivo* results suggests that while the *C*-terminal domain is not required for restoring AdoCbl synthesis in a *S. enterica cobD* strain, substitutions in this domain impact the L-Thr-P decarboxylase activity of the *N*-terminal domain. This suggests the possibility that a direct interaction between the *C*- and *N*-terminal domains is required for optimal activity of the enzyme. The phenotypic effect each substitution had on a *pduX* or *cobD* phenotype is summarized in Supplemental Figure S4.2.

***M. mazei* CobD contains iron.** The data hints at the likelihood that *Mm*CobD binds iron. We suspect the form of the iron is that of Fe-S center(s). We quantified the iron content of wild-type protein purified oxically, anoxically, and oxically purified protein incubated with $Fe(II)SO_4$ and $Na_2S$ to reconstitute any Fe-S clusters that may be present (see experimental procedures). Truncated and single residue substitution variant proteins we also analyzed. The results are summarized in Table 4.1. No Fe was detected in the *N*-terminal domain (*Mm*CobD$^{1\text{-}385}$). The Zn concentration was below the detectable level for ICP-OES, so ICP-MS was employed. Wild-type *Mm*CobD does not bind Zn at significant levels. Oxically purified wild-type protein binds between 2 and 5 Fe molecules per monomer of protein, while anoxically purified or anoxically reconstituted Fe-S samples bind an average of ~24 Fe per protein monomer. This was a unexpectedly high number of Fe for a single protein monomer. We have previously shown that *Mm*CobD is a dimer. There is the possibility that some of the iron atoms or iron sulfur clusters are shared between the two dimers. The *C*-terminus by itself, which we have previously shown to be a monomer in solution, only binds two Fe/monomer even when incubate anoxically with $Fe(II)SO_4$ and $Na_2S$ to reconstitute any potential Fe-S clusters. This suggest interactions between the *N*- and *C*-terminus or, the dimerization of *Mm*CobD is required to fully bind all the Fe molecules, or for the formation or stabilization of Fe-S

clusters in the protein. The variant proteins had lower levels of Fe to varying degrees. Interestingly, the *Mm*CobD$^{C458A}$ variant had the lowest level of Fe at 3 atoms per monomer. This variant when ectopically expressed in as *S. enterica cobD* strain significantly improved the growth to a level great than that of a *cobD* strain expressing *Mm*CobD$^{WT}$.

***M. mazei* CobD is somewhat redox sensitive.** After establishing *Mm*CobD sensitivity to oxygen (Fig. 4.3) we purified the protein and variants under anoxic conditions. The brown color (Figure 4.6 inset), improved purification with Fe(III)citrate supplementation (Fig. 4.2), and detection of high Fe to protein ratios (Table 4.1) led us to suspect the presence of Fe-S clusters in the protein. We analyzed the UV-Vis spectrum of the wild-type, truncated, and variant proteins (31 μM). The wild-type protein had a dominant $^{\lambda}$max at 420 nm which was unaffected by reduction with dithionite (2 mM), however a new $^{\lambda}$max appears at 312 nm after the addition of dithionite. This 312 nm $^{\lambda}$max is present in the spectra of all the variants and the *C*-terminal truncation (*Mm*CobD$^{G386-L497}$), but is absent in the *N*-terminal truncation (*Mm*CobDG$^{M1-W385}$). This suggests that the $^{\lambda}$max at ~312 is generated by redox sensitive features only in the *C*-terminus. The *Mm*CobD$^{M1-W385}$ protein has two prominent $^{\lambda}$max at 338 nm and 428 nm that are indicative of PLP (31-33). The bright yellow color of the purified protein (Fig. 4.6B, inset) confirms the presence of PLP and the absence of iron, which gives wild-type and *Mm*CobD$^{G386-L497}$ proteins their dark brown color. PLP is not reduced by dithionite and thus the spectra for *Mm*CobD$^{M1-W385}$ protein does not change with the addition of dithionite. However, the PLP derived $^{\lambda}$max at 428 nm in the *N*-terminus appears to be masking the $^{\lambda}$max at 407 nm that is derived from some Fe species in the *C*-terminus. This is apparent when comparing the spectra of the wild-type protein to the *C*-terminal domain and the *Mm*CobD$^{K234A}$ variant that is unable to bind PLP due to the alanine substitution of the catalytic lysine. There is a shift of the 420 nm $^{\lambda}$max to 407 nm in these two protein variants that lack PLP but retain Fe. For this reason, the presumably Fe derived feature at $^{\lambda}$max 407 nm in the *Mm*CobD$^{C434A}$ and *Mm*CobD$^{C458A}$ variants is not apparent due to the presence of PLP in those proteins. The much smaller intensity of the $^{\lambda}$max at 425 nm in these two variants is likely as a result of the reduced iron content due to the effect of the alanine substitutions. To try to identify the species of Fe chromophore that is present in

the protein, the spectra of the *Mm*CobD$^{K234A}$ variant is optimal because of the lack of PLP interference. This spectrum does not resemble the spectra of rubredoxin type $F^{3+}$ [1Fe-0S] clusters that typically have $^{\lambda}$max at 370 and 492 nm (34, 35). Nor does it resemble the spectra of ferredoxin type [2Fe-2S] clusters proteins, which have $^{\lambda}$max at 330, 422, and 464 nm (36), or other [2Fe-2S] clusters containing protein such as SoxR which have $^{\lambda}$max at 414, 448, and 548nm (37), or others (38-40). The spectra has a much higher resemblance to [4Fe-4S] cluster proteins such as IscA which have $^{\lambda}$max at 315 and 415 nm (41).

***M. mazei* CobD contain diamagnetic poly- or mononuclear Fe(II) centers.** Figure 4.7A shows the EPR spectrum of wild-type *Mm*CobD at 10 K. There are two EPR signals, one characterized by a $g_{xyz}$ values of g = 2.00, 2.00, 2.02 respectively, and a spin of 1/2, and second signal with a $g_{xyz}$ values value of 1.98, 2.00, 2.03, and a spin of 5/2. Spin quantitation vs. Cu(imidazole) standard was carried and double integration and normalization of the signals. The final quantitation indicated that the concentration of half integer spin (S=1/2 and S=5/2) Fe centers accounts for ~0.14% and ~1.26% per monomer of protein respectively. It was therefore concluded that there was only a small amount of the half integer species that might be impurities. This indicated that the Fe centers are mostly integer spin, either paramagnetic or diamagnetic. The EPR data rule out the presence of mononuclear Fe(III) centers, as these would exhibit stronger S=1/2 or S=5/2 signals. The data are consistent with mononuclear Fe(II) centers or any polynuclear iron clusters possessing an integer-spin ground state.

***Mm*CobD likely contains one or more [4Fe-4S]$^{2+}$.** MCD of *Mm*CobD$^{WT}$ revealed a broad diamagnetic MCD features with no temperature dependence and a feature at ~425 nm in CD and MCD corresponds to absorption feature seen in the room temperature UV-Vis spectra. The MCD data rule out the presence of mononuclear Fe(III) centers, as these would exhibit temperature-dependent signals in the visible/near-UV spectral region. The data are consistent with mononuclear Fe(II) centers possessing a coordination number of 4 or any polynuclear iron clusters possessing a diamagnetic, S=0 ground state. These data agree with the EPR data obtained. With the addition of dithionite (7 mM) no temperature-dependent features were present. The Fe-S clusters may be redox inactive. For the *Mm*CobD$^{C458A}$ variant all features were broad and temperature independent as with the wild-type protein. For the *Mm*CobD$^{C458A}$ variant No

temperature-dependent features were observed, only a protein signal at 280 nm. The *Mm*CobD$^{C458A}$ sample does not appear to have the same diamagnetic Fe-S cluster features. The feature at 420 nm in the room temperature absorption data could be solely from the PLP band. However, the broad diamagnetic MCD signals may not be intense enough to be visualized due to low protein sample concentration of the sample. The 420 nm absorption band is characteristic of $[4Fe4S]^{2+}$ cluster however, the band is equally intense in the spectrum of dithionite-treated *Mm*CobD$^{WT}$ and the variants containing fewer Fe atoms. While it is possible that the clusters are not reducible by dithionite, the fact that the intensity of the 420 nm absorption band remains unchanged in response to amino acid substitutions resulting in a decrease in Fe loading suggests that this band is due to the PLP cofactor.

**The *Mm*CobD decarboxylase activity is sensitive to divalent metals.** Both inhibition and enhancement of activity by metals of PLP-dependent enzymes or free PLP in solution is known to occur (42-46). This is the first demonstration of inhibition of CobD decarboxylase activity by metal ions. The inhibition or enhancement effect of metals on decarboxylation or transamination reactions varies with the enzyme and or substrate. In general, the metal ion inhibits the reaction by preventing the formation of Schiff Base between PLP and the substrate. In the case of *Se*CobD, Co(II) and Zn(II) inhibited the decarboxylation reaction (Supplemental figure S4.4). *Mm*CobD was inhibited by Zn(II) and Ni(II) and partially inhibited by Co(II). Cobalt is a known to inhibit some zinc finger proteins by the formation of Schiff bases complexes with the zinc coordinating histidine (47, 48). Both enzymes were inhibited by high concentrations of Mg(II) (10 mM). We are unable to access any enhancing effects by the metals on L-Thr-P decarboxylation activity. It is not clear why *Se*CobD and *Mm*CobD would be affected differently by different metal ion and if, or how, the presence or absence of the metal binding domain might affect the inhibition by metal ion, or what effect this inhibition might have on *Mm*CobD L-Thr kinase activity. We previously demonstrated that *Mm*CobD was unable to use Zn, Ni, Ca, Al, or Cu as divalent metals to facilitate the hydrolysis of ATP but was able to use Mg, Co, and Mn, however competitive inhibition by these metals was not assessed.

**DISCUSSION**

We previously uncovered a new enzymatic reaction performed by CobD from *M. mazei* (2). Here we further characterized the *C*-terminal domain and expand on the role it plays in the two enzymatic reactions localized to the *N*-terminus of *Mm*CobD.

*E. coli* is known to readily incorporate Zn(II) in place of Fe when overexpressing metalloproteins (49). The inability of *Mm*CobD$^{WT}$ to incorporate Zn and the tendency to generate a second low molecular weight protein species when limited for Fe (Fig. 4.2) demonstrates *Mm*CobD is not a 'zinc finger' protein as currently annotated but binds Fe instead. The size difference between the two bands might be the result of misfolded protein, or partial proteolysis due to failure to properly incorporate several Fe atoms need for structural stability, or to form one or more [4Fe-4S] cluster(s). There are 9 cysteine residues in the *C*-terminus of *Mm*CobD that could contribute to several Fe-S clusters of various species and combinations. There is also the possibility that the mass difference is a result of the loss of two [4Fe-4S] clusters (0.7 kDa), which could account for the mass difference. The low resolution of an SDS-PAGE gel makes quantifying such small differences in protein mass difficult.

The *Mm*CobD$^{K234A}$ variant, which lacks the ability to bind PLP, has 19 Fe/monomer with a very good standard deviation (SD) of 0.9. The reconstituted or anoxically purified wild-type proteins have an average of ~24 Fe/monomer, with rather poor SD of 3-7. The higher Fe content and poor SD may be as a result of insufficient buffer exchange to remove any unbound iron in these samples. No Fe is detectable in the *N*-terminus only protein sample. From this we can conclude that all the Fe is bound to the *C*-terminus only. The Fe quantity in the *Mm*CobD$^{K234A}$ variant of 19 ± 0.9 may be reflective of the true Fe content of the wild-type protein or there is the possibility that the absence of the PLP cofactor prevents the binding of 4 or more additional Fe molecules or one [4Fe-4S] cluster. This idea seems farfetched at first, but there are precedents, which will be discussed below.

We have shown the *C*-terminus of *Mm*CobD excretes influence on both the L-Thr-P decarboxylase and L-Thr kinase activities. There may be allosteric regulation of one domain by the other that is affected by the

substrates/products of the active sites/domains, similar to that observed in tryptophan synthase (50). Mutational analysis of putative Fe binding cysteines, which may be involved in forming the Fe-S clusters, indicate that the interplay between the two domains and activities are directly related to the presence of the Fe-S cluster containing *C*-terminus (Fig. 4.4, 4.5 & S4.3). The eukaryotic cystathionine β-synthase (CBS) contain both PLP and heme cofactors (51, 52). The PLP and heme sites of this enzyme communicate allosterically via a single α-helix (53). Mutational analysis of CBS resulted in the isolation of variants that are inhibited by or hyperactive (53, 54). This is similar to what we observed with the *Mm*CobD$^{C429A}$ variant, which was a less efficient decarboxylase. *Mm*CobD$^{C434A}$ had less efficient decarboxylase and kinase activities, while both activities were enhanced by the *Mm*CobD$^{C458A}$ variant (Fig. 4.4 & 4.5). The enhancement effect of *Mm*CobD$^{C458A}$ on both activities is particularly interesting. This result suggests the *Mm*CobD$^{C458A}$ protein was either a more efficient or more stable enzyme. However the effect on a *S. enterica pduX* strain suggested that the increased activity could be due to a higher affinity for L-Thr, or the enzyme might be more efficient at generating L-Thr-P, which the *N*-terminal decarboxylase domain can then convert to AP-P.

The 90% higher ATPase activity seen with anoxically purified and assayed *Mm*CobD is further evidence for a direct role by the [Fe-S] cluster containing *C*-terminus in affecting the enzymatic activities of the *N*-terminus. We know the role is not directly related to redox chemistry involvement in the decarboxylase or kinase reactions because the *Mm*CobD remains active when the *C*-terminus is removed. We believe the role is structural. Some disruptions to the *C*-terminus and presumably the [Fe-S] centers were more detrimental to the enzyme function than completely removing the *C*-terminus. Surprisingly, some disruptions introduced by single amino acid substitutions resulted in better than wild-type activity for both the decarboxylase activity (*Mm*CobD$^{H459A}$ and *Mm*CobD$^{C458A}$) and the kinase activities (*Mm*CobD$^{H459A}$) (Fig. 4.4, 4.5, & S4.3). Even more confounding is that *Mm*CobD$^{C458A}$ had a severe growth defect in a *pduX* strain despite having better than wild-type growth in a *cobD* strain. The contradictory effect of these two substitutions, C458 and H459, singles them out as particularly important

for modulating the kinase and decarboxylase activities. Of particular interest is the finding that the *Mm*CobD$^{C458A}$ variant, which had only 3 Fe molecules per monomer of protein, the lowest of the variants tested. Might this suggest that the presence of the Fe-S cluster(s) function as a kind of regulator of kinase and decarboxylase activities which slows them down? And when it is disrupted in a particular way the enzyme functions more efficiently? It does suggest interplay between the PLP containing *N*-terminus and the [4Fe-4S] cluster *C*-terminus that is reminiscent of other multi-domain, multi-cofactor containing enzymes such as cystathionine β-synthase (51, 52). We have seen that overexpression of PduX homologues results in a 10 h decrease in lag time and better than wild-type growth on ethanolamine (2). The reason for this dramatic growth improvement is unclear, but it does indicate that high levels of L-Thr kinase activity had a strong physiological effect on the cell outside of just AdoCba synthesis. We speculated that overexpression of an ATP consuming enzyme could be deleterious to the cellular energy balance, though this seems not to be the case with PduX. However, this may be the case with *Mm*CobD, or the very thing the [Fe-S] cluster domain is meant to modulate. What might this tell us about the physiological state of *M. mazei*? Perhaps this organism needs to tightly control the level of free L-Thr, or ATP consumption, or the biosynthesis of Cba within the cell? We may not be able to answer these questions directly in a heterologous system and will need to work in *M. mazei* to obtain the answer to these questions.

There is precedent for at least one iron sulfur cluster protein with kinase activity with the eukaryotic enzyme Asp1. Asp1 is a bifunctional phosphatase/kinase from *Schizosaccharomyces pombe* that regulates morphogenesis and has a [2Fe-2S] center. Wang *et. al* recently reported the [2Fe-2S] cluster in Asp1 inhibits the phosphatase activity and in turn increases kinase activity of this enzyme (55). Likewise, this modulation of two enzymatic activities by and Fe-S cluster is reminiscent of what has been observed with *Mm*CobD.

These data also allowed us to narrow down the regions and identify motifs of the *C*-terminus that are of particular importance for affecting the activities of the *N*-terminus. The effect of each substitution on

each enzymatic activity is shown in Supplemental Figure S4.2. *Mm*CobD has nine cysteine and five histadine residues that might coordinate Fe molecules or from [4Fe-4S] clusters, arranged in 4 cluster but only 3 of the 4 ($C^{424}X_2CXCX_4C^{434}$, $C^{455}X_2CHX_2H^{459}$, $C^{473}X_2H^{476}$) appear to be critical for influencing activity based on phenotypic analysis. The substitutions in the cluster $C^{412}X_4CH^{418}$ have little or no effect on the activity of *Mm*CobD *in vivo*. These three clusters bear a strong resemblance to a motif that have been proposed as low-potential [4Fe-4S] cluster binding motifs (56). Particularly striking is the two conserved prolines adjacent to cysteine residues and the conspicuous conserved glycine between two of the cysteine patches (Fig. S4.2). MCD data indicate that the clusters(s) in *Mm*CobD is a low-potential $[4Fe\text{-}4S]^{2+}$ cluster.

MCD data revels that *Mm*CobD redox inactive. However, the 407 nm $^{\lambda}$max in the UV-Vis spectra of the $MmCobD^{K234A}$ PLP free variant, and $MmCobD^{386\text{-}497}$ *C*-terminus protein appear to respond to dithionite reduction. Upon closer inspection one can see the apparent reduction is not the dramatic decrease of the $^{\lambda}$max that should occur with dithionite reduction. The MCD data are more sensitive and we can conclude that the [Fe-4S] cluster in *Mm*CobD is redox inactive, for which there are precedents (57-60). The slight or partial reduction might also hint at the possibility of more than one low-potential [Fe-S] cluster with different oxidation states ($^{2+}$, $^{1+}$, or $^{0}$). The EPR and MCD data were inconclusive. The signal from the PLP cofactor may be interfering with the [Fe-S] signal. However, the data are consistent with *Mm*CobD containing one or several low-potential $[4Fe\text{-}4S]^{2+}$ cluster(s) that are not reducible by dithionite. The roles played by the cluster(s) are unknown, but [4Fe-4S] clusters fulfilling purely a structural role have been identified in other proteins (61-64).

Although at this time the role of the [4Fe-4S] cluster in *Mm*CobD appears to be structural, we are exploring the possibility of a redox role elsewhere in the Cba biosynthetic pathway. The central cobalt ion in Cbas have to be reduced stepwise from Co(III) to Co(II) to Co(I) for the adenosyl group to be attached as the b-axial ligand. The first reduction can be carried out by free dihydroflavins (65-67) but the redox potential to go from Co(II) to Co(I) is too low (-610 mV) for known biological reductant (68, 69). This requires an ATP:co(I)rrinoid *adenosyltransferase*, (CobA, PduO, or EutT; EC 2.5.1.17) to raise the

reduction potential (70-73) and accept an electron from reduced flavodoxin A (FldA) to generate Co(I) (74-76). At which point CobA transfers the adenosyl group from ATP to Co(I) in the corrinoid (73, 77). This has been well studied in *S. enterica* however it is not known if other organisms use FldA for corrinoid reduction. Studies with CobA from *M. mazei* and FldA from *S. enterica* demonstrated poor coupling of the enzymes (78). Leaving the question open of how does *M. mazei* reduce Co for adenosylation to occur? There is the possibility of a dedicated corrinoid reductase. We are testing the possibility that the [4Fe-4S] cluster *C*-terminus of domain *Mm*CobD and the stand-alone homologues may function as a corrinoid reductase.

**ACKNOWLEDGMENTS**

This work was supported by NIH grant R37 GM40313 to J.C.E.-S. N.K.T. was supported in part by NIH grant F31 GM095230 and Advanced Opportunity Fellowship awarded by the Graduate School of the University of Wisconsin, Madison. We thank P. Renz for his gift of CNCby, and G. Gottschalk for his gift of *M. mazei* genomic DNA. Thanks also to A. Enriquez for photographing samples.

