Methylation reactions in the anaerobic biosynthesis of the lower ligand of Vitamin B_{12}

A thesis
Submitted in partial fulfilments of the requirements

Of the degree of
Doctor of Philosophy

By

Yamini Mathur
20152020

Indian Institute of Science Education and Research Pune

2022

CERTIFICATE

Certified that the work incorporated in the thesis entitled 'Methylation reactions in the anaerobic biosynthesis of the lower ligand of Vitamin B_{12}' submitted by Yamini Mathur was carried out by the candidate, under my supervision. The
work presented here or any part of it has not been included in any other thesis submitted previously for the award of any degree or diploma from any other University or institution.

Date: 5th September 2022

Dr. Amrita B. Hazra
(Signature of the supervisor)

DECLARATION

I declare that this written submission represents my ideas in my own words and where others' ideas have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Date: 5th September 2022

Yamini Mathur
20152020

Dedication

To my grandma, Shanti Devi Mathur.

In celebration of her lifelong perseverance to secure higher education for her daughters and granddaughters.

Acknowledgements

After the completion of my BSc degree, I right away joined IISER Pune for the integrated PhD program with little idea of what research really looks like on a day-to-day basis. Over the last seven years, IISER Pune community has provided me with the supportive environment to grow as a scholar. The journey wasn't the easiest, but it was a lot of fun!

First and foremost, I thank my thesis supervisor and mentor, Dr. Amrita Hazra. Our shared interest in metabolism, drawing out the complex pathways from memory, and of course fixing HPLC and glove bag together make us a strong team. On several tough parts of the project, our complementary prior training and perspectives led us to interesting discoveries and interpretations. I truly admire her problem-solving attitude in science and life, and I know I can always count on her advice. I am grateful to her for believing in me and helping me navigate through opportunities to learn new skills and share my work at global platforms.

My RAC members, Prof. Ruchi Anand (IIT Bombay) and Prof. Thomas Pucadyil (IISER Pune) have greatly contributed in the form of tips on troubleshooting, feedbacks on hypotheses and manuscripts, and finding my next steps in career. I will always be grateful for their constant encouragement and support.

Our labs, the C-125A in a nook of the chemistry department and the SFB-310 facing the gardens in the biology department, became my happy places because of the people with whom I shared these spaces. Our group always had a perfect mix of people with training and interests in different aspects of biology and chemistry which made research in the realm of biological chemistry extremely interesting. I would like to extend special thanks to Yashwant Kumar and Mrutyunjay Nair for sharing reagents, waiting for me during late night experiments, sharing tea and pizza, and discussing science- from lab to life as we see it. Rupali Sathe and I have been helping each other (literally) walk the balance between biology and chemistry. We have unboxed so many labware together and helped each other re-invent the wheel whenever necessary. I will always welcome her feedback and tips & tricks. I would also like to thank Ateek Shah for providing valuable chemistry and structural biology insights into the project. Undergraduate researchers Saswata Nayak and Lavisha Parab have helped me think clearer on communicating the project with their out-of-the-box questions.

My PhD thesis has significantly benefited from inputs from excellent BSMS and iPhD students at IISER Pune. Prathamesh Datar had helped to establish the routine protocols and build a foundation for the B_{12} project. Thereafter, Manjima Sathian, Abhinav Masih, Bhushan Kulkarni, Chinmaya Jena, and P Riya have contributed to various aspects of the project. I would especially like to mention the contributions by Sheryl Sreyas. I could not have asked for a better project partner for steamrolling to celebrating every little step forward. I also thank Vruta Gupte

and Aditya Bhattacharyya, for giving me a chance to explore more about my favourite cofactor - with B_{12} and SAM projects running, we have got the Batman and Robin of biochemistry with us! I am thankful to Aniket Vartak, for being an amazing colleague. His enthusiasm and new perspectives on the project give me high hopes and I am certain that the project will be in good hands. I am grateful to the Vibhor Gaikwad, S Rohan, and Shweta Nikhar for asking smart questions and keeping up a fun lab environment.

I sincerely thank Dr. Gayathri Pananghat for always helping us with troubleshooting, and also for trusting me with their brand new HPLC when ours was retired hurt; Dr. Nishad Matange introduced me to the concept of molecular evolution of proteins, which served as an important turning point for the project as well as allowed us to keep the project afloat during the pandemic. I'd also like to thank Dr Siddhesh Kamat for help with standardizing mass spectrometry methods; Dr. Harinath Chakrapani and his lab for reagents and discussions; Dr. Sudha Rajamani for sharing their HPLC and her encouraging words; Dr. Anand Krishnan for discussions on phylogenetic trees and undergraduate teaching; Dr Jeet Kalia, with whom I did my first lab rotation, for letting me try various parts of the project and encouraging me to ask stupid questions. I also thank Dr. Thomas Pucadyil and Dr. Nagaraj Balasubramanian for wonderful lab rotation projects that built the foundation for everything I have done so far and will do in the future. I am indebted to my mentor from undergraduate college, Dr. Aman Verma, for planting the interest in biochemistry and showing me the road to the research world.

I am fortunate to have worked with excellent former PhD students and postdoctoral fellows at IISER Pune. Debayan Sarkar has been being an amazing mentor and friend who patiently taught me the basics of experimental routines. Gregor Jose has helped me navigate through the frustrations of PhD and reminded me to celebrate little victories. I also thank Srishti Dar, Manish Kushwah, Siddhi Inchanalkar, Neha Deshpande, and Archana Pawar for providing a thorough training during my lab rotations. I also thank Mohan Reddy, Jyoti Baranwal, Srikant Harne, and Kunalika Jain, for help with various experimental methods and providing training on various instruments. I am grateful to Randhir Dhamdhere from Agilent Technologies for teaching the nitty-gritties of the HPLC instrumentation and maintenance; and Akanksha Singh from ScieX for providing the training on the LC-MS. My thesis work has greatly benefitted from institute facilities for LC-MS, NMR and I thank Saddam Sheikh and Nitin Dalvi for keeping them up-and-running. I would like to thank the administration and accounts staff at IISER Pune for their help with fellowships related matters.

I sincerely thank Prof. Michi Taga and her group at UC Berkeley, where the project was born, for their continued support with discussions and reagents. I am especially thankful to Kenny, Amanda, and Olga for sharing protocols, plasmids, and data all through the years. It is always a delight to share our progress with Taga lab and I'd miss receiving Michi's

knowledgeable insights. Also, thanks to Olga and Kris for welcoming me into their home during my travel, showing me around Berkeley, and introducing me to boba tea!

I sincerely thank Prof. Martin Warren and his group at University of Kent for welcoming me into their lab amidst a pandemic. Dr. Evelyne Deery taught me the cyanide-free cobamide purification which has opened new research avenues for our lab. I would always cherish our discussions and I truly admire how she has answers to every question related to B_{12}.

My friends have made sure I have plenty of good memories that I'll take from Pune. Aishwarya Venkataravi has been my support system and a buddy in all shenanigans, celebrations, and arts-n-craft throughout past seven years. Sagnik Ghosh has been enlightening my life with discussions and facts about everything under the sun and even beyond. Yashwant has always entertained my endless questions, unfunny jokes, and I am grateful to him for helping me face my fears- like the 18 hours long protein preps! My friends from college – Tanay Kingrani, Vertika Singh, Priyaneet Kaur, and Yasha Singh have been motivating me with their long-distance support.

I fail to gather words to describe the unwavering support from my parents, Manju and Kaushal, and my brother Vinayak. They inspire me with their love and I respect their efforts towards learning along with me. I ought to mention my dear cousins Divya, Priyanka, Pragya, Devesh, and Anurag for memorable road trips and good food!!

Lastly, I extent my gratitude to the various funding sources - IISER Pune for funding during the two years of MS program, CSIR for funding throughout the duration of my PhD. My visit to Prof Martin Warren's lab at University of Kent, Canterbury, UK was facilitated by the Newton Bhabha PhD placement program funded by Department of Biotechnology and the British Council. The National Science Foundation award to Amrita funded my travel and attendance at the West Coast Bacterial Physiologist (WCBP) meeting at Asilomar, California in 2017. CSIR-JRF contingency allowed me to attend the ICP-MS and ICP-AES training held at IIT Bombay in 2018 and the conference on Conflict and Cooperation in Cellular Populations held at inSTEM Bangalore in 2020. I received substantial financial aid for attending Gordon Research Conference on Chemistry and Biology of Tetrapyrroles 2022. I am grateful to Amrita's DBT Ramalingaswami reentry fellowship, and the DST SERB international travel scheme and the organizers of GRC for the funds. I also thank the various grants awarded to Amrita for supporting our research.

Sincerely,
Yamini Mathur
September 5th, 2022

List of Publications

Research article

CobT and BzaC catalyze the regiospecific activation and methylation of the 5-hydroxybenzimidazole lower ligand in anaerobic cobamide biosynthesis
Yamini Mathur, Sheryl Sreyas, Prathamesh M. Datar, Manjima B. Sathian, Amrita B. Hazra
Journal of Biological Chemistry 295 (31), 10522-10534 (**2020**)
DOI: https://doi.org/10.1074/jbc.RA120.014197 | Full text link
The article was selected as an Editor's pick, and is a part of a virtual issue 'Women in biological chemistry' by the Journal of Biological Chemistry

Commentary

Double agent indole-3-acetic acid: mechanistic analysis of indole-3-acetaldehyde dehydrogenase AldA that synthesizes IAA, an auxin that aids bacterial virulence
Ateek Shah[1], **Yamini Mathur[1]**, Amrita B. Hazra [[1] Shared first author]
Bioscience Reports 41 (8), (**2021**)
DOI: https://doi.org/10.1042/BSR20210598 | Full text link

Book chapter

Guardian of cobamide diversity: Probing CobT's role in lower ligand activation in the biosynthesis of Vitamin B_{12} and other cobamide cofactors
Yamini Mathur, Aniket R. Vartak, Amrita B. Hazra
Methods in Enzymology: B12 enzymes (668), *in press* (**2022**)
DOI: https://doi.org/10.1016/bs.mie.2022.01.001

Review

Methylations in B_{12} biosynthesis and catalysis
Yamini Mathur, Amrita B. Hazra
Current Opinion in Structural Biology 77:102490 (**2022**)
DOI: https://doi.org/10.1016/j.sbi.2022.102490

Research article (manuscript in preparation)
Functional differences in CobT homologs from varying gene neighborhoods
Yamini Mathur[1], Sheryl Sreyas[1], Amrita B. Hazra [[1] Shared first author]

Synopsis

Title: Methylations in the anaerobic biosynthesis of the lower ligand of vitamin B_{12}

Name: Yamini Mathur

Registration number: 20152020

Name of the thesis advisor: Dr. Amrita B. Hazra

Date of registration: 1st August 2017

Indian Institute of Science Education and Research Pune

Introduction

The vitamin B_{12} is a member of cobamide family of cofactors which are characterized by presence of a conserved corrin ring which coordinates with a cobalt ion, a variable and catalytically active upper ligand of cobalt ion, and a variable nucleotide loop that covalently host a lower ligand moiety (Roth, Lawrence, and Bobik 1996; Kennedy and Taga 2020). The lower ligand defines the identity of the cobamide and is chemically a derivative of benzimidazole, purine, or phenol. B_{12} is a cobamide with 5,6-dimethylbenzimidazole (DMB) as the lower ligand. In nature, only a subset of bacteria and archaea can make cobamides and the other organisms that require cobamide cofactors rely on obtaining it as a nutrient (Rodionov et al. 2003; Sokolovskaya, Shelton, and Taga 2020).

The biosynthesis of B_{12} and other cobamides is a modular pathway wherein the three individual components- the corrin ring, the upper ligand, and the lower ligand is synthesized independently and the final structure is assembled in the later stages of pathway, called the nucleotide loop assembly pathway (Mattes et al. 2017). The biosynthetic routes for corrin ring and the DMB lower ligand can either employ an oxygen-dependent pathway (aerobic) or proceed *via* the oxygen-independent (anaerobic) pathway. To date, the nucleotide loop assembly pathway is proposed to be identical in the aerobic and anaerobic pathway.

In my thesis, I study the anaerobic biosynthesis of DMB, the ligand of vitamin B_{12} and explore the initial reactions involved in the anaerobic nucleotide loop assembly. The anaerobic pathway for DMB biosynthesis was a longstanding question in the field of B_{12} biosynthesis (Renz 1993). In 2015, the discovery of the *bza* operon elucidated the genetic basis for how

anaerobic bacteria make DMB and three other benzimidazole derivatives which also exist as lower ligand in naturally occurring cobamides (Hazra et al. 2015). The *bza* operon in B_{12} producing obligate anaerobe *Eubacterium limosum* contains the *bzaA-bzaB-cobT-bzaC-bzaD-bzaE* genes regulated under a cobalamin riboswitch. Following the discovery of the pathway, the biochemical characterization of HBI synthase coded by the gene *bzaF*, a single gene homolog *of bzaA-bzaB* genes, led to thorough mechanistic analysis for rearrangement of AIR into 5-OHBza, the first dedicated intermediate in the pathway (Mehta et al. 2015; Gagnon et al. 2018). The biochemical nature and mechanism of the next enzymes of the pathway- the gene products of *cobT*, *bzaC*, *bzaD*, and *bzaE* remained to discovered. The CobT class of enzyme are α-phosphoribosyltransferases which were previously known to catalyze activation of lower ligands prior during the nucleotide loop assembly (Trzebiatowski and Escalante-semerena 1997; Zayas and Escalante-Semerena 2007). The gene product of the *bzaC* gene was proposed to be an O-methyltransferases, and the *bzaD*, and *bzaE* genes were proposed to code for two B_{12}-radical-SAM enzymes (Hazra et al. 2015). As per the predicted pathway, the BzaC enzyme catalyzes O-methylation using 5-hydroxybenzimidazole (5-OHBza) as substrate resulting in 5-methoxybenzimidazole (5-OMeBza) which then undergoes a C-methylation by BzaD to make 5-methoxy-6-methylbenzimidazole (5-OMe-6MeBza), and lastly BzaE catalyzes an unprecedented reaction involving removal of the methoxy group and installing a methyl group at C5 resulting in DMB. The benzimidazoles are rare biomolecules primarily found as cobamide lower ligands, and there are no reported reactions in literature which are equivalent to those catalyzed by BzaC, BzaD, and BzaE. We thus explore the function of CobT in the *bza* operon pathway and conduct the primary characterization of novel methyltransferases- BzaC, BzaD, and BzaE. Our experiments with these enzymes also lead us to interesting discoveries such as the uncanny presence of functional CobT homologs in eukaryotic microorganisms, the involvement of a domain of unknown function called DUF2284 *bza* operon from several anaerobes, and we also uncover the large diversity in the composition of *bza* operon genes across anaerobic bacteria- which may lead us to finding how bacteria make other benzimidazolyl lower ligands and ultimately help build tools to predict cobamide diversity in metagenomic data.

A summary of the chapters of this thesis describing our investigations of the enzymes of *bza* operon are as follows:

Chapter 1: Methylations in vitamin B12 biosynthesis and catalysis

In this chapter, we introduce the structure, function, and biosynthesis of vitamin B_{12} and its naturally occurring analogs called cobamides. We observe that the structure of B_{12} is decorated

with several methyl groups on individual structural components and we describe the origin and function of these methyl groups. We discuss the details of discovery and proposed mechanism for biosynthesis of DMB in bacteria through oxygen-dependent and oxygen-sensitive routes. Lastly, we present the following open questions pertaining to the *bza* operon enzymes CobT, BzaC, BzaD, and BzaE involved in the anaerobic biosynthesis of DMB:

a. What is the precise function of cobT in the bza operon?

Previous literature show that CobT activates the lower ligands prior to the cobamide assembly(Trzebiatowski and Escalante-semerena 1997; Zayas and Escalante-Semerena 2007). However, CobT homologs accommodate a wide range of substrates with varying substrate preference and orientation in which lower ligands attached to the ribose ring, and thus influence the diversity of cobamides produced by an organism (Hazra et al. 2013; Crofts et al. 2013; 2014). Notably, the *bza* operon pathway is predicted to make four benzimidazolyl molecules with subtle differences of one or two methyl groups, and all the intermediates can potentially act as CobT substrates. Thus, the inherent substrate preference of the *bza* operon CobT homolog will impact the final cobamide(s) resulting from the pathway. To uncover the contribution of CobT in the *bza* operon pathway, we aim to systematically study the reaction catalyzed by *bza* operon CobT with benzimidazoles found as intermediates in the anaerobic biosynthesis of DMB. We also discover that certain anaerobes have multiple *cobT* homologs in varying gene neighborhoods. While it is not uncommon for bacterial genomes to encode for multiple genes that lead to proteins of the same class, however, systematic studies to explore the functional differences among such homologs are rare in literature. We thus characterize the gene products of *cobT* encoded in different gene neighborhoods to reveal the how such homologs might contribute to cobamide diversity among anaerobic bacteria.

b. Reconstitution and mechanistic analysis of BzaC, BzaD, and BzaE methyltransferases.

Based on heterologous expression of *bza* operon and its synthetic variants, the gene products of *bzaC*, *bzaD*, and *bzaE* genes were predicted to catalyze three sequential and chemically distinct methylations (Hazra et al. 2015). Anaerobic bacteria can harbor none, one, two, or all three predicted methyltransferases, and thus these genes contribute to the diversity of benimidazolyl cobamides with lower ligands that vary on the basis of methyl substituents. Further, we find that two domain variants of BzaC homologs exists in nature wherein a fraction of homologs contain an additional domain of unknown function called the DUF2284. We begin with the reconstitution of the methyltransferase activity of BzaC, BzaD, and BzaE that would enable us to find the true substrates and cofactors utilized by the enzymes, which would be foundational to mechanistic analyses of the reactions, and

understanding how reaction proceeds under physiological conditions. Further, we conduct biochemical analysis of DUF2284 to understand the function of this domain in anaerobic DMB biosynthesis pathway. Mechanistic validation of these enzymes and the role of the cofactors involved will help in understanding anaerobic B_{12} biosynthesis and also lead the way for development of reliable tools to predict lower ligands formed in important microbial communities such as the human gut microbiome.

Chapter 2: Probing the functional variations in CobT encoded in different gene neighborhoods

In aerobes and facultative anaerobes, the *cobT* gene is typically localized with genes involves in the nucleotide loop assembly or corrin ring biosynthesis(Zayas and Escalante-Semerena 2007). However, the recent discovery of *bza* operon unveils presence of a *cobT* gene with the genes involved in the lower ligand biosynthesis (Hazra et al. 2015). Previous work from our group, also uncovered the existence of multiple *cobT* genes in *Eubacterium limosum* and *Moorella thermoacetica* (Datar and Hazra 2018). We began our studies with exploring the occurrence of multiple *cobT* genes in anaerobes and classify the *cobT* homologs into three groups based on the corresponding gene neighborhoods. We systematically study the differences in activity of the different CobT encoded in different gene neighborhoods, and we initiate molecular evolution experiments to find sequence level differences that render functional differences among the homologs. We analyze unique enzymatic features of CobT homolog encoded within the *bza* operon to precisely understand their role in the *bza* operon pathway.

Chapter 3: Identification and characterization of CobT homologs in eukaryotes

During the course of bioinformatics efforts to map the genomic distribution of bacterial *cobT*, a small fraction of eukaryotes that encode a putative *cobT* were discovered. This discovery was surprising because B_{12} biosynthesis is carried by only bacteria and archaea(Sokolovskaya, Shelton, and Taga 2020). To probe the relevance of this discovery, we experimental characterized of a subset of the eukaryotic CobT homologs to test whether these are indeed functional enzymes that activate lower ligands for cobamide biosynthesis. We then conducted a comprehensive comparative genomics analysis of other genes involved in cobamide assembly, uptake, and utilization in these eukaryotes. Collectively, these findings complement recent discoveries of cobamide assembly and remodeling in some marine algae (Baum et al. 2020), and thus challenges the known taxonomic limits of B_{12} biosynthesis enzymes. The work presented here provides a starting point for extensive explorations on physiological function of eukaryotic CobT enzymes as well as the biological significance of

eukaryotes empowered to attach lower ligands with respect to the natural environments and microbial communities they inhabit.

Chapter 4: Reconstitution of enzyme activity of *Mt*BzaC and understanding its role in the *bza* operon pathway

We begin our explorations on validation and mechanistic analyses of the methyltransferases of the *bza* operon, with *in vitro* characterization and reconstitution of BzaC, the predicted O-methyltransferase. In this chapter, I describe primary characterization of BzaC homologs from *Moorella thermoacetica* (*Mt*BzaC) which contains a dimerization domain and methyltransferase domain. We biochemically verify the function of the predicted domains and test a range of physiologically relevant substrates for *Mt*BzaC. We validate *Mt*BzaC as 5-hydroxybenzimidazole ribose: SAM O-methyltransferase (HBIR: SAM OMT). Finally, we combine our findings from characterization of *Moorella thermoacetica* CobT (*Mt*CobT) and *Mt*BzaC to identify a new intermediate and propose a revision to the previously proposed *bza* operon pathway.

Chapter 5: Investigations to decipher the function of DUF2284 and the last two methyltransferases in *bza* operon

The BzaC homolog from *Eubacterium limosum* (*El*BzaC) contains an additional C-terminal DUF2284 domain which is absent in the *Mt*BzaC homolog. We initiate characterization of DUF2284 using bioinformatic analysis of DUF2284 sequences and the comparative genomics of bza operon in all organism that harbor a *bzaC* gene. Analysis of the sequence allows us to biochemically characterize DUF2284 using *El*BzaC and two other mutant constructs as described in the chapter 5. Surprisingly, the comparative genomics analysis indicates that presence of DUF2284 is correlated with that of *bzaD* and/ or *bzaE*, suggesting a possible interaction of DUF2284 with the other two predicted methyltransferases BzaD and/or BzaE. We additionally conducted bioinformatic analysis to better understand the scope of biochemical nature and function of DUF2284 in bacteria and archaea. We then begin with biochemical characterization of BzaD with optimization of methods to obtain the predicted substrates and the enzymes. Our comparative genomics study unveils diversity in the *bza* operon composition among anaerobic bacteria, which strengthens the principle of using of the *bza* operon to predict lower ligand biosynthesis, enabling insights into cobamide diversity within microbial communities.

References:

Baum, Christoph, Riya C. Menezes, Aleš Svatoš, and Torsten Schubert. 2020. “Cobamide Remodeling in the Freshwater Microalga Chlamydomonas Reinhardtii.” FEMS Microbiology Letters 367 (20): 1–6. https://doi.org/10.1093/femsle/fnaa171.

Crofts, Terence S., Amrita B. Hazra, Jennifer LA Tran, Olga M. Sokolovskaya, Vadim Osadchiy, Omer Ad, Jeffrey Pelton, Stefan Bauer, and Michiko E. Taga. 2014. “Regiospecific Formation of Cobamide Isomers Is Directed by CobT.” Biochemistry 53 (49): 7805–15. https://doi.org/10.1021/bi501147d.

Crofts, Terence S., Erica C. Seth, Amrita B. Hazra, and Michiko E. Taga. 2013. “Cobamide Structure Depends on Both Lower Ligand Availability and CobT Substrate Specificity.” Chemistry and Biology 20 (10): 1265–74. https://doi.org/10.1016/j.chembiol.2013.08.006.

Datar, Prathamesh M., and Amrita B. Hazra. 2018. “Investigating the Regioselective Attachment of the Lower Ligand in Vitamin B12 Biosynthesis.” IISER Pune. http://dr.iiserpune.ac.in:8080/xmlui/handle/123456789/1023.

Gagnon, Derek M., Troy A. Stich, Angad P. Mehta, Sameh H. Abdelwahed, Tadhg P. Begley, and R. David Britt. 2018. “An Aminoimidazole Radical Intermediate in the Anaerobic Biosynthesis of the 5,6-Dimethylbenzimidazole Ligand to Vitamin B12.” Research-article. Journal of the American Chemical Society 140 (40): 12798–807. https://doi.org/10.1021/jacs.8b05686.

Hazra, Amrita B., Andrew W. Han, Angad P. Mehta, Kenny C. Mok, Vadim Osadchiy, Tadhg P. Begley, and Michiko E. Taga. 2015. “Anaerobic Biosynthesis of the Lower Ligand of Vitamin B12.” Proceedings of the National Academy of Sciences of the United States of America 112 (34): 10792–97. https://doi.org/10.1073/pnas.1509132112.

Hazra, Amrita B., Jennifer L.A. Tran, Terence S. Crofts, and Michiko E. Taga. 2013. “Analysis of Substrate Specificity in CobT Homologs Reveals Widespread Preference for DMB, the Lower Axial Ligand of Vitamin B12.” Chemistry and Biology 20 (10): 1275–85. https://doi.org/10.1016/j.chembiol.2013.08.007.

Kennedy, Kristopher J., and Michiko E. Taga. 2020. “Cobamides.” Current Biology 30 (2): R55–56. https://doi.org/10.1016/j.cub.2019.11.049.

Mattes, Theodoric A., Jorge C. Escalante-Semerena, Evelyne Deery, and Martin J. Warren. 2017. “Cobalamin Biosynthesis and Insertion.” In Encyclopedia of Inorganic and Bioinorganic Chemistry, R.A. Scott (Ed.), 1–24.

Mehta, Angad P., Sameh H. Abdelwahed, Michael K. Fenwick, Amrita B. Hazra, Michiko E. Taga, Yang Zhang, Steven E. Ealick, and Tadhg P. Begley. 2015. “Anaerobic 5-Hydroxybenzimidazole Formation from Aminoimidazole Ribotide: An Unanticipated Intersection of Thiamin and Vitamin B12 Biosynthesis.” Journal of the American Chemical Society 137 (33): 10444–47. https://doi.org/10.1021/jacs.5b03576.

Renz, P. 1993. “Biosynthesis of Vitamin B12 in Anaerobic Bacteria” 1121: 1117–21.

Rodionov, Dmitry A., Alexey G. Vitreschak, Andrey A. Mironov, and Mikhail S. Gelfand. 2003. “Comparative Genomics of the Vitamin B12 Metabolism and Regulation in Prokaryotes.” Journal of Biological Chemistry 278 (42): 41148–59. https://doi.org/10.1074/jbc.M305837200.

Roth, J R, J G Lawrence, and T A Bobik. 1996. “Cobalamin (Coenzyme B12): Synthesis and Biological Significance.” Annual Review of Microbiology 50: 137–81.

Sokolovskaya, Olga M., Amanda N. Shelton, and Michiko E. Taga. 2020. “Sharing Vitamins: Cobamides Unveil Microbial Interactions.” Science 369 (6499). https://doi.org/10.1126/science.aba0165.

Trzebiatowski, Jodi R, and Jorge C Escalante-semerena. 1997. “Purification and Characterization of CobT , the Nicotinate- Mononucleotide : 5 ,6-Dimethylbenzimidazole Phosphoribosyltransferase Enzyme from Salmonella Typhimurium.” Journal Biological Chemistry 272 (28): 17662–67.

Zayas, Carmen L., and Jorge C. Escalante-Semerena. 2007. “Reassessment of the Late Steps of Coenzyme B12 Synthesis in Salmonella Enterica: Evidence That Dephosphorylation of Adenosylcobalamin- 5′-Phosphate by the CobC Phosphatase Is the Last Step of the Pathway.” Journal of Bacteriology 189 (6): 2210–18. https://doi.org/10.1128/JB.01665-06.

Table of contents

List of figures

List of tables

Abbreviations

5’-methylthioadenosine	**MTA**
5-chlorobenzimidazole	**5-ClBza**
5-hydroxybenzimidazole	**5-OHBza**
5-methoxybenzimidazole	**5-OMeBza**
5-methoxybenzimidazole-riboside phosphate	**5-OMeBza-RP**
5-methyladenosine nucleosidase	**MTAN**
5-methylbenzimidazole	**5-MeBza**
5-OHBza-R methyltransferase	**HBIR-OMT**
6-hydroxybenzimidazole	**6-OMeBza**
6-hydroxybenzimidazole-riboside phosphate	**6-OHBza-RP**
adenine	**Ade**
α-adenine-riboside phosphate	**Ade-RP**
β-adenosine monophosphate	**AMP**
Aphanomyces astaci	***Aa***
Aphanomyces invadans	***Ai***
Area under curve	**AUC**
ArsAB	**Aromatic ribotide synthase**
B_{12}-dependent methionine synthase	**MetH**
B_{12}-dependent ribonucleotide reductase	**RNR II**
Basic Local Alignment Search Tool	**BLAST**
Bovine serum albumin	**BSA**
Cba	**cobamide**
Cbl or B_{12}	**cobalamin**
CH_3-H_4 folate	**methyl-tetrahydrofolate**
Chytriomyces confervae	***Cc***

Coenzyme B_{12} or Ado-Cbl	**Adenosylcobalamin**
Conserved Domain Database	**CDD**
Deoxyribonucleic acid	**DNA**
Desuldobulbous mediterraneous	***Dm***
Desulfofundulus thermosubsterraneus	***Dt***
Diode-Array Detection	**DAD**
Dithiothreitol	**DTT**
5,6-dimethylbenzimidazole	**DMB**
α-DMB-riboside phosphate	**DMB-RP**
Domain of unknown function	**DUF**
Dunaliella salina	***Ds***
Electrospray ionization	**ESI**
Eubacterium barkeri	***Eb***
Eubacterium limosum	***El***
Extracted ion chromatogram	**EIC**
Fluorescence Detector	**FLD**
tetrahydrofolate	**H_4 folate**
Hidden Markov Model	**HMM**
High resolution multiple reaction monitoring	**MRM-HR**
High-performance liquid chromatography	**HPLC**
Isopropyl β-D-1-thiogalactopyranoside	**IPTG**
Lactobacillus reuteri	***Lr***
Liquid chromatography–mass spectrometry	**LC-MS**
methylcobalamin	**Me-B_{12} or Me-Cbl**
methylmalonyl-CoA mutase	**MCM**
Moorella thermoacetica	***Mt***
MUltiple Sequence Comparison by Log-Expectation	**MUSCLE**

N^7-adenine riboside phosphate	**N^7-Ade-RP**
N^9-adenine riboside phosphate	**N^9-Ade-RP**
nicotinic acid mononucleotide	**NaMN**
Nitrilotriacetic acid	**NTA**
nicotiamide mononucleotide	**NMN**
Nuclear magnetic resonance	**NMR**
pseudocobalamin	**pB_{12}**
Phenylmethylsulfonyl fluoride	**PMSF**
Polymerase chain reaction	**PCR**
Restriction-free	**RF**
Revolutions per minute	**rpm**
Ribonucleic acid	**RNA**
S-adenosylhomocysteine	**SAH**
S-adenosylmethionine	**SAM**
Salmonella enterica CobT	***Se*CobT**
Sequence similarity network	**SSN**
Sodium dodecyl sulfate–polyacrylamide gel electrophoresis	**SDS-PAGE**
Specificity determining position	**SDP**
Ultraviolet	**UV**
Veillonella parvula CobT	***Vp*CobT**
cyanocobalamin	**Vitamin B_{12}**

Abstract

Vitamin B_{12} belongs to the cobamides family of cofactors required by several organisms including bacteria, archaea, protists, algae, and animals. Cobamides act as cofactors for methyl transfer and radical-based reaction in a wide range of biological processes. However, the biosynthesis of B_{12} and all other cobamides is limited to a subset of bacteria and archaea. The biosynthesis of cobamides involves a modular pathway in which the structural components - the tetrapyrrolic cobalt ion containing corrin ring, the upper ligand, and the lower ligand - are synthesized independently and assembled together in the later steps in the pathway. In B_{12}, the lower ligand is 5,6-dimethylbenzimidazole (DMB), however other cobamides contain a variety of lower ligands derived from benzimidazoles, purines, and phenols. The aim of my thesis was set to mechanistically investigating the recently discovered DMB biosynthesis *bza* operon in anaerobic bacteria. The first biosynthesis intermediate has been shown to be produced *via* a radical-SAM enzyme BzaAB/BzaF, and the remaining enzymes CobT, and BzaC, BzaD, and BzaE were to be characterized. The next gene *cobT* was predicted to encode a phosphoribosyltransferase that activates a wide range of lower ligands prior to cobamide assembly, and the genes *bzaCDE* were predicted to encode methyltransferases that yield distinctly methylated benzimidazole intermediates with DMB as the final product.

To begin, we characterize the *bza* operon CobT by examining differences in its enzyme activity and substrate preferences in comparison to other CobT homologs from different gene neighborhoods. Next, our in-depth biochemical analyses with CobT and BzaC from the anaerobe *Moorella thermoacetica* reveal that the activation of benzimidazole precedes the methylation steps, shedding light on previously unexplained observations of the highly regiospecific lower ligand attachment in anaerobic B_{12} biosynthesis. Further, we find a novel domain of unknown function (DUF) 2284 present in some BzaC homologs and our primary biochemical studies establish the DUF2284 as an iron-sulphur cluster binding domain. We also observe the patterns of co-occurrence of this domain with the putative methyltransferase genes *bzaD* and *bzaE* in a comparative genomics study which in turn reveals the diversity in composition of the *bza* operon across anaerobic bacteria. Finally, as a part of our bioinformatic investigations, we identify homologs of CobT in a handful of eukaryotes which typically do not produce B_{12}. We have initiated the studies with enzyme characterization and comparative genomics that will pave way for understanding cobamide metabolism in communities with microbial eukaryotes which have recently garnered attention in the B_{12} field. In summary, the insights we have gained from the bioinformatic, biochemical, and mechanistic explorations of the *bza* operon enzymes will aid in improving the industrial production of Vitamin B_{12} and other cobamides, as well as provide insights for metagenomic efforts to predict cobamide production and exchange in microbial communities.

Chapter 1

Methylations in vitamin B_{12} biosynthesis and catalysis

1.1 An introduction to B_{12}

Vitamin B_{12}, also known as cyanocobalamin, is an essential micronutrient required by organisms from all domains of life including bacteria, archaea, protists, animals, and some algae (Roth, Lawrence, and Bobik 1996; Croft, Warren, and Smith 2006). In active cofactor forms, methylcobalamin and adenosylcobalamin, B_{12} participate in complex methyl transfer and radical-based chemistry, respectively in a wide range of biological processes(Banerjee and Ragsdale 2003; Banerjee 1997; Giedyk, Goliszewska, and Gryko 2015). However, the biosynthesis of B_{12} is limited to a small subset of bacteria and archaea and rest all organisms rely on B_{12}-producing organisms for B_{12} as a nutrient (Sokolovskaya, Shelton, and Taga 2020; Rodionov et al. 2003).

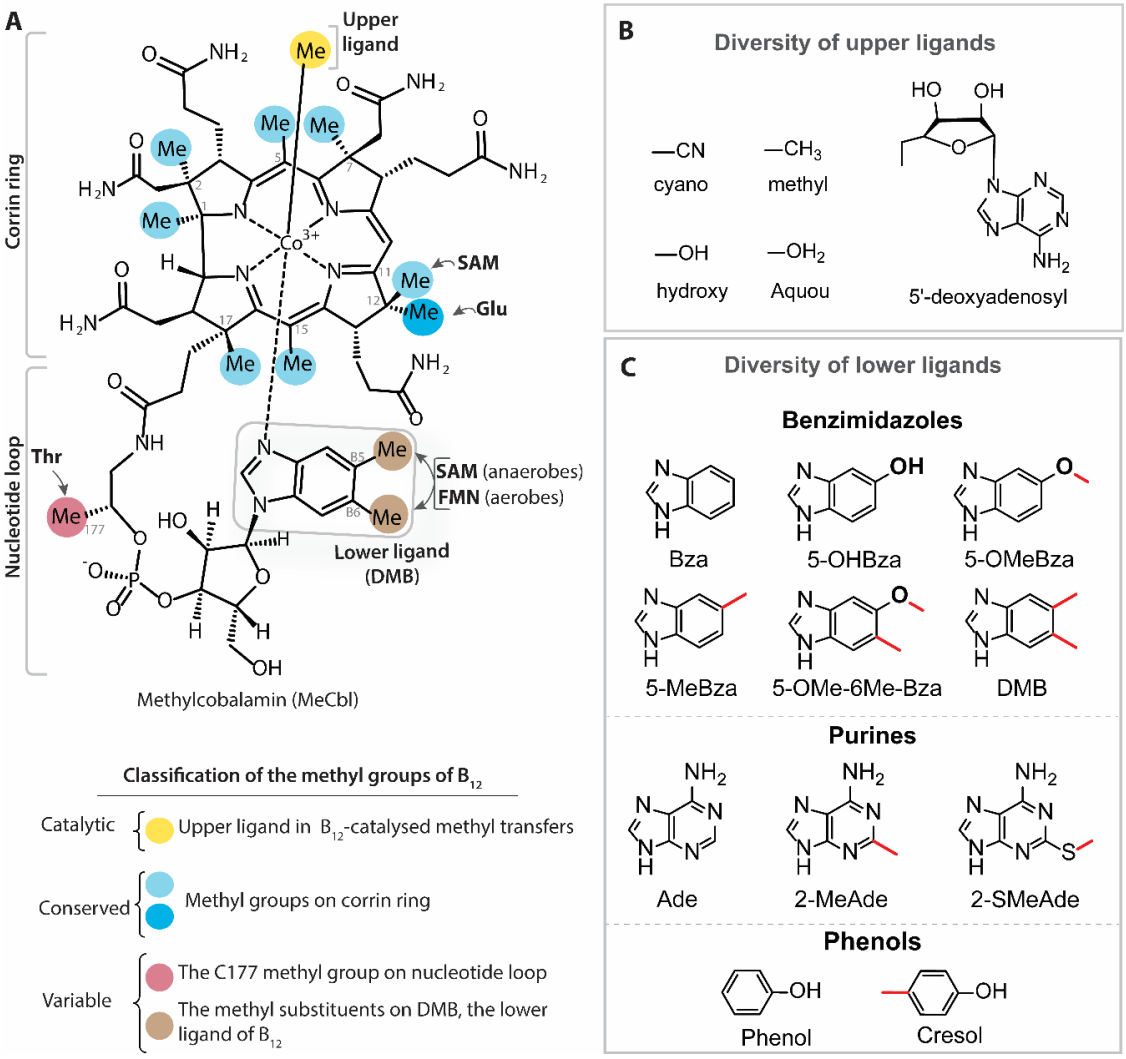


Figure 1.1 Structure of B_{12}. (**A**) The structure of methylcobalamin- an active cofactor form of B_{12}. B_{12} is characterized by a tetrapyrrolic corrin ring which contains a cobalt ion at its center. The cobalt

coordinates with two axial ligands called the upper ligand and the lower ligand, respectively. The lower ligand is covalently attached to the corrin ring via a nucleotide loop and variations in the chemical structure of lower ligand gives rise to a large family of cofactors called cobamides. There are multiple methyl groups that decorate the B_{12} structure which can be classified into three groups namely, catalytic, conserved, and variable based on their origin, function, and position. The catalytic methyl group is present as the upper ligand in the cofactor form called methylcobalamin which is participate in methyl transfer reactions. The conserved methyl groups of the corrin ring provide structural integrity and protect the cobalt reaction centre. The variable methyl groups of nucleotide loop and the lower ligand give contribute to diversity in the norcobamides and cobamides cofactors. Norcobamides are a differentiated from cobamides by the absence of the C177 methyl group on the nucleotide loop. **B)** The upper ligand of cobalt is a transient group which can either be a methyl, hydroxy, aquou, cyano, or 5-deoxyadenosyl (dAdo) group. The methyl and dAdo groups participate in methyl transfer and radical-based reactions respectively, whereas the cyano group is an artifact of B_{12} purification methods and in absence of methyl or dAdo groups the upper ligand position is often occupied by hydroxy or aquou groups. Since the upper ligand directly participates in the catalytic functioning of B_{12} as cofactor, the nature of upper ligand directly determines the physiological role of the cofactor. **C)** The lower ligand of B_{12} is 5,6-dimethylbenzimidazole moiety however, various microorganisms attach derivatives of benzimidazoles, purines, and phenols as lower ligands. The benzimidazole and purine derived lower ligands are attached to the corrin ring via an N-glycosidic bond involving a nitrogen from the five-membered rings of the ligand. The phenolyl lower ligands are connected to nucleotide loop via a O-glycosidic bond and do not coordinate with the central cobalt ion. Notably, several derivatives of the lower ligands within each class often differ by one or two methyl groups, thus underlining the contribution of methyl groups in increasing the diversity of cobamides. [Bza: benzimidazole; 5-OHBza: 5-hydroxybenzimidazole; 5-OMeBza: 5-methoxybenzimidazole; 5-MeBza: 5-methylbenzimidazole; 5-OMe-6-MeBza: 5-methoxy-6-methylbenzimidazole; 5-ClBza: 5-chlorobenzimidazole; DMB: 5,6-dimethylbenzimidazole; Ade: Adenine; 2-MeAde: 2-methyladenine; 2-SMeAde: 2-mercaptomethyladenine]

Often regarded as 'nature's beautiful cofactor', vitamin B_{12} belongs to a large family of cofactors called cobamides which are composed of a planar corrin ring with a central cobalt ion and two axial ligands called the upper ligand and lower ligand (Figure 1.1A)(Stubbe 1994; Roth, Lawrence, and Bobik 1996). The three structural components of B_{12} and other cobamides are described as follows:

1. The conserved corrin ring

Structurally, B_{12} is classified as one of the biological tetrapyrroles which also includes heme, chlorophylls, siroheme, bilins, and coenzyme F_{430} (Bryant, Hunter, and Warren 2020). In comparison to other tetrapyrroles, B_{12} has few notable unique features. First, the characteristic corrin ring of B_{12} is a contracted tetrapyrrolic ring containing 19 carbons, one less than other tetrapyrroles. Second, the central cobalt ion coordinates with two axial ligands which are absent in all other tetrapyrroles. Third, the peripheral acetate and propionate side chains of the corrin ring are neutralized by addition of terminal amide groups, and one of the propionate side chain is extended with an aminopropanol group to covalently attach the lower ligand. Lastly, the corrin framework is decorated by additional eight methyl groups that are added during the biosynthesis pathway, and possibly function to protect the reactive cobalt ion from water (Figure 1.1A).

2. The catalytic upper ligand

The upper ligand is often a transient group tethered via a cobalt-carbon bond that directly participates in the cobamide-mediated catalysis and determines the physiological

activity of the cofactor(Banerjee and Ragsdale 2003) (Figure 1.1A, 1.1B). The cobalt ion can traverse Co(I), Co(I), and Co(III) oxidation states, making cobamides suitable cofactors for a large range of biochemical reactions that involve the transfer of one or two electrons. When bound to the upper ligand, Co is present in the stable Co(III) form that can either transition to Co(II) radical upon a homolytic cleavage or can form a Co(I) nucleophile upon heterolytically cleaving the ligand (Giedyk, Goliszewska, and Gryko 2015).

In active co-factor forms, the upper ligand can be either a methyl group or a 5'-deoxyadenosine moiety for methyl transfers or radical based reactions, respectively (Figure 1.1B). A transient methyl upper ligand in methylcobalamin and other methylcobamides facilitates methyl transfer reactions from a methyl donor which can include S-adenosylmethionine (SAM), methyltetrahydrofolate (Me-THF), or metabolites such as methylated tertiary amines or methyl ethers (Figure 1.2, panel A). In methanogens and acetogens, the methylcobamides play a key role in trafficking the methyl groups to coenzyme M and tetrahydrofolate (Figure 1.2, panel A) (Matthews, Koutmos, and Datta 2008; Richter, Zepeck, and Kroutil 2015). Most prokaryotic and eukaryotic organisms utilize B_{12} or other cobamides as a cofactor for methionine synthase that regenerates methionine from homocysteine and Me-THF (Figure 1.2, panel A) (Banerjee 1997). The newly discovered class of B_{12}-radical-SAM (B_{12}-RSAM) enzymes combine the catalytic prowess of B_{12}, SAM, and Fe-S cluster to bring about extremely complex and elegant chemical transformations. The B_{12}-RSAM enzymes span a range of reactions including formation of C-C, C-P, C-N, and C-S bonds which result in intramolecular rearrangements, thioether bond formation as well as radical and non-radical mediated methylations. Some of the examples B_{12}-RSAM dependent methylations are seen in biosynthesis of antibiotics such as cabapenems and thiostrepton (Figure 1.2, panel A), methylhopanoids, and catabolism of steroids (Matthews, Koutmos, and Datta 2008; Elling et al. 2020; Knox et al. 2022; 2021; Fyfe et al. 2022). The B_{12}-RSAM enzymes also catalyze reactions which do not result in methylation of the substrate, such as the ring contraction reaction by OxsB in the biosynthesis of oxetanocin antibiotic (Zhong et al. 2021) (Figure 1.2C).

The B_{12} cofactor with adenosyl group (Adenosylcobalamin or AdoCbl) as the upper ligand primarily catalyze radical-based reactions such as intramolecular rearrangement of methylmalonyl-CoA, removal of 2'hydroxy group of ribonucleotides to form deoxyribonucleotides, dehydration of diol substrates (Banerjee 1997; Liu et al. 2018). The light-sensitivity of AdoCBl is employed in regulation of carotenoid biosynthesis in some soil bacteria (Padmanabhan et al. 2017) (Figure 1.2, panel B).

In certain exceptional reactions such as reductive dehalogenation, and quenosine reduction, a nascent cofactor is required to accommodate a reaction intermediate as an upper

ligand (Bridwell-Rabb, Li, and Drennan 2022; Bommer et al. 2014) (Figure 1.2D). Several other chemical groups such as hydroxy-, aquou (H_2O)-, nitroso-, glutathione are also found as the upper ligands under physiological conditions (Brown 2005). The commonly known cyano group is a non-physiological upper ligand added during standard B_{12} purification methods.

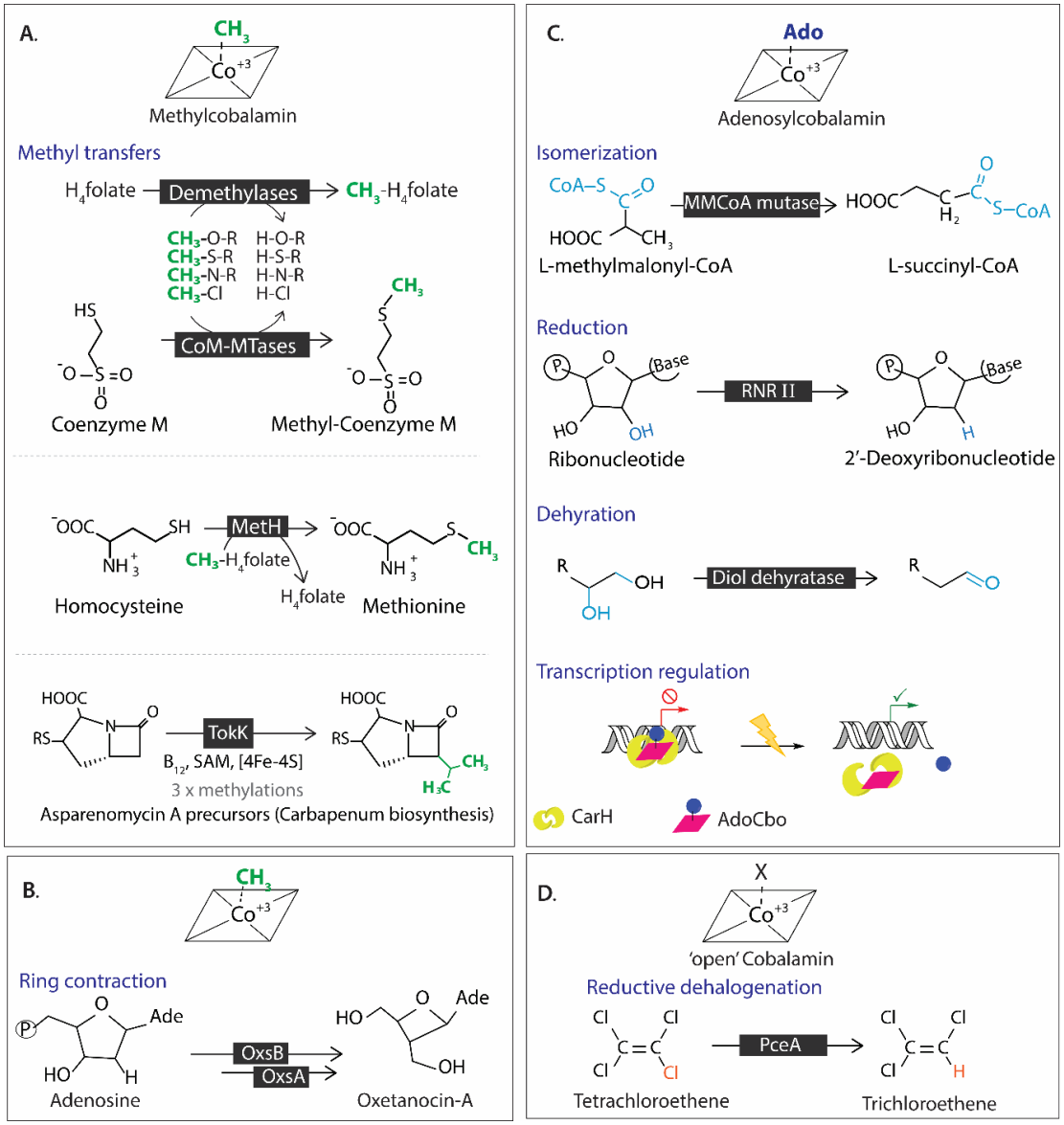


Figure 1.2 Diverse roles of B_{12}. A) Methylcobalamin (MeCbl), the active cofactor form with a methyl group as the upper ligand participates in reactions involving transfer of methyl groups. Some of the examples include methyl trafficiking in methanogenesis and acetogenesis, biosynthesis of methionine, and multiple methylations in biosynthesis of antibiotics such as carbapenems. The latter reaction, catalyzed by TokK is an example of a B_{12}-radical-SAM enzymes which involve B_{12}, SAM, and Fe-S clusters as cofactors. **B)** B_{12}-radical-SAM enzymes also participate in reactions such as ring contractions by OxsB. **C)** Adenosylcobalamin (AdoCbl) assist in catalysis of radical-based reactions such as isomerization of methylmalonyl-CoA, biosynthesis of deoxyribonucleotides, dehydration of diols, as well as photo regulation of carternoid biosynthesis pathway. **D)** Cobamides and norcobamides are involved in dehalogenation reactions wherein the upper ligand is typically a reaction intermediate.

3. The flexible and diverse nucleotide loop which contains the lower ligand

In the cobamide structure, the lower ligand is covalently attached to the corrin ring via a phosphoribosyl group and aminopropanol tail, and these three components are collectively called the nucleotide loop (Figure 1.1A). When the lower ligand is coordinated with the central cobalt ion, the cobamide is said to be in a cyclic 'base-on' state. However, when the lower ligand dissociates from cobalt or in the case of phenols which do not contain another heteroatom, the cobamide assumes a relaxed 'base-off' conformation (Banerjee and Ragsdale 2003). The base-on ↔ base-off transitions are often crucial for enzyme-cofactor interaction and the mechanism of catalysis.

The lower ligand contributes to a remarkable diversity in cobamides produced by microorganisms (Sokolovskaya, Shelton, and Taga 2020; Kennedy and Taga 2020; Erhard Stupperich 1988). Broadly, the classification of cobamides is based on the chemical identity of the lower ligands which are either derivatives of benzimidazoles, purines, or phenols (Figure 1.1C). Within each of these classes, the derivatives commonly vary by the number and location of the methyl groups. For example, the lower ligand in B_{12} is 5,6-dimethylbenzimidazole, however, cobamides with 5-methylbenzimidazole and benzimidazole lower ligands are also produced in nature (Brink et al. 1954; Kräutler, Kohler, and Stupperich 1988). Apart from methyl groups, the lower ligands are also seen to vary by hydroxy or sulfuryl groups (Erhard Stupperich 1988).

Additionally, the absence of a methyl group on C177 of the loop gives rise to a newly discovered class of B_{12} analogs, called the norcobamides (NorCba) (Figure 1A). This difference, however, is not particularly due to absence of a methylation and rather arises from incorporation of serine in place of threonine to generate the nucleotide loop. Norpseudocobalamin (nor-pB_{12}) (pseudocobalamin contains adenine as the lower ligand) is a native cofactor for the reductive dehalogenase PceA enzyme in the organohalide-respiring bacterium *Sulfospirillum multivorans*. Absence of the methyl group on C177 is proposed to can relax the structure and drive the base-off ↔ base-on equilibrium towards the base-off conformation.

It is interesting to note that the structure of B_{12} is heavily decorated with methyl groups which differ in their origin and function based on their location on the structure (Figure 1.1A). First, the cobalt ion can host a methyl group as an upper ligand during methyl transfer reaction. Second, the eight methyl groups on the corrin remain conserved across all cobamides, and lastly the variations in methylation on the nucleotide loop, primarily the lower ligand, contribute to the diversity of the cobamides cofactors.

A note on nomenclature of cobamides:

Cobamides are often also called corrinoids in literature. The lower ligand attached to the corrin ring dictates the nomenclature of the cobamides. The upper ligand attached in the cofactor form is added as a prefix. The B_{12} or cobalamin are cobamides that contains 5,6-dimethylbenzimidazole (DMB) as the lower ligand, and when a methyl or adenosyl group are present as the upper ligand, the cofactors are called methylcobalamin (MeCbl) or 5'-deoxyadenosylcobalamin (5'-dAdoCbl or coenzyme B_{12}). The cobamides with other lower ligands, such as 5-hydroxybenzimidazole (5-OHBza) are called 5-hydroxybenzimidazolylcobamide or [5-OHBza]Cba.

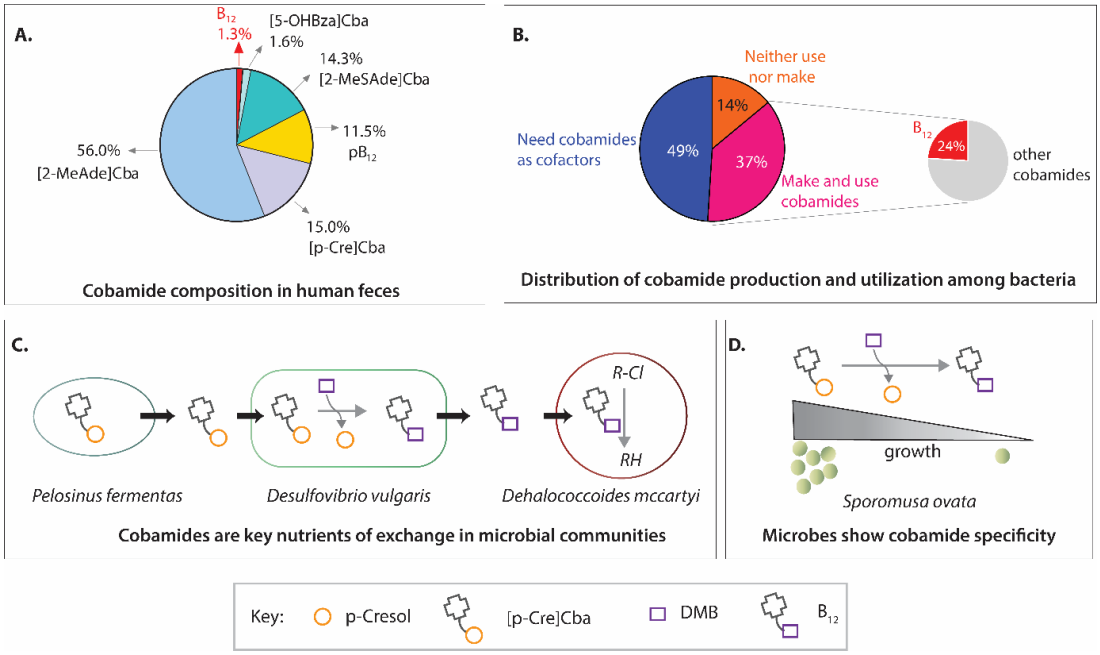


Figure 1.3 Vitamin B_{12} is a rare commodity. A) Percent abundance of cobamides in human feces which indirectly represent the cobamide composition in human gut. The B_{12} that is, the cobamide with DMB as the lower ligand is only 1.3% of the total cobamides, whereas purinyl cobamides such as 2-methyladeninylcobamine [2-MeAde]Cba, 2-methylmercaptoadeninylcobamide [2-MeSAde]Cba, and pseudocobalamin (pB12) form the majority fraction. (Adapted from data presented by (Allen and Stabler 2008)). **B)** A recent comparative genomics survey shows with 11,000 bacterial genomes show that 86% of genomes require cobamides as a cofactor for at least one function, only a small fraction of ~37% can make cobamides. Of the cobamide producers only 24% can make B_{12}, and the rest appear to make various other cobamides. (Adapted from data presented by (Shelton et al. 2018)). **C)** Microbial communities show complex exchange of cobamides in nature. For example, *Pelosinus fermentas* prodcues a p-cresolylcobamide which is acquired and remodeled by *Desulfovibrio vulgaris* to make B_{12}. The B_{12} is then acquired by *Dehalococcoides mccartyi* which supports reductive dehalogenation, and hence the growth of the organism. (Adapted from data presented by (Men et al. 2014)) **D)** Most organisms show an innate preference for a type of cobamide. For example, the anaerobic bacterium *Sporomusa ovata* produces p-cresolylcobamide as the native cobamide which supports its growth well. However, presence of DMB in the growth media redirects the cobamide biosynthesis pathway to make B_{12}, which ultimately proves detrimental to the growth of the organism. (Adapted from data presented by (Mok and Taga 2013)).