**FIGURES**

**Table 4.1. ICP-OES metal analysis of *Mm*CobD and variants.** Proteins were prepared aerobically or anaerobically as indicated. Where indicated Fe-S clusters were reconstituted anaerobically and non-protein associated contaminating metals removed as described in experimental procedures. ICP-MS was used to detect Zn. Values represent the concentration of protein or metal in nmol ± standard deviation of the mean of quadruplicate samples. Column 4 is the ratio of Fe molecules per monomer of protein. ND; none detected.

| Enzyme | [Protein] (nmol) | [Fe] (nmol) | Fe:Protein | [Zn] (µmol) | Zn:Protein |
|---|---|---|---|---|---|
| ***Mm*CobD$^{WT}$ (oxic)** | 2.7 ± 0.2 | 7 ± 0.6 | 2.6 ± 0.47 | 0.12 ± 0.01 | 0.04 ± 0.01 |
| ***Mm*CobD$^{WT}$ (oxic)** | 1.5 ± 0.6 | 2 ± 0.3 | 4.5 ± 0.35 | 0.04 ± 0.01 | 0.04 ± 0.01 |
| | | | | | |
| ***Mm*CobD$^{WT}$ (reconstituted)** | 0.81 ± 0.13 | 18.5 ± 1.5 | 23 ± 3 | | |
| ***Mm*CobD$^{WT}$ (reconstituted)** | 1.5 ± 0.6 | 35 ± 11 | 25 ± 7 | | |
| ***Mm*CobD$^{WT}$ (anoxic)** | 0.92 ± 0.11 | 24.5 ± 4.1 | 27 ± 4 | | |
| | | | | | |
| ***Mm*CobD$^{1-385}$ (oxic)** | 2.0 ± 0.1 | ND | - | | |
| **C458A (anoxic)** | 1.9 ± 0.3 | 5.0 ± 0.7 | 3 ± 0.04 | | |
| **C424A (anoxic)** | 1.4 ± 0.09 | 18.3 ± 1.2 | 13 ± 1.4 | | |
| **K234A (anoxic)** | 2.1 ± 0.2 | 39.3 ± 2.0 | 19 ± 0.9 | | |
| **S446A (anoxic)** | 1.8 ± 1.1 | 23.7 ± 6.2 | 16 ± 4.4 | | |
| | | | | | |
| ***Mm*CobD$^{386-497}$ (reconstituted)** | 2.3 ± 0.8 | 4.2 ± 1.5 | 2 ± 0.7 | | |

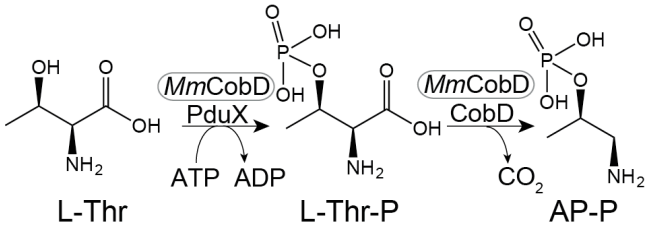



**Figure 4.1. AP-P synthesis. Schematic of the reaction performed by *Mm*CobD.** In *S. enterica* and other bacterial Cba producer this reaction is performed by two enzymes, PduX and CobD.

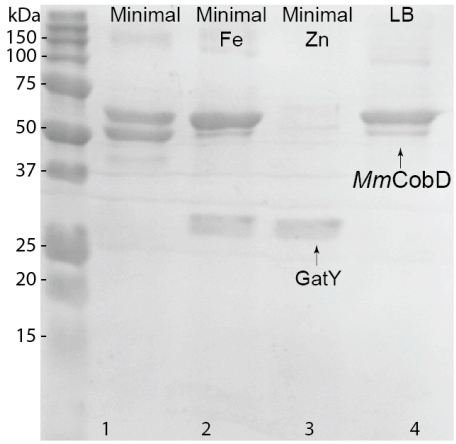



**Figure 4.2**. **The metal binding domain of *Mm*CobD is not a zinc finger.** SDS-PAGE gel of *Mm*CobD anoxically purified from cell grown in (lane 1) NCE minimal glycerol (22mM) medium supplemented with pyridoxine (1 mM); (lane 2) minimal medium supplemented with pyridoxine (1 mM) and Fe(III)-citrate (50 mM); (lane 3) minimal medium supplemented with pyridoxine (1 mM) and $ZnSO_4$ (50 mM) and; (lane 4) lysogenic broth (LB) rich medium.

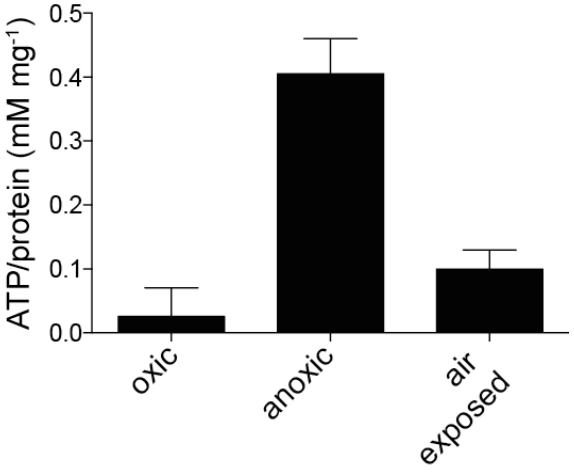



**Figure 4.3. Anoxically purified *Mm*CobD is a more active enzyme.** ATPase activity assay measured with ADP-glo Kinase Assay Kit (Promega). Reaction mixture contained HEPES buffer (50 mM, pH 8.5), $MgCl_2$ (1 mM), ATP (100 μM), L-Thr (300 μM), and purified protein (3 μM) incubated at 25°C for 1 h. Proteins were purified and incubated oxically or anoxically as indicated. No enzyme controls were subtracted to reduce background. Values are compared to a standard curve and converted mM ATP per μM protein, with the standard error mean of triplicate reactions represented by the error bars.

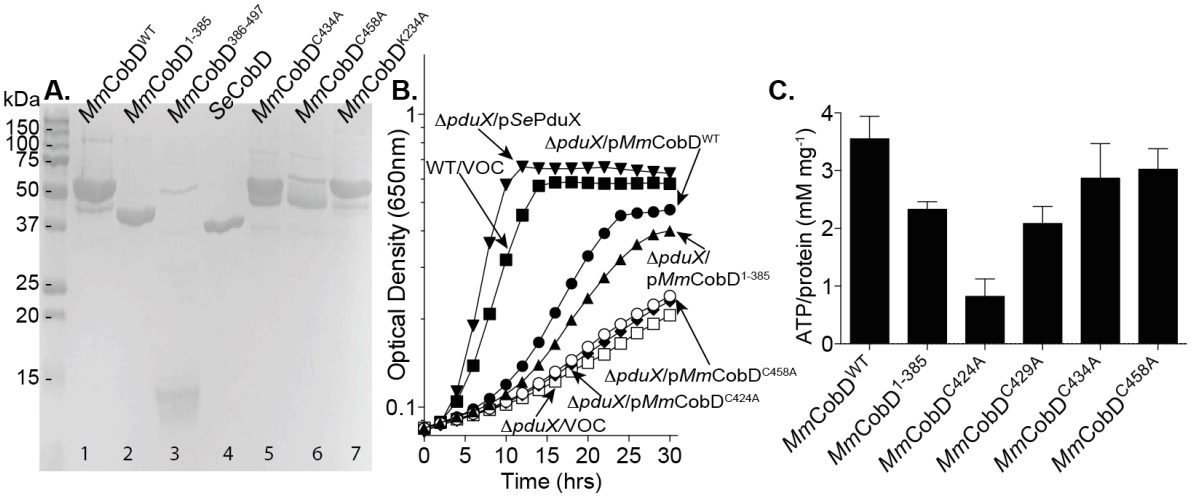



**Figure 4.4. Effect of mutations in the metal binding domain on the L-Thr kinase activity of *Mm*CobD. A.** SDS-PAGE gel of anaerobically purified *Mm*CobD (lane1), truncations (lanes 2, 3) and variants (lanes 5, 6, 7). *Se*CobD and *Mm*CobD$^{1\text{-}385}$ were purified aerobically. Only three representative variant proteins are included on the gel due to space restrictions. **B**. Cobalamin-dependent growth of *S. enterica pduX* strains with plasmids expressing wild-type or variant *Mm*CobD proteins. Cells were grown aerobically at 37ºC in NCE minimal medium with glycerol (22 mM) as the sole carbon and energy source, supplemented with Cby (1 nM), arabinose (250 μM), ampicillin (100 μg/mL), and $MgSO_4$ (1 mM). **C**. ATPase activity assay measured with ADP-glo Kinase Assay Kit (Promega). Reaction mixture contained HEPES buffer (50 mM, pH 6.8), TCEP (2 μM), $MgCl_2$ (1 mM), ATP (10 mM), L-Thr (50 mM), and purified protein (100 ng) incubated aerobically at 25°C for 1 h. Similar results were obtained with reactions incubated anaerobically.

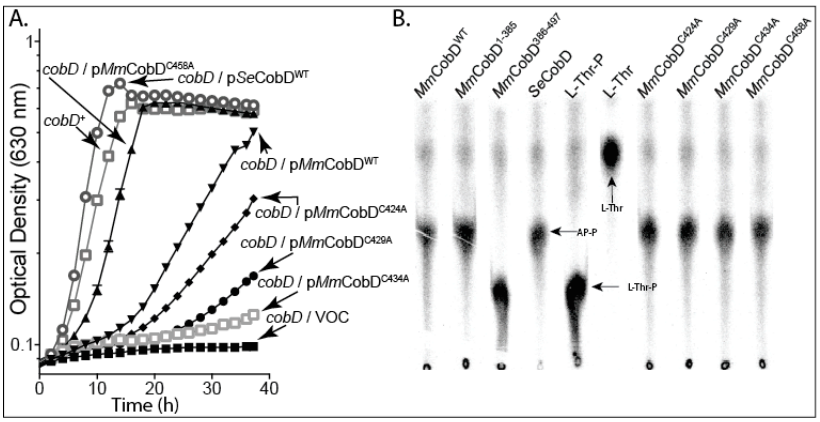



**Figure 4.5. Mutations in the metal binding domain affect *in vivo* decarboxylase activity to varying degrees. A.** Cobalamin-dependent growth assessment of *S. enterica* wild-type (*cobD*$^+$) and Δ*cobD* strains with plasmids expressing wild-type (WT) or point mutants within the metal binding domain of *M. mazei* CobD (*Mm*CobD). Included are CobD from *S. enterica* (*Se*CobD) as a positive control and a negative control Δ*cobD* strain expressing pBAD24 empty vector only control (VOC). Cells were grown aerobically at 37ºC in NCE minimal medium with glycerol (22 mM) as the sole carbon and energy source, supplemented with Cby (1 nM), arabinose (250 μM), ampicillin (100 μg/mL), and $MgSO_4$ (1 mM). **B.** Phosphorimage of the resolution of products and reactants by TLC of the L-Thr-P decarboxylation reaction with wild-type and variant *M. mazei* and *S. enterica* CobD proteins. AP-P, *1*-amino-*2*-propanol phosphate; L-Thr-P, L-Threonine-*O*-*3*-phosphate; L-Thr, L-threonine.

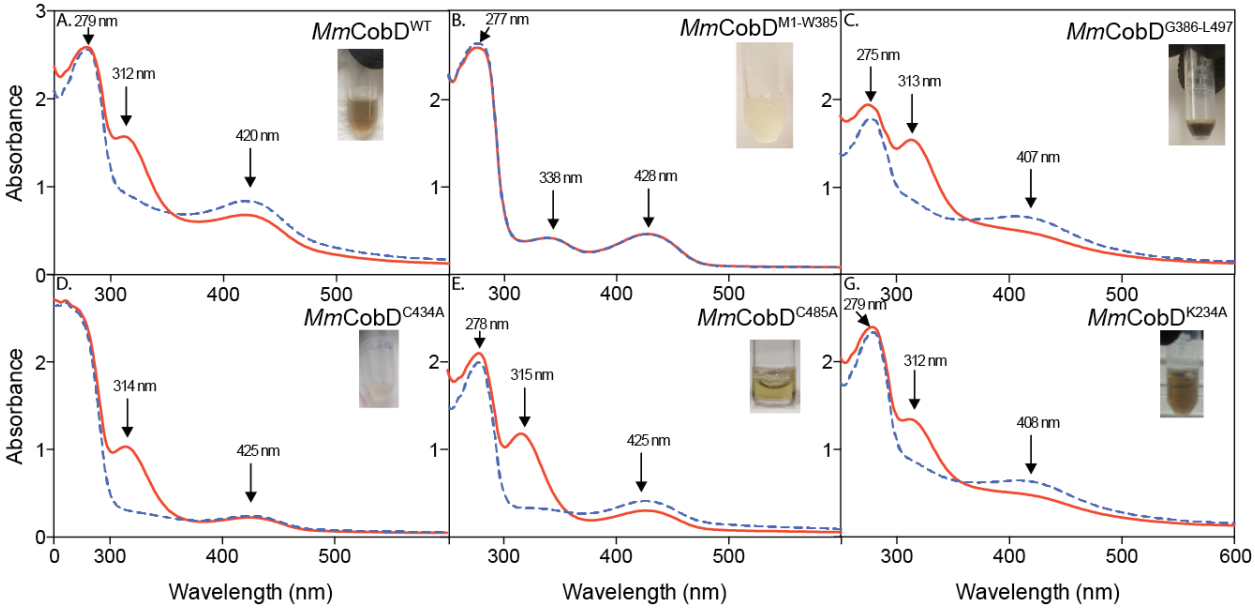



**Figure 4.6. UV-vis spectrum of *Mm*CobD protein and variants.** BluE dashed line represents the spectra of anaerobically purified protein (31 µM) under anaerobic conditions at 25°C. Solid red line represents the spectra of protein reduced with dithionite (2 mM). Insets of anaerobically purified proteins. $^{\lambda}$max for PLP are 338 nm and 428 nm. $^{\lambda}$max for the Fe species bound to protein are 312 nm when treated with dithionite and 408 nm when untreated.

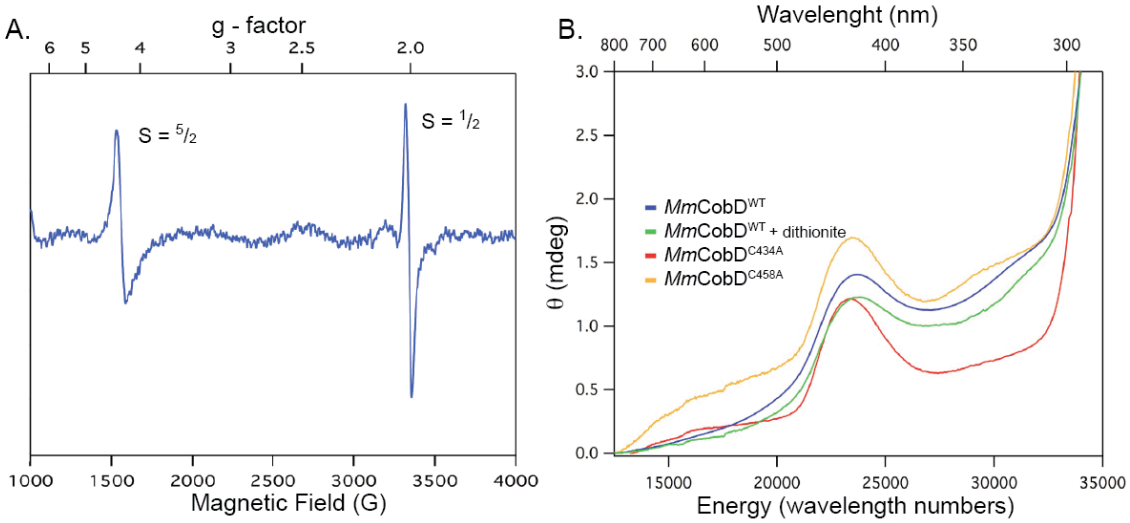



**Figure 4.7. Spectroscopic analysis of *Mm*CobD**. **A.** X-band EPR spectra collected at 15 K for wild-type *Mm*CobD. **B**. Representative MCD spectra for *Mm*CobD$^{WT}$, *Mm*CobD$^{C434A}$, *Mm*CobD$^{C458A}$ (31µM) +/- dithionite (7 mM). Wild-type protein and other variants tested have similar spectra.

## SUPPLEMENTAL FIGURES

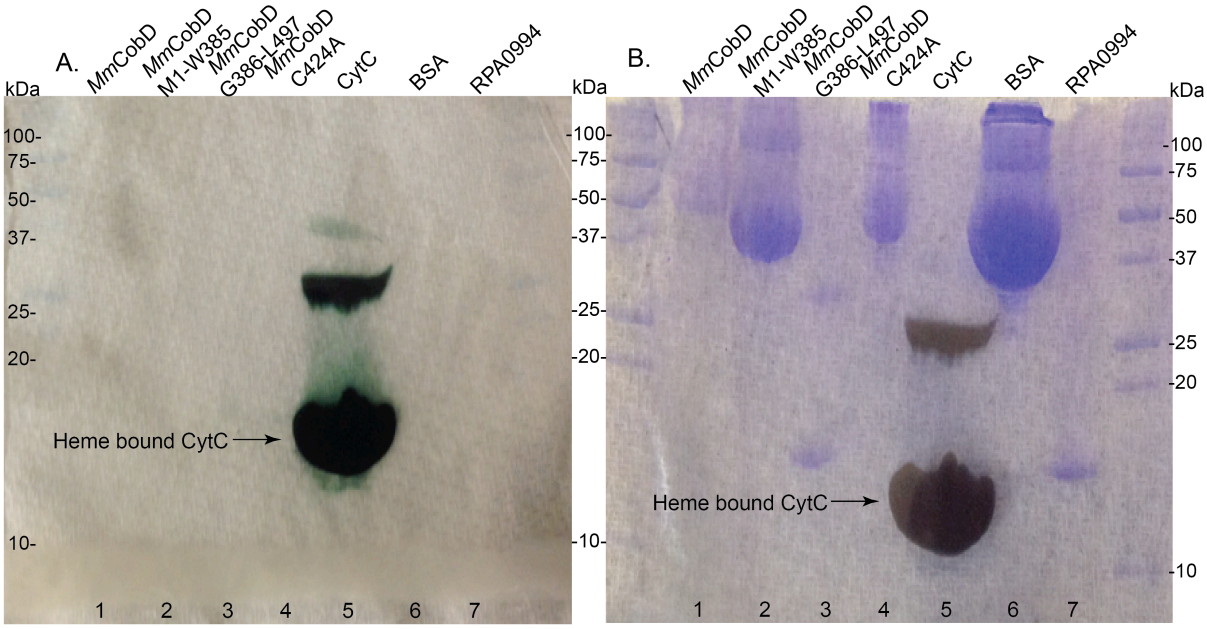



**Figure S4.1. Metal binding domain does not contain heme.** LDS-PAGE non-reducing gel stained with heme peroxidase assay (left) or Coomassie stain for proteins (right). Green band indicates the presence of heme. Blue bands are protein.

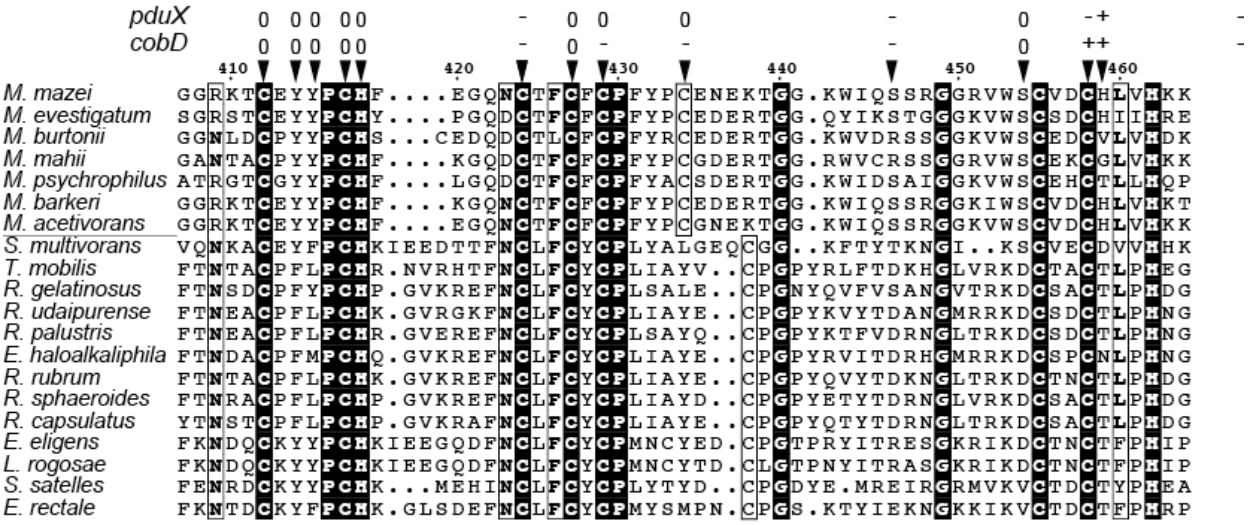



**Figure S4.2. Multiple sequence alignment of metal binding domains**. Sequences of the gene products for putative metal binding genes associated with cobalamin biosynthetic genes in the chromosome of various prokaryotes were aligned with the fused *C*-terminus metal binding domain of CobD from several *Methanosarcina* species. Black highlighted residues are conserved, boxed residues have similar chemical properties. Carrots symbols represent residues that were changed to alanine. The effect of the substitution when the variant protein encoding it was expressed on a *cobD* or *pduX* strain is marked with (0) for no change, (-) for a growth defect, or (+) for improvement of growth.