1.2 Vitamin B_{12} in human metabolism

Among a wide diversity of cobamides present in the nature, humans exclusively require B_{12} for two metabolic reactions (Banerjee 1997). One, as the cofactor for the methionine synthase B_{12} mediates the transfer of the methyl group from methyl-tetrahydrofolate (CH_3-H_4 folate) to homocysteine which produces methionine and H_4 folate, thus completing the methyl and folate cycles (Figure 1.2A). Two, as a cofactor for methylmalonyl-CoA mutase B_{12} catalyzes isomerization of methylmalonyl-CoA, a dead-end catabolite from oxidation of branched chain fatty acids and amino acids. The reaction results into succinyl-CoA which fluxes into central metabolism by entering the Krebs cycle (Figure 1.2B). The two B_{12}-dependent reactions are central to metabolism, and hence B_{12} has a profound effect on physical and mental health in humans. Genetic defects in B_{12} uptake and utilization, or deficiency of B_{12} in diet can result in disruption of folate and methyl cycles, and accumulation of homocysteine and methylmalonic acid which consequently can lead to pernicious anaemia, neural tube defects in infants, infertility, delayed recovery post-chemotherapy, and early onset of diabetes (O'Leary and Samman 2010; Green et al. 2017).

Analysis of cobamide composition in feces of human subjects reveal an abundance of purinyl cobamides- that is the cobamides with lower ligands such as adenine, 2-methyladenine, 2-methylmercaptoadenine (Figure 1.3A). The samples also contained a fairly high concentration of p-cresolylcobamides, however, B_{12} appears to be one of the least abundant cobamides with a surprisingly low concentration of 1.3% (Allen and Stabler 2008) . Even though humans gut microbes can produce cobamides, the lack of B_{12} absorption machinery in the gut restricts us exploiting the resource (Frye et al. 2021; Degnan, Taga, and Goodman 2014).

Since B_{12} is produced solely by microbes and is absent in plants, humans acquire it from meat, fish, dairy, and fermented products. Populations with predominantly vegetarian diets and low food security, such as in India, are at a higher risk for B_{12} deficiency related diseases. While supplements and fortified food can solve the problem, the lack of awareness and high costs for the supplements are deterrents for most people (Stabler and Allen 2004). A thorough understanding of genetics and biochemistry of natural B_{12} production is a pre-requisite to improving pharmaceutical production of the vitamin, as well as finding B_{12} producing microbes amenable for human consumption.

1.3 Biosynthesis of B_{12} and other cobamides involves multiple methylation steps

In nature, only a small subset of bacteria and archaea are capable of producing cobamides, and among those organisms capable of producing B_{12} as the cobamide is even rarer (Figure

1.3B). In a recent study with 11,000 bacterial genomes unveiled that of the 86% genomes that require cobamides, only 37% had the genetic potential to make a cobamide, and only 24% of cobamide producers (Shelton et al. 2018). Microbes that require cobamides for regular metabolic functions, often form complex interactions with cobamide producers within their communities and there are several examples underpinning the role of cobamide cofactors in modulating bacterial growth and significance in respective consortia (Figure 1.3C, 1.3D) (Sokolovskaya, Shelton, and Taga 2020; Seth and Taga 2014).

The *de novo* biosynthesis of B_{12} proceeds in three stages- (i) biosynthesis of the corrin ring, (ii) biosynthesis of the lower ligand, (iii) assembly of the complete cobamide. The modular design of the pathway allows incorporation of diverse lower ligands and hence, leads to a diverse set of cobamides. Depending upon the involvement of oxygen two routes exists for biosynthesis of corrin ring and the lower ligand, as described below.

i) Corrin ring biosynthesis

The corrin ring biosynthesis initiates with three primary metabolites- glutamine, succinyl-CoA, and glycine which yield a five-membered ring that are used to build the prototype molecule uroporphyrinogen III which is also the precursor for all other biological tetrapyrroles (Bryant, Hunter, and Warren 2020). Uroporphyrinogen III undergoes extrusion of C20 meso-carbon for biosynthesis of corrin ring, addition of a cobalt ion, amidation of the acetate and a propionate side chains, and a several methylations to produce adenosylcobyrinic acid, which acts as conserved backbone for all cobamides and norcobamides (Mattes et al. 2017). The order and mechanisms of each of these steps vary in the oxygen-dependent (aerobic) and the oxygen-sensitive (anaerobic) routes which have two major contrasts (Figure 1.4A). First, the aerobic route employs molecular oxygen as a reagent in the ring contraction step, whereas the same is executed via a radical-based mechanism in the anaerobic pathway. Second, the cobalt insertion precedes ring contraction in the anaerobic pathway, and whereas cobalt chelation occurs later in the aerobic pathway to save the metal ion from oxidative damage. The modifications to the peripheral side chains and the methylations occur in comparable order and mechanisms in the two routes. The aerobic pathway contains another intricacy- a protein called CobE acts as a chaperone for that hosts the intermediates from steps leading from precorrin-6B to hydrogenobyrinic acid (Deery et al. 2012).

The methyltransferases that accept Uroporphyrinogen-III derived substrates are classified under AdoMet (SAM)-dependent methyltransferases class-III. The methylations in corrin ring biosynthesis are interspersed throughout the pathway and some of these methyl groups are pre-requisites for ring contraction and other modifications from a mechanism point of view. Chemically, all the methylations are C-methylations and have been shown to be regiospecific reactions. S-adenosylmethionine (SAM) acts as a source for all the methyl groups except for

one methyl group on C12 which is derived from decarboxylation of an acetate side chain (Figure 1.4A).

Notably, the first methyltransferase in the pathway is an important juncture for pathway regulation and bioengineering B_{12} producing strains (Jiang et al. 2020). The enzyme uroporphyrinogen III: SAM methyltransferase (SUMT, called CobA and UroM) catalyzes first two methylations at C7 and C2. This is a common step in biosynthesis pathways for cobamides, siroheme, and bacterial heme d1. In some organisms such as *Salmonella enterica,* a tri-functional enzyme CysG catalyzes the bis-methylation as well as the NAD+ dependent dehydrogenation and ferrochelation required for siroheme biosynthesis (Stroupe et al. 2003; Vévodová et al. 2004). Further, a recent discovery of a riboswitch in the untranslated region upstream of *cobA* in Propionibacterium UF1 proves SUMT even more viable for B_{12} regulation for biotechnological applications (Li et al. 2020).

The methyltransferases in the pathway, show overall sequence and structure similarity. For example, the enzymes CobA/UroM, CbiL, and CbiF share overall homology but show few variations in the active sites to accommodate the increasingly methylated substrates as well as ensure accurate positioning of SAM for methylating the correct position on the ring. Further, since the B_{12} biosynthesis methyltransferases are exclusively present in a subset of bacteria, these enzymes have also garnered attention as drug targets (Kipkorir et al. 2021; Peinado et al. 2019). The shared overall structural design and substrate binding sites, in theory, allows the scope for designing efficient drugs. For instance, CobA was computationally identified as one of the potential targets for indole-derived drugs (Tha et al. 2020). Recently, a polyanionic inhibitor was shown to competitively bind to *Clostridium pseudotuberculosis* precorrin-4 methyltransferase CobM. Much finer details from a comparative structural and functional studies of methyltransferases involved in B_{12} biosynthesis will be required to realistically to use this pathway for design drugs against pathogens.

ii) The lower ligand biosynthesis

The biosynthesis of lower ligands proceeds separately from the corrin ring biosynthesis. Among the benzimidazolyl lower ligands, DMB is produced by both aerobic and anaerobic organisms via contrasting pathways as described in the next section. The aerobic is a single step reaction which leads to DMB as the sole product, however, the anaerobic pathway involves multi-steps and yield 5-OHBza, 5-OMeBza, 5-OMe-6MeBza, and DMB lower ligands. The two other benzimidazolyl lower ligands, 5-MeBza and Bza are naturally occurring lower ligands found in anaerobic bacteria however, the underlying pathway is yet to be discovered. The purinyl lower ligands adenine and 2-methyladenine are known to be sourced from cellular pool and degradation of modified tRNA, respectively. The phenolyl lower ligands, phenol and p-cresol,

are produced by anaerobic bacteria of the class negativicutes, however the origin of these molecules remains elusive till date.

iii) Nucleotide loop assembly (NLA) allows cobamide diversity

In the cobamide structure, the nucleotide loop refers to lower ligand, phosphoribosyl group and the linking aminopropanol tail. Nucleotide loop assembly (NLA) pathway combines the independently constructed corrin ring and the lower ligand moiety to build the final cobamide structure and involves CobU, CobT, CobS, and CobC (Figure 1.4B). The NLA in B_{12} biosynthesis pathway proceeds in the following manner: The enzyme CobD attached the aminopropanol-phosphate tail to the Adenosylcobyric acid, the final product of the corrin ring biosynthesis resulting in adenosylcobinamide-phosphate (AdoCbi-P). Next, the bifunctional enzyme CobU, (AdoCbi kinase/ AdoCbi-P guanylyltransferase) transfers a GMP moiety from GTP to AdoCbi-P resulting in an Adenoylcobinamide-GDP (AdoCbi-GDP) intermediate. Simultaneously, the enzyme CobT (NaMN: DMB a-phosphoribosyltransferase) activates the DMB into DMB-riboside phosphate (DMB-RP). The two activated intermediates, AdoCbi-GDP and the DMB-RP are then condensed by CobS (Cobalamin-5'-phosphate synthase). The reaction results in formation of a phosphodiester bond between 3'-OH of the ribotide and the phosphate of the aminopropanol phosphate linker, thus displacing the GMP moiety as a by-product. Finally, the 5'-phosphate of the α-LL-ribotide unit in the cobalamin-5'-phosphate has been to show to be removed by CobC (phosphatase), which yields adenosylcobalamin as final product.

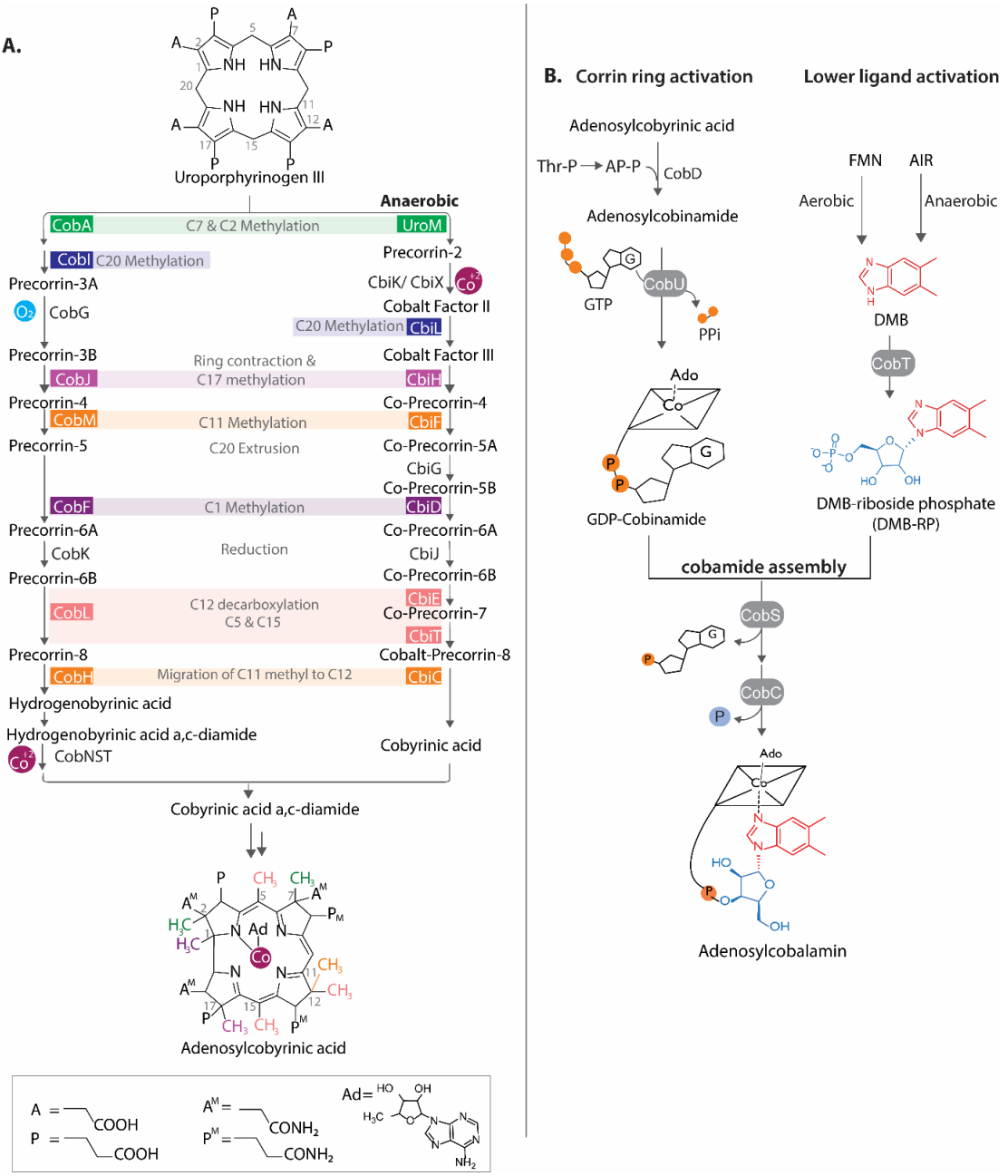


Figure 1.4 Biosynthesis of corrin ring and the cobamide assembly pathway. A) Methylations in biosynthesis of the corrin ring. The B_{12} biosynthesis in aerobes and anaerobes differ in (i) the mechanism for ring contraction and (ii) the order of cobalt chelation with respect to other changes to the ring (Deery et al. 2012; Moore et al. 2013). Notably, the order and mechanisms of methylations are comparable in both the pathways. The methyltransferase homologs from the aerobic and anaerobic pathways share sequence similarity in a range of 20-40% (Schubert et al. 1998). Briefly, nine methyl groups are added onto C1, C2, C5, C7, C11, C12, C15, C17, and C20 by SAM-dependent methyltransferases, the methyl group at C20 is lost during the ring contraction, the methyl group at C11 is migrated to C12, and decarboxylation of acetate side chain of C12 adds another methyl group. **B)** The cobamide assembly combines the activated corrin ring and the activated lower ligands which are independently produced via aerobic or anaerobic pathways. The enzyme CobD attaches the aminopropanol tail to the corrin ring, and the enzyme CobU transfers a GMP moiety to the cobinamide to yield GDP-cobinamide, the activated corrin ring. The lower ligand activation is catalyzed by the enzyme CobT, which can accommodate a wide range of benzimidazoles, purines, and phenols as substrates and thus, allows production of diverse cobamides. In B_{12} biosynthesis pathway, the enzyme CobS covalently attaches the lower ligand with the cobinamide producing cobalamin-5'-phosphate which undergoes a phosphate reaction catalyzed by CobC to finally result in cobalamin.

1.4 Biosynthesis of DMB in aerobes and anaerobes

The chemical origins of 5,6-dimethylbenzidazole (DMB)- the lower ligand of B_{12} was a missing piece of puzzle of B_{12} biosynthesis for a long time. Pioneering research in the lab of Paul Renz during 1980s-90s, established that aerobes and anaerobes source DMB through different mechanisms. The vitamin B_2 (riboflavin) was shown to be the precursor in aerobic bacteria *Propionibacterium freundenshii*. However, the labelling pattern of DMB in anaerobic bacterium *Eubacterium limosum* resembled that of purines, in a manner that the nitrogen atoms were derived from glutamine and glycine and the carbons were derived from tetrahydrofolate, threose or erythrose like sugar molecule, and methionine. In the recent years, the molecular origins of DMB in aerobic and anaerobic bacteria have been discovered to begin with flavin mononucleotide (FMN, which is phosphorylated riboflavin) and 5-aminoimidazole ribonucletide (AIR, which is also the precursor for purines). The details of the two pathways are as follows:

1.4.1. BluB upcycles Flavin Mononucleotide (FMN) to make DMB in an oxygen-dependent reaction

For DMB biosynthesis, the aerobic pathway rearranges an FMN molecule in presence of molecular oxygen as a reagent and the enzyme BluB, designated as 'Flavin destructase' (Taga and Larsen 2007, Gary and Escalante-Semerena 2007). In this reaction, the BluB enzyme orchestrates a systemic fragmentation and rearrangement of reduced FMN (FMNH2) to yield DMB. In order to achieve that, the bonds between the central six-membered ring of the isolloxazine is disrupted and the C1' of the ribityl tail is inserted between the two nitrogen atoms to create the five-membered ring of DMB (Taga and Larsen 2007, Renz P labelling studies with N15 and C1 FMNH2). Following the ring contraction, the remaining parts of the ribityl tail leaves as erythrose-4-phosphate. Consequently, DMB inherits its characteristic methyl groups from the C6 and C7 of FMN as a part of the six membered ring.

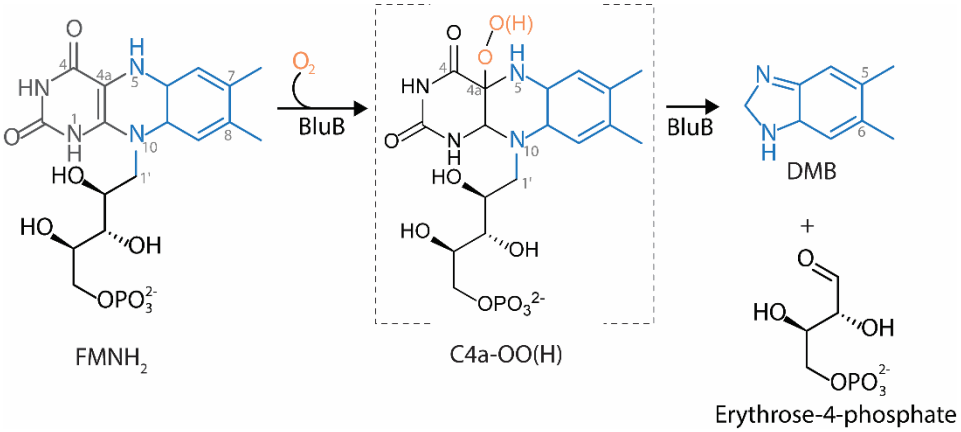


Figure 1.5 Aerobic Biosynthesis of DMB. **A)** In aerobic bacteria, DMB is derived from flavin mononucleotide (FMN) in a single oxygen-dependent step catalyzed by the enzyme BluB (also termed as FMN destructase), a member of the nitroreductase family. BluB orchestrates a systemic fragmentation and rearrangement of reduced FMN ($FMNH_2$) to yield DMB (Taga et al. 2007; Hazra, Ballou, and Taga 2018). The bonds between the central ring of the isolloxazine is disrupted and the C1' of the ribityl tail is inserted between the two nitrogen atoms to create the five-membered ring of DMB. The reaction proceeds through a C4a-peroxyflavin (C4a-OOH) intermediate however, the mechanism is yet under investigation and identifying the third product of the reaction will allow to solve this long-standing problem.

The *bluB* gene was identified following a curious observation wherein exogenous addition of DMB cured B_{12} auxotrophy in a *bluB* mutant of a plant root-dwelling bacterium *Sinorhizobium meliloti* (Campbell et al. 2006; Taga et al. 2007). A similar phenomenon was observed with *bluB* mutants of a photosynthetic bacterium *Rhodospirillum rubrum* (Gray and Escalante-Semerena 2007). Notably, despite the identification of *bluFEDCB* gene cluster the in *Rhodobacter capsulatus*, BluB managed to dodge its recognition as the DMB synthase for a long time due to its unexpected sequence similarity with the nitroreductases (Pollich and Klug 1995). BluB is a small protein of approximately 200 amino acids and shows high sequence similarity with two other FMN- binding families- iodotyrosinases and flavin nitroreductases (Figure 5)(Phatarphekar, Buss, and Rokita 2014; Taga et al. 2007) . The structural design of *S. meliloti* BluB highlights the specifications that ensures the execution of such a complex biochemical reaction. The active site of BluB is a deep electropositive pocket gated towards the outer periphery of the dimer by a flexible loop. The solvent protected active site allows confinement of $FMNH_2$ with molecular oxygen which in turn is oriented for attack on the C4' through two hydrogen bonds mediated by 2'-hydroxyl of $FMNH_2$ and another from a nearby peptide bond in the active site. Such an arrangement allows formation of a 4a-peroxyflavin intermediate (C4a-OO(H)) (Figure 1.5) (Collins et al. 2013; Hazra, Ballou, and Taga 2018). Currently, the field has three debating proposals for the mechanistic route from C4a-peroxyflavin to DMB. Systematic studies with active site mutants of BluB and use of flavin analogs to capture spectroscopic changes within milliseconds of reaction commencement have provided some insights into the reaction mechanism (Yu et al. 2012; Hazra, Ballou, and Taga 2018). Current state of understanding nudges towards the possibility of formation of a

dioxetane as an intermediate, however the validity of each predicted mechanisms needs further exploration. The rapid rate of reaction and accumulation of oxidized FMN in the catalytically impaired mutants has proven to be a challenge in the progress. Probing the fate of the pyrimidine in the isoalloxazine ring of FMN i.e. finding the third product of the reaction, can help in elucidation of the reaction mechanism (Taga et al. 2007; Hazra, Ballou, and Taga 2018). Interestingly, these studies brought attention to conservation of active site residues – Asp32 and Ser167 that reliably distinguish BluB from other closely related FMN-binding nitroreductases and iodotyrosinases (Taga et al. 2007; Yu et al. 2012; Hazra, Ballou, and Taga 2018). Identification of such signatures of BluB has enabled the further bioinformatic predictions that indicate the wide abundance of gene *bluB* among cobamide producers and its large contribution to B_{12} production in the environment (Shelton et al. 2018).

1.4.2 Anaerobic biosynthesis of DMB is a multi-step pathway which involves three distinct methylations

The anaerobic biosynthesis of DMB involves the *bzaA-bzaB-cobT-bzaC-bzaD-bzaE* genes, which synthesize DMB in the oxygen-sensitive pathway (Figure 1.6A) (Hazra et al. 2015). The function of each gene in the bza operon was discovered using heterologous expression of multiple constrcuts of the operon in a B_{12}-depdendent methionine auxotroph of *E.coli* (Figure 6B) (Hazra et al. 2015). Specifically, the gene products of *bzaA* and *bzaB* together (or their single gene homolog *bzaF*) produces 5-OHBza from 5-aminoimidazole ribotide (AIR), an intermediate in the purine biosynthesis pathway (Mehta et al. 2015) (Figure 1.6B, 1.6C). The first intermediate of the *bza* pathway, 5-OHBza is methylated by the gene product of *bzaC* to produce 5-OMeBza, which is followed by second methylation by the gene product of *bzaD* to produce 5-OMe-6-MeBza, and finally, the gene product of *bzaE* produces DMB (Figure 1.6B, 1.6C).

Interestingly, all the benzimidazole derivatives found as intermediates on the pathway, are also found as lower ligands in naturally occurring cobamides. Recent studies on anaerobic cobamide producers show that their genome contains either none, one, two, or all three of the *bza* methyltransferases depending on the cobamide they produce (Shelton et al. 2018; Hazra et al. 2015). For example, *Eubacterium limosum* and *Acetobacterium woodii* contain *bzaABCDE* and produce cobalamin, *Clostridium formicoaceticum* is reported to produce both [5-OMe-6-MeBza]Cba and cobalamin and is predicted to possess *bzaABCDE* (Hazra et al. 2015; E. Stupperich, Eisinger, and Schurr 1990), *Moorella thermoacetica* has *bzaABC* and produces 5-methoxybenzimidazoyl-cobamide ([5-OMeBza]Cba, Factor III_m) (Wurm, Renz, and Heckmann 1980; Irion and Ljungdahl 1965), and *Geobacter sulfurreducens* has only *bzaF* and produces 5-hydroxybenzimidazoyl-cobamide ([5-OHBza]Cba, Factor III) (Hazra et al. 2015) (Figure 1.6A). Hence, the three methyltransferases BzaC, BzaD, and BzaE appear to contribute

to the cobamide diversity that varies in their state of methylation of their lower ligands (Figure 1.6A, 1.6D).

Alongside the discovery of the *bza* operon, the first reaction in the pathway was validation through biochemical characterization of the enzyme BzaF from *Desulfuromonas acetoxidans* (Mehta et al. 2015). The enzyme BzaF shares ~ 40% identity with the enzyme ThiC (HMP-synthase) which facilitates the rearrangement of AIR to produce Hydroxymethylpyrimidine-phosphate, a precursor of vitamin B_1 biosynthesis. ThiC is an exceptional member of the radical SAM family which lacks the signature CX_3CX_nC consensus but retains the core TIM barrel fold of radical-SAM superfamily(Vey and Drennan 2011). The ThiC and BzaF contain a [4Fe-4S] cluster which initiates the reaction by generating 5'-deoxyadenosineradical (5'-dAdo) which initiates the rearrangement reaction (Gagnon et al. 2018). In *Eubacterium limosum*, the equivalent reaction is catalyzed by the gene products of *bzaA* and *bzaB*.

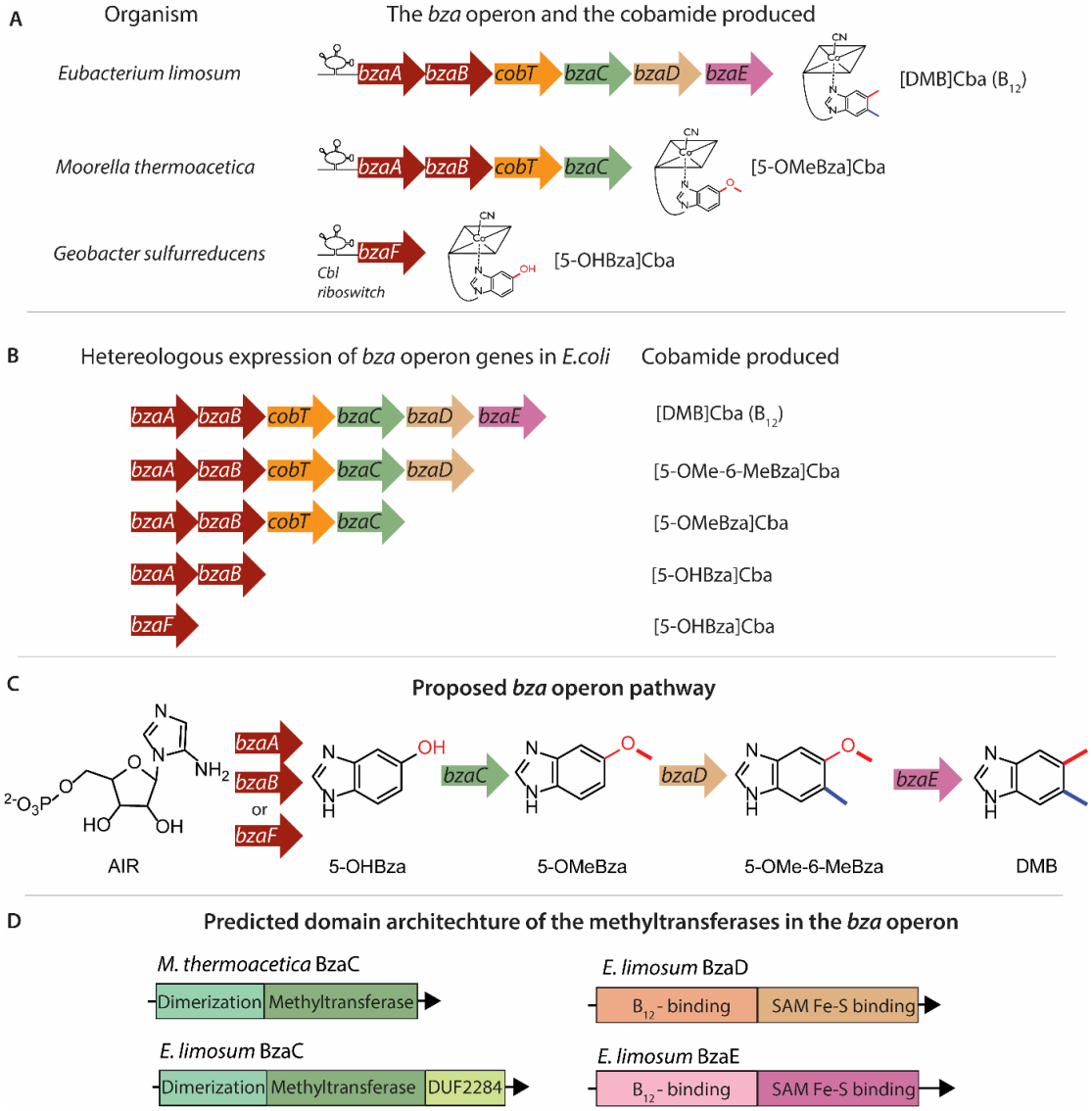


Figure 1.6 Anaerobic biosynthesis of DMB. A) The *bza* operon conducts anaerobic biosynthesis of benzimidazolyl lower ligand. The obligate anaerobe *Eubacterium limosum* contains genes *bzaA-bzaB-cobT-bzaC-bzaD-bzaE* in the *bza* operon which results in production of DMB as the lower ligand. *Moorella thermoacetica* has *bzaA-bzaB-cobT-bzaC* which results in biosynthesis of 5-OMeBza as the final lower ligand. *Geobacter sulfurreducens* encodes only *bzaF-cobT* and produces 5-OHBza as lower ligand. **B)** The function of *bza* operon genes was validated using a heterologous expression system in *E.coliΔmetE* strain which is a B_{12}-dependent methionine auxotroph. Expression of *bzaF* or *bzaA-bzaB* genes led to [5-OHBza]Cba, *bzaA-bzaB-cobT-bzaC* led to [5-OMeBza]Cba, *bzaA-bzaB-cobT-bzaC-bzaD* led to [5-OMe-6MeBza]Cba, and finally *bzaA-bzaB-cobT-bzaC-bzaD-bzaE* led to B_{12}. Thus, underlining the function of *bzaC*, *bzaD*, and *bzaE* gene products as methyltransferases. **C)** As per the *bza* operon pathway predicted by (Hazra et al. 2015), purine biosynthesis 5'-aminoimidazole ribonucleotide (AIR) is rearranged to 5-hydroxybenzimidazole (5-OHBza) in a radical-SAM mechanism by BzaF or BzaAB complex (Mehta et al. 2015; Gagnon et al. 2018a). Subsequently, the *bzaC* gene product was predicted to catalyze an O-methylation to produce 5-OMeBza, *bzaD* gene product was predicted to add a methyl group at C6, and lastly the *bzaE* gene product was predicted to catalyze an unprecedent demethoxylation and a methylation at C5. (Hazra et al. 2015). **D)** The predicted domain arhictechtures of the methyltransferases of the *bza* operon are shown. The two commonly found domain architectures of BzaC homologs differ by presence of additional domain of unknown function 2284 (DUF2284) at the C-terminal. The BzaD and BzaE show a domain architecture commonly associated with B_{12}-radical-SAM enzymes wherein the enzymes show a B_{12}-binding domain followed a SAM and Fe-S cluster binding domain. The presence of B_{12} binding domain, typically associated with enzymes that use B_{12} as cofactor, is intriguing and pose a chicken-and-egg like question.

The next gene in the *bza* operon is *cobT*, which encodes for an a-phosphoribosyltransferase (Trzebiatowski and Escalante-semerena 1997; Crofts et al. 2013; Cheong, Escalante-Semerena, and Rayment 2001; Hazra et al. 2013). The CobT class of enzymes from aerobes and facultative anaerobes can activate a range of derivatives of benzimidazoles, purines, and phenols. However, the precise function of CobT in the anaerobic biosynthesis of DMB remained unknown.

Finally, the *bza* operon codes for three methyltransferases which appear to catalyze three sequential but chemically distinct methylations (Figure 1.6C). The first methyltransferase- BzaC, is a predicted SAM-dependent O-methyltransferase (Figure 1.6D). We observe two domain variant of BzaC, one from Moorella thermoacetica which contains a dimerization domain and a methyltransferase domain; second from Eubacterium limosum which contains an additional domain of unknown function, DUF2284 (Figure 1.6D). The next two, BzaD and BzaE, are predicted to be B_{12}-binding radical-SAM methyltransferases (Figure 1.6D). The chemistry of the reactions involving BzaD and BzaE are unprecedented in biological chemistry and promise to reveal unique mechanisms of methylation.

Owing to the key role ascribed to cobamides in shaping microbial communities, the development of computational approaches to identify *bza* operon genes in metagenomic datasets to predict cobamide diversity in a community is starting to gain interest (Danchin and Braham 2017; Chittim, Irwin, and Balskus 2018). To confidently use the *bza* gene sequences as a proxy for the cobamide produced, the enzymatic activity and mechanism of each Bza enzyme need to be elucidated experimentally.

1.5 Open question addressed in the thesis

The anaerobic pathway for DMB biosynthesis was a longstanding question in the field of B_{12} biosynthesis, and the discovery of the *bza* operon in 2015 provided the genetic basis for how anaerobic bacteria make DMB and three other benzimidazole derivatives as lower ligand. Following the discovery of the pathway, the biochemical characterization of HBI synthase coded by the gene *bzaF*, which catalyzes rearrangement of AIR into 5-OHBza, the first dedicated intermediate in the pathway. Next, the biochemical nature and mechanism of the next enzymes of the pathway- the gene products of *cobT*, *bzaC*, *bzaD*, and *bzaE* remained to discovered. In my thesis, we address the following specific questions related to the *cobT*, *bzaC*, *bzaD*, and *bzaE* genes of the *bza* operon:

a. What is the precise function of cobT in the bza operon?

Previous literature show that CobT activates the lower ligands prior to the cobamide assembly(Trzebiatowski and Escalante-semerena 1997; Zayas and Escalante-Semerena 2007).

However, CobT homologs accommodate a wide range of substrates with varying substrate preference and orientation in which lower ligands attached to the ribose ring, and thus influence the diversity of cobamides produced by an organism (Hazra et al. 2013; Crofts et al. 2013; 2014). Notably, the *bza* operon pathway is predicted to make four benzimidazolyl molecules with subtle differences of one or two methyl groups, and all the intermediates can potentially act as CobT substrates. Thus, the inherent substrate preference of the *bza* operon CobT homolog will impact the final cobamide(s) resulting from the pathway. To uncover the contribution of CobT in the *bza* operon pathway, we aim to systematically study the reaction catalyzed by *bza* operon CobT with benzimidazoles found as intermediates in the anaerobic biosynthesis of DMB. We also discover that certain anaerobes have multiple *cobT* homologs in varying gene neighborhoods. While it is not uncommon for bacterial genomes to encode for multiple genes that lead to proteins of the same class, however, systematic studies to explore the functional differences among such homologs are rare in literature. We thus characterize the gene products of *cobT* encoded in different gene neighborhoods to reveal the how such homologs might contribute to cobamide diversity among anaerobic bacteria.

b. Reconstitution and mechanistic analysis of BzaC, BzaD, and BzaE methyltransferases

Based on heterologous expression of *bza* operon and its synthetic variants, the gene products of *bzaC*, *bzaD*, and *bzaE* genes were predicted to catalyze three sequential and chemically distinct methylations (Hazra et al. 2015). Anaerobic bacteria can harbor none, one, two, or all three predicted methyltransferases, and thus these genes contribute to the diversity of benimidazolyl cobamides with lower ligands that vary on the basis of methyl substituents. Further, we find that two domain variants of BzaC homologs exists in nature wherein a fraction of homologs contain an additional domain of unknown function called the DUF2284. We begin with the reconstitution of the methyltransferase activity of BzaC, BzaD, and BzaE that would enable us to find the true substrates and cofactors utilized by the enzymes, which would be foundational to mechanistic analyses of the reactions, and understanding how reaction proceeds under physiological conditions. Further, we conduct biochemical analysis of DUF2284 to understand the function of this domain in anaerobic DMB biosynthesis pathway. Mechanistic validation of these enzymes and the role of the cofactors involved will help in understanding anaerobic B_{12} biosynthesis and also lead the way for development of reliable tools to predict lower ligands formed in important microbial communities such as the human gut microbiome.

Chapter 2: Probing the functional variations in CobT encoded in different gene neighborhoods

In aerobes and facultative anaerobes, the *cobT* gene is typically localized with genes

involves in the nucleotide loop assembly or corrin ring biosynthesis(Zayas and Escalante-Semerena 2007). However, the recent discovery of *bza* operon unveils presence of a *cobT* gene with the genes involved in the lower ligand biosynthesis (Hazra et al. 2015). Previous work from our group, also uncovered the existence of multiple *cobT* genes in *Eubacterium limosum* and *Moorella thermoacetica* (Datar and Hazra 2018). We began our studies with exploring the occurrence of multiple *cobT* genes in anaerobes and classify the *cobT* homologs into three groups based on the corresponding gene neighborhoods. We systematically study the differences in activity of the different CobT encoded in different gene neighborhoods, and we initiate molecular evolution experiments to find sequence level differences that render functional differences among the homologs. We analyze unique enzymatic features of CobT homolog encoded within the *bza* operon to precisely understand their role in the *bza* operon pathway.

Chapter 3: Identification and characterization of CobT homologs in eukaryotes

During the course of bioinformatics efforts to map the genomic distribution of bacterial *cobT*, a small fraction of eukaryotes that encode a putative *cobT* were discovered. This discovery was surprising because B_{12} biosynthesis is carried by only bacteria and archaea(Sokolovskaya, Shelton, and Taga 2020). To probe the relevance of this discovery, we experimental characterized of a subset of the eukaryotic CobT homologs to test whether these are indeed functional enzymes that activate lower ligands for cobamide biosynthesis. We then conducted a comprehensive comparative genomics analysis of other genes involved in cobamide assembly, uptake, and utilization in these eukaryotes. Collectively, these findings complement recent discoveries of cobamide assembly and remodeling in some marine algae (Baum et al. 2020), and thus challenges the known taxonomic limits of B_{12} biosynthesis enzymes. The work presented here provides a starting point for extensive explorations on physiological function of eukaryotic CobT enzymes as well as the biological significance of eukaryotes empowered to attach lower ligands with respect to the natural environments and microbial communities they inhabit.

Chapter 4: Reconstitution of enzyme activity of *Mt*BzaC and understanding its role in the *bza* operon pathway

We begin our explorations on validation and mechanistic analyses of the methyltransferases of the *bza* operon, with *in vitro* characterization and reconstitution of BzaC, the predicted O-methyltransferase. In this chapter, I describe primary characterization of BzaC homologs from *Moorella thermoacetica* (*Mt*BzaC) which contains a dimerization domain and methyltransferase domain. We biochemically verify the function of the predicted domains and

test a range of physiologically relevant substrates for *Mt*BzaC. We validate *Mt*BzaC as 5-hydroxybenzimidazole ribose: SAM O-methyltransferase (HBIR: SAM OMT). Finally, we combine our findings from characterization of *Moorella thermoacetica* CobT (*Mt*CobT) and *Mt*BzaC to identify a new intermediate and propose a revision to the previously proposed *bza* operon pathway.

Chapter 5: Investigations to decipher the function of DUF2284 and the last two methyltransferases in *bza* operon

The BzaC homolog from *Eubacterium limosum* (*El*BzaC) contains an additional C-terminal DUF2284 domain which is absent in the *Mt*BzaC homolog. We initiate characterization of DUF2284 using bioinformatic analysis of DUF2284 sequences and the comparative genomics of bza operon in all organism that harbor a *bzaC* gene. Analysis of the sequence allows us to biochemically characterize DUF2284 using *El*BzaC and two other mutant constructs as described in the chapter 5. Surprisingly, the comparative genomics analysis indicates that presence of DUF2284 is correlated with that of *bzaD* and/ or *bzaE*, suggesting a possible interaction of DUF2284 with the other two predicted methyltransferases BzaD and/or BzaE. We additionally conducted bioinformatic analysis to better understand the scope of biochemical nature and function of DUF2284 in bacteria and archaea. We then begin with biochemical characterization of BzaD with optimization of methods to obtain the predicted substrates and the enzymes. Our comparative genomics study unveils diversity in the *bza* operon composition among anaerobic bacteria, which strengthens the principle of using of the *bza* operon to predict lower ligand biosynthesis, enabling insights into cobamide diversity within microbial communities.

Publications arising from this chapter:

Some parts of the chapter are communicated as review-
Methylations in B_{12} biosynthesis and catalysis, **Yamini Mathur**, Amrita B. Hazra, Current Opinion in Structural Biology (*in revision*)

References:

Baum, Christoph, Riya C. Menezes, Aleš Svatoš, and Torsten Schubert. 2020. "Cobamide Remodeling in the Freshwater Microalga Chlamydomonas Reinhardtii." FEMS Microbiology Letters 367 (20): 1–6. https://doi.org/10.1093/femsle/fnaa171.

Crofts, Terence S., Amrita B. Hazra, Jennifer LA Tran, Olga M. Sokolovskaya, Vadim Osadchiy, Omer Ad, Jeffrey Pelton, Stefan Bauer, and Michiko E. Taga. 2014. "Regiospecific Formation of Cobamide Isomers Is Directed by CobT." Biochemistry 53 (49): 7805–15. https://doi.org/10.1021/bi501147d.

Crofts, Terence S., Erica C. Seth, Amrita B. Hazra, and Michiko E. Taga. 2013. “Cobamide Structure Depends on Both Lower Ligand Availability and CobT Substrate Specificity.” Chemistry and Biology 20 (10): 1265–74. https://doi.org/10.1016/j.chembiol.2013.08.006.

Datar, Prathamesh M., and Amrita B. Hazra. 2018. “Investigating the Regioselective Attachment of the Lower Ligand in Vitamin B12 Biosynthesis.” IISER Pune. http://dr.iiserpune.ac.in:8080/xmlui/handle/123456789/1023.

Gagnon, Derek M., Troy A. Stich, Angad P. Mehta, Sameh H. Abdelwahed, Tadhg P. Begley, and R. David Britt. 2018. “An Aminoimidazole Radical Intermediate in the Anaerobic Biosynthesis of the 5,6-Dimethylbenzimidazole Ligand to Vitamin B12.” Research-article. Journal of the American Chemical Society 140 (40): 12798–807. https://doi.org/10.1021/jacs.8b05686.

Hazra, Amrita B., Andrew W. Han, Angad P. Mehta, Kenny C. Mok, Vadim Osadchiy, Tadhg P. Begley, and Michiko E. Taga. 2015. “Anaerobic Biosynthesis of the Lower Ligand of Vitamin B12.” Proceedings of the National Academy of Sciences of the United States of America 112 (34): 10792–97. https://doi.org/10.1073/pnas.1509132112.

Hazra, Amrita B., Jennifer L.A. Tran, Terence S. Crofts, and Michiko E. Taga. 2013. “Analysis of Substrate Specificity in CobT Homologs Reveals Widespread Preference for DMB, the Lower Axial Ligand of Vitamin B12.” Chemistry and Biology 20 (10): 1275–85. https://doi.org/10.1016/j.chembiol.2013.08.007.

Kennedy, Kristopher J., and Michiko E. Taga. 2020. “Cobamides.” Current Biology 30 (2): R55–56. https://doi.org/10.1016/j.cub.2019.11.049.

Mattes, Theodoric A., Jorge C. Escalante-Semerena, Evelyne Deery, and Martin J. Warren. 2017. “Cobalamin Biosynthesis and Insertion.” In Encyclopedia of Inorganic and Bioinorganic Chemistry, R.A. Scott (Ed.), 1–24.

Mehta, Angad P., Sameh H. Abdelwahed, Michael K. Fenwick, Amrita B. Hazra, Michiko E. Taga, Yang Zhang, Steven E. Ealick, and Tadhg P. Begley. 2015. “Anaerobic 5-Hydroxybenzimidazole Formation from Aminoimidazole Ribotide: An Unanticipated Intersection of Thiamin and Vitamin B12 Biosynthesis.” Journal of the American Chemical Society 137 (33): 10444–47. https://doi.org/10.1021/jacs.5b03576.

Renz, P. 1993. “Biosynthesis of Vitamin B12 in Anaerobic Bacteria” 1121: 1117–21.

Rodionov, Dmitry A., Alexey G. Vitreschak, Andrey A. Mironov, and Mikhail S. Gelfand. 2003. “Comparative Genomics of the Vitamin B12 Metabolism and Regulation in Prokaryotes.” Journal of Biological Chemistry 278 (42): 41148–59. https://doi.org/10.1074/jbc.M305837200.

Roth, J R, J G Lawrence, and T A Bobik. 1996. “Cobalamin (Coenzyme B12): Synthesis and Biological Significance.” Annual Review of Microbiology 50: 137–81.

Sokolovskaya, Olga M., Amanda N. Shelton, and Michiko E. Taga. 2020. “Sharing Vitamins: Cobamides Unveil Microbial Interactions.” Science 369 (6499). https://doi.org/10.1126/science.aba0165.

Trzebiatowski, Jodi R, and Jorge C Escalante-semerena. 1997. “Purification and Characterization of CobT , the Nicotinate- Mononucleotide : 5 ,6-Dimethylbenzimidazole Phosphoribosyltransferase Enzyme from Salmonella Typhimurium.” Journal Biological Chemistry 272 (28): 17662–67.

Zayas, Carmen L., and Jorge C. Escalante-Semerena. 2007. “Reassessment of the Late Steps of Coenzyme B12 Synthesis in Salmonella Enterica: Evidence That Dephosphorylation of Adenosylcobalamin- 5′-Phosphate by the CobC Phosphatase Is the Last Step of the Pathway.” Journal of Bacteriology 189 (6): 2210–18. https://doi.org/10.1128/JB.01665-06.

Chapter 2

Probing the enzyme function of CobT homologs encoded in different gene neighborhoods

2.1 Introduction

2.1.1 CobT activates lower ligands prior for the cobamide assembly

The CobT enzyme catalyzes the formation of a unique α-glycosidic bond between C1' of the ribose-phosphate and one of the heteroatoms of the lower ligand substrates (Trzebiatowski and Escalante-semerena 1997; Cheong, Escalante-Semerena, and Rayment 2001). The resulting product of this reaction is a α-riboside phosphate derivative of the lower ligand, called the activated lower ligand, which enters the last steps of cobamide assembly (Zayas and Escalante-Semerena 2007) (Figure 2.1). Thus, in biosynthesis of B_{12}, the CobT enzyme activates 5,6-dimethylbenzimidazole (DMB **1**) to produce α-DMB-riboside phosphate (DMB-RP **2**) (Figure 2.1A). CobT can also activate adenine (Ade **3**) to α-adenine-riboside phosphate (Ade-RP **4**) which is an epimer of the ribonucleotide β-adenosine monophosphate (AMP). The attachment of Ade-RP **4** leads to biosynthesis of pseudocobalamin (pB_{12}) (Figure 2.1B). A subclass of CobT called ArsAB is a heterodimeric complex which activates phenolic lower ligands by creating an α-O-glycosidic bond (Chan and Escalante-Semerena 2011) (Figure 2.1C).

2.1.2 CobT: one of the guardians of cobamide diversity in the microbial world

The range of cobamides produced by microorganism depends on two major factors. First, the availability of lower ligands which can either be produced or acquired by the organism. Second, lower ligand substrates that can be activated and attached to the corrin ring by the nucleotide loop assembly pathway (Crofts et al. 2013; Keller et al. 2014). Once a lower ligand substrate is produced or acquired, CobT acts as first line of screening for whether that molecule can enter the cobamide assembly. Typically, CobT homologs can activate multiple substrates, however, the range and relative preference of substrates vary across the homologs (Hazra et al. 2013). Previously, bacterial homologs of CobT have been shown to activate various benzimidazoles, purines, and phenols (Hazra et al. 2013; Crofts et al. 2013) (Figure 2.2A). However, substrates preference and promiscuity of a CobT homolog determine the structure of cobamides produced by an organism. For example, *Salmonella enterica* produces psuedocobalamin as the native cobamide. However, due to higher affinity of *Salmonella*

enterica CobT (*Se*CobT) for DMB over adenine, the organism predominantly makes B_{12} when DMB is present in growth media (Anderson et al. 2008; Hazra et al. 2013).

The physiological source of the phosphoribosyl group is still unknown and under in vitro conditions, CobT utilizes pyridyl biomolecules α-NMN, α-NaMN, α-NAD and α-NaAD as phosphoribosyl donors (Figure 2.2B). Previous studies have attempted to understand role of all four cofactors as CobT co-substrate. The rate of activity for *Propionibacterium freudenreichii* CobT (*Pf*CobT) with NaMN is 7 folds higher than its activity with NMN (Fyfe and Friedmann 1969; Trzebiatowski and Escalante-Semerena 1997). The activity of *Clostridium sticklandii* CobT (*Cs*CobT) is 860 and 90 folds higher with NaMN than with NMN to activate benzimidazole and adenine respectively (Fyfe and Friedmann 1969). In the same study, NaAD was shown to support *Cs*CobT-catalyzed activation of benzimidazole 100-folds better than NAD. Similar cofactor preference patterns were recorded with *Se*CobT where the apparent $K_{m(SeCobT:NMN)}$ was 44-folds higher than $K_{m(SeCobT:NaMN)}$, however the net catalysis efficiency (k_{cat}) remains comparable between the phosphoribosyl donors (Trzebiatowski & Escalante-Semerena, 1997). The structure of *Se*CobT bound to the products, DMB-RP and nicotinate shows no apparent structural reasons for why CobT shows higher affinity for NaMN over NMN (Cheong et al., 1999). Later, *Se*CobT additionally shown to transfer ADP-ribosyl group from NaAD and NAD to DMB to produce 5,6-dimethylbenzimidazole-α-dinucleotide (α-DAD) which also fluxes into the cobamide biosynthesis pathway (Figure 2.2B). However, the K_m value for *Se*CobT with NaMN is 17 times higher and the catalytic efficiency is ~400-fold higher than that of NAD (Maggio-Hall & Escalante-Semerena, 2003). Later, the *Sinorhizobium meliloti* CobT (*Sm*CobU) was also shown to transfer phosphoribosyl from NaMN, NMN and ADP-ribosyl from NaAD and NAD resulting in DMB-RP and α-DAD respectively (Hazra et al., 2013). In general, CobT homologs tested with other benzimidazolyl substrates and adenine showed that only NaMN and NMN gave products for all the bases (Hazra et al., 2013).

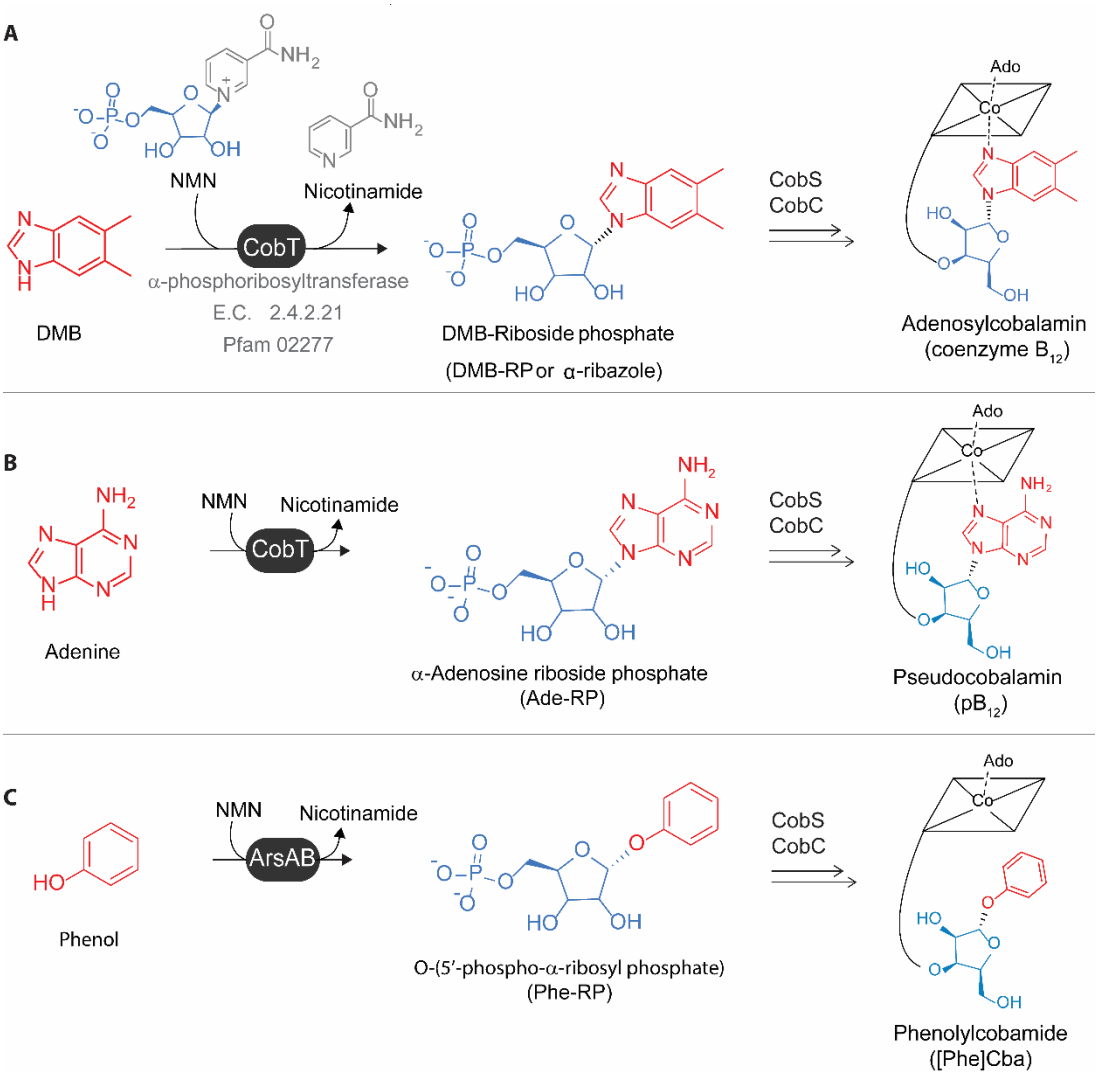


Figure 2.1 Lower ligand activation by CobT. A) In vitamin B_{12} biosynthesis pathway, as studied in aerobes and facultative anaerobes, the CobT enzyme catalyzes formation of an α-glycosidic bond between the nitrogen of DMB and the C1 of a ribose moiety resulting in DMB-riboside phosphate (DMB-RP or α -ribazole) as the product. The DMB-RP, the activated lower ligand, is utilized as substrate for last two steps of cobalamin assembly catalyzed by enzymes CobS and CobC. The phosphosribosyl group can be derived from either Nicotiamide mononucleotide (NMN, shown here) or Nicotinic acid mononucleotide (NaMN). CobT is classified as DMB: NaMN phosphoribosyltransferase under the protein family pf02277 and E.C. number 2.4.2.21. **B)** CobT can also activate adenine as a substrate resulting in an a-adenosine monophosphate (Ade-RP), which is an epimer of the ribonucleotide b-adenosine monophosphate (AMP). The cobamide with adenine as a lower ligand is called pseudocobalamin (pB_{12}) as it cannot be utilized as cofactor by humans, however, pB_{12} is a native and functional cobamide cofactor for several microorganisms. **C)** A subclass of CobT enzymes form heterodimeric ArsAB (Aromatic ribotide synthase) complex which activate phenolic lower ligands by creating an α-O-glycosidic bond. The phenolic cobamides lack a second heteroatom to coordinate with cobalt ion and hence, always remain in the base-off conformation.