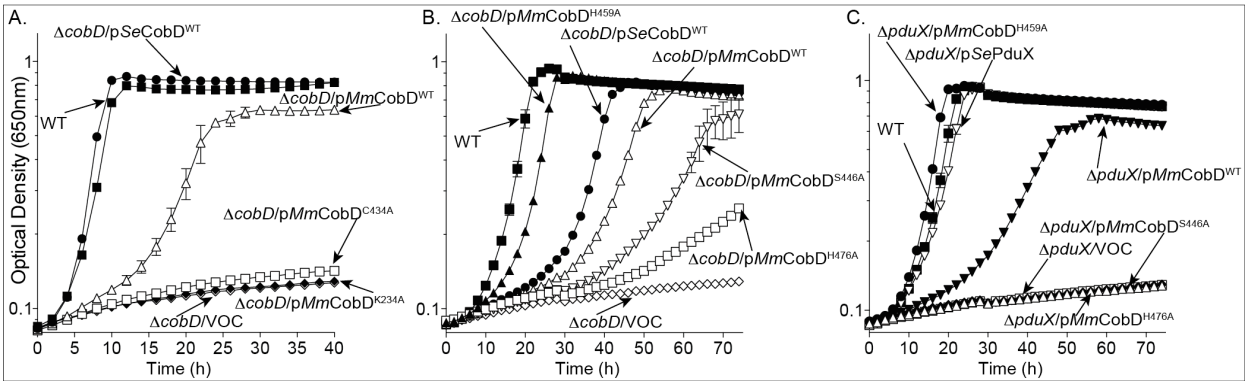



**Figure S4.3. Growth phenotype of representative *Mm*CobD mutants**. Cobalamin-dependent growth of *S. enterica* wild-type, **A**. & **B**. *cobD,* and **C**. *pduX* strains with plasmids expressing wild-type or variant *Mm*CobD proteins. Cells were grown aerobically at 37ºC in NCE minimal medium with glycerol (22 mM) as the sole carbon and energy source, supplemented with Cby (15 nM), arabinose (100 μM), ampicillin (100 μg/mL), and $MgSO_4$ (1 mM).

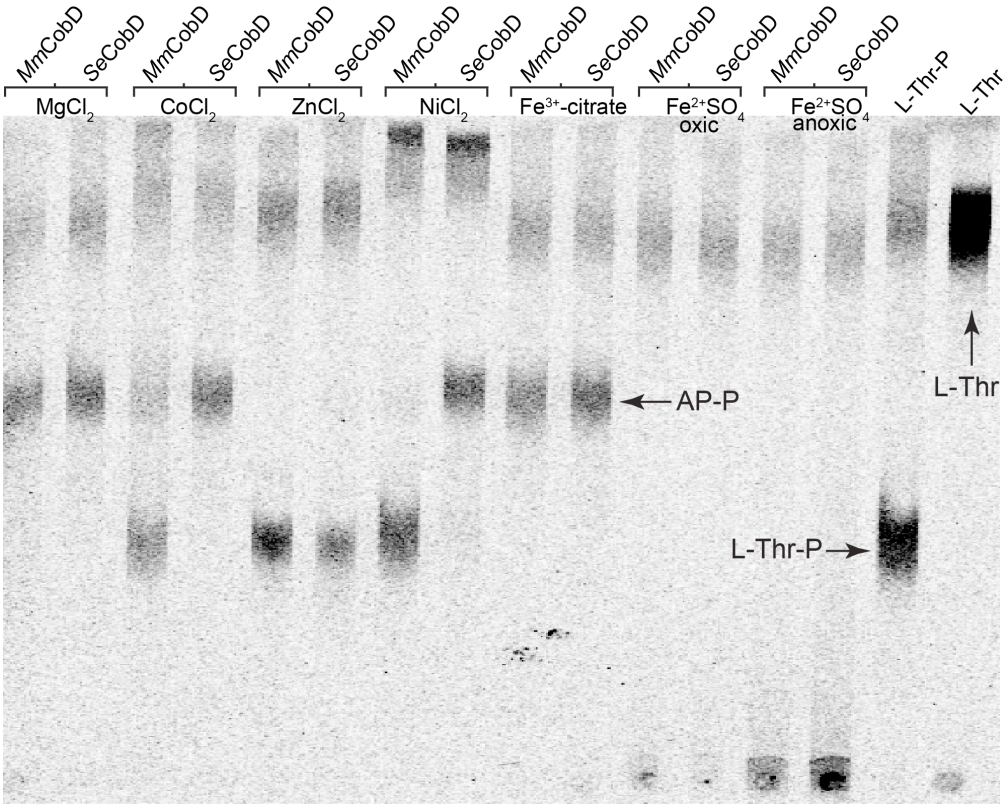



**Figure S4.4**. ***Mm*CobD decarboxylase activity is sensitive to some divalent metals.** Phosphorimage of the resolution of products and reactants by TLC of the *Mm*CobD or *Se*CobD L-Thr-P decarboxylation reaction incubated with 1mM of the indicated metal.

**Table S4.1. Strains and plasmids list**. *S. enterica* strains are derivatives of serovar Typhimurium strain LT2. Strains and plasmids were constructed during the course of this work unless stated otherwise.

| Strain | Relevant genotype | Reference / Source |
|---|---|---|
| ***Salmonella enterica*** | | |
| JE7088 | Δ*metE2702 ara-9* | Laboratory collection |
| **Derivatives of JE7088** | | |
| JE21644 | / pBAD24 *bla*$^+$ | |
| JE21557 | / pTEV5 *bla*$^+$ | |
| JE12656 | Δ*pduX*516 | |
| JE12941 | Δ*cobD*1371 | Laboratory collection |
| JE2216 | Δ*cobD*1371 | Laboratory collection |
| JE14935 | Δ*cobD*1371 / pBAD24 *bla*$^+$ | |
| JE21558 | Δ*cobD*1371 / pTEV5 *bla*$^+$ | |
| JE21631 | Δ*cobD*1371/ pCOBD6 *bla*$^+$ | |
| JE18501 | Δ*cobD*1371/ p*Mm*COBD7 *bla*$^+$ | |
| JE21621 | Δ*cobD*1371/ p*Mm*COBD9 *bla*$^+$ | |
| JE18967 | Δ*cobD*1371/ p*Mm*COBD13 *bla*$^+$ | |
| JE19195 | Δ*cobD*1371/ p*Mm*COBD17 *bla*$^+$ | |
| JE19196 | Δ*cobD*1371/ p*Mm*COBD18 *bla*$^+$ | |
| JE19197 | Δ*cobD*1371/ p*Mm*COBD19 *bla*$^+$ | |
| JE21351 | Δ*cobD*1371/ p*Mm*COBD21 *bla*$^+$ | |
| JE21338 | Δ*cobD*1371/ p*Mm*COBD22 *bla*$^+$ | |
| JE21339 | Δ*cobD*1371/ p*Mm*COBD23 *bla*$^+$ | |
| JE21340 | Δ*cobD*1371/ p*Mm*COBD24 *bla*$^+$ | |
| JE21341 | Δ*cobD*1371/ p*Mm*COBD25 *bla*$^+$ | |
| JE21342 | Δ*cobD*1371/ p*Mm*COBD26 *bla*$^+$ | |
| JE21343 | Δ*cobD*1371/ p*Mm*COBD27 *bla*$^+$ | |
| JE21537 | Δ*cobD*1371/ p*Mm*COBD28 *bla*$^+$ | |
| JE21538 | Δ*cobD*1371/ p*Mm*COBD29 *bla*$^+$ | |
| JE21539 | Δ*cobD*1371/ p*Mm*COBD30 *bla*$^+$ | |
| JE21540 | Δ*cobD*1371/ p*Mm*COBD31 *bla*$^+$ | |
| JE21541 | Δ*cobD*1371/ p*Mm*COBD32 *bla*$^+$ | |
| JE21776 | Δ*cobD*1371/ p*Mm*COBD33 *bla*$^+$ | |
| JE21870 | Δ*cobD*1371/ p*Mm*COBD34 *bla*$^+$ | |
| JE21980 | Δ*cobD*1371/ p*Mm*COBD35 *bla*$^+$ | |
| JE22090 | Δ*cobD*1371/ p*Mm*COBD36 *bla*$^+$ | |
| JE18765 | Δ*pduX*516 / p*Mm*COBD7 *bla*$^+$ | |
| JE21619 | Δ*pduX*516 / p*Mm*COBD9 *bla*$^+$ | |
| JE18970 | Δ*pduX*516 / p*Mm*COBD13 *bla*$^+$ | |
| JE19204 | Δ*pduX*516 / p*Mm*COBD17 *bla*$^+$ | |
| JE21336 | Δ*pduX*516 / p*Mm*COBD18 *bla*$^+$ | |
| JE19205 | Δ*pduX*516 / p*Mm*COBD19 *bla*$^+$ | |
| JE21290 | Δ*pduX*516 / p*Mm*COBD21 *bla*$^+$ | |
| JE21291 | Δ*pduX*516 / p*Mm*COBD22 *bla*$^+$ | |
| JE21292 | Δ*pduX*516 / p*Mm*COBD23 *bla*$^+$ | |
| JE21293 | Δ*pduX*516 / p*Mm*COBD24 *bla*$^+$ | |
| JE21294 | Δ*pduX*516 / p*Mm*COBD25 *bla*$^+$ | |
| JE21295 | Δ*pduX*516 / p*Mm*COBD26 *bla*$^+$ | |
| JE21337 | Δ*pduX*516 / p*Mm*COBD27 *bla*$^+$ | |

| | | |
|---|---|---|
| JE21532 | Δ*pduX*516 / p*Mm*COBD28 *bla*$^{+}$ | |
| JE21533 | Δ*pduX*516 / p*Mm*COBD29 *bla*$^{+}$ | |
| JE21534 | Δ*pduX*516 / p*Mm*COBD30 *bla*$^{+}$ | |
| JE21535 | Δ*pduX*516 / p*Mm*COBD31 *bla*$^{+}$ | |
| JE21536 | Δ*pduX*516 / p*Mm*COBD32 *bla*$^{+}$ | |
| JE21590 | Δ*pduX*516 / p*Mm*COBD33 *bla*$^{+}$ | |
| JE21869 | Δ*pduX*516 / p*Mm*COBD34 *bla*$^{+}$ | |
| JE21979 | Δ*pduX*516 / p*Mm*COBD35 *bla*$^{+}$ | |
| JE22080 | Δ*pduX*516 / p*Mm*COBD36 *bla*$^{+}$ | |
| ***Escherichia coli*** | | |
| C43(DE3) | F$^{-}$ ompT gal hsd$S_B$ (rB$^{-}$mB) [dcm] [Ion] | (27) |
| DH5α | F$^{-}$/ endA1 hsdR17(rk$^{-}$, mk$^{+}$) glnV44 thi-1 recA1 gyrA96 (Nx$^{R}$) relA1 U169 deoR (Φ80-dlacZ M15 Δ(lacZYA-argF) phoAsupE44 relA1 | (18) |
| **Plasmid** | | |
| pCOBD6 | *S. enterica cobD*$^{+}$ in pT7-7 *bla*$^{+}$ | Laboratory collection |
| p*Mm*COBD7 | *M. mazei cobD*$^{1\text{-}385}$ in pBAD24 *bla*$^{+}$ | |
| p*Mm*COBD9 | *M. mazei cobD*$^{1\text{-}385}$ in pTEV5 *bla*$^{+}$ | |
| p*Mm*COBD13 | *M. mazei cobD*$^{386\text{-}497}$ in pBAD24 *bla*$^{+}$ | |
| p*Mm*COBD17 | *M. mazei cobD*$^{+}$ in pBAD24 *bla*$^{+}$ | |
| p*Mm*COBD18 | *M. mazei cobD*$^{+}$ pTEV5 *bla*$^{+}$ | |
| p*Mm*COBD19 | *M. mazei cobD*$^{386\text{-}497}$ pTEV5 *bla*$^{+}$ | |
| p*Mm*COBD21 | *M. mazei cobD*$^{H396A}$ pTEV5 *bla*$^{+}$ | |
| p*Mm*COBD22 | *M. mazei cobD*$^{C429A}$ pTEV5 *bla*$^{+}$ | |
| p*Mm*COBD23 | *M. mazei cobD*$^{C427A}$ pTEV5 *bla*$^{+}$ | |
| p*Mm*COBD24 | *M. mazei cobD*$^{C424A}$ pTEV5 *bla*$^{+}$ | |
| p*Mm*COBD25 | *M. mazei cobD*$^{S454A}$ pTEV5 *bla*$^{+}$ | |
| p*Mm*COBD26 | *M. mazei cobD*$^{C458A}$ pTEV5 *bla*$^{+}$ | |
| p*Mm*COBD27 | *M. mazei cobD*$^{H418A}$ pTEV5 *bla*$^{+}$ | |
| p*Mm*COBD28 | *M. mazei cobD*$^{Y414A}$ pTEV5 *bla*$^{+}$ | |
| p*Mm*COBD29 | *M. mazei cobD*$^{C434A}$ pTEV5 *bla*$^{+}$ | |
| p*Mm*COBD30 | *M. mazei cobD*$^{S446A}$ pTEV5 *bla*$^{+}$ | |
| p*Mm*COBD31 | *M. mazei cobD*$^{H459A}$ pTEV5 *bla*$^{+}$ | |
| p*Mm*COBD32 | *M. mazei cobD*$^{H476A}$ pTEV5 *bla*$^{+}$ | |
| p*Mm*COBD33 | *M. mazei cobD*$^{Y414S/Y415A}$ pTEV5 *bla*$^{+}$ | |
| p*Mm*COBD34 | *M. mazei cobD*$^{C412A}$ pTEV5 *bla*$^{+}$ | |
| p*Mm*COBD35 | *M. mazei cobD*$^{C417A}$ pTEV5 *bla*$^{+}$ | |
| p*Mm*COBD36 | *M. mazei cobD*$^{K234A}$ pTEV5 *bla*$^{+}$ | |
| pPDU15 | *S. enterica pduX*$^{+}$ in pBAD30 *bla*$^{+}$ | Laboratory collection |
| pPDU23 | *S. enterica pduX*$^{+}$ in pTEV5 *bla*$^{+}$ | Laboratory collection |
| pBAD24 | cloning/complementation vector *bla*$^{+}$ | (19) |
| pBAD30 | cloning/complementation vector *bla*$^{+}$ | (19) |
| pTEV5 | cloning/overexpression vector, *N*-terminal rTEV cleavable $His_6$ tag *bla*$^{+}$ | (20) |

## REFERENCES


1. **Escalante-Semerena JC.** 2007. Conversion of cobinamide into adenosylcobamide in bacteria and archaea. J Bacteriol **189:**4555-4560.

2. **Tavares NK.** 2016. Biosynthesis of the Aminopropanol Moiety of Cobamides by Bacterial and Archaeal Enzymes. PhD. University of Georgia.

3. **Cheong CG, Escalante-Semerena JC, Rayment I.** 2002. Structural studies of the L-threonine-*O*-3-phosphate decarboxylase (CobD) enzyme from *Salmonella enterica*: the apo, substrate, and product-aldimine complexes. Biochemistry **41:**9079-9089.

4. **Cheong CG, Bauer CB, Brushaber KR, Escalante-Semerena JC, Rayment I.** 2002. Three-dimensional structure of the L-threonine-*O*-3-phosphate decarboxylase (CobD) enzyme from *Salmonella enterica*. Biochemistry **41:**4798-4808.

5. **Brushaber KR, O'Toole GA, Escalante-Semerena JC.** 1998. CobD, a novel enzyme with L-threonine-*O*-3-phosphate decarboxylase activity, is responsible for the synthesis of (*R*)-1-amino-2-propanol *O*-2-phosphate, a proposed new intermediate in cobalamin biosynthesis in *Salmonella typhimurium* LT2. J Biol Chem **273:**2684-2691.

6. **Fan C, Fromm HJ, Bobik TA.** 2009. Kinetic and functional analysis of L-threonine kinase, the PduX enzyme of *Salmonella enterica*. J Biol Chem **284:**20240-20248.

7. **Fan C, Bobik TA.** 2008. The PDUX enzyme of *Salmonella enterica* is an L-threonine kinase used for coenzyme $B_{12}$ synthesis. J Biol Chem **283:**11322-11329.

8. **Peariso K, Zhou ZS, Smith AE, Matthews RG, Penner-Hahn JE.** 2001. Characterization of the zinc sites in cobalamin-independent and cobalamin-dependent methionine synthase using zinc and selenium X-ray absorption spectroscopy. Biochemistry **40:**987-993.

9. **Drennan CL, Huang S, Drummond JT, Matthews RG, Ludwig ML.** 1994. How a protein binds $B_{12}$: A 3.0Å X-ray structure of $B_{12}$-binding domains of methionine synthase. Science **266:**1669-1674.

10. **Hall DA, Vander Kooi CW, Stasik CN, Stevens SY, Zuiderweg ER, Matthews RG.** 2001. Mapping the interactions between flavodoxin and its physiological partners flavodoxin reductase and cobalamin-dependent methionine synthase. Proc Natl Acad Sci USA **98:**9521-9526.

11. **Taylor RT, Weissbach H.** 1973. $N^5$-methylenetetrahydrofolate-homocysteine methyltransferases, p 121-165. *In* Boyer PD (ed), The Enzymes, vol 9. Academic Press, Inc., New York.

12. **Datsenko KA, Wanner BL.** 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci USA **97:**6640-6645.

13. **Berkowitz D, Hushon JM, Whitfield HJ, Jr., Roth J, Ames BN.** 1968. Procedure for identifying nonsense mutations. J Bacteriol **96:**215-220.

14. **Balch WE, Wolfe RS.** 1976. New approach to the cultivation of methanogenic bacteria: 2-mercaptoethanesulfonic acid (HS-CoM)-dependent growth of *Methanobacterium ruminantium* in a pressurized atmosphere. Appl Environ Microbiol **32:**781-791.

15. **Bertani G.** 1951. Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. J Bacteriol **62:**293-300.

16. **Bertani G.** 2004. Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. J Bacteriol **186:**595-600.

17. **Raleigh EA, Lech K, Brent R.** 1989. Selected topics from classical bacterial genetics, p 1.4. *In* Ausubel FA, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (ed), Current Protocols in Molecular Biology, vol 1. Wiley Interscience, New York.

18. **Woodcock DM, Crowther PJ, Doherty J, Jefferson S, De Cruz E, Noyer-Weidner M, Smith SS, Michael MZ, Graham MW.** 1989. Quantitative evaluation of *Escherichia coli* host strains for tolerance to cytosine methylation in plasmid and phage recombinants. Nucl Acids Res **17:**3469-3478.

19. **Guzman LM, Belin D, Carson MJ, Beckwith J.** 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J Bacteriol **177:**4121-4130.

20. **Rocco CJ, Dennison KL, Klenchin VA, Rayment I, Escalante-Semerena JC.** 2008. Construction and use of new cloning vectors for the rapid isolation of recombinant proteins from *Escherichia coli*. Plasmid **59:**231-237.

21. **O'Toole GA, Rondon MR, Escalante-Semerena JC.** 1993. Analysis of mutants of defective in the synthesis of the nucleotide loop of cobalamin. J Bacteriol **175:**3317-3326.

22. **Palmer LD, Downs DM.** 2013. The thiamine biosynthetic enzyme ThiC catalyzes multiple turnovers and is inhibited by *S*-adenosylmethionine (AdoMet) metabolites. J Biol Chem **288:**30693-30699.

23. **Francis RT, Jr., Becker RR.** 1984. Specific indication of hemoproteins in polyacrylamide gels using a double-staining process. Anal Biochem **136:**509-514.

24. **Thomas PE, Ryan D, Levin W.** 1976. An improved staining procedure for the detection of the peroxidase activity of cytochrome P-450 on sodium dodecyl sulfate polyacrylamide gels. Anal Biochem **75:**168-176.

25. **Sasse J.** 1991. Detection of proteins, p 10.16.11-10.16.18. *In* Ausubel FA, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (ed), Current Protocols in Molecular Biology, vol 1. Wiley Interscience, New York.

26. **Stoll S, Schweiger A.** 2006. EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. J Magn Reson **178:**42-55.

27. **Miroux B, Walker JE.** 1996. Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. J Mol Biol **260:**289-298.

28. **Harris CE, Kobes RD, Teller DC, Rutter WJ.** 1969. The molecular characteristics of yeast aldolase. Biochemistry **8:**2442-2454.

29. **Kobes RD, Simpson RT, Vallee RL, Rutter WJ.** 1969. A functional role of metal ions in a class II aldolase. Biochemistry **8:**585-588.

30. **Thomas PE, Ryan D, Levin W.** 1976. An improved staining procedure for the detection of the peroxidase activity of cytochrome P-450 on sodium dodecyl sulfate polyacrylamide gels. Anal Biochem **75:**168-176.

31. **Beattie AE, Gupta SD, Frankova L, Kazlauskaite A, Harmon JM, Dunn TM, Campopiano DJ.** 2013. The pyridoxal 5'-phosphate (PLP)-dependent enzyme serine palmitoyltransferase (SPT): effects of the small subunits and insights from bacterial mimics of human hLCB2a HSAN1 mutations. Biomed Res Int **2013:**194371.

32. **Vedavathi M, Girish KS, Kumar MK.** 2006. A novel low molecular weight alanine aminotransferase from fasted rat liver. Biochemistry (Mosc) **71 Suppl 1:**S105-112.

33. **Maenpuen S, Sopitthummakhun K, Yuthavong Y, Chaiyen P, Leartsakulpanich U.** 2009. Characterization of Plasmodium falciparum serine hydroxymethyltransferase-A potential antimalarial target. Mol Biochem Parasitol **168:**63-73.