Another feature of CobT that influences the structure of cobamide produced is the regioselectivity of lower ligand activation (Crofts et al. 2014). In case of benzimidazoles and purines, either of the two Nitrogens in the five membered rings can participate in glycosidic bond (Figure 2.3). Thus, asymmetric substrates such as 5-hydroxybenzimidazole (5-OHBza **8**), 5-methoxybenzimidazole (5-OMeBza **11**), and adenine (Ade **3**) can yield two isomeric riboside

phosphate products which in turn can lead to two isomeric cobamides (Figure 2.3A, 2.3B). Interestingly, CobT homologs show differences in ratio of the two products and hence, the regioselectivity of the reaction depends on the intrinsic nature of the proetin. For example, with 5-OMeBza as a substrate the *Se*CobT is regioselective for production of 5-methoxybenzimidazole-riboside phosphate (5-OMeBza-RP **12**) and the *Veillonella parvula* CobT (*Vp*CobT) shows regioselective formation of 6-OMeBza-RP (**13**). In agreement to the *in vitro* finding, *S. enterica* preferentially produces 5-methoxybenzimidazolylcobamide ([5-OMeBza]Cba) and the *V. parvula* favors production of 6-methoxybenzimidazolylcobamide ([6-OMeBza]Cba) when 5-OMeBza is fed to the respective organisms. Further, the regioselectivity of CobT homologs is sensitive to subtle differences in the substrate structure. For instance, *Se*CobT reacts with 5-OHBza to yield nearly equal amounts of the 5-hydroxybenzimidazole-riboside phosphate (5-OHBza-RP **9**) and 6-hydroxybenzimidazole-riboside phosphate (6-OHBza-RP **10**), in contrast to its regioselective activation of 5-OMeBza. The regioselectivity observed *in vitro* with CobT homologs agrees with the regioselectivity of CobT function under physiological conditions, and thus biochemical characterization of CobT homologs pose as reliable tools to study molecular drivers of cobamide diversity.

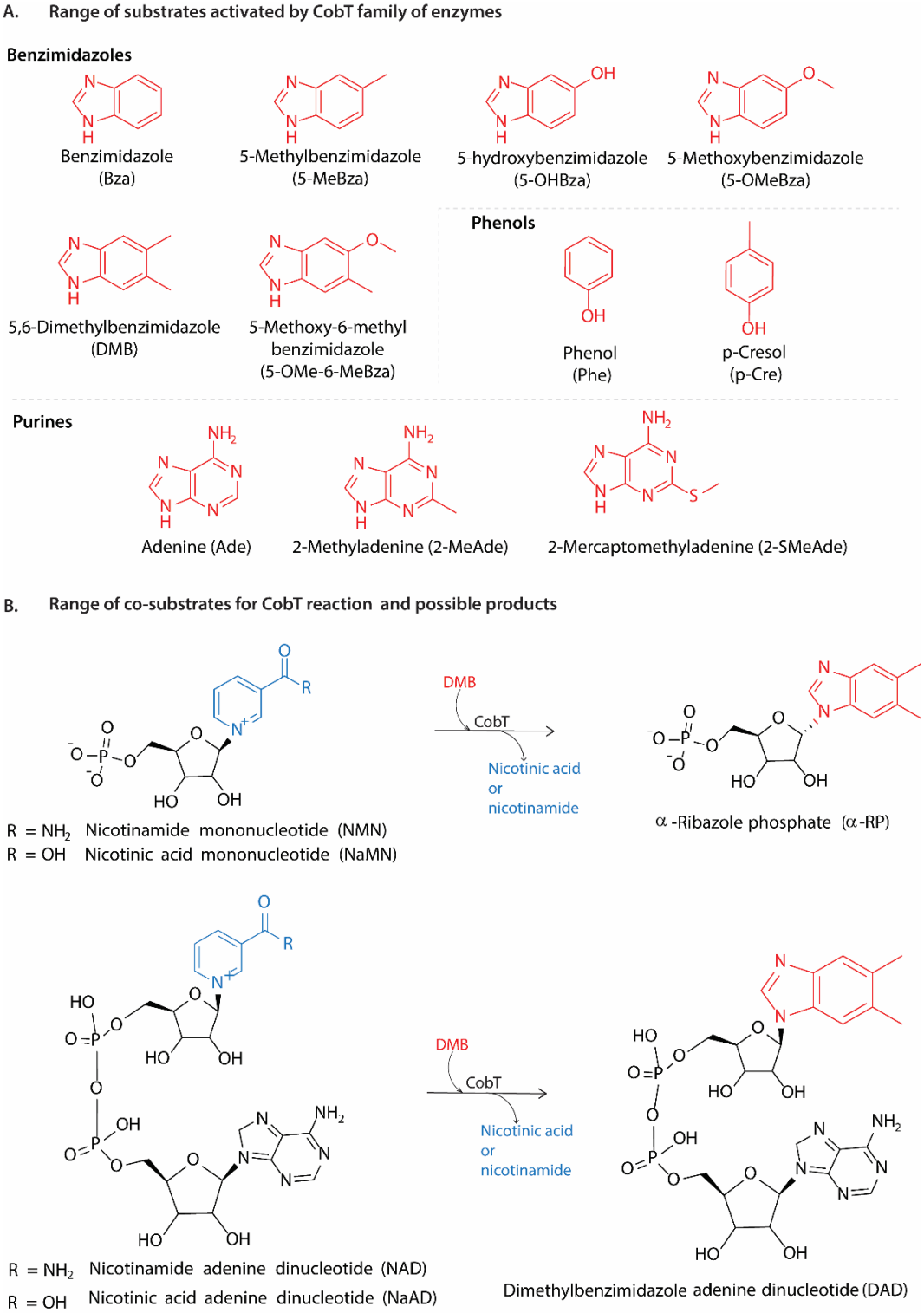


Figure 2.2 Diversity of substrates accommodated by CobT. The CobT enzymes activate a wide range of benzimidazoles, purines, and phenols. All these molecules are in turn found as lower ligands of naturally occurring cobamides. In this study, we tested CobT homologs with benzimidazole (Bza), 5-hydroxybenzimidazole (5-OHBza **9**), 5-methoxybenzimidazole (5-OMeBza **11**), 5-methylbenzimidazole (5-MeBza), 5,6-dimethylbenzimidazole (DMB **1**), and adenine (Ade **3**).

To sum up, the substrate specificity and regioselectivity of CobT, and the availability of the lower ligands cumulatively determine the diversity of cobamides that can be produced

by any microorganism (Cheong, Escalante-Semerena, and Rayment 2001; Chan et al. 2014; Hazra et al. 2013; Crofts et al. 2013; 2014; Chan and Escalante-Semerena 2011). The CobT homologs constitute a family of closely related versatile proteins which exhibit widespread differences in substrate specificity and regioselectivity. From previous studies on CobT homologs from diverse bacteria using *in vitro* assays as well as heterologous expression in bacterial systems, we distill the following two key points that inform our studies with this class of enzymes:

1. CobT homologs activate more than one substrate, and DMB is a widely preferred substrate
2. Most CobT homologs show differences in the regioselectivity with 5-OHBza and 5-OMeBza, which differ by only one methyl group. No homolog has been found to catalyze regiospecific activation of both the substrates.

The molecular factors and the signatures in the protein sequences that confer the functional variability among the CobT homologs is not yet extensively explored.

2.1.3 CobT is widely present in bacteria and archaea

In nature, microbes can either conduct *de novo* biosynthesis of cobamides or can salvage the structural components from its environment to assemble cobamides (Seth and Taga 2014). Several organisms use a third mode called remodeling wherein the organism imports a cobamide and replaces the lower ligand (Men et al. 2015). The lower ligand activation and cobamide assembly steps catalyzed by CobT and CobS respectively are central to all modes of cobamide biosynthesis. Hence, the *cobT* and *cobS* genes are widespread across bacterial and archaeal species (Ma, Tyrell, and Beld 2019; Sokolovskaya, Shelton, and Taga 2020; Shelton et al. 2018). The subclass ArsAB that activate phenolic lower ligands is limited to the bacterial class Negativiticutes of the phylum Firmicutes (Chan and Escalante-Semerena 2011; Shelton et al. 2018). An exceptional small group of microbes such as *Listeria innocua* import DMB-α-riboside using a transporter CblT which is subsequently phosphorylated by CblS to form DMB-RP and thus, it produces B_{12} in a CobT-independent manner (Gray and Escalante-Semerena 2010).

2.1.4 Anaerobes often have multiple *cobTs* within the same genome but at different locus

In organisms such as *Escherichia coli*, which is a cobamide salvager, the *cobT* gene is a part of the the *cobUTSC* operon. In cobamide producers such as *S. enterica* (makes pB_{12}), *and S. meliloti* (makes B_{12}) have *cobT* adjacent to other cobamide biosynthesis genes. The *arsA* and *arsB* genes are typically always found adjacent to one another and contains an overlap of about

8 base pairs. However, the gene neighborhood of an archaeal CobT from *Methanocaldococcus jannaschii* does not contain any B_{12}-related gene (data not shown here).

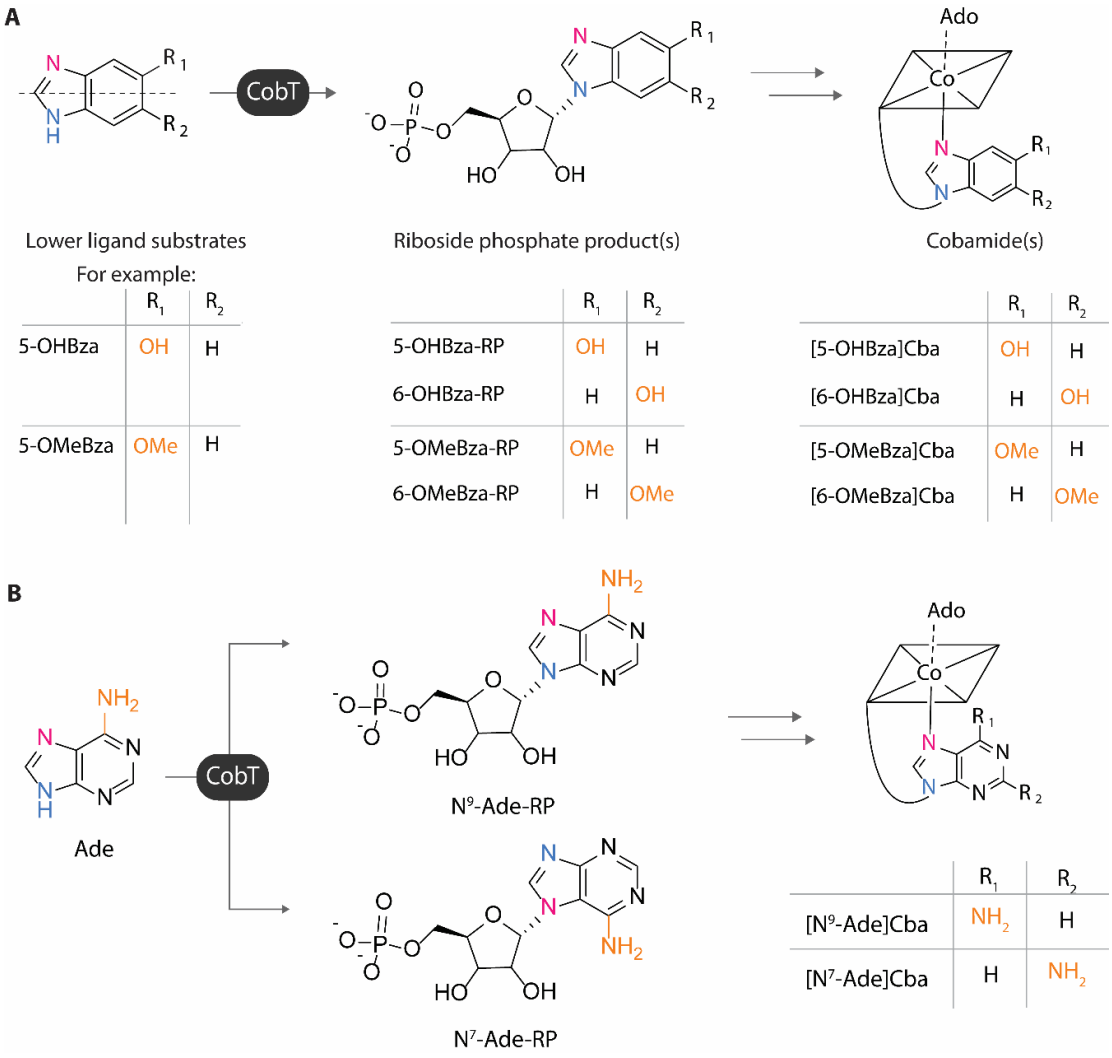


Figure 2.3 The regioselectivity of CobT reaction determines the cobamide structure. In benzimidazoles and purines, either of the two nitrogen in the five membered rings can participate in the reaction due to resonance and hence, localization of the hydrogen to be abstracted by the enzyme. As a result, from each asymmetric substrate two isomeric products that can be formed, each of which can be attached to the corrin ring thus resulting in two isomeric cobamides. **A)** Reaction scheme for activation of asymmetric benzimidazole derivatives such as 5-OHBza and 5-OMeBza. 5-OHBza can result in 5-hydroxybenzimidazole riboside phosphate (5-OHBza-RP) and 6-hydroxybenzimidazole riboside phosphate (6-OHBza-RP) as products. 5-OMeBza can result in 5-methoxybenzimidazole riboside phosphate (5-OMeBza-RP) and 6-methoxybenzimidazole riboside phosphate (6-OMeBza) as products. Each of the CobT products are known to be attached to the final cobamide. Thus, the regioselectivity of the CobT reaction determines the orientation in which the functional group of the lower ligand is attached, which in turn influences the biochemical properties of cobamide as a cofactor. **B)** Reaction scheme for activation of adenine with expected CobT products, N^9-adenine riboside phosphate (N^9-Ade-RP), and N^7-adenine riboside phosphate (N^7-Ade-RP) and cobamides resulting from each one of them.

In anaerobic bacteria, the *bza* operon was found to contain a *cobT* gene, however the precise reason for such localization remained unknown (Hazra et al. 2015). It should be noted that, aerobic pathway makes only one benzimidazole derivative i.e. DMB as the lower ligand

(Taga et al. 2007). On the contrary, the *bza* operon pathway makes four distinct benzimidazole derivatives as intermediates which all potentially can serve as physiological substrates for CobT (Hazra et al. 2015). Understanding the substrate preference and regioselectivity of *bza* operon CobT will aid in determine the true physiological substrate as well provide a better detail in the pathway mechanism. Further, we also find that certain anaerobic bacteria contain additional *cobT* genes often present in different gene neighborhoods. This observation opens a new avenue for exploring the function of CobT homologs encoded in varying gene neighborhoods.

In this study, we initiated bioinformatic and biochemical analyses to understand how the genomic context influences the function of CobT homologs encoded in different gene neighborhoods and explore why certain organisms have multiple copies of CobT (some parts of this study is reported in MS theses Prathamesh M Data 2018, Sheryl Sreyas 2020, and P Riya, 2022)(Sreyas and Hazra 2020; Riya and Hazra 2022; Datar and Hazra 2018). Here, we find three patterns in gene neighborhood of CobT homologs that is the *cobT* gene can be localized with the *bza* operon, the *cob* operon and/or with genes not related to B_{12} metabolism. We classify the CobT homologs from these genomic contexts as CobT1, CobT2, and CobT3, respectively. We biochemically characterized a subset of 10 homologs comprising and checked for substrate promiscuity, regioselectivity and preference of each homolog. To further understand the molecular basis of the functional differences in CobT homologs, we undertake a bioinformatic study to determine the amino acid residues that are show conservation among the three classes of CobT.

2.2 Materials and methods

The genomic DNA for *Eubacterium limosum* and *Eubacterium barkeri*, and the *Moorella thermoacetica cobT2* gene was a gift from Michiko E. Taga at University of California, Berkeley. The genomic DNA for *Desulfobacterium vacuolatum*, *Desulfotomaculum thermosubterraneum*, *Desulfosarcina cetonica*, and *Desulfobulbous mediterraneous* were purchased from RIKEN, Japan BRC through the National BioResource Project of the MEXT/AMED, Japan. DMB, 5-OMeBza, 5-OMeBza, 5-ClBza, Bza, Ade were purchased from Merck Sigma. 5-OHBza was synthesized in house as described in chapter 3. Molecular biology kits and enzymes required for cloning were purchased from Takara, Promega, and HiMedia as per availability. All antibiotics and media for bacterial cultures were purchased from TCI and HiMedia respectively. Resins and reagents for protein purification were procured from BioRad, GE Healthcare, and Merck Sigma.

2.3.1 Bioinformatics methods

We curated dataset for 124 CobT sequences by first finding CobT in organisms known to produce or utilize cobamides. To do so, a list of organisms was prepared by intensive

literature search, and then we used the CobT sequences from *Salmonella enterica*, *Sinorhizobium meliloti* (called *Sm*CobU in literature), *Eubacterium limosum* and *Moorella thermoacetica* as query sequences to run a BLASTp against the target list of organisms (S. S. F. Altschul et al. 1990). Next, we conducted similar search for CobT against all organisms that possess *bza* operon. We also added CobT sequences from randomly selected organisms for which much information on cobamide metabolism is not otherwise available. We then manually recorded the gene neighborhood of each CobT hit using NCBI GenBank and wherever necessary the identity of the gene was assigned using prediction from the Conserved Domain Database (CDD) (Benson et al. 2005; Marchler-Bauer and Bryant 2004). The gene neighborhood diagrams were prepared using Clinker (Gilchrist and Chooi 2021). Next, the curated set of CobT protein sequences were subjected to a phylogeny analysis which is described elsewhere (Sreyas and Hazra 2020). Additionally, a sequence similarity network (SSN) was generated using CobT sequences from bacteria, archaea, and cyanobacteria. Nicotinamide phosphoribosyltransferase sequences were used as a control group. The SSN was finalized with sequence lengths within the range of 300 to 400 amino acids and networks were calculated at alignment scores of 40, 76, 104, and 113. The networks were arranged in ylayouts organic format using CytoScape and nodes of the bacterial sequences were colored based on the gene neighborhood of the corresponding sequence (Gerlt 2017).

2.3.2. Prediction of Specificity determining positions (SDPs)

To calculate SDPs, a multiple sequence alignment (MSA) of 174 CobT protein sequences was generated using MAFFT on JalView and the output was converted to GDE format (Katoh et al. 2002; Procter et al. 2021; Kalinina et al. 2004). The sequences were sorted into three groups namely CobT1, CobT2, CobT3 corresponding to the three gene neighborhood categories as described above. The sequences were then subjected to SDP calculation on SDPpred1.0 (version 1 is now unavailable. Link to v2: http://www.bioinf.fbb.msu.ru/SDPpred/) with 1000 shuffling. The results above the Z-score were mapped onto the sequence alignment. The same MSA was also used for an analysis on ConSurf server which allows projection of sequence conservation across homologs onto a known 3D structure of the protein (Ashkenazy et al. 2016). The conservation pattern for CobT sequences across diverse bacteria was projected onto the structure of *Salmonella enterica* CobT bound to DMB-RP (PDB 1L4E) (Cheong, Escalante-Semerena, and Rayment 2001).

2.3.3 Cloning, overexpression, and purification of the protein

The candidate *cobT* genes were amplified from the genomic DNA of the interest and inserted into the pET28a vector between the BamHI and NdeI sites using restriction free cloning. The pET28a multiple cloning site provides an N-terminal 6x-histidine tag and IPTG

inducible protein overexpression under the T7 promotor. Briefly, the gene is amplified with a pair of primers that add 20 and 22 bp complementary to the sites of insertion in pET28a at the 5'- and 3'- end of the gene, respectively and the resulting amplicon is called a megaprimer. Next, we prepare three PCR cocktails containing varying molar ratios of megaprimer :vector of 0:1, 5:1, and 10: 1 and subject the mix to a 2-step RF-PCR as follows- 2 minutes at 95°C, 30 cycles at 95°C for 15 seconds and 68°C for 8 mins, final extension at 68°C for 20 mins, and infinite hold at 4°C. We treat the PCR mix with DpnI which removes the methylated parent plasmid and then transform the reaction product in *E.coli* DH5α strain. Transformants are selected on LB agar plates containing 25 μg/ml kanamycin containing 0.2% Glucose and then screened for gene insertion using colony PCR using primers for T7 promoter and T7 terminator regions. The gene sequences were verified by DNA sequencing.

A plasmid harboring the *cobT* gene of interest was transformed into *E.coli* BL21(DE3) cells and transformants were selected on LB agar containing 25 μg/ml kanamycin. A 10 ml of overnight starter culture in LB broth containing 25 μg/ml kanamycin was used to inoculate the secondary culture of 2L LB broth containing 25μg/ml kanamycin and incubated at 37°C till $O.D._{600nm}$ reaches 0.6. Next, protein overexpression was induced by addition of 250 μM IPTG and culture is incubated at 25μC for 12 hours. Cells were harvested by centrifugation at 5000 rpm at 4°C for 10 mins and stored in -80°C until further use.

The cell pellet from 2L culture was resuspended in 10mL of lysis buffer (50 mM phosphate, 300 mM NaCl, 5 mM imidazole, pH 8.0) along with 100 μM of protease inhibitor PMSF. Next, cell suspension was subjected to ultra-sonication on ice at 60% amplitude with '1s on- 3s off' cycle for 10 minutes. The cell lysate was centrifuged at 40,000g for 15 min at 4°C. The clarified lysate was loaded on a Ni-NTA column pre-equilibrated with lysis buffer. Next, to remove non-specifically bound proteins, the column was washed with 10 column-volumes of wash buffer (50 mM Potassium phosphate buffer, 300 mM NaCl, 50 mM Imidazole, pH 8.0). Finally, the 6x-His-CobT was eluted by passing elution buffer (50 mM Potassium phosphate buffer, 300 mM NaCl, 250 mM Imidazole, pH 8.0), and the eluent was collected in fractions of 1 ml each. The elution of protein was monitored using a qualitative and inexpensive Bradford assay- on a paraffin layer, aliquot multiple drops of 10μL of Bradford reagent and add 5μL of each elution fraction. The blue coloration of the mix indicates presence of protein in the fraction. We then analyze an aliquot of the protein fractions that show highest coloration in Bradford assay on SDS-PAGE gel. Lastly, we pool the fractions with highest concentration and purity of the purified protein.

To remove excess imidazole and salt from the protein, we desalt the pooled eluent using 10 kDa cut-off desalting columns. First, the desalting column was equilibrated with 12mL of desalting buffer (50mM phosphate, 150 mM NaCl and 7μM β-mercaptoethanol, pH 8.0). Then, 3 ml of protein was loaded onto the column and was allowed to flow through completely. Next, 3.5 mL of precooled desalting buffer was added on the column and the eluent which contained desalted protein was collected. The desalted protein was stored with 12% glycerol at -80 °C. The purified CobT shows long shelf with activity retained even after a year of storage at -80 °C. The purified protein was analyzed on SDS-PAGE gel (Figure 2.6).

To calculate the concentration of the purified protein, we dilute an aliquot 5 to 10 times and record its absorbance (A) at 280nm. The theoretical molar extinction coefficient (ε) was estimated using the ExPaSY Protparam tool (https://web.expasy.org/protparam/) (Wilkins et al. 2005) and the protein concentration was calculated using the Beer-Lambert equation [A= εcl; A= Absorbance, ε= molar extinction coefficient, c=concentration, l= path length of the cuvette]. Additionally, the concentration was measured via Bradford assay using Bovine serum albumin (BSA) as standard as described previously (Ernst and Zor 2010).

2.3.4 Reconstitution of CobT activity

The *in vitro* enzyme assay with CobT was set up as described in literature with slight modifications (Hazra et al. 2013; Crofts et al. 2014). Briefly, to buffer constituting 50 mM Tris-Cl pH 8.0 and 1 mM $MgCl_2$, we add 500 μM DMB, 10 μM purified enzyme and 2 mM NMN in the final volume of 250 μL. Controls without enzyme, NMN, and DMB respectively were also set up alongside. The reaction mixtures were incubated for 48 hours at 25°C.

To determine the substrate promiscuity, we test CobT activity with 5-OHBza, 5-OMeBza, 5-MeBza, Bza, and Adenine in reactions with 500μM substrate, 2mM NMN, 10μM purified CobT in 50mM Tris-Cl containing 10mM $MgCl_2$ at 25°C for 48 hours. For each substrate, we set up controls lacking the NMN, enzyme, and the substrate. Subsequently, the reactions were quenched with 1% formic acid and the precipitated enzyme was removed by centrifugation at 14,000 rpm for 20 minutes at 4°C. The supernatant of each sample was collected in fresh tube, neutralized by addition of 25 mM of NaOH and analyzed on reverse phase HPLC and LC-MS.

2.3.5 Analysis of CobT reaction using HPLC analysis coupled with UV-Vis and MS

HPLC analysis was conducted on Agilent 1260 infinity II UHPLC system paired with a UV-Vis DAD and FLD detectors. We used Agilent Zorbax Eclipse Plus- C18 XDB column 5μ 250mm x 4.60 mm column or Agilent Zorbax Eclipse Plus- C18 XDB column 5μ 150mm x 4.60 mm column as per availability and requirements of the reaction. The mobile phase was

composed of ammonium acetate, 10mM, pH 6.5 (solvent A) and methanol (solvent B). A flow rate of 0.5mL/min and the column temperature of 25°C were maintained throughout the analysis. For all HPLC and LCMS analysis the gradient was applied as follows: 0-2 min 100% A; 2-15 min from 100% A to 70% B; 15- 20 min hold at 70% B; 20.5-25 min to 100% B; 25.5-35min 100% solvent A. The chromatograms were recorded at 250 nm, 260 nm and 280 nm and UV-Vis spectra for all peaks were recorded from 190 nm to 390 nm at a step size of 2 nm.

The LC-MS analysis was conducted on ScieX X500R-QTOF Mass spectrometer system attached to Exion-LC series UHPLC. The MS analyses were conducted in +ESI mode with +4500 V ion spray voltage, at a medium de-clustering potential of 80V and a low collision energy of 5V at 400 °C. A two part MS experiment was optimized first, a TOF-MS for a mass range of 100Da to 1000Da and second, a targeted MRM-HR analysis using the following precursor ion mass and fragment ion mass for the analyte of interest as shown in table 1. All LC-MS data was subsequently processed on the ScieX-OS Explorer. The extracted ion chromatogram (EIC) was obtained using the expected mass with a width of ±0.02. The mass spectra of the reactants and products were derived from the total ion chromatogram (TIC).

Compound	Precursor ion mass $[M+H]^+$ (m/z)	Fragment mass $[M+H]^+$ (m/z)
NMN	334.0333	123.0553
Nicotinamide	123.0553	
DMB **1**	147.0917	123.0553
DMB-R	279.1339	
DMB-RP **2**	357.0857	
5-OHBza	135.0553	135.0553
5-OHBza-R and 6-OHBza-R	267.0975	
5-OHBza-RP and 6-OHBza-RP	347.0693	
5-OMeBza	149.0709	149.0709
5-OMeBza-R and 6-OMeBza-R	281.1132	
5-OMeBza-RP and 6-OMeBza-RP	361.0795	
Adenine **3**	136.0618	136.0618
Ade-R	268.104	
N^9-Ade-RP, N^7-Ade-RP **4, 5**	346.0558	

Table 2.1: The mass/charge (m/z) values used to monitor the CobT reaction.

2.3.6 Analysis of reaction

The reaction products are labile due to spontaneous removal of the phosphate group, thus the reaction turnover was monitored through consumption of substrate. Standard curves were generated for all the substrates in the range of 50 μM to 1.2 μM concentration. The end-

point assays were monitored on HPLC and area under the curve for substrates in the controls and the reactions were calculated using Agilent OpenLab CDS software. When required, the concentration of product formed was calculated by subtracting the concentration of remaining substrate at time point from the concentration of substrate in the no enzyme control or from zero-minute aliquot of the reaction mix. To calculate the regioselectivity, the area under the curve for each CobT product was determined at 280 nm or 260 nm as applicable.