34. **Gupta N, Bonomi F, Kurtz DM, Jr., Ravi N, Wang DL, Huynh BH.** 1995. Recombinant Desulfovibrio vulgaris rubrerythrin. Isolation and characterization of the diiron domain. Biochemistry **34:**3310-3318.

35. **Das A, Coulter ED, Kurtz DM, Ljungdahl LG.** 2001. Five-gene cluster in Clostridium thermoaceticum consisting of two divergent operons encoding rubredoxin oxidoreductase-rubredoxin and rubrerythrin-type A flavoprotein-high-molecular-weight rubredoxin. Journal of Bacteriology **183:**1560-1567.

36. **Vidakovic M, Fraczkiewicz G, Dave BC, Czernuszewicz RS, Germanas JP.** 1995. The Environment of [2Fe-2S] Clusters in Ferredoxins - the Role of Residue-45 Probed by Site-Directed Mutagenesis. Biochemistry **34:**13906-13913.

37. **Tan G, Lu J, Bitoun JP, Huang H, Ding H.** 2009. IscA/SufA paralogues are required for the [4Fe-4S] cluster assembly in enzymes of multiple physiological pathways in Escherichia coli under aerobic growth conditions. Biochem J **420:**463-472.

38. **Jung YS, Gao-Sheridan HS, Christiansen J, Dean DR, Burgess BK.** 1999. Purification and biophysical characterization of a new [2Fe-2S] ferredoxin from Azotobacter vinelandii, a putative [Fe-S] cluster assembly/repair protein. J Biol Chem **274:**32402-32410.

39. **Yuvaniyama P, Agar JN, Cash VL, Johnson MK, Dean DR.** 2000. NifS-directed assembly of a transient [2Fe-2S] cluster within the NifU protein. Proc Natl Acad Sci U S A **97:**599-604.

40. **Iwasaki T, Isogai Y, Iizuka T, Oshima T.** 1995. Sulredoxin: a novel iron-sulfur protein of the thermoacidophilic archaeon Sulfolobus sp. strain 7 with a Rieske-type [2Fe-2S] center. J Bacteriol **177:**2576-2582.

41. **Qian L, Zheng C, Liu J.** 2013. Characterization of iron-sulfur cluster assembly protein IscA from Acidithiobacillus ferrooxidans. Biochemistry (Mosc) **78:**244-251.

42. **Zabinski RF, Toney MD.** 2001. Metal ion inhibition of nonenzymatic pyridoxal phosphate catalyzed decarboxylation and transamination. J Am Chem Soc **123:**193-128.

43. **Casasnovas R, Frau J, Ortega-Castro J, Donoso J, Munoz F.** 2013. C-H activation in pyridoxal-5'-phosphate and pyridoxamine-5'-phosphate Schiff bases: effect of metal chelation. A computational study. J Phys Chem B **117:**2339-2347.

44. **Longenecker JB, Snell EE.** 1957. The comparative activities of metal ions in promoting pyridoxal-catalyzed reactions of amino acids. Journal of the American Chemical Society **79:**141-145.

45. **Metzler DE, Snell EE.** 1952. Some transamination reactions involving vitamin-B6. Journal of the American Chemical Society **74:**979-983.

46. **Matsuo Y.** 1957. Formation of Schiff Bases of Pyridoxal Phosphate - Reaction with Metal Ions. Journal of the American Chemical Society **79:**2011-2015.

47. **Heffern MC, Kurutz JW, Meade TJ.** 2013. Spectroscopic elucidation of the inhibitory mechanism of Cys2His2 zinc finger transcription factors by cobalt(III) Schiff base complexes. Chemistry **19:**17043-17053.

48. **Hurtado RR, Harney AS, Heffern MC, Holbrook RJ, Holmgren RA, Meade TJ.** 2012. Specific inhibition of the transcription factor Ci by a cobalt(III) Schiff base-DNA conjugate. Mol Pharm **9:**325-333.

49. **Sobota JM, Imlay JA.** 2011. Iron enzyme ribulose-5-phosphate 3-epimerase in *Escherichia coli* is rapidly damaged by hydrogen peroxide but can be protected by manganese. Proc Natl Acad Sci U S A **108:**5402-5407.

50. **Dunn MF.** 2012. Allosteric regulation of substrate channeling and catalysis in the tryptophan synthase bienzyme complex. Arch Biochem Biophys **519:**154-166.

51. **Kery V, Bukovska G, Kraus JP.** 1994. Transsulfuration depends on heme in addition to pyridoxal 5'-phosphate. Cystathionine beta-synthase is a heme protein. J Biol Chem **269:**25283-25288.

52. **Singh S, Madzelan P, Banerjee R.** 2007. Properties of an unusual heme cofactor in PLP-dependent cystathionine beta-synthase. Nat Prod Rep **24:**631-639.

53. **Yadav PK, Xie P, Banerjee R.** 2012. Allosteric communication between the pyridoxal 5'-phosphate (PLP) and heme sites in the H2S generator human cystathionine beta-synthase. J Biol Chem **287:**37611-37620.

54. **Smith AT, Su Y, Stevens DJ, Majtan T, Kraus JP, Burstyn JN.** 2012. Effect of the disease-causing R266K mutation on the heme and PLP environments of human cystathionine beta-synthase. Biochemistry **51:**6360-6370.

55. **Wang H, Nair VS, Holland AA, Capolicchio S, Jessen HJ, Johnson MK, Shears SB.** 2015. Asp1 from Schizosaccharomyces pombe binds a [2Fe-2S](2+) cluster which inhibits inositol pyrophosphate 1-phosphatase activity. Biochemistry **54:**6462-6474.

56. **Hagen WR, Vanoni MA, Rosenbaum K, Schnackerz KD.** 2000. On the iron-sulfur clusters in the complex redox enzyme dihydropyrimidine dehydrogenase. European Journal of Biochemistry **267:**3640-3646.

57. **Ren B, Duan X, Ding H.** 2009. Redox control of the DNA damage-inducible protein DinG helicase activity via its iron-sulfur cluster. J Biol Chem **284:**4829-4835.

58. **Chevallet M, Dupuis A, Issartel JP, Lunardi J, van Belzen R, Albracht SP.** 2003. Two EPR-detectable [4Fe-4S] clusters, N2a and N2b, are bound to the NuoI (TYKY) subunit of NADH:ubiquinone oxidoreductase (Complex I) from Rhodobacter capsulatus. Biochim Biophys Acta **1557:**51-66.

59. **Hans M, Bill E, Cirpus I, Pierik AJ, Hetzel M, Alber D, Buckel W.** 2002. Adenosine triphosphate-induced electron transfer in 2-hydroxyglutaryl-CoA dehydratase from Acidaminococcus fermentans. Biochemistry **41:**5873-5882.

60. **Schurmann P, Stritt-Etter AL, Li J.** 1995. Reduction of ferredoxin:thioredoxin reductase by artificial electron donors. Photosynth Res **46:**309-312.

61. **Kuo CF, McRee DE, Fisher CL, O'Handley SF, Cunningham RP, Tainer JA.** 1992. Atomic structure of the DNA repair [4Fe-4S] enzyme endonuclease III. Science **258:**434-440.

62. **Fromme JC, Verdine GL.** 2003. Structure of a trapped endonuclease III-DNA covalent intermediate. EMBO J **22:**3461-3471.

63. **Agnelli P, Dossena L, Colombi P, Mulazzi S, Morandi P, Tedeschi G, Negri A, Curti B, Vanoni MA.** 2005. The unexpected structural role of glutamate synthase [4Fe-4S](+1,+2) clusters as demonstrated by site-directed mutagenesis of conserved C residues at the N-terminus of the enzyme beta subunit. Arch Biochem Biophys **436:**355-366.

64. **Vollmer SJ, Switzer RL, Debrunner PG.** 1983. Oxidation-reduction properties of the iron-sulfur cluster in Bacillus subtilis glutamine phosphoribosylpyrophosphate amidotransferase. J Biol Chem **258:**14284-14293.

65. **Mera PE, Escalante-Semerena JC.** 2010. Dihydroflavin-driven adenosylation of 4-coordinate Co(II) corrinoids: are cobalamin reductases enzymes or electron transfer proteins? J Biol Chem **285:**2911-2917.

66. **Fonseca MV, Escalante-Semerena JC.** 2000. Reduction of cob(III)alamin to cob(II)alamin in *Salmonella enterica* Serovar Typhimurium LT2. J Bacteriol **182:**4304-4309.

67. **Moore TC, Mera PE, Escalante-Semerena JC.** 2014. the EutT enzyme of *Salmonella enterica* is a unique ATP:Cob(I)alamin adenosyltransferase metalloprotein that requires ferrous ions for maximal activity. J Bacteriol **196:**903-910.

68. **Olteanu H, Wolthers KR, Munro AW, Scrutton NS, Banerjee R.** 2004. Kinetic and thermodynamic characterization of the common polymorphic variants of human methionine synthase reductase. Biochemistry **43:**1988-1997.

69. **Brown KL.** 2005. Chemistry and enzymology of vitamin $B_{12}$. Chem Rev **105:**2075-2149.

70. **Stich TA, Brooks AJ, Buan NR, Brunold TC.** 2003. Spectroscopic and computational studies of Co(3+)-corrinoids: spectral and electronic properties of the B(12) cofactors and biologically relevant precursors. J Am Chem Soc **125:**5897-5914.

71. **Stich TA, Buan NR, Brunold TC.** 2004. Spectroscopic and computational studies of Co(2+)corrinoids: spectral and electronic properties of the biologically relevant base-on and base-off forms of Co(2+)cobalamin. J Am Chem Soc **126:**9735-9749.

72. **Park K, Mera PE, Escalante-Semerena JC, Brunold TC.** 2008. Kinetic and spectroscopic studies of the ATP:corrinoid adenosyltransferase PduO from *Lactobacillus reuteri*: substrate specificity and insights into the mechanism of Co(II)corrinoid reduction. Biochemistry **47:**9007-9015.

73. **Stich TA, Buan NR, Escalante-Semerena JC, Brunold TC.** 2005. Spectroscopic and computational studies of the ATP:Corrinoid adenosyltransferase (CobA) from *Salmonella enterica*: Insights into the mechanism of adenosylcobalamin biosynthesis. J Am Chem Soc **127:**8710-8719.

74. **Buan NR, Escalante-Semerena JC.** 2005. Computer-assisted docking of flavodoxin with the ATP:Co(I)rrinoid adenosyltransferase (CobA) enzyme reveals residues critical for protein-protein interactions but not for catalysis. J Biol Chem **280:**40948-40956.

75. **Hoover DM, Jarrett JT, Sands RH, Dunham WR, Ludwig ML, Matthews RG.** 1997. Interaction of *Escherichia coli* cobalamin-dependent methionine synthase and its physiological partner flavodoxin: binding of flavodoxin leads to axial ligand dissociation from the cobalamin cofactor. Biochemistry **36:**127-138.

76. **Fonseca MV, Escalante-Semerena JC.** 2001. An in vitro reducing system for the enzymic conversion of cobalamin to adenosylcobalamin. J Biol Chem **276:**32101-32108.

77. **Fonseca MV, Buan NR, Horswill AR, Rayment I, Escalante-Semerena JC.** 2002. The ATP:co(I)rrinoid adenosyltransferase (CobA) enzyme of *Salmonella enterica* requires the 2'-OH Group of ATP for function and yields inorganic triphosphate as its reaction byproduct. J Biol Chem **277:**33127-33131.

78. **Buan NR, Rehfeld K, Escalante-Semerena JC.** 2006. Studies of the CobA-type ATP:Co(I)rrinoid adenosyltransferase enzyme of *Methanosarcina mazei* strain Gö1. J Bacteriol **188:**3543-3550.

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

## SUMMARY AND CONCLUSIONS

**Overview**. The early and late steps of cobamide (Cba) biosynthesis has been well studies in model organisms such as *Salmonella enterica*. However, many gaps remain, in particular missing steps of the pathway in archaea and bacteria that use the late cobalt insertion pathway. Work on enzymes that incorporate a variety of different nucleosides into the lower ligand has been ongoing (1-12). Only recently has work been done on the variations in the linker between the corrin ring and the nucleotide (13, 14). The linker that connects the corrin ring and lower ligand base to form a complete corrinoid is typically 1-amino-2-propanol phosphate (AP-P). In *Salmonella enterica* AP-P is synthesized by two enzymes, PduX (EC 2.7.1.177), a L-Threonine (L-Thr) kinase, and CobD (EC 4.1.1.81), a L-Threonine phosphate decarboxylase. If L-Ser is used as the substrate, ethanolamine phosphate (EA-P) is incorporated as the linker (13, 14). L-Thr-P has been generated as a linker in the laboratory (13) but has never been isolated in nature. PduX is the only identified L-Thr kinase, but homologues for this enzyme are absent from many AdoCba producers. Of note is the absence of PduX homologues from organisms that use the late Co insertion pathway.

***RsBluE* is the L-Thr kinase for as specialized subgroup of Rhodobacterales**. We have identified a homologue of PduX that is restricted to a subset of Rhodobacterales. The *bluE* gene in *Rhodobacter sphaeroides* 2.4.1 encodes the L-Thr kinase enzyme. We have demonstrated *in vivo* and *in vitro* the L-Thr kinase activity of BluE. This is first L-Thr kinase found in a AdoCba producer that uses the late cobalt insertion pathway. Within the genome of *R. sphaeroides* we have also identified BluF, a homologue of CobC (EC 2.7.8.26), a adenosylcobalamin phosphate (AdoCbl-P) phosphatase. The BluF enzyme is widespread in AdoCba producers and phylogenic analysis has revealed it to be the missing

AdoCbl-P phosphatase for the aerobic/late cobalt insertion pathway. The characterization of BluE and BluF is ongoing.

***Methanosarcina mazei* CobD is a bifunctional metalloprotein.** We have recently found that CobD from the methanogenic archaeum *M. mazei* encodes both L-Thr kinase and L-Thr-P decarboxylase activities. *Mm*CobD has one or more $[4Fe\text{-}4S]^{2+}$ cluster(s) that appear to function as a regulatory domain for the two enzymatic activities. We are currently investigating sequence, structural, spectroscopic, and enzymatic differences between the *S. enterica* and *M. mazei* CobD enzymes to elucidate the apparently novel ATP binding site/sequence and residues responsible for the L-Thr kinase activity. Many AdoCba producers lack homologues of *pduX* and it has been an open question as to how these organisms generate L-Thr-P. Phylogenetic and sequence analysis of CobD from these organisms is underway and may reveal motifs that can be used to identify and separate CobDs that encode only the decarboxylase activity and those that also encode the kinase activity.

**Other putative metalloproteins involved in AdoCba biosynthesis.** Bioinformatics analysis of AdoCba biosynthetic gene clusters uncovered the presence of loci that encode putative [Fe-S] cluster containing, ATP binding enzymes of unknown function within the genomes of hyperthermophilic archaea and bacteria. *In vitro* assessment of these enzymes has shown them to have ATPase activity, which we hypothesize, could be linked to L-Thr kinase activity (unpublished data). These ATPases are only active after the reconstitution of the [Fe-S] cluster under anaerobic conditions and at temperatures between 60 and 70ºC. We have not been able to assess the function of these enzymes *in vivo* due to the high temperature required for activity, and a lack of a genetically tractable hyperthermophilic model organism with AdoCba-dependent growth conditions. The role of the [Fe-S] cluster is not known.

**Enzymatic diversity for the synthesis of nucleotide linker of AdoCbas**. The enzymes described above represent three new classes of L-Thr kinases involved in AdoCba biosynthesis. We have filled in several gaps in the AdoCba biosynthetic pathways of archaea, hyperthermophiles, and the late cobalt insertions pathway. Characterization of these enzymes adds to the array of diversity, which has been uncovered for the late steps of the AdoCba biosynthetic pathway in prokaryotes.

**FUTURE DIRECTIONS**

**Is CobD from other organisms bifunctional?** We have discovered a secondary function encoded in the CobD enzyme from *M. mazei,* which utilizes the late insertion pathway. This work opens the question of whether or not *Mm*CobD is unique in its ability to both phosphorylate L-Thr and decarboxylase the product to produce AP-P, or if this is a more general and widespread enzymatic function found in CobDs from organisms that lack PduX homologues? Preliminary unpublished data suggest that the latter is the case. CobD from *Sulfurospirillum multivorans*, like the *M. mazei* enzyme, has both enzymatic activities. Like *Mm*CobD, CobD from *S. multivorans* (*Smul*CobD) has a longer *N*-terminus, relative to *S. enterica* and CobD from other organisms that possess PduX or BluE. *Smul*CobD has specificity for L-Ser-P over L-Thr-P and produces ethanolamine phosphate (EA-P) linker. Recent publications on *Smul*CobD have speculated that the longer *N*-terminus is involved in specificity for L-Ser-P over L-Thr-P (14); we suspect the extended *N*-terminus of *Smul*CobD and other CobDs is responsible for all or part of the ATPase/L-Thr kinase function. This can easily be explored by removal of the *N*-termini from several CobDs with an extended *N*-terminus and assaying the enzymes for L-Thr kinase activity. Additionally, we will construct chimeric proteins by fusing the *N*-termini of *Mm*CobD or *Smul*CobD to CobD from *S. enterica* to assess whether the addition of this region conveys L-Thr function onto the *Se*CobD enzyme. Random mutagenesis of *Mm*CobD can be used as a approach to identify what is likely a novel ATP binding domain and the residues responsible for L-Thr recognition and binding. Structural analysis of this enzyme in comparison to *Se*CobD will offer insight into the mechanism of actions and these studies are currently underway.

**How do the metal binding proteins that are not fused to CobD interact with CobD or other protins, and what function(s) do they have?** So far we have identified L-Thr kinase activity for *Smul*CobD and *Mm*CobD, both of which are derived from organisms that utilize the early cobalt insertion pathway. We are in the process of testing CobD from other organism such as *Rhodopseudomonas palustris*. *R. palustris* also possesses a gene that encodes a putative metal-binding protein with sequence homology to the *C*-terminus of *Mm*CobD. We suspect that the presence of this genetic marker may be a used as an

indicator for CobDs that encode the L-Thr kinase function. CobD from *R. palustris* also has an extended *N*-terminus. The gene that encodes the *R. palustris* metal-binding protein is not fused to any other proteins but it exists as an open reading frame clustered with other AdoCba synthesis genes. How do these proteins interact with CobD within these organisms? Do they play the same regulatory role as the *C*-terminal [4Fe-4S] cluster-containing domain of *Mm*CobD? These metal binding proteins are also found fused to other proteins involved in the early steps of the late Co insertion pathway such as CbiA, CbiD, CbiH, CbiZ and in at least one instance to a component of the Cba transport system BtuD.

**Is the metal binding domain of *Mm*CobD involved in redox chemistry? Is it the elusive Cba reductase?** The function of the 'zinc finger'/metal binding proteins in AdoCba biosynthesis remains unknown. These [Fe-S] proteins may be involved in redox chemistry. A reductase enzyme is required to reduce the cobalt in Cba and Cba precursors from Co(II) to Co(I) (15). The flavin reductase FldA is responsible for this cobalt reduction in *S. enterica* (16) but other dedicated Cba cobalt reductases have not been identified in any organism. The [Fe-S] cluster containing proteins might fulfill this role. We will be testing this hypothesis in the near future. We have also identified other putative [Fe-S] cluster proteins without sequence homology to these 'zinc fingers' clustered with AdoCba biosynthetic genes in the chromosomes of hyperthermophilic bacteria and archaea of the order Thermotogales, and the genera *Thermococcus*, and *Pyrococcus*. We are investigating these putative Fe-S clusters containing proteins as potential AdoCba cobalt reductases.

**Other Cba associated genes with unknown and unexplored functions.** In addition to the AdoCba associated genes mentioned above, we and others have encountered several AdoCba associated genes whose function(s) remain uncharacterized. Among them are *cblX, cblY, cbiY, cobW,* and *cblZ*. Our lab has cloned these genes, but we have not rigorously investigated their function. A methodical genetic and biochemical exploration could lead to new function/pathways for AdoCba biosynthesis in other organisms. A good starting point might be a straightforward bioinformatics analysis of genomes of organisms containing these genes. This approach was fruitful for identifying the function for the BluE and *Mm*CobD kinase functions.

**Candidate L-Thr kinases in other AdoCba producers.** Further bioinformatics analysis of available sequenced genomes has revealed that there are many bacterial and archaeal AdoCba producers that do not have genes that encode homologues of either PduX or BluE. Neither *pduX* nor *bluE* are present in other purple photosynthetic bacteria such as *R. palustris* and *Rhodospirullum rubrum*. This confirms that BluE is the L-Thr kinase of the Rhodobacterales only, and not of purple photosynthetic bacteria or late Co insertion pathway organisms in general. This would suggest that there is at least one other yet-to-be-identified non-orthologous enzyme for the synthesis of L-Thr-P in these organisms. For instance, *Bacillus megaterium* lacks homologues of *pduX* and *bluE* and encodes all the AdoCba production genes in 2 large operons. All the genes within these 2 operons have assigned functions with the exception of *cbiY* (locus BMD_2995) and *cobW* (locus BMD_1961). The function of CobW is unknown but it is annotated as having an ATP binding domain. The function of CbiY has been investigated but remains unknown (17). Heterologous expression of CbiY does not complement a *S. enterica pduX* strain (unpublished). However, this may be a result of ineffective expression in *S. enterica*. The L-Thr kinase activity of CbiY has not been investigated *in vitro*. *cbiY* homologues are found associated with AdoCba synthesis genes in a number of organisms such as *Bacillus anthracis, Brevibacillus brevis, Mycobacterium tuberculosis,* and *Staphylococcus aureus*. However, they may be misannotations due to the sequence similarity to another AdoCba synthesis gene, the DMB synthase, BluB (6, 7).