2.3.7 Competition assays

In the pairwise competition assays with CobTs, we set up reactions with all combinations of the DMB, 5-OHBza, 5-OMeBza, and Ade lower ligands i.e for comparing preference among four substrates, we set up a series of six reactions. The stepwise method for competition between each pair of substrates is described as follows:

In a fresh microcentrifuge tube, we prepare a substrate pre-mix which is a equimolar mixture of the substrates that are competed say, A and B. Then, we prepare three fresh microcentrifuge tubes for acid-quenching the reactions by adding 2μL of 90% formic acid in each tube. In a fresh tube, the we add the reaction constituents - 50mM Tris-Cl, 10mM MgCl2, 1 mM each of substrate A and substrate B (using an equimolar substrate mix), 2 mM NMN and lastly, 5 μM of purified enzyme in a final volume of 500 μL on ice using pre-chilled reagents. We then promptly aliquot 150 μL as the 'zeroth time point' sample from the reaction mix and transfer to one of the acid-treated tubes. The rest of the reaction mix is incubated in a dry bath set at 25°C and 150 μL aliquots are transferred for quenching after 1 minutes, 30 minutes (or 3 hours) and an endpoint at 24 hours of incubation. The quenched reaction aliquots are centrifuged 14,000 rpm at 4°C for 20 mins and the 50 μL of supernatant is subjected to the HPLC analysis. The experiment is repeated with all combinations of substrates.

Quantitation and data analyses

To quantitate the reaction progress, we determine the depletion of substrate using the standard curves generated using area under curve on HPLC. The percent conversion is calculated by decrease in concentration of the substrate at a time point where activity is just beginning to be observed as compared to the 1 minute (a proxy for zeroth minute).

The relative preference is calculated as follows:

In a pairwise competition between substrates *A* and *B* in presence of enzyme *E*:

$$A + B \xrightarrow{E} A' + B'$$

The molar ratio (fold-preference), R_{AB} for substrate *A* w.r.t. to substrate *B* is calculated as follows:

$R_{AB} = \frac{[A']}{[B']}$; where $[A']$ and $[B']$ are the concentrations of products.

Next, molar ratio percent preference (R^{100}) which is the reciprocal of the molar ratio R_{AB} multiplied by hundred, i.e. for a A versus B competition, $R^{100}_{BA} = \frac{[B']}{[A']} \times 100$. This factor tells the moles of B' formed per hundred moles of A' when the substrates A and B are competed.

We then calculate the molar ratio percent preference for each combination of the competed substrates. Hence, in a scenario where a competition assay is conducted with four substrates, we calculate the preference of substrate *A* w.r.t. substrates B, C, D; we also calculate the $R^{100}_{CA} = \frac{[C']}{[A']} \times 100$, $R^{100}_{DA} = \frac{[D']}{[A']} \times 100$ values.

Next, we derive the relative percent preference for substrate A w.r.t. to the rest of substrates collectively, which paints a picture for how the enzyme uptakes substrate A when a mixture of all four substrates would be present.

$$\text{relative \% preference for A } (\mathrm{P}^A) = \frac{R_{AA}}{(R_{AA} + R_{BA} + R_{CA} + R_{DA})} \times 100$$

$$\because R_{AA} = 1;\ \mathrm{P}^A = \frac{1}{(1 + R_{BA} + R_{CA} + R_{DA})} \times 100$$

$$or\ \mathrm{P}^A = \frac{1}{(1 + \frac{[B']}{[A']} + \frac{[C']}{[A']} + \frac{[D']}{[A']})} \times 100$$; where $[B']$, $[C']$, $[D']$ are the concentrations of products for B, C, and D in individual competitions with A.

Similarly, to calculate the preference of B w.r.t. A, C, D substrates, we determine $R^{100}_{AB} = \frac{[A']}{[B']} \times 100$, $R^{100}_{CB} = \frac{[C']}{[B']} \times 100$, and $R^{100}_{DB} = \frac{[D']}{[B']} \times 100$ where $[A']$, $[B']$, $[C']$, $[D']$ are concentrations of products of A, B, C, D from individual A+B, B+C, and B+D competition reactions.

The relative percent preference for substrate B w.r.t. A, C, D would be:

$$or\ \mathrm{P}^B = \frac{1}{(\frac{[A']}{[B']} + 1 + \frac{[C']}{[B']} + \frac{[D']}{[B']})} \times 100.$$

The molar ratios, molar ratio percent preference, and the relative percent preference values are calculated for substrates C and D respectively. The P^A, P^B, P^C, P^D values add up to a 100 and can be analyzed as parts of whole values.

Note: Most CobT homologs showed poor turnover with 5-OHBza and Ade in the competition assays, such that the, the decrease in the area under the curve for substrate peak occurred within the error range. This often resulted in ΔAUC values in negative numbers. To resolve the negative values in our calculation, we monitored the formation of any product with the substrate, and we adjusted the negative values such that:

i. If concentration of product was estimated using its AUC against a standard curve using 1μM to 10 μM range of respective substrate.

ii. If no product was observed at A260 for Ade or A280 or fluorescence (λ_{ex}280: λ_{em}312) for 5-OHBza, we adjusted the product concentration as 1000 folds lesser than the product of the competitor substrate, which signifies a infinitesimally low preference for these substrates.

2.4 Results

2.4.1 Classification of CobT sequences based on gene neighborhood

Using previously characterized CobT homologs from bacteria and archaea as query sequences for BLASTp searches, we curated a dataset of 166 sequences. The archaeal CobT sequences, as also reported previously, show only 6-23 % identity with bacterial homologs, and form a distant clade upon phylogeny analysis (Sreyas and Hazra 2020). Since, the archaeal cobamide biosynthesis genes are only partially identified and annotated, we focussed on examining gene localization of *cobT* genes in bacterial genomes. Additionally, studying a small dataset of sequences allowed us to systematically probe minor differences in highly similar bacterial CobT homologs that give rise to functional differences.

Thus, we identified CobT homologs from 85 unique anaerobic bacteria using BLASTp searches and manually determines the gene neighborhood for each hit (Sreyas and Hazra 2020). We find that several of these organisms contain multiple genes encoding a CobT. Based on the genomic context observed for each *cobT* gene, we classify CobT homologs in three groups:

1. *cobT1*: the gene is localized within the *bza* operon which is responsible for biosynthesis of lower ligand (Figure 2.4A)
2. *cobT2*: the gene is localized either within the *cob* operon which is responsible for nucleotide loop assembly or is present near other cobamide biosynthesis genes (Figure 2.4B)
3. *cobT3*: gene is present in a gene neighborhood devoid of any B_{12} related genes, and do not show any patterns in conservation of this gene localization (Figure 2.4C)

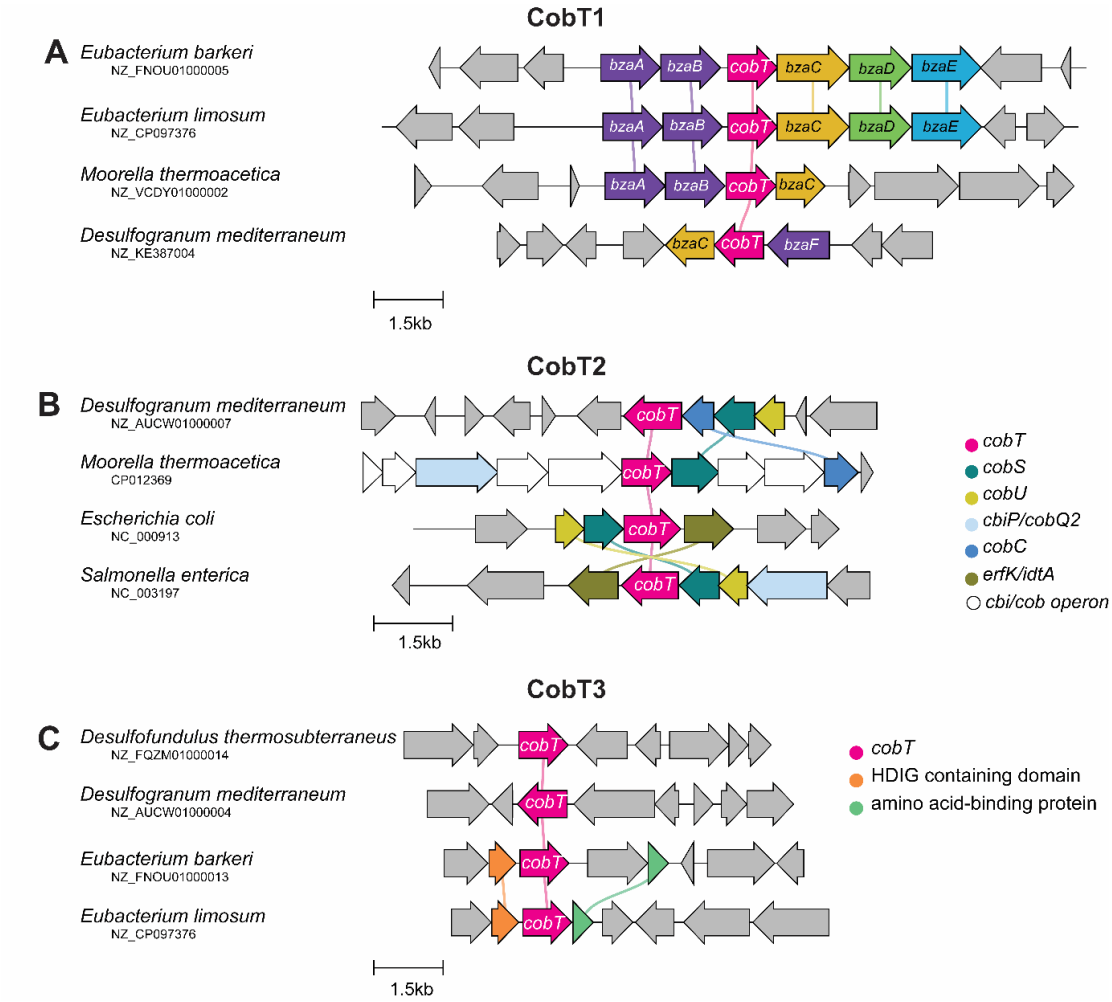


Figure 2.4 The patterns in genomic location of *cobT* in bacteria. The gene localization of CobT show three predominant patterns. **A)** In the *bza* operon, which is responsible for biosynthesis of lower ligands in anaerobes, *cobT* gene mostly follows the *bzaAB* or *bzaF* genes. We classify the CobT homologs encoded within the *bza* operon as **'CobT1'**. **B)** Typically, *cobT* gene is localized with the genes of cobamide assembly *cobS* and *cobU*, as seen in facultative anaerobes *Escherichia coli* and *Salmonella enterica*. We classify the *cob* operon *cobT* as the **'CobT2**'. In certain anaerobes too, a *cobT* gene is found in proximity of *cobS* and *cobU* (such as *Desulfogranum mediteeraneum*), and in some a *cobT* gene is found in a gene cluster coding for entire cobamide biosynthesis pathway (such as *Moorella thermoacetica*). **C)** Several organisms also contain a *cobT* gene in non-conserved neighborhood which does not contain any B_{12}-related genes in +/- ten genes. We classify all the *cobTs* from random neighborhoods as **'CobT3'.** The gene maps were drawn using Clinker. All *cobT* genes are shown in pink. The connecting line between the genes indicate that the translated proteins show >40% seuqence identity**.** Genes that occur in gene clusters from more than two organism can colored, and are annotated within the gene arrow in panel A and are described as legend for panel B and C. The alphanumeric identifier indicate the genome assembly/contig in NCBI GenBank corresponding to the gene clusters.

We also observe certain *cobT* homologs that additionally contain an alcohol dehydrogenase domain or an aminotransferase domain. We excluded such sequences from our study asthese proteins may or may not be relevant to the B_{12} biosynthesis pathway. CobT enzymes can create unique a-glycosidic bond and hence, elucidation of function of novel domain architectures with CobT domain might shed light onto more roles of this class in other metabolic pathways.

On a phylogeny tree, most of the CobT1 sequences appear to show higher conservation away from CobT2 and CobT3 which seem to be separated in multiple clades in agreement with taxonomic diversity (Figure 2.5A, 2.5B). A quantitative assessment of genomic localization of CobT homologs among anaerobic bacteria shows that most organisms that contain a *bza* operon CobT additionally contains in the three categories, indicate that most of the organisms (55%) that encode a *cobT* gene within the *bza* operon also has an additional *cobT* gene which is more commonly located in non-B_{12} related neighborhood. On the other hand, the occurrence of *cobT2*, that is *cobT* within the gene locus that encode for *cbi* and *cob* genes, is higher among the anaerobes that lack *bza* operon. The gene neighborhood of *cobT3* genes have no other B_{12}-biosynthesis genes are situated in the limits of ± five genes. In certain organisms, *cobT3* is placed under a cobalamin riboswitch and hence the function of *cobT3* in cobamide assembly pathways remains plausible (Figure 2.5C).

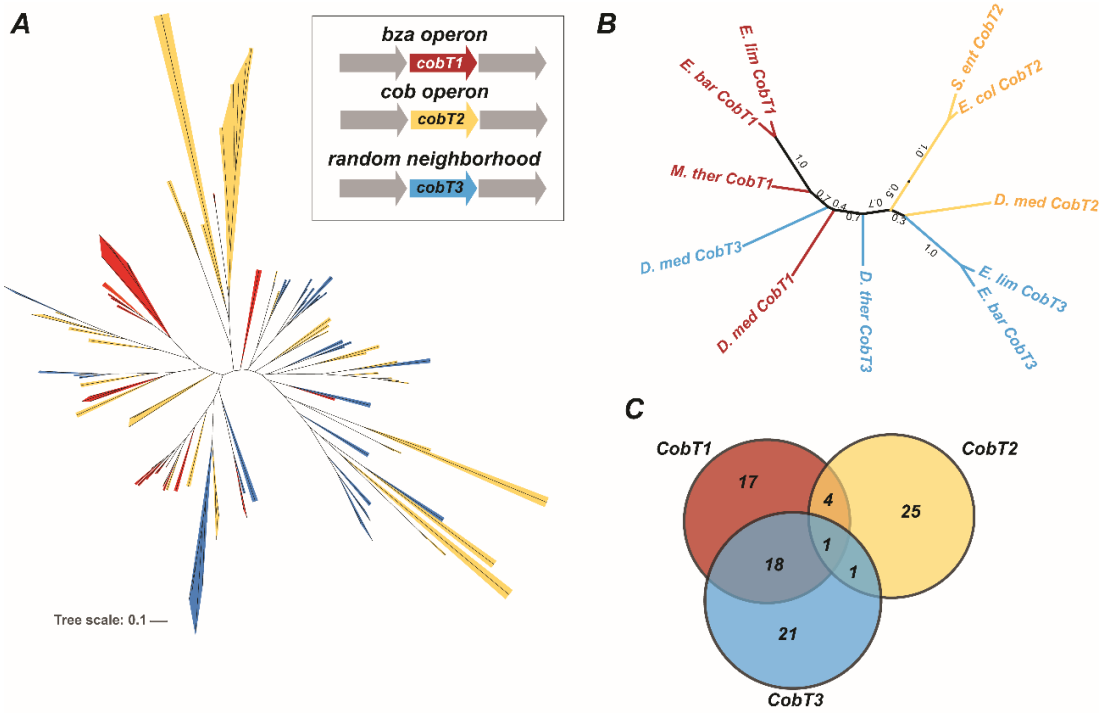


Figure 2.5. Phylogeny and occurrence of bacterial CobT homologs. A) An unrooted phylogeny tree with 174 CobT homologs from diverse bacterial taxa. The colored ranges indicate the gene neighborhood of the respective homolog- the CobT1 homologs which are bza operon are shown in red highlight, the CobT2 homologs which are encoded alongside *cob*/*cbi* genes are shown in yellow highlight, and the CobT3 homologs which occur in non-B_{12} related gene neighborhoods are shown in blue highlights. The CobT1 homologs sort in two major clades whereas the CobT2 and CobT3 sequence appear to be interspersed throughout the tree. **B)** A tree with previously characterized Salmonella enterica CobT (*Se*CobT) and the ten CobT homologs that we characterized in this study. **C)** The Venn diagram represents the occurrence and distribution of CobT homologs among 87 anaerobic bacteria. The numbers indicate the count of organisms having one or two or all three types of CobT homologs. Among the anaerobes that contain more than one *cobT* gene, the combination of *cobT1* and *cobT3* appear to be most prominent in our dataset. The organisms that lack a *bza* operon, either have a *cobT2* or *cobT3*, with an exception of only one organism.

We shortlisted the *cobT* homologs from the following organisms for biochemical characterization:

1. *Eubacterium limosum (El)* and *Eubacterium barkeri (Eb)* which contain a *cobT1* and *cobT3* and produce DMB using the full *bza* operon i.e. *bzaA-bzaB-cobT-bzaC-bzaD-bzaE* genes
2. *Moorella thermoacetica* (*Mt*) which produces [5-OMeBza]Cba and contain a *cobT1* and *cobT2*
3. *Desuldobulbous mediterraneous (Dm)*, a unique genome that has all three- *cobT1*, *cobT2*, and *cobT3* and its *bza* operon consists of *bzaF-cobT-bzaC* genes
4. *Desulfofundulus thermosubsterraneus* (*Dt*) which the *bza* operon lacks a *cobT* genes and the genome contains a single *cobT3* gene under a B_{12}-riboswitch.

Upon looking at the sequences of each one carefully, we found that the *Moorella thermoacetica cobT2* gene has a loss of function mutation of active site residues, and hence we did not clone this homolog for further study. We cloned the rest of 9 *cobT* homologs and purified the resulting proteins for further biochemical characterization (Figure 2.6).

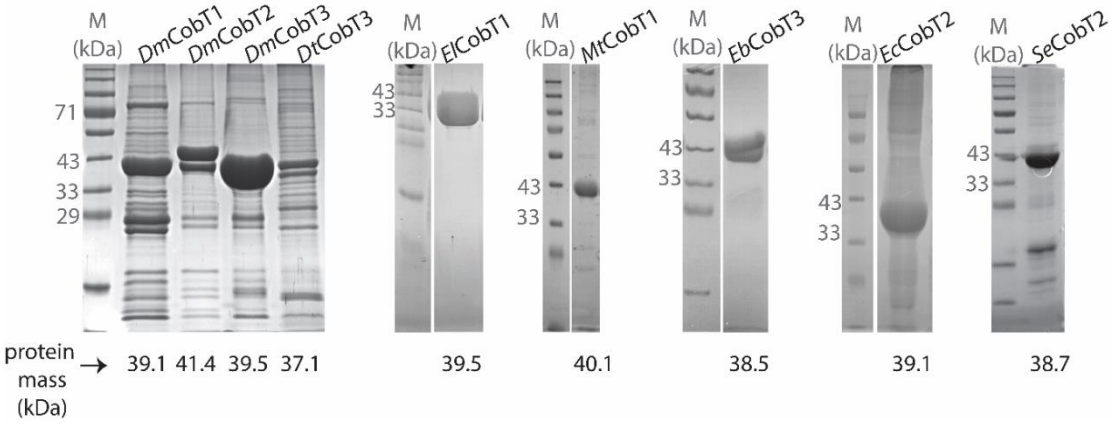


Figure 2.6 SDS-PAGE gels showing CobT homologs purified using Ni-NTA affinity chromatography.

2.4.2 *In vitro* assessment of the phosphoribosyltransferase activity

To test whether all the candidate CobT homologs from varying gene neighborhoods result in functional proteins, we reconstituted the phosphoribosyltransferase using DMB, which is known to be the most preferred lower ligand substrate (Figure 2.7A). The HPLC chromatogram for reconstitution for *Mt*CobT1, *Dm*CobT1, *Dm*CobT2, *Dm*CobT3, *Dt*CobT3, and *Ebr*CobT3 with DMB (Figure 2.7B, peak at 25 min) shows formation of the expected product DMB-RP (Figure 2.7B, peak at 19 min) with varying net activity. We also see efficient conversion with *Eb*CobT1, *El*CobT1, and *El*CobT3 which are reported earlier from our lab (Datar and Hazra 2018; Sreyas and Hazra 2020). *Dm*CobT1 shows exceptionally low activity <1% conversion with DMB as discussed later. Nonetheless, all the tested CobT1, CobT2, and CobT3

representative homologs are functional phosphoribosyltransferases. We then tested the range of substrate promiscuity and we find *El*CobT1, *Ebr*CobT1, *Mt*CobT1, *El*CobT1, and *Ebr*CobT3, efficiently activate Bza, 5-MeBza, and 5-ClBza, and adenine. Similar experiments with *Dm*CobT1, *Dm*CobT2, *Dm*CobT3 are currently underway.

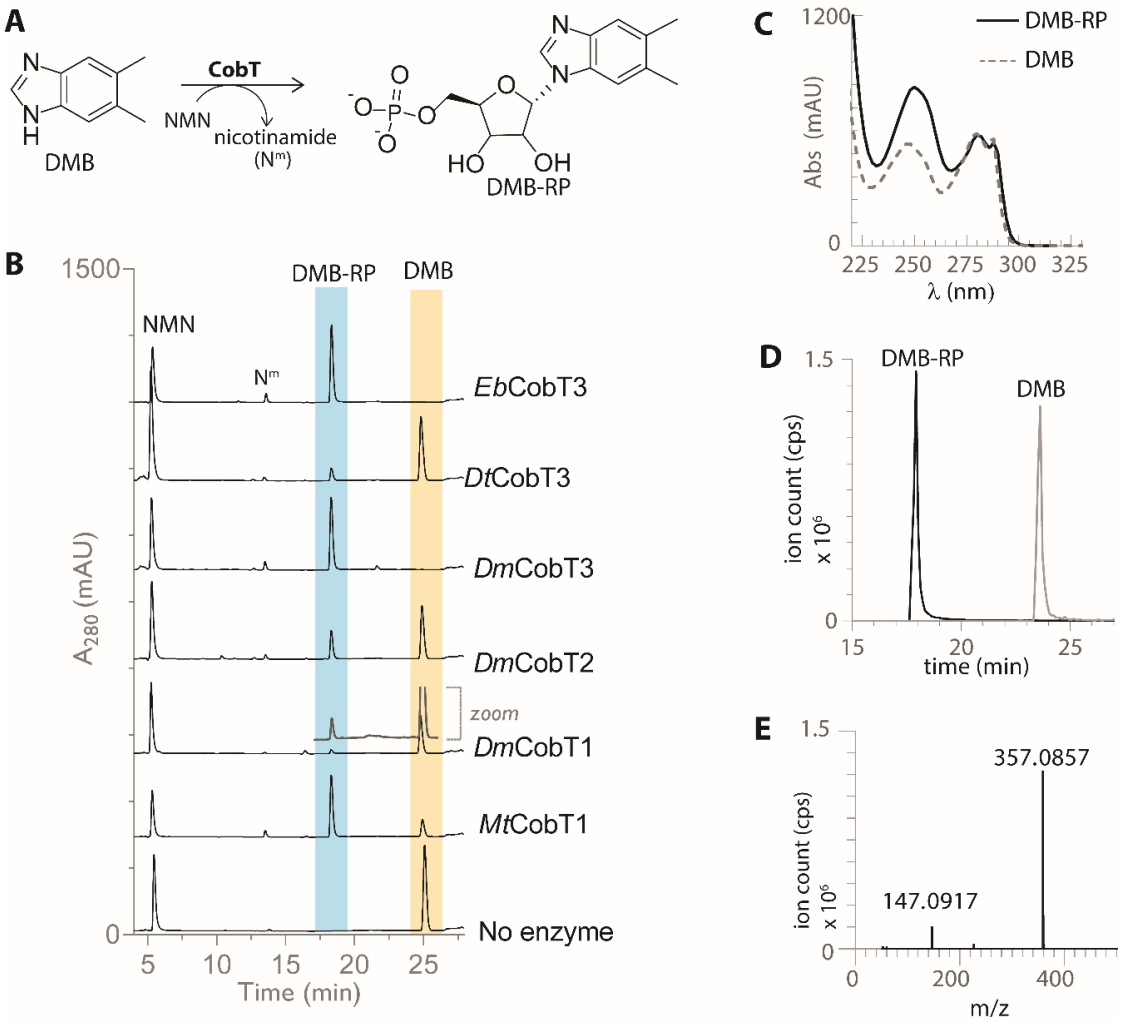


Figure 2.7 Reconstitution of CobT activity with DMB as the substrate. A) Reaction scheme for CobT reaction with 5,6-dimethylbenzimidazole (DMB) and nicotinamide mononucleotide (NMN) as substrates. The reaction is known to yield DMB-riboside phosphate (DMB-RP) and nicotinamide (N^m) as products. **B)** HPLC chromatogram for the no enzyme control and reaction traces for reconstitution of *Mt*CobT1, *Dm*CobT1, *Dm*CobT2, *Dm*CobT3, *Dt*CobT3, and *Eb*CobT3. The peaks at 25 min and 19 min correspond to the substrate DMB and the product DMB-RP respectively. **C)** UV-Vis spectra of DMB show a $\lambda_{280} > \lambda_{250}$ whereas DMB-RP exhibit $\lambda_{280} < \lambda_{250}$, a feature used to characterize the substrate and product peaks. **D)** The extracted ion chromatogram (XIC) for the reaction using m/z values 357.0857 and 149.0917 corresponding to DMB-RP and DMB, which correspond to the product and the substrate of the reaction respectively. **E)** The mass spectra of the product peak at 19 min showing peaks at 357.085 corresponding to the expected mass of the product DMB-RP. We also observe minor peaks at 149.091 and 279.133 for fragments corresponding to the riboside form (DMB-R), and base (DMB) as well.

2.4.3 The three kinds of CobT vary in regioselectivity

We examine the regioselectivity activation of 5-OHBza and 5-OMeBza by the CobT1, CobT2, and CobT3 homologs.

We use the 5-OHBza and 5-OMeBza as substrates to test the regioselectivity of the following reasons:

a) 5-OHBza and 5-OMeBza appear as intermediates on the *bza* operon and these molecules differ by only one methyl group. Hence, these molecules allow us to probe differences in enzyme activity arising from subtle changes in substrate structure.
b) the isomeric products from both substrates can be separated on the C-18 columns using reverse phase HPLC and show distinct UV-Vis spectra which allows for precise analysis of regioselectivity of reaction

We observe that *El*CobT1, *Ebr*CobT1, and *Mt*CobT1 show a single peak upon reaction with 5-OHBza (Figure 2.8B, peak at 9 mins) which we confirm as 5-OHBza-RP isomer using the characteristic UV-Vis and mass spectra (Datar and Hazra 2018; Sreyas and Hazra 2020) . On the contrary, the CobT1 homologs show two distinct product peaks at 14 and 14.5 min with 5-OMeBza as the substrate (Figure 2.9B) (Datar and Hazra 2018; Sreyas and Hazra 2020) . Form the distinct UV-Vis and identical mass spectra we confirm the peak at 14 min as 5-OMeBza-RP isomer and the 14.5 min as 6-OMeBza-RP isomer. The regiospecific activation of 5-OHBza by CobT1 homologs (Figure 2.8D) is an unprecedented observation since most of the CobT homologs from aerobic and facultative anaerobic bacteria show regioselectivity for 5-OMeBza instead. Interestingly, both 5-OHBza and 5-OMeBza appear as intermediates of the *bza* operon pathway in which the CobT1 homologs participate. The observation that 5-OHBza is activated by CobT1 in a regiospecific manner but the 5-OMeBza yields two isomeric products led to re-evaluate the function of the methyltransferases as described in chapter 3.

The *Desulfobulbous mediterraneous* CobT2 (*Dm*CobT2) show complementary regioselectivity with 5-OHBza and 5-OMeBza. That is, *Dm*CobT2 activates 5-OHBza to produce 6-OHBza-RP in higher yield, and instead makes more of 5-OMeBza-RP with 5-OMeBza (Figure 2.8B, 2.8D and Figure 2.9B, 2.9D). The contrasting regioselectivity of *Dm*CobT2 with 5-OHBza and 5-OMeBza indicates that enzyme can distinguish the two substrates based on a methyl group. Previously, CobT2 equivalent homologs from aerobes and other anaerobes that lack bza operon have been shown to activate 5-OMeBza in a regiospecific manner, but the regioselectivity for 5-OHBza varies across the homologs (Crofts et al. 2014; Datar and Hazra 2018).