**Putative thermophilic L-thr kinases.** Bioinformatic analysis of Thermatogales has found that these anaerobic extreme thermophiles cluster all their AdoCba biosynthetic genes into a single large operon (18). They possess homologues of all the AdoCba synthesis genes except CobC and PduX. Within these operons are two genes of unknown function that do not share sequence identity with any known AdoCba biosynthetic genes. ORF Tmel_0856 from *Thermosipho melanesiensis* BI429 encodes an enzyme that has cobalamin phosphate phosphatase activity (data unpublished), which presents the likelihood that the second ORF, Tmel_0716, encodes the missing L-Thr kinase enzyme. ORF Tmel_0716 has a predicted ATP binding domain as well as a [4Fe-4S] cluster motif. BLAST searches reveal homologous proteins in AdoCba biosynthetic operons of most AdoCba producing Thermotogales as well as extreme thermophilic

archaea of the genera *Pyrococus* and *Thermoccus*. This gene may represent a new class of L-Thr kinase restricted to extreme thermophiles.

**Other putative 'zinc ribbon' proteins associated with AdoCba biosynthetic genes.** Within the Actinomycetes is gene annotated as a 'zinc ribbon' protein adjacent to *cobC*. This gene is much longer (738bp) than the other AdoCba associated 'zinc finger' encoding genes (~300bp). The product of this gene may be yet another 'zinc finger' variant that is restricted to the Actinomycetes and may be worth future investigation as a potential L-Thr/L-Ser kinase. Additionally, among the other families of AdoCba producing bacteria there is a AdoCba associated protein named *CblX* (19). This small (8.5 kDa) protein is also annotated as having a zinc ribbon motif and is found in *Sinorhizobium, Mesorhizobium, Agrobacterium, Magnetospirillum, Methylobacterium, Caulobacter* and some strains of *Rhodobacter*. *cblX* is often found associated with *cblY*, neither of which have characterized functions. Within these organisms c*blX* may be another candidate for the missing L-Thr kinase or α-ribazole phosphate, as suggested by Rodionov *et. al.* (19).

**REFERENCES**


1. **Chan CH, Escalante-Semerena JC.** 2011. ArsAB, a novel enzyme from Sporomusa ovata activates phenolic bases for adenosylcobamide biosynthesis. Mol Microbiol **81:**952-967.

2. **Crofts TS, Seth EC, Hazra AB, Taga ME.** 2013. Cobamide structure depends on both lower ligand availability and CobT substrate specificity. Chem Biol doi:10.1016/j.chembiol.2013.08.006.

3. **Hazra AB, Han AW, Mehta AP, Mok KC, Osadchiy V, Begley TP, Taga ME.** 2015. Anaerobic biosynthesis of the lower ligand of vitamin B12. Proc Natl Acad Sci U S A **112:**10792-10797.

4. **Hazra AB, Tran JL, Crofts TS, Taga ME.** 2013. Analysis of substrate specificity in CobT homologs reveals widespread preference for DMB, the Lower axial ligand of vitamin B12. Chem Biol doi:10.1016/j.chembiol.2013.08.007.

5. **Mok KC, Taga ME.** 2013. Growth inhibition of *Sporomusa ovata* by incorporation of benzimidazole bases into cobamides. J Bacteriol **195:**1902-1911.

6. **Taga ME, Larsen NA, Howard-Jones AR, Walsh CT, Walker GC.** 2007. BluB cannibalizes flavin to form the lower ligand of vitamin B12. Nature **446:**449-453.

7. **Gray MJ, Escalante-Semerena JC.** 2007. Single-enzyme conversion of $FMNH_2$ to 5,6-dimethylbenzimidazole, the lower ligand of $B_{12}$. Proc Natl Acad Sci U S A **104:**2921-2926.

8. **Nussbaumer C, Imfeld M, Worner G, Arigoni D.** 1981. Biosynthesis of vitamin B12: Mode of incorporation of factor III into cobyrinic acid. Proc Natl Acad Sci USA **78:**9-10.

9. **Ohlenroth K, Friedmann HC.** 1968. Formation of vitamin B12 5'-phosphate during bacterial conversion of factor A to vitamin B12. Biochim Biophys Acta **170:**465-467.

10. **Dion HW, Calkins DG, Pfiffner JJ.** 1952. Hydrolysis products of pseudovitamin B12. J Amer Che Soc **74:**1108-1108.

11. **Dion HW, Calkins DG, Pfiffner JJ.** 1954. 2-Methyladenine, an hydrolysis product of pseudovitamin B12d. J Amer Che Soc **76:**948-949.

12. **Kräutler B, Fieber W, Osterman S, Fasching M, Ongania K-H, Gruber K, Kratky C, Mikl C, Siebert A, Diekert G.** 2003. The cofactor of tetrachloroethene reductive dehalogenase of *Dehalospirillum multivorans* is Norpseudo-B12, a new type of natural corrinoid. Helvetica Chimica Acta **86:**3698-3716.

13. **Zayas CL, Claas K, Escalante-Semerena JC.** 2007. The CbiB protein of *Salmonella enterica* is an integral membrane protein involved in the last step of the de novo corrin ring biosynthetic pathway. J Bacteriol **189:**7697-7708.

14. **Keller S, Treder A, SH VR, Escalante-Semerena JC, Schubert T.** 2016. The SMUL_1544 gene product governs norcobamide biosynthesis in the tetrachloroethene-respiring bacterium Sulfurospirillum multivorans. J Bacteriol doi:10.1128/JB.00289-16.

15. **Mera PE, Escalante-Semerena JC.** 2010. Dihydroflavin-driven adenosylation of 4-coordinate Co(II) corrinoids: are cobalamin reductases enzymes or electron transfer proteins? J Biol Chem **285:**2911-2917.

16. **Buan NR, Escalante-Semerena JC.** 2005. Computer-assisted docking of flavodoxin with the ATP:Co(I)rrinoid adenosyltransferase (CobA) enzyme reveals residues critical for protein-protein interactions but not for catalysis. J Biol Chem **280:**40948-40956.

17. **Collins HF, Biedendieck R, Leech HK, Gray M, Escalante-Semerena JC, McLean KJ, Munro AW, Rigby SE, Warren MJ, Lawrence AD.** 2013. *Bacillus megaterium* has both a functional BluB protein required for DMB synthesis and a related flavoprotein that forms a stable radical species. PLOS ONE **8:**e55708.

18. **Swithers KS, Petrus AK, Secinaro MA, Nesbo CL, Gogarten JP, Noll KM, Butzin NC.** 2012. Vitamin B(12) synthesis and salvage pathways were acquired by horizontal gene transfer to the Thermotogales. Genome Biol Evol **4:**730-739.

19. **Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS.** 2003. Comparative genomics of the vitamin $B_{12}$ metabolism and regulation in prokaryotes. J Biol Chem **278:**41148-41159.

# APPENDIX A

## THE GENOME OF *RHODOBACTER SPHAEROIDES* STRAIN 2.4.1 ENCODES FUNCTIONAL COBINAMIDE SALVAGING SYSTEMS OF ARCHAEAL AND BACTERIAL ORIGINS[4]

Acknowledgements: The work described in this chapter was performed in collaboration with Michael J. Gray. I constructed the plasmids containing *cbiZ* or *cbiS* alleles from *Streptomyces coelicolor*, *Methanocaldococcus jannaschii*, *Pyrococcus furiosus*, *Ferroplasma acidarmanus*, and *Aeropyrum pernix,* and grathered related growth data of strains carrying these plasmids. Michael J. Gray performed all other experiments.

[4] Gray M.J., Tavares N.K., and Escalante-Semerena J.C. 2008. *Mol. Micro.* 70:824-836. Reprinted here with permission from the publisher.

**ABSTRACT**

Bacteria and archaea use distinct pathways for salvaging exogenous cobinamide (Cbi), a precursor of adenosylcobalamin (AdoCbl, coenzyme $B_{12}$). The bacterial pathway depends on a bifunctional enzyme with kinase and guanylyltransferase activities (CobP in aerobic AdoCbl synthesizers) to convert AdoCbi to AdoCbi-GDP via an AdoCbi-P intermediate. Archaea lack CobP, and use a different strategy for the synthesis of AdoCbi-GDP. Archaea cleave off the aminopropanol (AP) group of AdoCbi using the CbiZ AdoCbi amidohydrolase to generate adenosylcobyric acid (AdoCby), which is converted to AdoCbi-P by the CbiB synthetase, and to AdoCbi-GDP by the CobY guanylyltransferase. We report phylogenetic, *in vivo,* and *in vitro* evidence that the genome of *Rhodobacter sphaeroides* encodes functional enzymes for Cbi salvaging systems of both bacterial and archaeal origin. Products of the reactions were identified by high performance liquid chromatography, UV-visible spectroscopy, and bioassay. The *cbiZ* genes of several bacteria and archaea restored Cbi salvaging in a strain of *Salmonella enterica* unable to salvage Cbi. Phylogenetic data led us to conclude that CbiZ is an enzyme of archaeal origin that was horizontally transferred to bacteria. Reasons why some bacteria may contain both types of Cbi salvaging system are discussed.

**INTRODUCTION**

Adenosylcobalamin (AdoCbl, coenzyme $B_{12}$) is a complex cobalt-containing cyclic tetrapyrrole coenzyme with both upper and lower axial ligands coordinating the central cobalt atom. *De novo* synthesis of AdoCbl requires a great deal of genetic information ($\geq$25 genes) that can be found in bacteria and archaea [reviewed in (1-3)]. The AdoCbl biosynthetic pathway has been best studied in *Pseudomonas denitrificans* and *Salmonella enterica* serovar Typhimurium LT2 (hereafter serovar Typhimurium) (3, 4). In contrast, studies of AdoCbl biosynthesis in archaea have been limited (5-11).

Many organisms salvage incomplete corrinoids (e.g., cobinamide, Cbi, a stable AdoCbl precursor) from their environments. Although Cbi is not an intermediate of the *de novo* synthesis pathway (12, 13), it can be converted into one via a process known as Cbi salvaging.

Bacteria and archaea use different pathways to salvage Cbi (Fig. A.1). In bacteria, Cbi salvaging starts with the attachment of 5'-deoxyadenosine (the upper ligand) to the corrin ring yielding adenosylcobinamide (AdoCbi) (14, 15). The latter is then phosphorylated to AdoCbi-phosphate (AdoCbi-P), a true *de novo* intermediate that can be ultimately converted to AdoCbl (12). The phosphorylation of AdoCbi is catalyzed by a bifunctional NTP:AdoCbi kinase (EC 2.7.7.62), GTP:AdoCbi-P guanylyltransferase (EC 2.7.1.156) enzyme (CobP in *P. denitrificans;* CobU in *S. enterica*), which is conserved among AdoCbl-producing bacteria (16, 17).

Previous work from our laboratory showed that archaea lack the bifunctional NTP:AdoCbi kinase, GTP:AdoCbi-P guanylyltransferase enzyme, and use an alternative pathway to salvage Cbi (Fig. A.1) (6, 7). In the archaeal pathway AdoCbi is converted to adenosylcobyric acid (AdoCby) by an AdoCbi amidohydrolase enzyme called CbiZ or CbiS, when CbiZ is fused to an α-ribazole-phosphate phosphatase (CobZ, EC 3.1.3.73) domain (8-11).

The conversion of AdoCby to AdoCbl proceeds via the same biochemical reactions found in bacteria, except that a monofunctional guanylyltransferase enzyme named CobY converts AdoCbi-P to AdoCbi-GDP (6, 11).

Putative orthologues of *cbiZ* are present in approximately 78% of the available archaeal genomes (36/46 sequenced genomes), none of which contain an orthologue of *cobP*. In contrast, only 38 of the 570 (7%) bacterial genomes sequenced to date encode predicted *cbiZ* orthologues. All the bacterial genomes predicted to encode CbiZ-like proteins also synthesize proteins orthologous to CobP (18).

In this paper, we examine the phylogenetic distribution of CbiZ in archaea and bacteria, and report evidence that supports the conclusion that CbiZ is an archaeal function that was horizontally transferred to bacteria. We also demonstrate, using an *in vivo* complementation system that predicted *cbiZ* genes from several archaea and bacteria encode functional Cbi amidohydrolases. Finally, we show that the genome of the photosynthetic α-proteobacterium *Rhodobacter sphaeroides* (19) encodes functional CobP and CbiZ enzymes. The function of the CbiZ and CobP proteins was demonstrated *in vivo* and *in vitro*.

This is the first demonstration of AdoCbi amidohydrolase (CbiZ) activity in any bacterium. These findings raised two questions: i) why do some bacteria contain two apparently redundant Cbi salvaging pathways? and ii) how is Cbi salvaged in organisms that have both systems?

## RESULTS

***cbiZ* orthologues are broadly distributed among archaea and bacteria.** We retrieved the sequences of 93 putative CbiZ orthologues, including 38 proteins from 36 archaea (out of a total 46 available archaeal genomes), and 55 proteins from 38 bacteria (out of a total 570 available bacterial genomes).

The phylogenetic relationships among *cbiZ* genes were inferred by maximum parsimony analysis (Fig. 5.2A) using the DNAPARS application of PHYLIP version 3.66 (20). For comparison, 16S rRNA gene sequences from these 74 species were also examined (Fig A.2B).

Most archaea contained only a single *cbiZ* allele, but two methanogenic archaea (*Methanocorpusculum labreanum* and *Methanospiriilum hungatei*) contained two copies each. Most of the bacteria we analyzed also contained a single *cbiZ* allele, but *Pelobacter propionicus* and *Bacillus* sp. B14905 contained two each, and the three species of *Dehalococcoides* contained between three and eight copies of *cbiZ*, some of which were identical in amino acid sequence, suggesting that some of these copies arose by gene duplications. In most instances, multiple *cbiZ* alleles in a single organism clade together, but the two *cbiZ* genes of *M. hungatei* were distantly related. According to the 16S phylogeny, *M. hungatei* is most closely related to *M. labreanum*, *Methanoculleus marisnigri*, and *Methanoregula boonei* (Fig. A.2B). One of the *cbiZ* alleles of *M. hungatei* (locus tag Mhun0271), however, clades with *cbiZ* genes from the more distantly related methanogens *Methanococcoides burtonii*, *Methanobrevibacter smithii*, and *Methanosphaera stadtmanae* (Fig. A.2A). Mhun0271 is flanked by predicted transposases in the *M. hungatei* genome (18).

Notably, the archaeal and bacterial *cbiZ* alleles do not cluster by domain. For example, a well-supported clade (top of Fig. A.2A) includes *cbiZ* alleles from six methanogenic archaea as well as bacteria from several different taxa. Similarly, the *cbiZ* genes of the actinobacteria (including *S. coelicolor* and *M.*

*avium*) and α-proteobacteria (including *Paracoccus denitrificans* and *Rhodospirillum rubrum*) appeared to be more closely related to the *cbiZ* genes of the extremely halophilic archaea than to any other bacterial clades.

**Diverse *cbiZ* alleles complement Cbi salvaging mutant strains of serovar Typhimurium.** To determine whether the putative *cbiZ* genes of representative archaeal and bacterial species encoded functional Cbi amidohydrolases, we performed complementation experiments with serovar Typhimurium strains defective in Cbi salvaging. The serovar Typhimurium strains used in these experiments lacked MetE, the Cbl-independent methionine synthase, and therefore depended on the Cbl-dependent methionine synthase (MetH) to make methionine (21). To block *de novo* synthesis of the corrin ring we grew all strains under oxic conditions. Under such conditions, in medium devoid of methionine, growth depended on corrinoid salvaging.

Genes encoding putative CbiZ proteins from different sources were cloned into the plasmid pBAD24, placing the *cbiZ* genes under control of the arabinose-inducible $P_{BAD}$ promoter (22). Plasmids were constructed containing predicted *cbiZ* alleles from the bacteria *Rhodobacter sphaeroides*, *Rhodospirillum rubrum*, *Bacillus halodurans*, *Streptomyces coelicolor*, and *Dehalococcoides etheneogenes*, and from the archaea *Methanosarcina mazei*, *Methanocaldococcus jannaschii*, *Ferroplasma acidarmanus*, *Pyrococcus furiosus*, and *Aeropyrum pernix*. Plasmids harboring putative *cbiZ* alleles were transformed into serovar Typhimurium strain JE8312 (Δ*cobU* Δ*ycfN*) / pCOBY38 *cobY*$^{+}$). This strain lacked both CobU and YcfN, and was therefore unable to salvage Cbi (16, 23), hence it was a Cbl auxotroph. Strain JE8312 also carried a plasmid encoding CobY, the AdoCbi-P guanylyltransferase from *Methanosarcina mazei* (6), which allowed the cell to synthesize AdoCbl from Cby (Fig. A.1). Plasmid pBAD24 (22) was used as negative control. Resulting strains were grown aerobically in medium supplemented with $(CN)_2$Cby, $(CN)_2$Cbi, or CNCbl, either with or without the addition of the inducer (*i.e.*, 2.5 mM arabinose) (Fig. A.3).

In the absence of arabinose induction, the *cbiZ* genes from *R. sphaeroides*, *B. halodurans*, *M. jannaschii*, *F. acidarmanus*, and *P. furiosus* supported growth of strain JE8312 on Cbi, indicating that plasmid-encoded *cbiZ* genes directed the synthesis of enzymes that converted Cbi to Cby, which was ultimately converted to Cbl and allowed growth (Fig. A.3, black tracings). Complementation in the absence of induction indicated that residual transcription generated enough protein to support growth and indicated that sufficient activity was associated with the CbiZ enzymes even when the latter were present at low level. In some cases (*i.e.*, *R. rubrum*, *M. mazei*), addition of at least 2.5 mM arabinose was needed to support growth on Cbi, but this high concentration of inducer had a deleterious effect on Cbi or Cby salvaging when the cell carried plasmids expressing *cbiZ* genes from *R. sphaeroides* or *B. halodurans* (Fig. A.3, gray tracings). Cbi salvaging was not observed when the *cbiZ* genes from *S. coelicolor* or *D. etheneogenes* or the *cbiS* gene from *A. pernix* were present in strain JE8312, regardless of the concentration of arabinose (data not shown).

**Cbi salvaging in serovar Typhimurium strains lacking the native Cbi salvaging system required AdoCbi-P synthetase (CbiB) function and a functional CbiZ enzyme.** It was important to determine whether the *cbiZ* genes we cloned encoded proteins with the expected Cbi amidohydrolase, or whether they had a new Cbi kinase activity. To distinguish between these possibilities we tested whether CbiZ-dependent Cbi salvaging required CbiB function as observed in the archaea *M. mazei* and *Halobacterium* sp. NRC-1 (8, 9). For this purpose we used a serovar Typhimurium strain with a mutation in the gene encoding the L-Thr *O*-3-phosphate decarboxylase (CobD) enzyme (12, 24). As predicted by the pathway shown in Figure A.1, inactivation of *cobD* prevents salvaging of Cbi, because AdoCby is not converted to AdoCbi-P in the absence of aminopropanol-phosphate (AP-P). The pathway in Figure A.1 also indicates that the effect of the absence of CobD is correctible by the addition of aminopropanol (AP) to the medium (12, 24).

Plasmids harboring putative *cbiZ* genes were transformed into serovar Typhimurium strain JE7864 (*cbiP cobD cobU* / pCOBY38 *cobY*$^{+}$). In addition to lacking CobD, this strain lacked CobU to prevent Cbi salvaging by direct phosphorylation. The presence of plasmid pCOBY38 allowed strain JE7864 to

synthesize Cbl from Cby, but not from Cbi. Plasmid pBAD24 was used as negative control in these experiments.

Derivatives of strain JE7864 carrying plasmids encoding *cbiZ* genes from different organisms were grown aerobically in medium supplemented with $(CN)_2$Cbi, $(CN)_2$Cbi plus AP, or CNCbl, either with or without the addition of 2.5 mM arabinose to induce expression of *cbiZ* (Fig. A.4). None of the *cbiZ* alleles restored Cbi-dependent growth of strain JE7864 in the absence of exogenous AP.

In the presence of exogenous AP without arabinose induction, Cbi-dependent growth was restored by the presence of *cbiZ* alleles from *R. sphaeroides*, *B. halodurans*, *M. jannaschii*, *F. acidarmanus*, and *P. furiosus*. At high levels of inducer (2.5 mM arabinose), growth was restored by c*biZ* genes from *R. sphaeroides*, *R. rubrum*, *M. mazei*, *B. halodurans*, *M. jannaschii*, *F. acidarmanus*, and *P. furiosus*; the onset of growth was delayed in the presence of *cbiZ* from *B. halodurans* and *R. sphaeroides*. Cbi-dependent growth was not observed at any concentration of inducer when the *cbiZ* genes from *S. coelicolor* or *D. etheneogenes* or the *cbiS* gene from *A. pernix* were present in strain JE7864 (data not shown).