The CobT3 homologs appear to be a mix bag of varying enzyme properties. The *Eubacterium barkeri* CobT3 (*Eb*CobT3) and *Eubacterium limosum* CobT3 (*El*CobT3) show preferential formation of 6-RP isomer both 5-OHBza and 5-OMeBza. (Figure 2.8B, 2.8D and

Figure 2.9B, 2.9D). The *Desulfobulbous mediterraneous* CobT3 (*Dm*CobT3) and *Desulfofundulus thermosubsterraneus* CobT3 (*Dt*CobT3) on the other hand, show regioselectivity in favor of the 5-RP orientation with 5-OHBza and 5-OMeBza. The *Dt*CobT3 is the only CobT encoded by *Desulfofundulus thermosubsterraneus*, and the regioselectivity similar to CobT1 warrants further exploration of its role in the lower ligand biosynthesis.

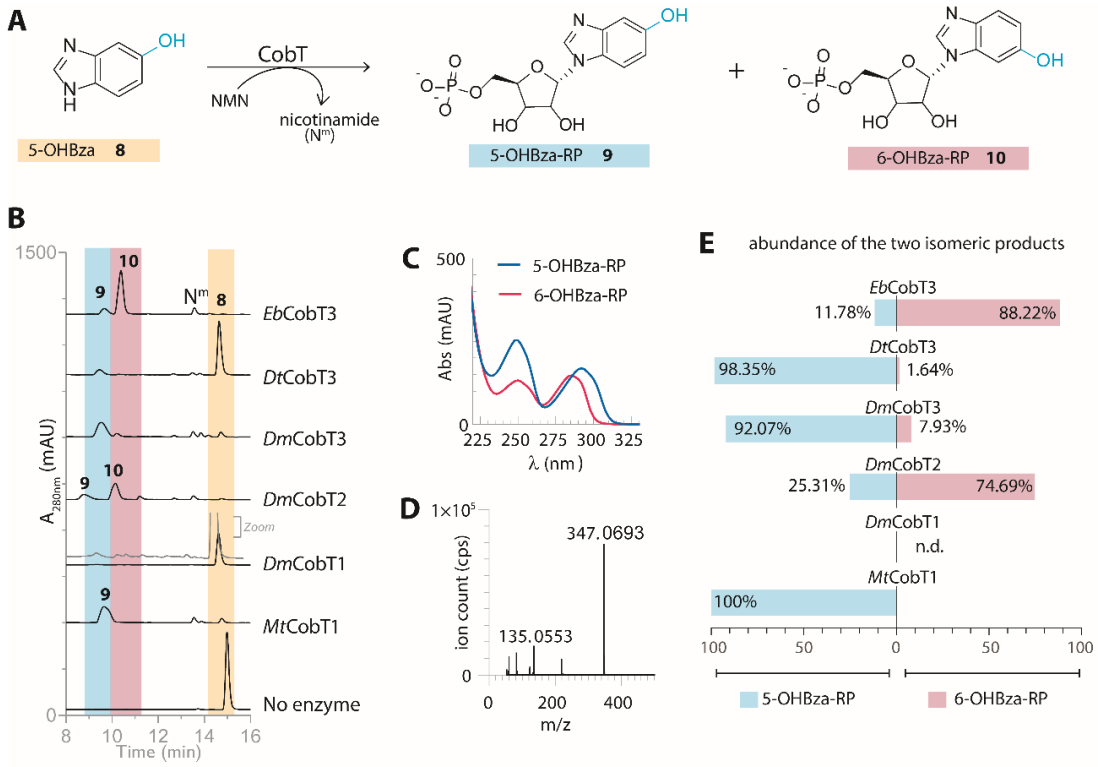


Figure 2.8 Differences in regioselectivity CobT homologs with 5-OHBza. A) CobT reaction with 5-OHBza **8** can yield two isomeric products, 5-OHBza-riboside phosphate (5-OHBza-RP **9**) and 6-OHBza-riboside phosphate (6-OHBza-RP **10**). **B)** HPLC chromatogram for 5-OHBza (peak at 15 min) activation by *Mt*CobT1, *Dm*CobT1, *Dm*CobT2, *Dm*CobT3, *Dt*CobT3, and *Eb*CobT3. **C)** The characteristic UV-Vis spectra for 5-OHBza-RP shows $\lambda_{250} > \lambda_{280}$ and 6-OHBza-RP shows $\lambda_{250} < \lambda_{280}$. The spectra is used to assign the products as 5-OHBza-RP **9** or 6-OHBza-RP **10** in the HPLC chromatogram. **D)** A representative mass spectrum for the product showing the identifying fragment of 135.0553 corresponding to the lower ligand base. **E)** Percent formation of 5-OHBza-RP **9** (blue bars, left) or 6-OHBza-RP **10** (pink bars, right) by each homolog is shown to represent the intrinsic regioselectivity of the homologs. The values could not be calculated for *Dm*CobT1 as due to negligible reaction yield.

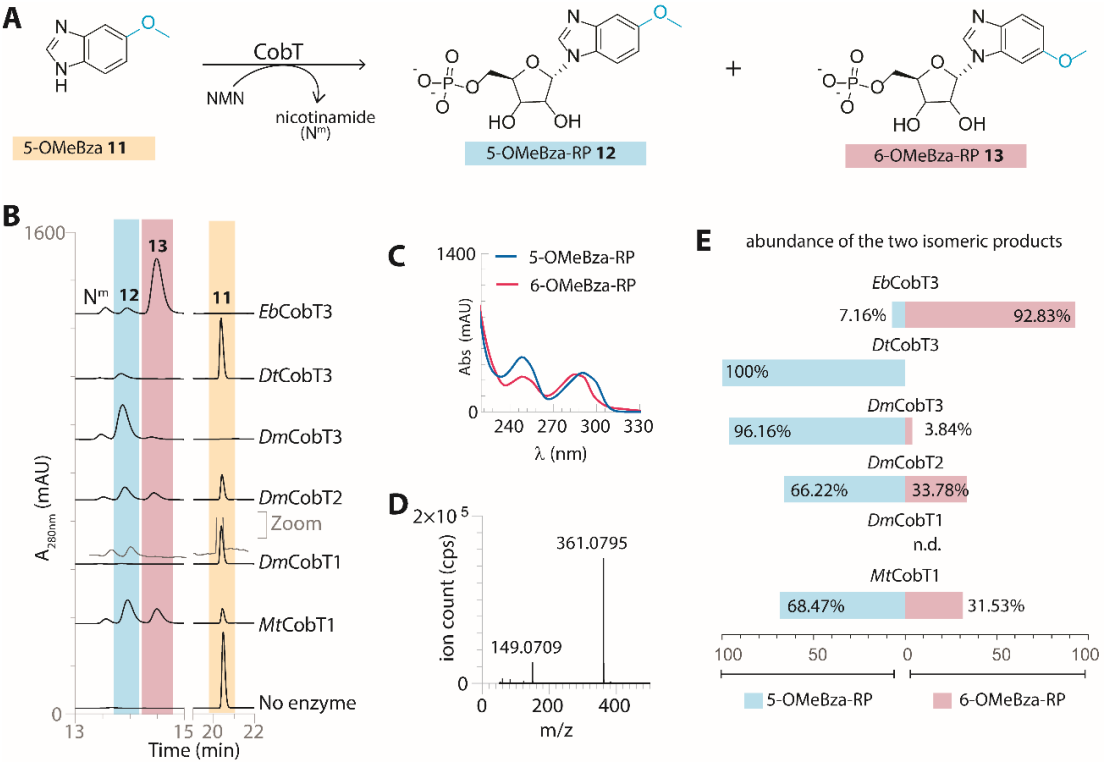


Figure 2.9 Differences in regioselectivity CobT homologs with 5-OMeBza. A) CobT reaction with 5-OMeBza **11** can yield two isomeric products, 5-OHBza-riboside phosphate (5-OMeBza-RP **12**) and 6-OHBza-riboside phosphate (6-OMeBza-RP **13**). **B)** HPLC chromatogram for 5-OMeBza (peak at 21 min) activation by MtCobT1, DmCobT1, DmCobT2, DmCobT3, DtCobT3, and EbCobT3. **C)** The characteristic UV-Vis spectra for 5-OMeBza-RP **12** shows $\lambda_{250} > \lambda_{280}$ and 6-OMeBza-RP **13** shows $\lambda_{250} < \lambda_{280}$. The spectra are used to assign the products as 5-OMeBza-RP **12** or 6-OHBza-RP **13** in the HPLC chromatogram. **D)** A representative mass spectrum for the product showing the identifying fragment of 149.0709 corresponding to the lower ligand base. **E)** Percent formation of 5-OMeBza-RP **12** (blue bars, left) or 6-OMeBza-RP **10** (pink bars, right) by each homolog is shown to represent the intrinsic regioselectivity of the homologs. The values could not be calculated for DmCobT1 as due to negligible reaction yield.

2.4.4 The one which prefers adenine - *Desulfobulbous mediterraneous CobT1*

Of all the tested homologs, the *Dm*CobT1 homolog was the most difficult one to clone, overexpress, and purify. It shows poor protein expression and stability which makes it difficult to characterize the protein extensively. In our primary characterization the *Dm*CobT1 appears to be a functional enzyme which activates DMB however with very poor activity. Since, *Dm*CobT1 is encoded within the *bza* operon, we expected the enzyme to show regiospecific activation and substrate preference for 5-OHBza. However, the enzyme shows negligible activity with 5-OHBza and 5-OMeBza which could not be detected on the HPLC coupled with UV-Vis detector. Interestingly, *Dm*CobT1 show up to be an odd one as it shows uncanny high activity with adenine as a substrate (Figure 2.10A, 2.10B). Previously, *Mt*CobT1, *Ebr*CobT1, *E.coli* CobT, *Se*CobT, *Lactobacillus reuteri* (*Lr*CobT) have been shown to activate adenine as well. However, adenine is the least preferred substrate for all the CobTs as compared to benzimidazoles derivatives. We conducted a pilot competition assay with *Dm*CobT1 using the

substrates DMB, 5-OHBza, 5-OMeBza, and Ade. The competition assay allows us to find physiologically relevant substrate preference both qualitatively and quantitatively.

We find that *Dm*CobT1 shows 40%, 80%, and 60 % turnover of adenine in presence of DMB, 5-OHBza, and 5-OMeBza respectively while the conversion of the benzimidazoles remain <1% (Figure 2.10C). Further, we observe two distinct product peaks for adenine which we confirm N^9-Ade-RP and the N^7-Ade-RP through characteristic UV-Vis spectra. Further replicates of this experiment and competition assays to probe substrate preference in *Dm*CobT2, *Dm*CobT3, and *Dt*CobT3 are currently underway.

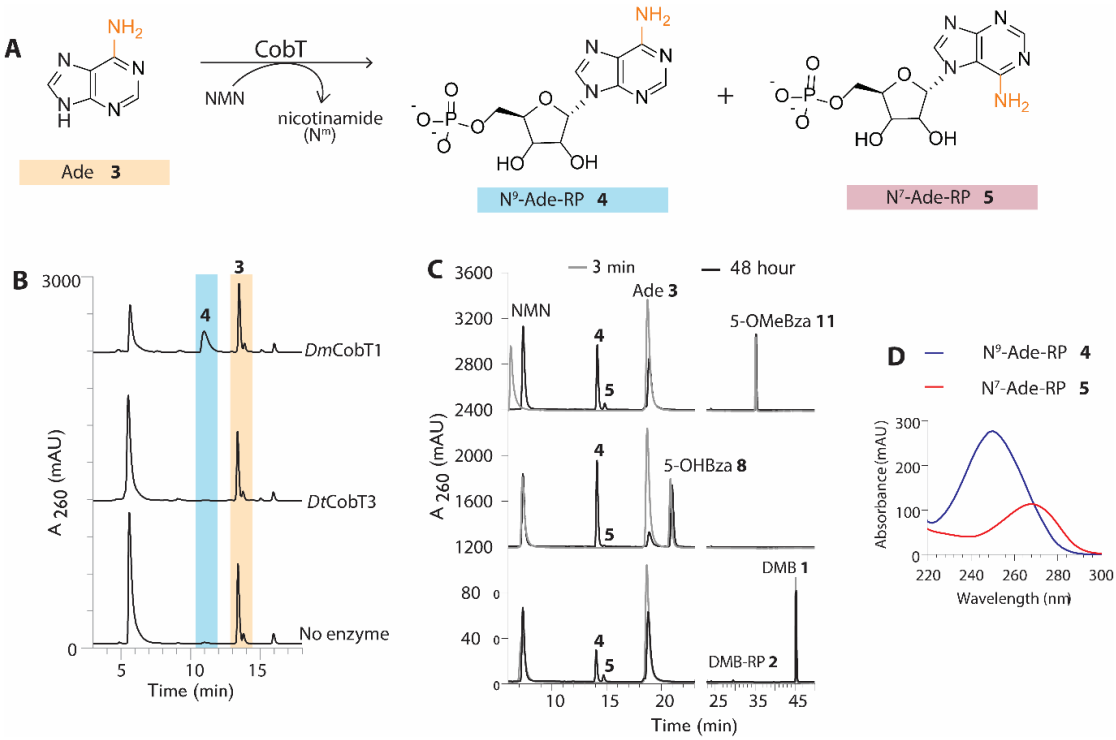


Figure 2.10 *Dm*CobT1 shows preference for adenine. A) Reaction scheme for activation of adenine (Ade **3**) to yield two isomeric products N^9-Ade-RP **4**, which is a α–epimer of b-AMP, and N^7-Ade-RP **5**. **B)** HPLC chromatogram for reconstitution of *Dm*CobT1 with adenine as compared to that of *Dt*CobT3 as an example. *Dm*CobT1 successfully activates ade **3** (peak at 14 mins) to yield N^9-Ade-RP **4 as** product (top trace, peak at 10 min). **D)** We ascertain the nature of isomer using the distinctive UV-Vis spectra for the two isomers. **C)** HPLC traces for competition assay to test substrate preference of *Dm*CobT1 among the combinations of DMB **1** and ade **3** (bottom traces), 5-OHBza **8** and ade **4** (middle traces), and 5-OMeBza **11** and ade **4** (top traces). The 3-minute trace represent the zeroth time point and the 48-hour trace is the reaction end-point. In all traces, we observe peaks corresponding to N^9-Ade-RP **4** and N^7-Ade-RP **5.** No product for 5-OHBza **8** and 5-OMeBza **11** were observed, and DMB-RP **2** was seen in very low yield. **D)** The UV-Vis spectra of N^9-Ade-RP **4** shows λ_{max} at 250 nm whereas and N^7-Ade-RP **5** show λ_{max} at 250 nm, thus allowing identification of the two product peaks.

2.4.5 Analysis of CobT sequences to find molecular reasons for functional differences

We generated a sequence similarity network with curated dataset with CobT homologs encoded in varying gene neighborhood in diverse taxa, archaeal and cyanobacterial CobT

homologs, and nicotinamide phosphoribosyltransferases as a control group (Figure 11 A). At a low score 46 (35% identity), the SSN appear as three clusters corresponding to the control group, archaeal and archaea-like CobT sequences, and the bacterial CobT sequences (data not shown). With an increase in the cut-off score to 98 (55% identity, data not shown), there appears to be a taxonomic segregation of bacterial sequences and subjecting the network to a higher cut-off value at 113 (60% identity), we observe that CobT1 homologs get separated into 3 notable groups (Figure 2.11A). The first cluster is predominantly CobT1 sequences from Firmicutes, second with CobT1 from Proteobacteria. The third cluster appears to be interesting and informative- the CobT1 homologs from organisms that contain only *bzaF* gene in the lower ligand biosynthesis pathway show higher similarity with CobT2 and CobT3 homologs. Notably, most of the CobT3 homologs seen on this cluster are from organisms that lack a *bza* operon. The one-to-one comparison of all sequences at the large scale indicates – 1) sequence conservation of CobT agrees with our classification based on gene neighborhood, 2) CobT1 sequences appear to be highly conserved and specialized homologs.

We then probed the sequence level differences that govern the functional differences among the CobT1, CobT2, and CobT3 homologs. We first project the patterns of conservation among homologs on previously solved *Se*CobT structure (PDB IL4E) using the ConSurf (Figure 2.11B). As expected, the catalytically essential residues are highly conserved across all homologs, and highly variable regions are present on the loops at the periphery of the structure. The active site of *Se*CobT exists at the dimerization interface, in the way that the phosphoribosyl donor is buried within the deep conserved pocket at the C-terminal of each monomer, and the lower ligand substrate lies at the hydrophobic dimerization interface. More so, the C5 and C6 of the six-membered ring of DMB-RP bound to C-terminal pocket of one monomer faces the N-terminal loop of the other monomer.

Next, to find residues that are vary across the CobT1, CobT2, and CobT3 groups but may show conservation within a group, we conducted the specificity determining positions (SDPs) analysis using SDPpred tool (Kalinina et al. 2004). The SDPs are defined as the amino acid sites in a multiple sequence alignment that show patterns in conservation that concur with divergence of subfamilies. We map the predicted SDPs with high scores on the *Se*CobT sequence and structure to understand the possible molecular contribution of these sites. Overall, we find most of the predicted SDPs lie close to the conserved residues known to bind the phosphoribosyldonor, or the substrate. Some of these sites also lie within the variable region of the dimerization interface, and these sites may or may not impact the protein function. Notably, two of the predicted SDP residues Leucine30 (L30) and Lysine31 (K31) in proximity of the C5 and C6 of DMB and 5-OMeBza substrates bound to *Se*CobT (PDB 1L4E, 1JHP).

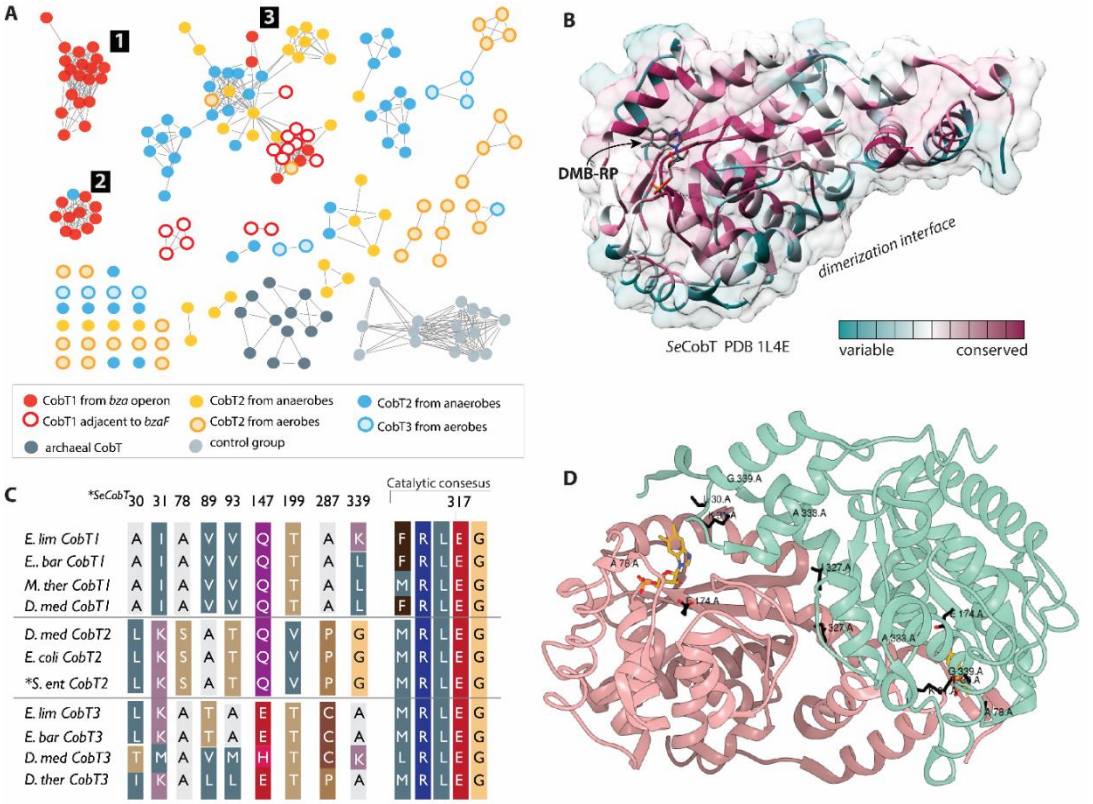


Figure 2.11 What makes CobT1, CobT2, and CobT3 homologs function differently? A) A sequence similarity network constructed with CobT homologs. Each circular node represents a protein sequence and the connecting edge corresponds to the similarity between two nodes. The nodes are colored based on the gene neighbourhood and physiology of the organism. The cluster-1 is predominantly CobT1 from firmicutes and cluster-2 is mostly CobT1 from proteobacteria. Cluster-3 show shared similarities between anaerobic CobT1, CobT2, and CobT3 homologs. Note that CobT1 from anaerobes that only code for a bzaF gene show higher similarity with other CobT homolog from anaerobic bacteria. **B)** Structure of Salmonella enterica CobT (*Se*CobT) bound to DMB-RP (PDB 1L4E). The color gradient is generated using ConSurf (Ashkenazy et al. 2016) and indicates conservation of each amino acid site across 174 bacterial CobT homologs. The core structural folds including the active site of the protein, as expected, is highly conserved and the six-membered ring of the substrate faces the surface of the monomer. **C)** Specificity determining positions predicted by comparing the sequences of CobT1, CobT2, and CobT3 homologs using SDPpred (Kalinina et al. 2004). The predicted sites show conservation that correlates with the classification of the homologs based on the genomic context. The catalytic glutamine (E317 in *Se*CobT) is conserved across all homologs. **D)** The dimeric *Se*CobT marked with the predicted SDP. Of all the predicted SDPs, only three sites, L30, K31, and E174 lie in close proximity of the substrate. Specifically, the L30 and K31 lie close to the C5 and C6 of the 6-membered ring of DMB. Since, most benzimidazolyl CobT substrates differ by hydroxy, methyl, or methoxy functional groups at the C5 or C6, the L30 and K31 pose as reasonable starting points to explore molecular basis of the regioselectivity.

The L30 side chain is at 5.5 Å from C5 of DMB and at 4.2 Å from C5 of 5-OMeBza. Since the functional groups that differentiate the benzimidazole lower ligands are at C5 and C6 of the benzimidazole scaffold, understanding the residues that surround C5 and C6 of the benzimidazole ring can hint towards molecular basis of CobT's sensitivity towards methyl groups in 5-OHBza and 5-OMeBza. Among the CobT1 homologs, we find that the Alanine30 (A30) and Isoleucine31 (I31) are conserved in place of L30 and K31 of CobT2 homologs. The CobT3 group, however, does not show patterns similar to CobT2 at these sites. Either I30 or

L30 and K31 are found in CobT3 homologs characterized in this study with an exception of the *Dt*CobT3 which has residues T30 and M31. To further validate the role of these residues in determining regiospecificity, mutants L30A, L30T, and K31I were created in our lab. Characterization of these mutants to study with respect to regioselectivity towards 5-OHBza and 5-OMeBza are underway.

From an overall comparison, the residues at the dimerization interface, specially the N-terminal loop that interacts with the benzimidazole ring appears to play key role in substrate recognition and regioselectivity of the reaction. Further mutagenesis studies with the residues shortlisted from the ConSurf and SDPpred analysis would be required to experimental validate the bioinformatic predictions. Unpinning the residues responsible for varying substrate preference and regioselectivity among CobT homologs will aid in predicting activity of other CobT homologs which would be a valuable insight into predicted cobamides that can be produced by an organism.

2.5 Discussion

From comparing the differences in the regioselectivity exhibited by the CobT homologs in activation of 5-OHBza and 5-OMeBza we find, that CobT1 and CobT2 homologs are capable of distinguishing the additional methyl group. The distinguishing feature among is the regiospecific activation of 5-OHBza by CobT1 homologs. CobT3 homologs, on the other hand, retain the regioselectivity to make either 5-RP or 6-RP and same orientation for 5-OHBza and 5-OMeBza is dictated. Even though the CobT homologs are tightly conserved class of proteins, these homologs seemingly have adapted to employ different mechanisms for substrate recognition and orientation. More importantly, homologs encoded within the same genome show contrasting regioselectivity which under physiological conditions can lead to different cobamide structures. Previous work with isomeric cobamides [5-OMeBza]Cba and [6-OMeBza]Cba show that a subtle difference that is the location of the functional group of lower ligand potentially influences the 'base-on'↔ 'base-off' dynamics which in turn impacts the interaction between the cobalt ion and the upper ligand. Thus, we need to question the physiological relevance of containing multiple CobT genes and understand how organisms regulate the different CobT homologs to produce cobamides fit for biochemical function is important. The *cobT1* and *cobT2* homologs are often regulated under B_{12}-riboswitches along with other genes involved in lower ligand biosynthesis and cobamide assembly. From quantitation of genomic context of *cobT* homologs (Figure 2.6C) it is apparent that most anaerobes show mutual exclusivity with *cobT1* and *cobT2* which can be important for achieving finer control over cobamide structure. The evidence of truncated CobT2 in *Moorella thermoacetica* which is otherwise embedded within the B_{12} biosynthesis locus of >20 genes,

supports the preference for CobT1 homolog which catalyzes regiospecific activation of 5-OHBza, the first intermediate in the *bza* operon pathway. As we'll see in the next chapter, this allows *Moorella thermoacetica* to produce [5-OMeBza]Cba as the sole cobamide.

However, we do not see such mutual exclusivity with *cobT1* and *cobT3,* and instead a large proportion of organisms contain this combination. The significance of this observation, and the function and regulation of *cobT3* genes requires extensive *in vivo* experimentation with organisms such as *Eubacterium limosum.* Notably, *Desulfobulbous mediterraneous* is a unique find with four hits for CobT within the same genome, three of these are single domain CobT homologs that we characterize here and the fourth one is a fusion protein with alcohol dehydrogenase domain. The *bza* operon of *D. mediterraneous* contains *bzaF-cobT-bzaC* genes and is predicted to make 5-OMeBza as lower ligand. The unusual preference for adenine exhibited by *Dm*CobT1 from the *bza* operon requires further verification *in vitro* as well as *in vivo*. To this end, we are cloning the operon in pTH1227 vector for validation in an engineered B_{12}-dependent strain of *E.coli* which was used for discovery of the *bza* operon. Further, there are no currently available bacterial strains that can selectively overproduce pseudocobalamin for laboratory or commercial use. Hence, testing the effect of overexpression of *Dm*CobT1 in a cobamide producing strain can be beneficial for improving pseudocobalamin production. Lastly, the occurrence of multiple genes that encode for homologs of one protein family is not unusual for prokaryotic and eukaryotic genomes. In literature, several studies spanning functional differences among isoforms of eukaryotic enzymes have been reported, however comparative analyses of such homologs from prokaryotic genomes are only beginning to be conducted. Our approach with comprehensive biochemical and bioinformatic analyses focusing on key functional properties of the enzyme family can also be used to study physiological function and relevance of multiple homologs encoded within same genome.

Publications arising from this chapter:

Book chapter

Guardian of cobamide diversity: Probing CobT's role in lower ligand activation in the biosynthesis of Vitamin B_{12} and other cobamide cofactors
Yamini Mathur, Aniket R. Vartak, Amrita B. Hazra
Methods in Enzymology: B12 enzymes (668), *in press* (**2022**)
DOI: https://doi.org/10.1016/bs.mie.2022.01.001

Research article (manuscript in preparation)

Functional differences in CobT homologs from varying gene neighborhoods
Yamini Mathur[1], Sheryl Sreyas[1], Amrita B. Hazra [[1] Shared first author]

References:

Altschul, SF Stephen F., Warren Gish, Webb Miller, Eugene W. EW Myers, and David J. DJ Lipman. 1990. “Basic Local Alignment Search Tool.” Journal of Molecular Biology 215 (3): 403–10. https://doi.org/10.1016/S0022-2836(05)80360-2.

Anderson, Peter J., Jozsef Lango, Colleen Carkeet, Audrey Britten, Bernhard Kräutler, Bruce D. Hammock, and John R. Roth. 2008. “One Pathway Can Incorporate Either Adenine or Dimethylbenzimidazole as an α-Axial Ligand of B12 Cofactors in Salmonella Enterica.” Journal of Bacteriology 190 (4): 1160–71. https://doi.org/10.1128/JB.01386-07.