**Protein extracts enriched for *R. sphaeroides* CbiZ hydrolyze corrinoids *in vitro*.** To confirm the *in vivo* results reported above, and to compare the activity of a bacterial CbiZ to the reported activities of archaeal CbiZ enzymes (9, 11), *R. sphaeroides* CbiZ was over-produced using serovar Typhimurium strain JE10789 (*cobU ycfN* / pCbiZ$_{Rs}$), induced with 1.5 mM arabinose. Unlike at higher arabinose concentrations (*e.g.* 2.5 mM, see Fig. A.3), at this level of induction, pCbiZ$_{Rs}$ does not have a deleterious effect on Cbi or Cby salvaging (data not shown).

Corrinoids were incubated with protein extract enriched with CbiZ, and amidohydrolase activity was monitored by HPLC. The identity of Cby generated by this system was confirmed by comparison of the retention time (5.2 min) and the UV-vis spectrum of authentic Cby (11), and by bioassay of the HPLC-purified product (data not shown). The highest specific activity was observed with AdoCbi as a substrate ($880 \pm 63$ pmol Cby $mg^{-1}$ $min^{-1}$). Specific activities of the enzyme when $(CN)_2$Cbi or CNCbl was used as substrate were $120 \pm 16$ and $15 \pm 2$ pmol Cby $mg^{-1}$ $min^{-1}$, respectively. No AdoCbl hydrolysis was

detectable. Substantially higher specific activities have been reported for some purified archaeal CbiZ proteins (9, 11), but direct comparisons to the CbiZ-enriched protein extract used here are difficult.

**The *cobP* gene of *R. sphaeroides* complemented the AdoCbi kinase and AdoCbi-P guanylyltransferase defects of a serovar Typhimurium *cobU* strain.** We selected *Rhodobacter sphaeroides* 2.4.1 as a model organism to study in more detail the corrinoid salvaging functions of a bacterium containing both the bacterial and archaeal Cbi salvaging systems. Bacteria that synthesize the NTP:AdoCbi kinase and GTP:AdoCbi-P guanylyltransferase (CobP/CobU) enzyme can efficiently salvage Cbi (16, 17). So, why would a bacterium that synthesizes CobP (or CobU) need CbiZ? One possible explanation may be that the former might lack the Cbi kinase activity needed to salvage Cbi. To test this possibility, plasmids pCobP$_{Rs}$ (encoding *R. sphaeroides cobP*) and pCobU$_{Se}$ (encoding serovar Typhimurium *cobU*) were transformed into serovar Typhimurium strain JE0824 (*cbiP cobU*), and the resulting strains were tested for their ability to salvage either Cby or Cbi. *R. sphaeroides cobP* restored growth of JE0824 on both Cby and Cbi (Fig. A.5), indicating that CobP from *R. sphaeroides* had kinase and guanylyltransferase activities.

***R. sphaeroides* CobP converts AdoCbi into AdoCbi-GDP *in vitro*.** To further confirm the *in vivo* results reported above, *R. sphaeroides* CobP and serovar Typhimurium CobU were over-produced using plasmids pCobP$_{Rs}$ and pCobU$_{Se}$. To eliminate background activity plasmids were introduced into serovar Typhimurium strain JE8268 (*cobU ycfN*), which lacked Cbi kinase and Cbi-P guanylyltransferase activities.

To test the conversion of AdoCbi to AdoCbi-GDP, AdoCbi was incubated with GTP and protein extract enriched with either CobP or CobU; synthesis of AdoCbi-GDP was monitored by HPLC. Extracts enriched with either CobU or CobP accumulated a corrinoid that eluted 6.5 min post-injection, a time that was identical to that of authentic Cbi-GDP. The UV-vis spectrum of the compound eluting at 6.5 min was identical to the reported spectrum of Cbi-GDP (25). Under these conditions, the *R. sphaeroides* CobP-enriched cell-free extract accumulated 43 ± 3 pmol Cbi-GDP min$^{-1}$ mg$^{-1}$ protein, and the serovar Typhimurium CobU-enriched extract accumulated 536 ± 32 pmol Cbi-GDP min$^{-1}$ mg$^{-1}$ protein.

To test AdoCbi kinase activity alone, AdoCbi was incubated with ATP and cell-free extract enriched with either CobP or CobU, and Cbi-P accumulation was monitored by HPLC. Extracts enriched with either CobU or CobP accumulated a corrinoid that eluted 7.8 min post-injection, a time that was identical to that of authentic Cbi-P (data not shown). The *R. sphaeroides* CobP-enriched extract accumulated 5.7 ± 3 pmol Cbi-P $min^{-1}$ $mg^{-1}$ protein. The serovar Typhimurium CobU-enriched extract accumulated 74 ± 5 pmol Cbi-P $min^{-1}$ $mg^{-1}$ total protein.

To test for AdoCbi-P guanylyltransferase activity, AdoCbi-P was incubated with GTP and cell-free extract enriched with either CobP or CobU, and Cbi-GDP accumulation was monitored by HPLC. The *R. sphaeroides* CobP-enriched extract accumulated 116 ± 14 pmol Cbi-GDP $min^{-1}$ $mg^{-1}$ protein. The serovar Typhimurium CobU-enriched extract accumulated 360 ± 117 pmol Cbi-GDP $min^{-1}$ $mg^{-1}$ total protein.

The ratio of AdoCbi-P guanylyltransferase activity to AdoCbi kinase activity was 20 for the *R. sphaeroides* CobP-enriched extract, as compared to 5 for the serovar Typhimurium CobU-enriched extract. AdoCbi-P guanylyltransferase activity to AdoCbi kinase activity ratios of 14 and 2 have been reported for purified *Pseudomonas denitrificans* CobP and serovar Typhimurium CobU enzymes, respectively (16, 17).

## DISCUSSION

**The distribution of CbiZ among prokaryotes suggests a history of horizontal gene transfer.** The simplest interpretation of the CbiZ phylogenetic data is that the CbiZ enzyme originated among the Archaea, and that at some points the *cbiZ* gene was transferred to different bacterial lineages. In most cases, only a subset of each bacterial taxon contains *cbiZ*. For example, among the actinobacteria *S. coelicolor*, *Thermobifida fusca*, *Saccharopolyspora erythraea*, *Salinospora tropica*, *Salinospora arenicola*, *M. avium*, and *M. avium paratuberculosis* all appear to encode CbiZ proteins, but none of the other 48 available actinobacterial genomes do, including 15 other strains of *Mycobacteria* (18).

Most of the known archaeal genomes contain at least one *cbiZ* orthologue. The relationships among archael CbiZ enzymes roughly parallel the 16S rRNA phylogeny (Fig. A.2B), with a few exceptions, such

as the CbiZ protein synthesized by the crenarchaeon *Ferroplasma acidarmanus*. $CbiZ_{Fa}$ clusters among CbiZ proteins from the euryarchaeal *Pyrobaculum* species (Fig. A.2A), again, suggestive of horizontal gene transfer. It is interesting to note that among the archaea which do not encode a CbiZ homologue, some (*e.g. Sulfolobus acidocaldaricus*, *Pyrobaculum arsenaticum*; have close relatives with CbiZ (Fig. A.2), and one (*Methanothermobacter thermoautotrophicus*) is predicted to encode a full set of AdoCbl synthetic enzymes (6). Whether or not archaea that lack CbiZ can salvage Cbi from their environment remains an open question.

The group of bacteria whose genomes encode CbiZ include pathogens such as *Leptospira interrogans* (26), *Porphyromonas gingivalis* (27), *Ochrobacterium anthropi* (28, 29), *M. avium* and *M. avium paratuberculosis* (30), and organisms of environmental relevance such as a group of thermophilic and alkaliphilic *Bacillus* and *Geobacillus* species (31, 32), the antibiotic-producing actinomycete *S. coelicolor* (1), and the tetrachloroethene-reducing *Dehalococcoides* species (33, 34). The role of Cbl and Cbi salvaging in the metabolism of these organisms has not been investigated.

**Why is the *cbiZ* gene often associated with Cbl transport genes in bacteria?** The *cbiZ* genes of many bacteria are found in genetic loci close to genes encoding homologues of components of the corrinoid transport (Btu) system (35-39). For example, in *R. sphaeroides*, *cbiZ* is the last gene in an apparent five-gene operon also encoding homologues of *btuB*, *btuF*, *btuC*, and *btuD*. The *cbiZ* genes of *R. rubrum*, *Chlorobium tepidum*, *Chlorobium limicola*, *Porphyromonas gingivalis*, *Prosthecochloris aestuarii*, *Pelobacter propinicus*, and some of the *Dehalococcoides cbiZ* alleles are also found in close proximity to apparent *btu* operons. Even more strikingly, nearly all of the *cbiZ* genes of *Bacillus* or *Geobacillus* species are fused to genes for *btuD*, and the *cbiZ* gene of the propionate-degrading sulfate reducer *Syntrophobacter fumaroxidans* (40) is fused to a homologue of *btuF*. We hypothesize that these associations reflect a strategy for the assembly of nucleotide loops with specific lower ligands.

**CbiZ genes vary in their level of activity in a heterologous expression system.** The ability of *cbiZ* genes to restore Cbi salvaging in serovar Typhimurium strains unable to convert Cbi into Cbl fell into four categories depending on the level of induction (Figs. A.3, A.4): 1) complementation both in the

presence and absence of inducer, 2) complementation only in the absence of inducer, 3) complementation only in the presence of inducer, and 4) unable to complement at any concentration of inducer. These results suggest that, while the majority of genes annotated as *cbiZ* do indeed encode Cbi amidohydrolases, these enzymes vary in their properties. It is difficult to determine whether variation in complementation is due to differences in enzyme activity or to differences in expression levels. It is striking, however, that the *cbiZ* alleles from the hyperthermophilic archaea *M. jannaschii* and *P. furiosus*, which have optimal growth temperatures of 85 and 100°C, respectively (41, 42) are active in serovar Typhimurium at 37°C even in the absence of inducer. It is also notable that the CbiZ protein of *R. sphaeroides* and the BtuD-CbiZ fusion protein of *B. halodurans* inhibit the conversion of Cby to Cbl when induced, suggesting that these enzymes either hydrolyze intermediates of the Cbl biosynthesis pathway (i.e., AdoCbi-P or AdoCbi-GDP), or interfere with the function of the native AdoCbl biosynthetic machinery in unknown ways. Further work is needed to examine the properties of these enzymes, and to understand how the deleterious effects of these enzymes are prevented in the organisms that synthesize them.

It is unknown whether the three *cbiZ* alleles whose activity was not detectible *in vivo* (*S. coelicolor*, *D. etheneogenes* DET0242, and *A. pernix*) are pseudogenes, are genes encoding enzymes with functions other than Cbi amidohydrolases, or are simply genes that do not function or are not expressed in serovar Typhimurium. We note that the lack of activity of *D. etheneogenes* DET0242 CbiZ may be due to the absence of a conserved asparagine residue. However, the *S. coelicolor* and *A. pernix* CbiZ sequences appear unremarkable. Further studies, either with purified enzymes or in the natural host species, are necessary to determine the functional characteristics of these enzymes.

**Why would any organism have dual Cbi salvaging systems?** One of the most intriguing unanswered questions in the $B_{12}$ field is why cobamides with different lower ligands are commonly found in nature (43). It is known that some cobamide-dependent enzymes are sensitive to the identity of the lower ligand (44, 45). Retention of CbiZ in an organism that already synthesizes a bifunctional CobP (CobU)-like enzyme (e.g., *R. sphaeroides*) may be a response to a strong positive selection exerted by the existence in the cell of an essential Ado-cobamide-dependent enzyme that requires a specific lower ligand to function.

In these organisms, the CbiZ amidohydrolase activity would ensure the recycling of the valuable corrin ring and its conversion to the specific cobamide the organism needs as a coenzyme. We speculate that organisms where such positive selection is strongest would fuse CbiZ to the Btu transport system to ensure that every corrinoid molecule entering the cell is stripped of its nucleotide loop. Experiments are currently underway to test this hypothesis.

**Other questions that need to be addressed.** In a bacterium with two Cbi salvaging pathways, is one preferentially used over the other? And if so, under what conditions? Mutational and physiological analyses of *R. sphaeroides* are needed to determine the exact pathway of Cbi salvaging in this organism. The fusion of CbiZ enzymes with proteins of the Btu corrinoid-specific transport system is also intriguing, hence it would be important to determine whether CbiZ proteins that are not fused interact with Btu proteins. Ultimately, structural analyses of CbiZ proteins are needed to provide a context for the mechanistic analysis of the CbiZ enzyme.

**Acknowledgments.** This work was supported by PHS grant R01-GM40313 from the National Institute of General Medical Sciences (to J.C.E.-S.). We thank Cameron Currie, Eric Caldera, and Jarrod Scott (UW-Madison) for assistance with phylogenetic analyses. We also thank Stephen Zinder (Cornell University) for the gift of *D. etheneogenes* 195 genomic DNA, Tim Donohue (UW-Madison) for providing *R. sphaeroides* strain 2.4.1, and Kyle Hasenstein for technical assistance.

## EXPERIMENTAL PROCEDURES

**Phylogenetic analysis.** Sequences of predicted CbiZ proteins (members of pfam01955 (46)) were retrieved from the Integrated Microbial Genomes database (18), whereas 16S rRNA genes from each organism encoding a CbiZ orthologue were obtained from the GreenGenes database (47). Since CbiZ domains can be fused to other domains, CbiZ sequences were trimmed to remove non-homologous sequence fused to the CbiZ core. 16S DNA sequences and trimmed CbiZ protein sequences were aligned using ClustalX 2.0 (48); the CbiZ protein alignment was then converted into a DNA alignment. Maximum parsimony trees were generated for each dataset with PHYLIP version 3.66 (20).

**Bacterial strains and growth conditions.** *E. coli* strains were grown at 37°C in lysogenic broth (LB, Difco) (49, 50). All *Salmonella enterica* sv. Typhimurium LT2 strains used in these studies were derived from strain TR6583 (*metE205 ara-9*). Serovar Typhimurium strains were grown at 37°C in LB, nutrient broth (NB, Difco), or no-carbon essential (NCE) minimal medium (51) containing $MgSO_4$ (1 mM), glycerol (22 mM), and trace minerals (52). For growth curves of serovar Typhimurium, starter cultures were grown aerobically overnight in NB containing appropriate antibiotics and used to inoculate fresh medium (5% v/v inoculum). Growth curves were obtained using an ELx808 Ultra Microplate Reader (Bio-Tek Instruments) in a total volume of 200 µl per well. When present in the medium, ampicillin was at 100 µg $ml^{-1}$ and kanamycin was at 50 µg $ml^{-1}$. When added, corrinoids were at 15 nM, and they were in their cyano form. Adenosylated corrins and $(CN)_2$Cbi-GDP were generated as described (6, 13). Dicyanocobyric acid [$(CN)_2$Cby] was a gift from Paul Renz (Universität-Hohenheim, Stuttgart, Germany). All other chemicals were purchased from Sigma.

**Genetic and molecular techniques.** DNA manipulations were performed using described methods (53). Restriction and modification enzymes were purchased from Fermentas (Ontario, Canada) or Promega (Madison, WI) and used according to the manufacturer's instructions. All DNA manipulations were performed in *E. coli* DH5α (54, 55). Plasmid DNA was isolated using the Wizard Plus SV Plasmid Miniprep kit (Promega). PCR products were purified with the QiaQuick PCR purification kit (Qiagen). DNA sequencing reactions were performed using non-radioactive BigDye® protocols (ABI PRISM; Applied Biosystems) and resolved at the Biotechnology Center of the University of Wisconsin-Madison.

**Construction of bacterial and archaeal *cbiZ* plasmids.** The identity of all of the *cbiZ* genes whose clone is described below was confirmed by sequencing with primers pBAD5' and pBAD3'.

**pCbiZ$_{Rs}$**. The *Rhodobacter sphaeroides cbiZ* gene coding sequence plus 3 bp of 5' and 15 bp of 3' sequence was amplified using primers Rs_cbiZ_EcoRI_5' and Rs_cbiZ_XbaI_3', and the resulting product was cloned into the *Eco*RI and *Xba*I sites of plasmid pBAD24 (22) to yield plasmid pCbiZ$_{Rs}$.

**pCbiZ$_{Rr}$**. The *Rhodospirillum rubrum cbiZ* coding sequence was amplified using primers Rr_cbiZ_EcoRI_5' and Rr_cbiZ_PstI_3', and the resulting product was cloned into the *Eco*RI and *Pst*I sites of plasmid pBAD24 to yield plasmid pCbiZ$_{Rr}$.

**pCbiZ$_{Bh}$**. The *Bacillus halodurans btuD-cbiZ* coding sequence was amplified using primers Bh_btuD_EcoRI_5' and Bh_btuD_PstI_3', and the resulting product was cloned into the *Eco*RI and *Pst*I sites of plasmid pBAD24 to yield plasmid pCbiZ$_{Bh}$.

**pCbiZ$_{Sc}$**. The *Streptomyces coelicolor cbiZ* coding sequence was amplified using primers Sc_cbiZ_NcoI_5' and Sc_cbiZ_HindIII_3', and the resulting product was cloned into the *Nco*I and *Hin*dIII sites of plasmid pBAD24 to yield plasmid pCbiZ$_{Sc}$.

**pCbiZ$_{De}$**. The *Dehalococcoides etheneogenes cbiZ* coding sequence (locus tag DET0242) was amplified using the primer pairs De_0242_KpnI_5' and De_0242_XbaI_3' and the resulting product was cloned into the *Kpn*I and *Xba*I sites of plasmid pBAD24 to yield plasmid pCbiZ$_{De}$.

**pCbiZ$_{Mm}$**. The *Methanosarcina mazei cbiZ* coding sequence was amplified using primers Mm_cbiZ_NcoI_5' and Mm_cbiZ_SphI_3', and the resulting product was cloned into the *Nco*I and *Sph*I sites of plasmid pBAD24 to yield plasmid pCbiZ$_{Mm}$.

**pCbiZ$_{Pf}$**. The *Pyrococcus furiosus cbiZ* coding sequence was amplified using primers Pf_cbiZ_KpnI_5' and Pf_cbiZ_HindIII_3', and the resulting product was cloned into the *Kpn*I and *Hin*dIII sites of plasmid pBAD24 to yield plasmid pCbiZ$_{Pf}$.

**pCbiZ$_{Mj}$**. The *Methanocaldococcus jannaschii cbiZ* coding sequence was amplified using primers Mj_cbiZ_NcoI_5' and Mj_cbiZ_HindIII_3', and the resulting product was cloned into the *Nco*I and *Hin*dIII sites of plasmid pBAD24 to yield plasmid pCbiZ$_{Mj}$. **pCbiZ$_{Fa}$**. The *Ferroplasma acidarmanus cbiZ* coding sequence was amplified using primers Fa_cbiZ_KpnI_5' and Fa_cbiZ_HindIII_3', and the resulting product was cloned into the KpnI and HindIII sites of plasmid pBAD24 to yield plasmid pCbiZ$_{Fa}$.

**pCbiS$_{Ap}$.** The *Aeropyrum pernix cbiS* coding sequence was amplified using primers Ap_cbiS_EcoRI_5' and Ap_cbiS_XbaI_3', and the resulting product was cloned into the *Eco*RI and *Xba*I sites of plasmid pBAD24 to yield plasmid pCbiS$_{Ap}$.

**Construction of a plasmid carrying the *cobP* gene of *R. sphaeroides*.** The *R. sphaeroides cobP* coding sequence plus 205 bp of 5' and 15 bp of 3' sequence was amplified using primers Rs_cobP_EcoRI_5' and Rs_cobP_XbaI_3', and the resulting product was cloned into the *Eco*RI and *Xba*I sites of plasmid pBAD24 to yield plasmid pCobP$_{Rs}$. The identity of the insert was confirmed by sequencing with primers pBAD5' and pBAD3'.

**Preparation of cell-free extracts enriched in CbiZ$_{Rs}$, CobP$_{Rs}$, and CobU$_{Se}$.** The wild-type alleles of the *R. sphaeroides cbiZ* and *cobP* genes and the serovar Typhimurium *cobU* genes were over-expressed using plasmids pCbiZ$_{Rs}$, pCobP$_{Rs}$, and pCobU$_{Se}$, respectively, in serovar Typhimurium strain JE8268 (*cobU ycfN*). Over-production strains were grown overnight in 5 ml of LB broth containing ampicillin, sub-cultured into 400 ml of fresh medium containing ampicillin and 1.5 mM arabinose, and incubated at 37°C with shaking for 6 h. Cells were harvested by centrifugation at 4ºC (15 min at 5,000 x g) in a Beckman-Coulter Avanti J-25I centrifuge. Cells were resuspended in 4 ml of HEPES buffer (50 mM, pH 7.5) containing NaCl (100 mM ) and dithiothreitol (DTT, 5 mM). Cells were broken by sonication (5 min) with a Fisher Scientific Sonic Dismembrator 550 working at half duty. Cell lysates were clarified by centrifugation (30 min at 18,000 x g at 4°C). Soluble proteins were dialyzed against two liters of the re-suspension buffer in a Slidalyzer (Pierce) cassette (MWCO 10,000) with two buffer changes, with 10% (v/v) glycerol added to the buffer for the final dialysis step. Protein extracts were flash frozen in liquid $N_2$ and stored at -80°C until used. As a negative control, the same procedure was also used to prepare protein extract from cells containing the control plasmid pBAD24. Extracts were examined by SDS-PAGE (56) and Coomassie Blue staining (57), but over-expressed proteins were produced at too low an abundance to be seen (data not shown).

**Assay of corrinoid amidohydrolase (CbiZ) activity.** Corrinoid amidohydrolase activity assays were performed as described with slight modifications (9). Briefly, assays were performed in 200 µl volumes

containing 50 µg of protein extract, HEPES buffer (50 mM, pH 7.5), dithiothreitol (5 mM), and corrinoid (30 µM). Reactions were incubated at 30°C for 15 min in the dark and then heat-inactivated at 80°C for 20 min. Precipitated protein was removed by centrifugation at room temperature (5 min @ 16,100 x g). KCN was added to a final concentration of 0.1 M, corrinoids were converted to their cyano forms under strong light for 10 min. Samples containing cyano-corrinoids were filtered using 0.2 µm Spin-X columns (Corning) before they were resolved by HPLC (see below).

**Assay of NTP:adenosylcobinamide kinase / GTP:adenosylcobinamide-phosphate guanylyltransferase activity.** NTP:adenosylcobinamide kinase / GTP:adenosylcobinamide-phosphate guanylyltransferase activity assays were performed as described with slight modifications (13). Briefly, assays were performed in 200 µl volumes containing 25 or 50 µg of protein extract, Tris-HCl (50 mM, pH 8.5), Tris(2-carboxyethyl)-phosphine hydrochloride (1.25 mM, TCEP-HCl), $MgCl_2$ (10 mM), glycerol (680 mM), NaCl (0.1 M), GTP (100 µM), and AdoCbi (30 µM). Reactions were incubated at 30°C for 30 min in the dark followed by heat inactivation at 80°C for 20 min. Precipitated protein was removed by centrifugation at room temperature (5 min @ 16,100 x g). KCN was added to a final concentration of 0.1 M, corrinoids were converted to their cyano forms under strong light for 10 min, samples were filtered using 0.2 µm Spin-X columns (Corning), and corrinoids were resolved by HPLC (see below).

**HPLC analysis.** Corrinoids were separated by HPLC using a modification of the System I of Blanche *et al.* (58). Authentic $(CN)_2$Cby, dicyanocobinamide [$(CN)_2$Cbi], $(CN)_2$Cbi-GDP, and cyanocobalamin (CNCbl) were used as standards. Corrinoids were resolved using a Beckman Coulter System Gold® 126 HPLC system equipped with a Beckman Coulter System Gold® 508 autosampler and a 150 x 4.6 mm Alltima HP C18 AQ column (Alltech). Corrinoids were detected using a photodiode array detector that acquired data in the 200 - 600 nm range. The column was equilibrated at 1 ml $min^{-1}$ with 77% solvent A [potassium phosphate (0.1 M, pH 6.5), KCN (10 mM)] and 23% solvent B [potassium phosphate (50 mM, pH 8), KCN (5 mM), acetonitrile (50% v/v)]. The column was developed with a 43.2 min linear gradient to 47% A / 53% B. A second linear gradient developed the column to 100% B over 5 min, and after 5 min, a third linear gradient returned the column to 23% B over 5 min.

The area under peaks containing CN-corrinoids was integrated and CN-corrinoids were quantified by comparison to a standard curve of authentic $(CN)_2$Cbi subjected to the HPLC protocol described above. The detection limit was 1 pmol of $(CN)_2$Cbi.

The identity of corrinoids was established by bioassay using serovar Typhimurium strains TR6583 (*cob*$^+$), JE8126 (*cbiB*), and JE8312 (*cobU ycfN* / pCOBY38 *cobY*$^+$) as indicator strains. A 2-µl sample of HPLC-purified corrinoid was spotted onto a 0.7% (w/v) agar overlay containing cells of the indicator strain on NCE glycerol minimal medium. As controls, 5 pmol of authentic $(CN)_2$Cby, $(CN)_2$Cbi, and CNCbl were also spotted onto the overlays. Plates were incubated aerobically at 37°C for 24 h. Under aerobic conditions, *de novo* corrin ring biosynthesis is blocked in serovar Typhimurium, making growth dependent on exogenously provided incomplete (e.g., Cbi, Cby) or complete corrinoids (e.g., Cbl) (21). Growth of strain TR6583 indicated the presence of a corrinoid in the extract. In strain JE8126 (*cbiB cobU*$^+$), AdoCbi-P synthesis is blocked, preventing growth on Cby. However, strain JE8126 would grow if Cbi were present in the sample. In contrast, the genetic background of strain JE8312 (*cobU ycfN* / pCOBY38 *cobY*$^+$) would support growth on Cbi, but would support growth on Cby if a functional CbiZ enzyme were made by the cell. In the absence of CbiZ, any growth of strain JE8312 indicated the presence of a corrinoid beyond Cbi in the extract, e.g., Cbl.

**FIGURES**

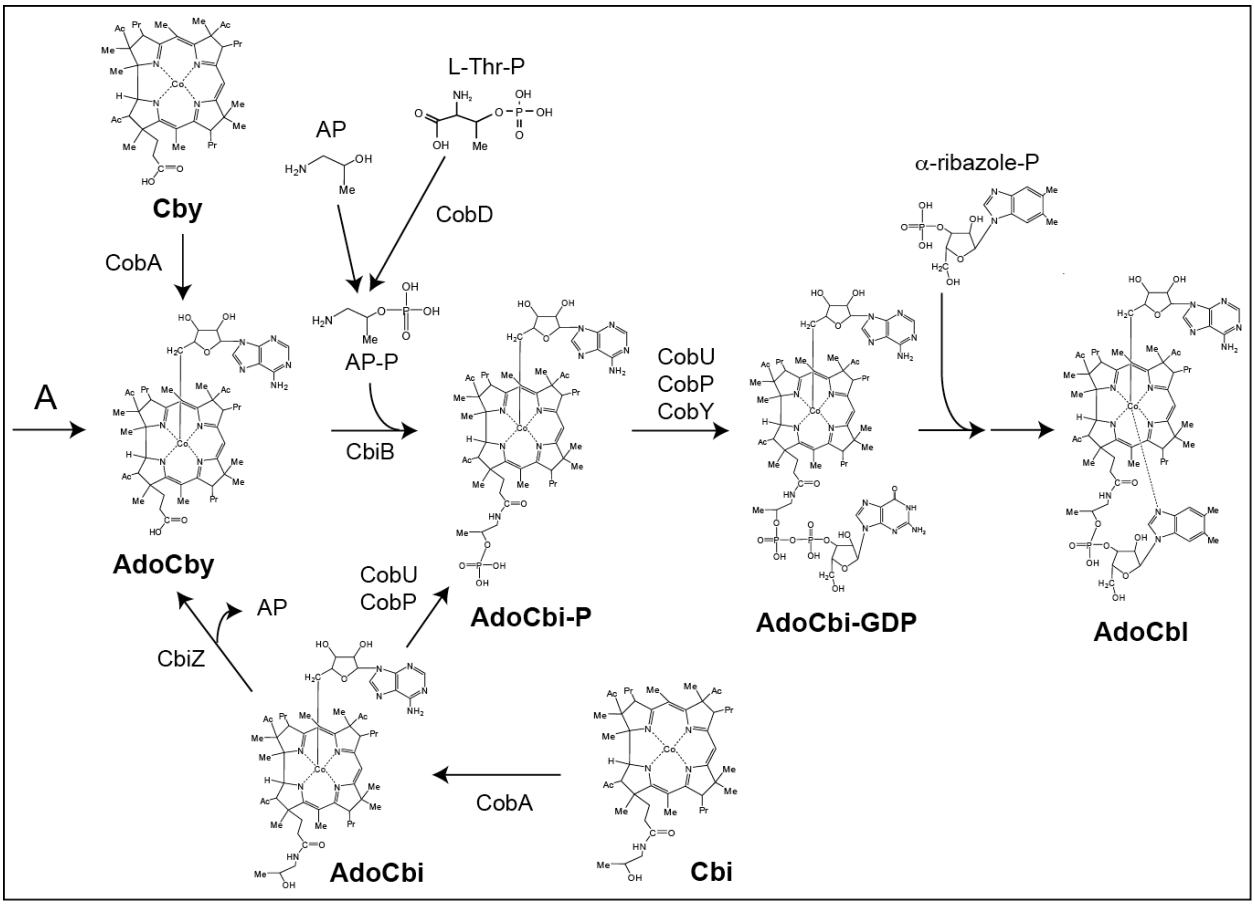



**Figure A.1. Abbreviated view of the cobinamide salvaging pathways in bacteria and archaea.** Corrin ring-containing intermediates are in bold text. The letter **A** indicates the *de novo* corrin ring biosynthesis pathway. Abbreviations: AP, 1-amino-2-propanol; Cby, cobyric acid; AdoCby, adenosylcobyric acid; Cbi, cobinamide; AdoCbi, adenosylcobinamide; AdoCbl, adenosylcobalamin; CbiB, adenosylcobinamide-phosphate synthetase; CobD, threonine-phosphate decarboxylase; CobU and CobP, NTP:adenosylcobinamide kinase, GTP:adenosylcobinamide-phosphate guanylyltransferase; CobY, GTP:adenosylcobinamide-phosphate guanylyltransferase; CbiZ, adenosylcobinamide amidohydrolase, CobA, ATP:corrinoid adenosyltransferase. Functional groups are indicated as follows: Me, methyl; Ac, acetamide; Pr, propionamide.

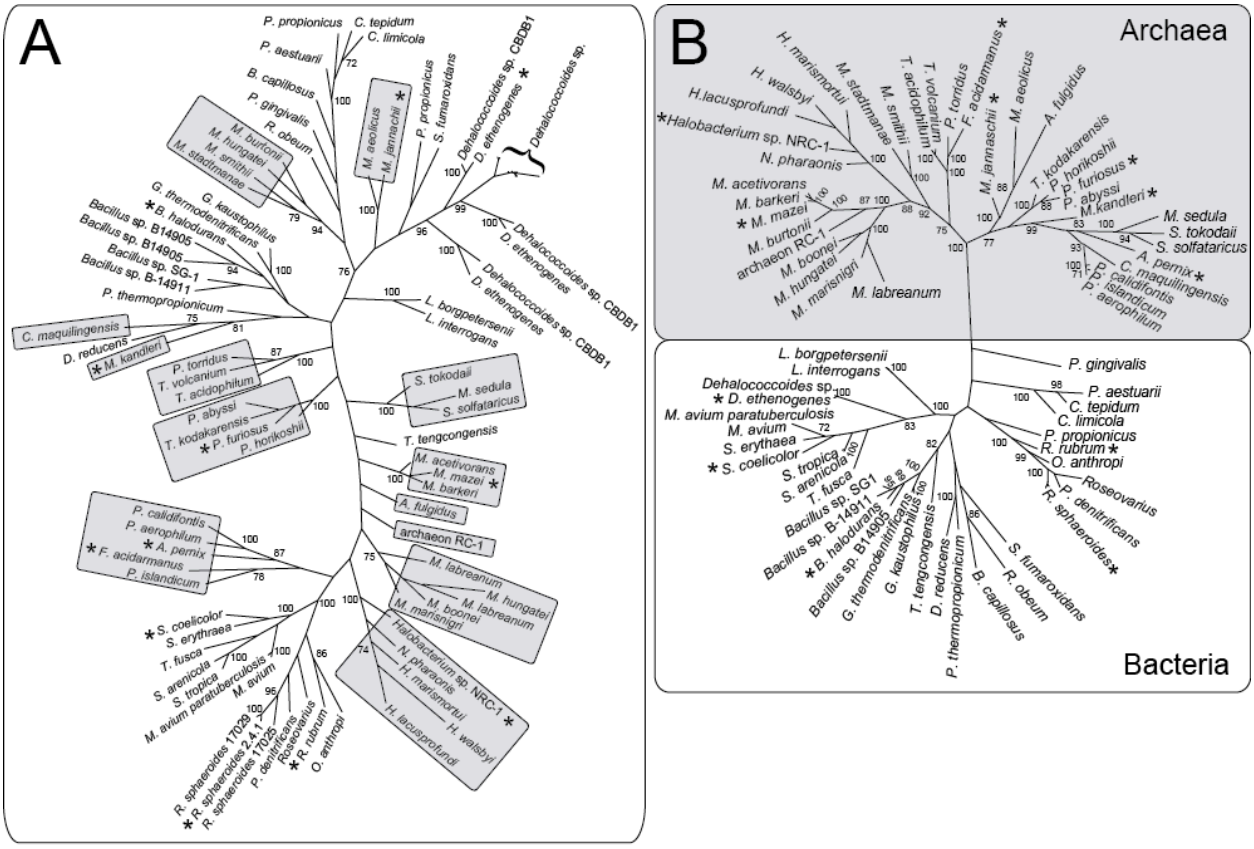



**Figure A.2. Phylogenetic distribution of *cbiZ* orthologues among archaea and bacteria.**

**A. Phylogenetic distribution of *cbiZ* homologues.** Maximum parsimony phylogeny of *cbiZ* homologues, labeled with the species of origin for each gene. Genes from archaeal species are indicated with grey boxes. Nodes with bootstrap support greater than 70% are labeled. Genes whose function have been directly examined in this or previous studies (9, 11) are indicated with an asterix.

**B. 16S rRNA phylogeny of organisms whose genomes contain *cbiZ* homologues.** Maximum parsimony phylogeny of organisms whose genomes contain homologues of *cbiZ*, inferred from the 16S rRNA gene. Nodes with bootstrap support greater than 70% are indicated. Organisms whose *cbiZ* homologues have been directly examined in this or previous studies (9, 11) are indicated with an asterix.

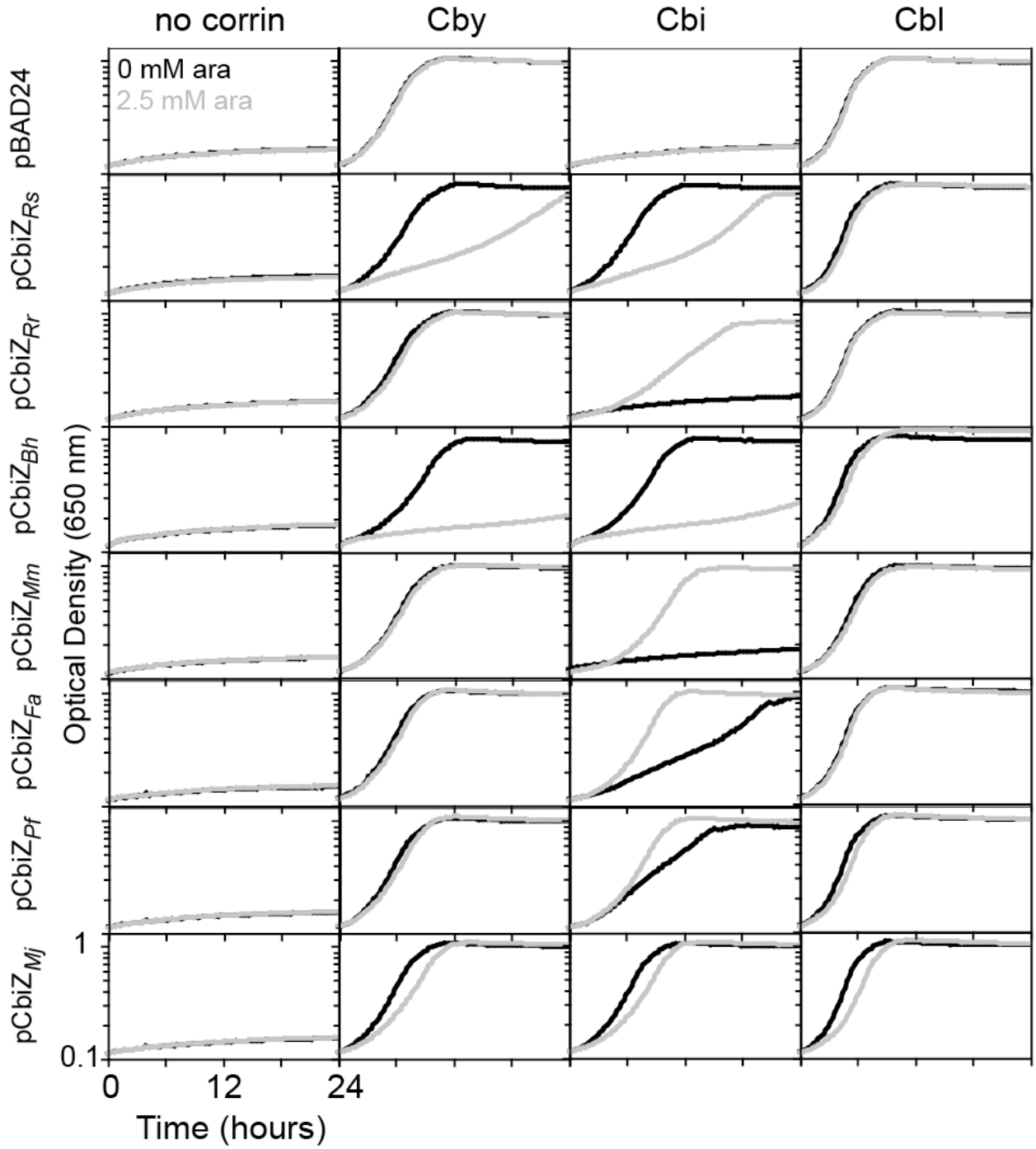



**Figure A.3. Complementation of Cbi salvaging in serovar Typhimurium by *cbiZ* alleles from bacteria and archaea.** Corrinoid dependent aerobic growth of serovar Typhimurium JE8312 (*metE205 ara-9 cob1315* (Δ*cobU*) *ycfN112* (Δ*ycfN*) / pCOBY38 [*cobY*$^{+}$ *kan*$^{+}$]) derivatives in NCE containing glycerol (22 mM ), $MgSO_4$ (1 mM), trace minerals, ampicillin (100 µg µl$^{-1}$), and kanamycin (50 µg µl$^{-}$). Optical density at 650 nm was measured for 24 h at 37°C. Corrinoids were added at 15 nM. Arabinose (2.5 mM) was added as indicated. Plasmids used were: vector, pBAD24; *R. sphaeroides cbiZ*, pCbiZ$_{Rs}$; *R. rubrum cbiZ*, pCbiZ$_{Rr}$; *B. halodurans btuD-cbiZ*, pCbiZ$_{Bh}$; *M. mazei cbiZ*, pCbiZ$_{Mm}$; *M. jannaschii cbiZ*, pCbiZ$_{Mj}$; *P. furiosus cbiZ*, pCbiZ$_{Pf}$; *F. acidarmanus cbiZ*, pCbiZ$_{Fa}$. Growth curves were obtained using an ELx808 Ultra Microplate reader (Bio-Tek Instruments). Each growth curve was performed in triplicate.

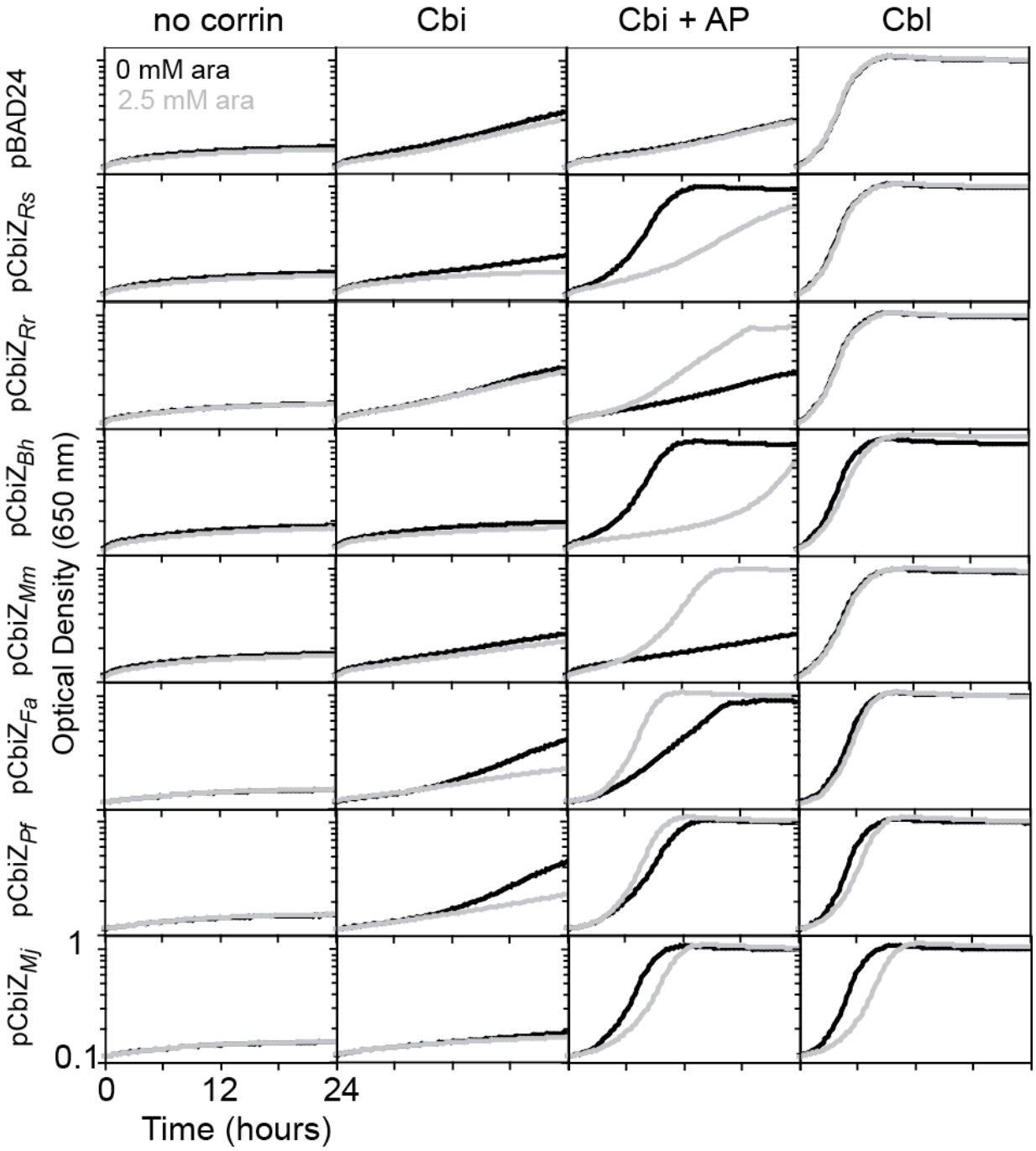



**Figure A.4. CbiZ-dependent Cbi salvaging in serovar Typhimurium requires CbiB function.** Corrinoid dependent aerobic growth of serovar Typhimurium JE7864 (*metE205 ara-9 cbiP236*::Tn*10*d(*tet*$^{+}$) *cobD1302*::Tn*10*d(*cat*$^{+}$) *cobU330* / pCOBY38 [*cobY*$^{+}$ *kan*$^{+}$]) derivatives in NCE containing glycerol (22 mM ), $MgSO_4$ (1 mM), trace minerals, ampicillin (100 µg µl$^{-1}$), and kanamycin (50 µg µl$^{-}$). Optical density at 650 nm was measured for 24 h at 37°C; corrinoids were added at 15 nM. Arabinose (2.5 mM) was added as indicated. 1-Amino-2-propanol (AP, 10 mM) was added as indicated. Plasmids used were: vector, pBAD24; *R. sphaeroides cbiZ*, pCbiZ$_{Rs}$; *R. rubrum cbiZ*, pCbiZ$_{Rr}$; *B. halodurans btuD-cbiZ*, pCbiZ$_{Bh}$; *M. mazei cbiZ*, pCbiZ$_{Mm}$; *M. jannaschii cbiZ*, pCbiZ$_{Mj}$; *P. furiosus cbiZ*, pCbiZ$_{Pf}$; *F. acidarmanus cbiZ*, pCbiZ$_{Fa}$. Growth curves were obtained using an ELx808 Ultra Microplate reader (Bio-Tek Instruments). Each growth curve was performed in triplicate.

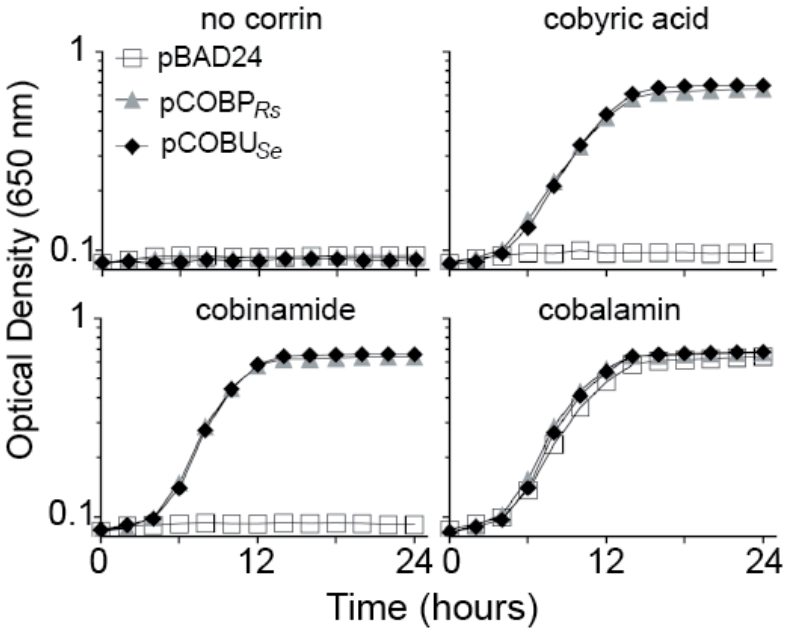



**Figure A.5. *R. sphaeroides cobP* complements both the cobinamide kinase and cobinamide-phosphate guanylyltransferase defects of an serovar Typhimurium *cobU* mutant**. Corrinoid dependent aerobic growth of serovar Typhimurium JE0824 (*metE205 ara-9 cbiP*236::Tn*10*d(*tet*$^{+}$) *cobU330*) derivatives at 37°C in NCE medium containing glycerol (22 mM), $MgSO_4$ (1 mM), and ampicillin (100 μg μl$^{-1}$). When added, corrinoids were present at 15 nM. Plasmids used were: vector, pBAD24; *R. sphaeroides cobP*, pCobP$_{Rs}$; serovar Typhimurium *cobU*, pCobU$_{Se}$. Error bars represent one standard deviation. Growth curves were obtained using an ELx808 Ultra Microplate reader (Bio-Tek Instruments). Each growth curve was performed in triplicate.

**REFERENCES**


1. **Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, Bateman A, Brown S, Chandra G, Chen CW, Collins M, Cronin A, Fraser A, Goble A, Hidalgo J, Hornsby T, Howarth S, Huang CH, Kieser T, Larke L, Murphy L, Oliver K, O'Neil S, Rabbinowitsch E, Rajandream MA, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek A, Woodward J, Barrell BG, Parkhill J, Hopwood DA.** 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature **417:**141-147.

2. **Roessner CA, Scott AI.** 2006. Fine-tuning our knowledge of the anaerobic route to cobalamin (vitamin $B_{12}$). J Bacteriol **188:**7331-7334.

3. **Escalante-Semerena JC.** 2007. Conversion of cobinamide into adenosylcobamide in bacteria and archaea. J Bacteriol **189:**4555-4560.

4. **Blanche F, Cameron B, Crouzet J, Debussche L, Thibaut D, Vuilhorgne M, Leeper FJ, Battersby AR.** 1995. Vitamin $B_{12}$: How the problem of its biosynthesis was solved. Angew Chem Int Ed Engl **34:**383-411.

5. **Brindley AA, Raux E, Leech HK, Schubert HL, Warren MJ.** 2003. A story of chelatase evolution: identification and characterization of a small 13-15-kDa "ancestral" cobaltochelatase (CbiXS) in the archaea. J Biol Chem **278:**22388-22395.

6. **Thomas MG, Escalante-Semerena JC.** 2000. Identification of an alternative nucleoside triphosphate: 5'- deoxyadenosylcobinamide phosphate nucleotidyltransferase in *Methanobacterium thermoautotrophicum* ΔH. J Bacteriol **182:**4227-4233.

7. **Woodson JD, Peck RF, Krebs MP, Escalante-Semerena JC.** 2003. The *cobY* gene of the archaeon *Halobacterium* sp. strain NRC-1 is required for de novo cobamide synthesis. J Bacteriol **185:**311-316.

8. **Woodson JD, Zayas CL, Escalante-Semerena JC.** 2003. A new pathway for salvaging the coenzyme $B_{12}$ precursor cobinamide in archaea requires cobinamide-phosphate synthase (CbiB) enzyme activity. J Bacteriol **185:**7193-7201.

9. **Woodson JD, Escalante-Semerena JC.** 2004. CbiZ, an amidohydrolase enzyme required for salvaging the coenzyme $B_{12}$ precursor cobinamide in archaea. Proc Natl Acad Sci USA **101:**3591-3596.

10. **Zayas CL, Woodson JD, Escalante-Semerena JC.** 2006. The *cobZ* gene of *Methanosarcina mazei* Gö1 encodes the nonorthologous replacement of the α-ribazole-5'-phosphate phosphatase (CobC) enzyme of *Salmonella enterica*. J Bacteriol **188:**2740-2743.

11. **Woodson JD, Escalante-Semerena JC.** 2006. The *cbiS* gene of the archaeon *Methanopyrus kandleri* AV19 encodes a bifunctional enzyme with adenosylcobinamide amidohydrolase and alpha-ribazole-phosphate phosphatase activities. J Bacteriol **188:**4227-4235.

12. **Brushaber KR, O'Toole GA, Escalante-Semerena JC.** 1998. CobD, a novel enzyme with L-threonine-*O*-3-phosphate decarboxylase activity, is responsible for the synthesis of (*R*)-1-amino-2-propanol *O*-2-phosphate, a proposed new intermediate in cobalamin biosynthesis in *Salmonella typhimurium* LT2. J Biol Chem **273:**2684-2691.

13. **Thomas MG, Thompson TB, Rayment I, Escalante-Semerena JC.** 2000. Analysis of the adenosylcobinamide kinase / adenosylcobinamide phosphate guanylyltransferase (CobU) enzyme of *Salmonella typhimurium* LT2. Identification of residue H46 as the site of guanylylation. J Biol Chem **275:**27376-27386.

14. **Fonseca MV, Escalante-Semerena JC.** 2000. Reduction of cob(III)alamin to cob(II)alamin in *Salmonella enterica* Serovar Typhimurium LT2. J Bacteriol **182:**4304-4309.

15. **Fonseca MV, Buan NR, Horswill AR, Rayment I, Escalante-Semerena JC.** 2002. The ATP:co(I)rrinoid adenosyltransferase (CobA) enzyme of *Salmonella enterica* requires the 2'-OH Group of ATP for function and yields inorganic triphosphate as its reaction byproduct. J Biol Chem **277:**33127-33131.

16. **O'Toole GA, Escalante-Semerena JC.** 1995. Purification and characterization of the bifunctional CobU enzyme of *Salmonella typhimurium* LT2. Evidence for a CobU-GMP intermediate. J Biol Chem **270:**23560-23569.

17. **Blanche F, Debussche L, Famechon A, Thibaut D, Cameron B, Crouzet J.** 1991. A bifunctional protein from *Pseudomonas denitrificans* carries cobinamide kinase and cobinamide phosphate guanylyltransferase activities. J Bacteriol **173:**6052-6057.

18. **Markowitz VM, Korzeniewski F, Palaniappan K, Szeto E, Werner G, Padki A, Zhao X, Dubchak I, Hugenholtz P, Anderson I, Lykidis A, Mavromatis K, Ivanova N, Kyrpides NC.** 2006. The integrated microbial genomes (IMG) system. Nucleic Acids Res **34:**D344-348.

19. **Mackenzie C, Eraso JM, Choudhary M, Roh JH, Zeng X, Bruscella P, Puskas A, Kaplan S.** 2007. Postgenomic adventures with *Rhodobacter sphaeroides*. Annu Rev Microbiol **61:**283-307.

20. **Felsenstein J.** 2005. PHYLIP (Phylogeny Inference Package) version 3.6 *Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.*

21. **Jeter RM, Olivera BM, Roth JR.** 1984. *Salmonella typhimurium* synthesizes cobalamin (vitamin $B_{12}$) de novo under anaerobic growth conditions. J Bacteriol **159:**206-213.

22. **Guzman LM, Belin D, Carson MJ, Beckwith J.** 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J Bacteriol **177:**4121-4130.

23. **Otte MM, Woodson JD, Escalante-Semerena JC.** 2007. The thiamine kinase (YcfN) enzyme plays a minor but significant role in cobinamide salvaging in *Salmonella enterica*. J Bacteriol **189:**7310-7315.

24. **Grabau C, Roth JR.** 1992. A *Salmonella typhimurium* cobalamin-deficient mutant blocked in 1-amino-2-propanol synthesis. J Bacteriol **174:**2138-2144.

25. **Chowdhury S, Thomas MG, Escalante-Semerena JC, Banerjee R.** 2001. The Coenzyme B12 Analog 5'-Deoxyadenosylcobinamide-GDP Supports Catalysis by Methylmalonyl-CoA Mutase in the Absence of Trans-ligand Coordination. J Biol Chem **276:**1015-1019.

26. **Nascimento AL, Verjovski-Almeida S, Van Sluys MA, Monteiro-Vitorello CB, Camargo LE, Digiampietri LA, Harstkeerl RA, Ho PL, Marques MV, Oliveira MC, Setubal JC, Haake DA, Martins EA.** 2004. Genome features of Leptospira interrogans serovar Copenhageni. Braz J Med Biol Res **37:**459-477.

27. **Nelson KE, Fleischmann RD, DeBoy RT, Paulsen IT, Fouts DE, Eisen JA, Daugherty SC, Dodson RJ, Durkin AS, Gwinn M, Haft DH, Kolonay JF, Nelson WC, Mason T, Tallon L, Gray J, Granger D, Tettelin H, Dong H, Galvin JL, Duncan MJ, Dewhirst FE, Fraser CM.** 2003. Complete genome sequence of the oral pathogenic bacterium *Porphyromonas gingivalis* strain W83. J Bacteriol **185:**5591-6601.

28. **Zuo Y, Xing D, Regan JM, Logan BE.** 2008. Isolation of the exoelectrogenic bacterium *Ochrobactrum anthropi* YZ-1 by using a U-tube microbial fuel cell. Appl Environ Microbiol **74:**3130-3137.

29. **Song S, Ahn JK, Lee GH, Park YG.** 2007. An epidemic of chronic pseudophakic endophthalmitis due to *Ochrobacterium anthropi*: clinical findings and managements of nine consecutive cases. Ocular Imm and Inflam **15:**429-437.

30. **Li L, Bannantine JP, Zhang Q, Amonsin A, May BJ, Alt D, Banerji N, Kanjilal S, Kapur V.** 2005. The complete genome sequence of *Mycobacterium avium* subspecies paratuberculosis. Proc Natl Acad Sci U S A **102:**12344-12349.

31. **Takami H, Takaki Y, Chee GJ, Nishi S, Shimamura S, Suzuki H, Matsui S, Uchiyama I.** 2004. Thermoadaptation trait revealed by the genome sequence of thermophilic *Geobacillus kaustophilus*. Nucleic Acids Res **32:**6292-6303.

32. **Takami H, Nakasone K, Takaki Y, Maeno G, Sasaki R, Masui N, Fuji F, Hirama C, Nakamura Y, Ogasawara N, Kuhara S, Horikoshi K.** 2000. Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*. Nucleic Acids Res **28:**4317-4331.

33. **Kube M, Beck A, Zinder SH, Kuhl H, Reinhardt R, Adrian L.** 2005. Genome sequence of the chlorinated compound-respiring bacterium Dehalococcoides species strain CBDB1. Nat Biotechnol **23:**1269-1273.

34. **Seshadri R, Adrian L, Fouts DE, Eisen JA, Phillippy AM, Methe BA, Ward NL, Nelson WC, Deboy RT, Khouri HM, Kolonay JF, Dodson RJ, Daugherty SC, Brinkac LM, Sullivan SA, Madupu R, Nelson KE, Kang KH, Impraim M, Tran K, Robinson JM, Forberger HA, Fraser CM, Zinder SH, Heidelberg JF.** 2005. Genome sequence of the PCE-dechlorinating bacterium *Dehalococcoides ethenogenes*. Science **307:**105-108.

35. **Cherezov V, Yamashita E, Liu W, Zhalnina M, Cramer WA, Caffrey M.** 2006. In meso structure of the cobalamin transporter, BtuB, at 1.95 Å resolution. J Mol Biol **364:**716-734.

36. **Kandt C, Xu Z, Tieleman DP.** 2006. Opening and closing motions in the periplasmic vitamin $B_{12}$ binding protein BtuF. Biochemistry **45:**13284-13292.

37. **Karpowich NK, Huang HH, Smith PC, Hunt JF.** 2003. Crystal structures of the BtuF periplasmic-binding protein for vitamin B12 suggest a functionally important reduction in protein mobility upon ligand binding. J Biol Chem **278:**8429-8434.

38. **Hvorup RN, Goetz BA, Niederer M, Hollenstein K, Perozo E, Locher KP.** 2007. Asymmetry in the structure of the ABC transporter binding protein complex BtuCD-BtuF. Science **317:**1387-1390.

39. **Borths EL, Poolman B, Hvorup RN, Locher KP, Rees DC.** 2005. In vitro functional characterization of BtuCD-F, the *Escherichia coli* ABC transporter for vitamin B12 uptake. Biochemistry **44:**16301-16319.

40. **Harmsen HJ, Van Kuijk BL, Plugge CM, Akkermans AD, De Vos WM, Stams AJ.** 1998. *Syntrophobacter fumaroxidans* sp. nov., a syntrophic propionate-degrading sulfate-reducing bacterium. Int J Syst Bacteriol **48 Pt 4:**1383-1387.

41. **Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, Blake JA, FitzGerald LM, Clayton RA, Gocayne JD, Kerlavage AR, Dougherty BA, Tomb JF, Adams MD, Reich CI, Overbeek R, Kirkness EF, Weinstock KG, Merrick JM, Glodek A, Scott JL, Geoghagen NS, Venter JC.** 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. Science **273:**1058-1073.

42. **Jones WJ, Leigh JA, Mayer F, Woese CR, Wolfe RS.** 1983. *Methanococcus jannaschii* sp. nov., an extremely thermophilic methanogen from a submarine hydrothermal vent. Arch Microbiol **136:**254-261.

43. **Renz P.** 1999. Biosynthesis of the 5,6-dimethylbenzimidazole moiety of cobalamin and of other bases found in natural corrinoids, p 557-575. *In* Banerjee R (ed), Chemistry and Biochemistry of B12. John Wiley & Sons, Inc., New York.

44. **Lengyel P, Mazumder R, Ochoa S.** 1960. Mammalian Methylmalonyl Isomerase and Vitamin B(12) Coenzymes. Proc Natl Acad Sci U S A **46:**1312-1318.

45. **Barker HA, Smyth RD, Weissbach H, Toohey JI, Ladd JN, Volcani BE.** 1960. Isolation and properties of crystalline cobamide coenzymes containing benzimidazole or 5, 6-dimethylbenzimidazole. J Biol Chem **235:**480-488.

46. **Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer EL, Bateman A.** 2006. Pfam: clans, web tools and services. Nucleic Acids Res **34:**D247-251.

47. **DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL.** 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol **72:**5069-5072.

48. **Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG.** 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res **25:**4876-4882.

49. **Bertani G.** 1951. Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. J Bacteriol **62:**293-300.

50. **Bertani G.** 2004. Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. J Bacteriol **186:**595-600.

51. **Berkowitz D, Hushon JM, Whitfield HJ, Jr., Roth J, Ames BN.** 1968. Procedure for identifying nonsense mutations. J Bacteriol **96:**215-220.

52. **Balch WE, Wolfe RS.** 1976. New approach to the cultivation of methanogenic bacteria: 2-mercaptoethanesulfonic acid (HS-CoM)-dependent growth of *Methanobacterium ruminantium* in a pressurized atmosphere. Appl Environ Microbiol **32:**781-791.

53. **Ausubel FA, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl.** 1989. Current protocols in molecular biology. Greene Publishing Associates & Wiley Interscience, New York, N.Y.

54. **Raleigh EA, Lech K, Brent R.** 1989. Selected topics from classical bacterial genetics, p 1.4. *In* Ausubel FA, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (ed), Current Protocols in Molecular Biology, vol 1. Wiley Interscience, New York.

55. **Woodcock DM, Crowther PJ, Doherty J, Jefferson S, De Cruz E, Noyer-Weidner M, Smith SS, Michael MZ, Graham MW.** 1989. Quantitative evaluation of *Escherichia coli* host strains

for tolerance to cytosine methylation in plasmid and phage recombinants. Nucl Acids Res **17:**3469-3478.

56. **Laemmli UK.** 1970. Cleavage and structural proteins during the assembly of the head of bacteriophage T4. Nature **227:**680-685.
57. **Sasse J.** 1991. Detection of proteins, p 10.16.11-10.16.18. *In* Ausubel FA, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (ed), Current Protocols in Molecular Biology, vol 1. Wiley Interscience, New York.

58. **Blanche F, Thibaut D, Couder M, Muller JC.** 1990. Identification and quantitation of corrinoid precursors of cobalamin from *Pseudomonas denitrificans* by high-performance liquid chromatography. Anal Biochem **189:**24-29.