Ashkenazy, Haim, Shiran Abadi, Eric Martz, Ofer Chay, Itay Mayrose, Tal Pupko, and Nir Ben-Tal. 2016. “ConSurf 2016: An Improved Methodology to Estimate and Visualize Evolutionary Conservation in Macromolecules.” Nucleic Acids Research 44 (W1): W344–50. https://doi.org/10.1093/NAR/GKW408.

Benson, Dennis A., Ilene Karsch-Mizrachi, David J. Lipman, James Ostell, and David L. Wheeler. 2005. “GenBank.” Nucleic Acids Research 33 (DATABASE ISS.): 34–38. https://doi.org/10.1093/nar/gki063.

Chan, Chi Ho, and Jorge C. Escalante-Semerena. 2011. “ArsAB, a Novel Enzyme from Sporomusa Ovata Activates Phenolic Bases for Adenosylcobamide Biosynthesis.” Molecular Microbiology 81 (4): 952–67. https://doi.org/10.1111/j.1365-2958.2011.07741.x.

Chan, Chi Ho, Sean A. Newmister, Keenan Talyor, Kathy R. Claas, Ivan Rayment, and Jorge C. Escalante-Semerena. 2014. “Dissecting Cobamide Diversity through Structural and Functional Analyses of the Base-Activating CobT Enzyme of Salmonella Enterica.” Biochimica et Biophysica Acta - General Subjects 1840 (1): 464–75. https://doi.org/10.1016/j.bbagen.2013.09.038.

Cheong, Cheom-gil Gil, Jorge C. Escalante-Semerena, and Ivan Rayment. 2001. “Structural Investigation of the Biosynthesis of Alternative Lower Ligands for Cobamides by Nicotinate Mononucleotide: 5,6-Dimethylbenzimidazole Phosphoribosyltransferase from Salmonella Enterica.” Journal of Biological Chemistry 276 (40): 37612–20. https://doi.org/10.1074/jbc.M105390200.

Crofts, Terence S., Amrita B. Hazra, Jennifer LA Tran, Olga M. Sokolovskaya, Vadim Osadchiy, Omer Ad, Jeffrey Pelton, Stefan Bauer, and Michiko E. Taga. 2014. “Regiospecific Formation of Cobamide Isomers Is Directed by CobT.” Biochemistry 53 (49): 7805–15. https://doi.org/10.1021/bi501147d.

Crofts, Terence S., Erica C. Seth, Amrita B. Hazra, and Michiko E. Taga. 2013. “Cobamide Structure Depends on Both Lower Ligand Availability and CobT Substrate Specificity.” Chemistry and Biology 20 (10): 1265–74. https://doi.org/10.1016/j.chembiol.2013.08.006.

Datar, Prathamesh M., and Amrita B. Hazra. 2018. “Investigating the Regioselective Attachment of the Lower Ligand in Vitamin B12 Biosynthesis.” IISER Pune. http://dr.iiserpune.ac.in:8080/xmlui/handle/123456789/1023.

Ernst, Orna, and Tsaffrir Zor. 2010. “Linearization of the Bradford Protein Assay.” Journal of Visualized Experiments, no. 38: e1918. https://doi.org/10.3791/1918.

Gerlt, John A. 2017. “Genomic Enzymology: Web Tools for Leveraging Protein Family Sequence-Function Space and Genome Context to Discover Novel Functions.” Biochemistry 56 (33): 4293–4308. https://doi.org/10.1021/acs.biochem.7b00614.

Gilchrist, Cameron L.M., and Yit Heng Chooi. 2021. “Clinker & Clustermap.Js: Automatic Generation of Gene Cluster Comparison Figures.” Bioinformatics 37 (16): 2473–75. https://doi.org/10.1093/bioinformatics/btab007.

Gray, Michael J., and Jorge C. Escalante-Semerena. 2010. “A New Pathway for the Synthesis of α-Ribazole-Phosphate in Listeria Innocua.” Molecular Microbiology 77 (6): 1429–38. https://doi.org/10.1111/j.1365-2958.2010.07294.x.

Hazra, Amrita B., Andrew W. Han, Angad P. Mehta, Kenny C. Mok, Vadim Osadchiy, Tadhg P. Begley, and Michiko E. Taga. 2015. “Anaerobic Biosynthesis of the Lower Ligand of Vitamin B12.” Proceedings of the National Academy of Sciences of the United States of America 112 (34): 10792–97. https://doi.org/10.1073/pnas.1509132112.

Hazra, Amrita B., Jennifer L.A. Tran, Terence S. Crofts, and Michiko E. Taga. 2013. “Analysis of Substrate Specificity in CobT Homologs Reveals Widespread Preference for DMB, the Lower Axial Ligand of Vitamin B12.” Chemistry and Biology 20 (10): 1275–85. https://doi.org/10.1016/j.chembiol.2013.08.007.

Kalinina, Olga V., Pavel S. Novichkov, Andrey A. Mironov, Mikhail S. Gelfand, and Aleksandra B. Rakhmaninova. 2004. “SDPpred: A Tool for Prediction of Amino Acid Residues That Determine Differences in Functional Specificity of Homologous Proteins.” Nucleic Acids Research 32 (WEB SERVER ISS.): 424–28. https://doi.org/10.1093/nar/gkh391.

Katoh, Kazutaka, Kazuharu Misawa, Kei Ichi Kuma, and Takashi Miyata. 2002. “MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform.” Nucleic Acids Research 30 (14): 3059–66. https://doi.org/10.1093/nar/gkf436.

Keller, Sebastian, Markus Ruetz, Cindy Kunze, Bernhard Kräutler, Gabriele Diekert, and Torsten Schubert. 2014. “Exogenous 5,6-Dimethylbenzimidazole Caused Production of a Non-Functional Tetrachloroethene Reductive Dehalogenase in Sulfurospirillum Multivorans.” Environmental Microbiology 16 (11): 3361–69. https://doi.org/10.1111/1462-2920.12268.

Ma, Amy T., Breanna Tyrell, and Joris Beld. 2019. “Specificity of Cobamide Remodeling, Uptake and Utilization in Vibrio Cholerae.” Molecular Microbiology 113 (1): 89–102. https://doi.org/10.1111/mmi.14402.

Marchler-Bauer, Aron, and Stephen H. Bryant. 2004. “CD-Search: Protein Domain Annotations on the Fly.” Nucleic Acids Research 32 (Web server issue): W327–31. https://doi.org/10.1093/nar/gkh454.

Men, Yujie, Erica C. Seth, Shan Yi, Terence S. Crofts, Robert H. Allen, Michiko E. Taga, and Lisa Alvarez-Cohen. 2015. “Identification of Specific Corrinoids Reveals Corrinoid Modification in Dechlorinating Microbial Communities.” Environmental Microbiology 17 (12): 4873–84. https://doi.org/10.1111/1462-2920.12500.

Procter, James B, G Mungo Carstairs, Ben Soares, Kira Moura, T Charles Ofoegbu, Daniel Barton, Lauren Lui, et al. 2021. “Alignment of Biological Sequences with Jalview James.” In Multiple Sequence Alignment: Methods and Protocols, Methods in Molecular Biology, 2231:203–23.

Riya, P, and Amrita B. Hazra. 2022. “A Study on the Functional and Genomic Context of CobT in Eukaryotes and the Molecular Factors Governing Its Regioselectivity.” IISER Pune. http://dr.iiserpune.ac.in:8080/xmlui/handle/123456789/6933.

Seth, Erica C., and Michiko E. Taga. 2014. “Nutrient Cross-Feeding in the Microbial World.” Frontiers in Microbiology 5 (JULY): 1–6. https://doi.org/10.3389/fmicb.2014.00350.

Shelton, Amanda N., Erica C. Seth, Kenny C. Mok, Andrew W. Han, Samantha N. Jackson, David R. Haft, and Michiko E. Taga. 2018. “Uneven Distribution of Cobamide Biosynthesis and Dependence in Bacteria Predicted by Comparative Genomics.” ISME Journal 13 (3): 789–804. https://doi.org/10.1038/s41396-018-0304-9.

Sokolovskaya, Olga M., Amanda N. Shelton, and Michiko E. Taga. 2020. “Sharing Vitamins: Cobamides Unveil Microbial Interactions.” Science 369 (6499). https://doi.org/10.1126/science.aba0165.

Sreyas, Sheryl, and Amrita B. Hazra. 2020. “Probing the Mechanism and Genomic Context of CobT, the Lower Ligand Activating Enzyme in Vitamin B12 Biosynthesis.” IISER Pune. http://dr.iiserpune.ac.in:8080/xmlui/handle/123456789/4812.

Taga, Michiko E., Nicholas A. Larsen, Annaleise R. Howard-Jones, Christopher T. Walsh, and Graham C. Walker. 2007. “BluB Cannibalizes Flavin to Form the Lower Ligand of Vitamin B12.” Nature 446 (7134): 449–53. https://doi.org/10.1038/nature05611.

Trzebiatowski, Jodi R, and Jorge C Escalante-semerena. 1997. “Purification and Characterization of CobT , the Nicotinate- Mononucleotide : 5 ,6-Dimethylbenzimidazole Phosphoribosyltransferase Enzyme from Salmonella Typhimurium.” Journal Biological Chemistry 272 (28): 17662–67.

Wilkins, M.R., E. Gasteiger, A. Bairoch, J.C. Sanchez, K.L. Williams, R.D. Appel, and D.F. Hochstrasser. 2005. “Protein Identification and Analysis Tools on the ExPASy Server.” In John M. Walker (Ed): The Proteomics Protocols Handbook, Humana Press, 571–607.

Zayas, Carmen L., and Jorge C. Escalante-Semerena. 2007. “Reassessment of the Late Steps of Coenzyme B12 Synthesis in Salmonella Enterica: Evidence That Dephosphorylation of Adenosylcobalamin- 5′-Phosphate by the CobC Phosphatase Is the Last Step of the Pathway.” Journal of Bacteriology 189 (6): 2210–18. https://doi.org/10.1128/JB.01665-06.

Chapter 3

Identification and characterization of CobT homologs in eukaryotes

3.1 B_{12} production and utilization in nature

The complete biosynthesis of B_{12} and other cobamides is limited to a small subset of bacteria and archaea (Table 1) (Roth, Lawrence, and Bobik 1996; Sokolovskaya, Shelton, and Taga 2020). Most of other bacteria, archaea, and eukaryotes that require cobamides as cofactors rely on acquiring these molecules from their environment and hence cobamides, along with several other cofactors act as molecules of exchange among species (Sokolovskaya, Shelton, and Taga 2020; Seth and Taga 2014). A recent computational survey of 11,000 bacterial genomes show that while 86% of genomes encode for at least one B_{12} requiring enzyme, only 37% genomes encode the B_{12} biosynthesis pathway (Shelton et al. 2018). Several microbiological studies have shown that archaeal species produce a variety of cobamides, however, we still do not completely understand the steps and enzymes involved in cobamide biosynthesis in archaea (Woodson and Escalante-Semerena 2004; Scherer et al. 1984; Rodionov et al. 2003). The eukaryotes, on the other hand, have been shown to lack any *de novo* biosynthesis pathway for B_{12} and the utilization of B_{12} varies significantly among the various phyla (Table 1). Most animals including humans require B_{12} for methionine synthesis and catabolism of branched chain fatty acids and amino acids (Banerjee and Ragsdale 2003). Several protists have been thought to require B_{12} for these two functions, however the B_{12}-dependent functions in these phyla is not extensively explored. Plants have been known to neither can use nor make B_{12} or any other cobamides (Roth, Lawrence, and Bobik 1996; Katherine E. Helliwell, Wheeler, and Smith 2013). Earlier, algae and fungi were also thought to live a B_{12}-independent life, however advances in the last decade has provided abundant evidences that algae require B_{12} for methionine synthesis as well as they form sustainable mutualistic relations with B_{12}-producing bacteria to acquire this vitamin (Croft et al. 2005; Cooper et al. 2019). However, very little is known about B_{12} requirements among fungi. Fungi is a widespread phylum with species inhabiting diverse ecological niches and consortia. The model experimental fungal species such as *Saccharomyces cerevisae*, *Schizosaccharomyces pombe* do not require B_{12} cofactor and this observation was in theory extrapolated to all fungal species for a long time (Y. Zhang et al. 2009). Interestingly, a recent computational study challenged that this school of thought (Orłowska, Steczkiewicz, and Muszewska 2021). Probing the genome sequences for early-diverging fungi show that all non-Dikarya fungal lineages code genes that utilize B_{12} as cofactor and genes that can load the upper ligand of

cobalt ion. Further experimental validation of this observation and analysis to understand B_{12} acquisition will add valuable insights into role of fungi in cobamide diversity and exchange in the microbial world.

	Make	Use	Import	Remodel
Bacteria	✓	✓	✓	✓
Archaea	✓	✓	✓	✓
Eukarya				
Protist	✗	✓	✓	?
Fungi	✗	?	?	?
Algae	✗	✓	✓	✓
Plants	✗	✗	✗	✗
Animals	✗	✓	✓	✗

Table 3.1: **Production and utilization B_{12} and other cobamides in the nature.** Only a small subset of bacteria and archaea code for complete biosynthesis of cobamides. All other organisms that require cobamide cofactors rely on obtaining this vitamin from their environment through sophisticated import mechanisms. Among eukaryotes, the B_{12} dependencies vary across kingdoms and new observations in physiological studies and computational explorations are continuously expanding the field's understanding on B_{12} metabolism among eukaryotes. While plants have been known to neither require nor make cobamides. Protists and animals, including humans, seem to specifically require B_{12} among all other cobamides. Though previously thought to not require B_{12}, algae have been shown to utilixe, acquire, remodel, and even assimilate in high quantities. Recently, early diverging fungi were predicted to utilize B_{12} for primary metabolism.

The B_{12} metabolism in microscopic eukaryotes is not limited to acquisition and utilization of the vitamin. Recently, a methionine auxotroph of marine alga *Chlamydomonas reinhardtii* was shown to uptake and subsequently remodel various cobamides and norcobamides to make B_{12}, provided that DMB is also available in the media (Baum et al. 2020). Cobamide remodeling is a mode of cobamide biosynthesis wherein the organism uptakes a cobamide from its environment, hydrolyze its nucleotide loop to remove the lower ligand and riboside moiety, and then re-purpose the cobinamide backbone to attach a different lower ligand and thus make a new cobamide (Seth and Taga 2014).

The observations supporting the cobamide remodeling in algal species had raised questions for finding pathway and the enzymes responsible for this phenomenon. One of the traditional routes to find the pathway would be to set up genetic screens using mutants, however this is a tedious method owing to lack of challenges in growing axenic cultures and lack of standardized methods for genetic manipulation for most of the eukaryotic organisms. The recent advances in bioinformatic tools and genome sequencing has enabled us find homologs

of known and then experimentally validate their function. Here, we use a similar approach to find cobamide assembly pathway in eukaryotic genomes and biochemically characterize the sequences that show high similarity. In this chapter, we describe the discovery of CobT homologs among a subset of eukaryotic genomic and validation of few representative sequences as phosphoribosyltransferases. This study provides a missing link to study remodeling pathway in previously studied algae as well as expands the scope of finding cobamide dependencies and assembly in other eukaryotic microorganisms.

3.2 Methods

3.2.1 Sequence analysis

All multiple sequence alignments were generated using MUSCLE2.0 using JalView and were later visualized for representation using either BioEdit (Edgar 2004; Procter et al. 2021; Hall 1999).

3.2.2 Sequence Similarity Network

To generate a sequence similarity network, we compiled a dataset comprising of previously characterized bacterial CobTs, archaeal CobTs, and the candidate eukaryotic CobTs. The UniProt accession IDs were uploaded on the EFI-SSN server (Gerlt 2017). An SSN was finalized at a cut-off of e-value of 50 for all sequences within 100 to 600 amino acid length. The network was processed in Cytoscape3.0 and formatted in ylayouts- organic layout.

3.2.3 Gene synthesis, cloning, overexpression, and protein purification

From the analysis of candidate CobT sequences, we selected *cobT* genes from fungal species *Aphanomyces astaci*, *Aphanomyces invadans*, *Chytriomyces confervae,* and an algal species *Dunaliella salina* for further characterizations. The gene sequences were optimized for overexpression in *E.coli* BL21(DE3). The genes were synthesized and subsequently cloned into pET28a between the NdeI and BamHI sites. The gene synthesis and cloning were carried out by GenScript, BioTech Desk, India. The resulting plasmids were transformed into *E.coli* BL21(DE3) and the subsequent steps of protein overexpression and purification were conducted as described in section 2.3.3.

3.2.4 Reconstitution with DMB, and other substrates

The enzymatic activity of purified candidate eukaryotic CobT proteins was tested with substrates DMB, 5-OHBza, 5-OMeBza, 5-MeBza, Bza, 5-ClBza, and Adenine as described in section 2.3.4. The end-point reactions were monitored using HPLC coupled with UV-Vis spectroscopy and products were confirmed by LCMS analysis as described in 2.3.5

3.2.5 Comparative genomics using Hidden Markov Models

To identify genes involved in B_{12} assembly, remodeling, uptake, and utilization we conducted a comparative genomics study with eukaryotic organisms that show presence of a candidate CobT. We used Hidden Markov Models (HMMs) (Finn, Clements, and Eddy 2011; Potter et al. 2018) as a tool to detect the proteins of interest in the translated proteomes available for these organisms on UniProt.

a. B_{12} assembly genes: CobS , CobU
b. B_{12} transport and uptake: BtuF, BtuC, BtuD, BtuG
c. B_{12} utilization: CblA, CblB, CblC, CblD, MeaA, MetH, RNRII
d. B_{12} remodel: CbiZ and CobD
e. B_{12}-independent methionine synthase, MetE

The Pfam database (Finn et al. 2014)hosts HMM and the information regarding the number of sequences used to develop HMM and the cut-off scores is also provided along with the model. We first downloaded the HMM for the protein families corresponding each protein of interest. Next, we conducted HMMsearch on HMM on EBI (https://www.ebi.ac.uk/Tools/hmmer/search/hmmsearch) for each Pfam against 'Reference proteomes' to search for all organisms under UniProt or directly within the proteome of particular organism by selecting the 'restrict by taxonomy' option. The cut-off value for each Pfam was set to that suggested by Pfam for each model. From the results, we note the bit-score, domain organization, and UniProt accession numbers.

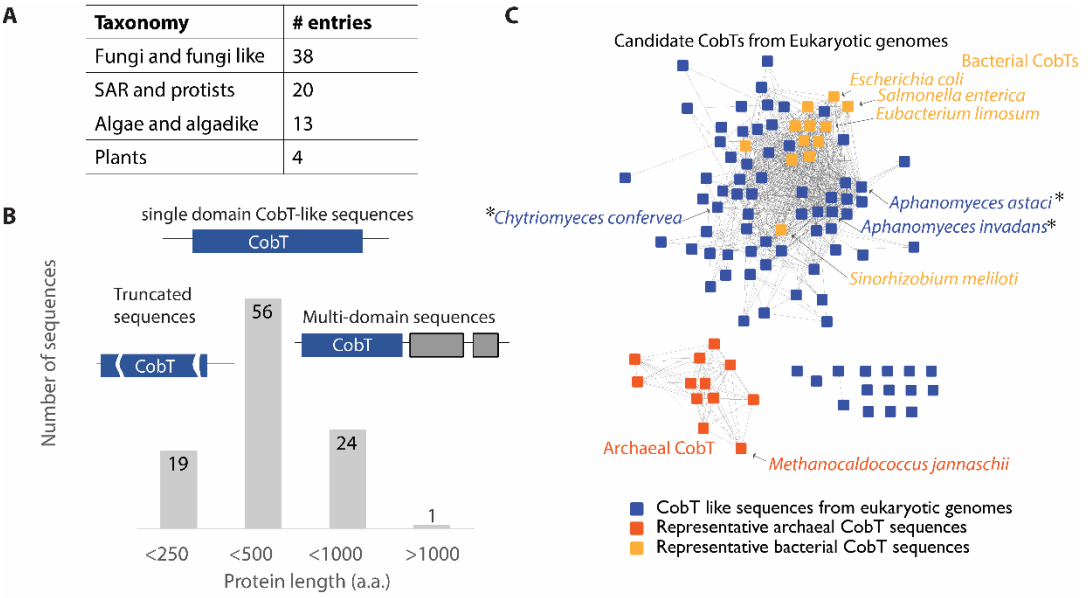


Figure 3.1 Identification of eukaryotic CobT homologs. A) Occurrence of CobT sequences in eukaryotic genomes. The UniProt database contain ~101 sequences of the CobT protein family. A gross distribution of the sequences into fungi, algae, protists, and plants is shown here. **B)** Distribution of eukaryotic CobT sequences w.r.t. to sequence length. The database show both single domain and multidomain sequences corresponding to CobT. The single domain sequences <250 amino acid residues show large and abberant truncations in the sequence which are likely to destroy enzyme function. In this

study, we focus on single domain CobT sequences that lie within the range of 350-450 amino acid residues. **C**) The sequence similarity network shows that the predicted eukaryotic CobT sequences are likely to be homologous to bacterial CobT homologs as opposed to archaeal CobT homologs. The nodes corresponding to CobT homologs which are previously characterized are labelled and the candidate CobT homologs tested in this study are marked with an asterisk (*).

3.4 Results

3.4.1 Inspecting the eukaryotic CobT-like sequences

The CobT family of proteins are classified under protein family Pfam02277 and a subclass annotated as TIGR3003 contains all CobT sequences of archaeal and cyanobacterial origins. While curating a dataset for bacterial CobTs encoded in different gene neighborhoods, we noticed the entries labelled CobT and CobT-like proteins that direct to eukaryotic genomes. To systematically address this observation, we filtered the entire Pfam02277 database hosted on UniProt comprising of ~22,00 entries to isolate all sequences that are belong to eukaryotes and we obtained sequences that we refer to as candidate eukaryotic CobT sequences. We first segregated sequences as based on number and classification of predicted domains. Of the total 100 sequences, 56 sequences are predicted to have a single domain corresponding to the CobT family of proteins. Interestingly, multidomain sequences often had additional domains such as CobS, and CobU which are also involved in B_{12} biosynthesis. A few multidomain sequences show CobT with domains not usually found within B_{12} biosynthesis enzymes. (Figure 3.1).

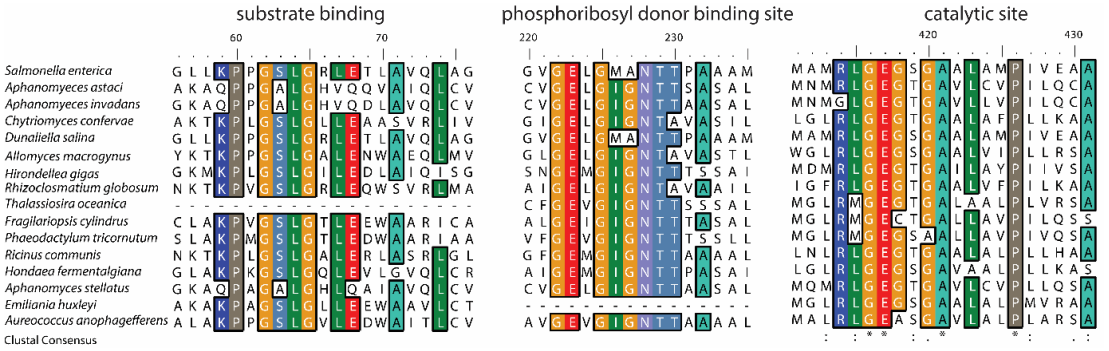


Figure 3.2 Sequence analysis for candidate eukaryotic CobT homologs. Sequence alignment of a subset of eukaryotic CobT homologs with *Salmonella enterica* CobT (*Se*CobT) as a representative of bacterial CobT sequences. We monitor presence and conservation of residues on the substrate binding loop at the dimerization interface, the active site pocket that binds the phosphoribosyl donor, and the consensus containing the catalytic glutamine residue.

For this project, we began exploring the single domain CobT-like sequences in eukaryotes. We manually examined these sequences for residues known to be essential for catalytic function in bacterial homologs of CobT (Figure 3.2). To do so, the eukaryotic CobT-like sequences were aligned with bacterial CobTs from *Salmonella enterica*, *Sinorhizobium meliloti*, and *Eubacterium limosum*. We removed sequences that showed deleterious mutations

in the cofactor binding site and the catalytic site residues. Further, certain sequences show long insertions of multiple either glutamine, alanine, or glycine residues and in some instances a string of 10-12 amino acids were found to be inserted between the cofactor binding sequence and catalytic residues. We then conducted a BLASTp search for each of these candidate eukaryotic CobT sequences and the sequences that showed >90% similarity with any bacterial CobT sequences were removed from further analysis on account of errors in annotation. The resulting subset of 24 CobT-like sequences have conserved catalytic residues and show minimal insertions and deletions of residues within the sequence. To further validate the activity of CobT-like sequences, we selected three fungal homologs from *Aphanomyces astaci* (*Aa*CobT), *Aphanomyces invadans* (*Ai*CobT), *Chytriomyces confervae* (*Cc*CobT), and one algal homolog from *Dunaliella salina* (*Ds*CobT) based on conservation of sequences. The genes sequences corresponding the *Aa*CobT, *Ai*CobT, *Cc*CobT, and *Ds*CobT were synthesized and cloned into an overexpression vector for overexpression in *E.coli* BL21(DE3) strain. The proteins were purified using Ni-NTA affinity chromatography (Figure 3.2) and tested for phosphoribosyltransferase activity.

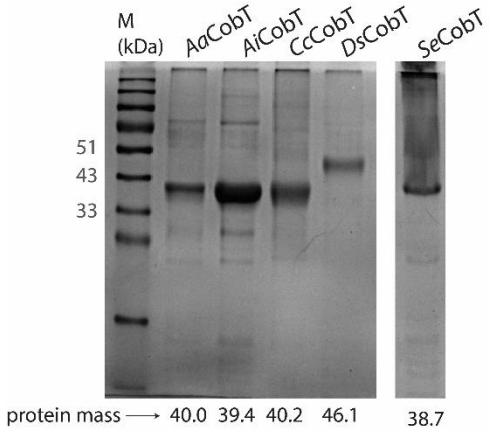


Figure 3.3 Purification of eukaryotic CobT homologs. 12% SDS-PAGE gel showing purified 6x-His tagged eukaryotic CobT-like proteins from Aphanomyces astaci (AaCobT), Aphanomyces invadans (AiCobT), Chytriomyces confervae (CcCobT), and one algal homolog from Dunaliella salina (DsCobT).

3.4.2 Eukaryotic CobT-like sequences are functional CobT homologs with nuances in substrate promiscuity and regioselectivity

A previous study with bacterial CobT homologs of diverse taxonomic descent show that despite a wide range of substrates that are activated by CobT family of enzymes, DMB is the most preferred substrate (Hazra et al. 2013). We tested the *in vitro* activity of the eukaryotic CobT-like proteins with 500 μM DMB as a substrate and 1 mM NMN as the phosphoribosyl donor in an end-point assay for 48-hour incubation (Figure 3.4A, 3.4B). The HPLC chromatogram of the reaction show the peak for substrates NMN and DMB at 5 minutes and 27 minutes respectively and the peaks for the products nicotinamide and DMB-RP product at

15 minutes and 21 minutes. The enzymes show 100% conversion with DMB which validates that *Aa*CobT, *Ai*CobT, *Cc*CobT, and *Ds*CobT as functional phosphoribosyltransferases.

Next, we tested the range of substrates and regioselectivity of reaction for the each of the eukaryotic CobTs with 5-OHBza, 5-OMeBza, and adenine. We invariably observe 100% conversion of 5-OHBza and 5-OMeBza by all four enzymes. The *Aa*CobT show regioselective formation of 5-OHBza-RP and shows a notable flip in regioselectivity with 5-OMeBza with higher production of 6-OMeBza-RP, which indicates that the homolog sensitive to the presence of the methyl group (Figure 3.4C, 3.4D). This methylation sensitive change in regioselectivity is reminiscent of bacterial CobT1 homologs described in chapter 2. On the contrary, the *Ai*CobT, CcCobT, and *Ds*CobT show similar regioselectivity for both 5-OHBza and 5-OMeBza (Figure 3.4C, 3.4D). That is, whereas the *Ai*CobT and *Cc*CobT show higher yield for 6-OHBza-RP and 6-OMeBza-RP with 5-OHBza and 5-OMeBza respectively. Similarly, *Ds*CobT is regioselective for 5-OHBza-RP and 5-OMeBza-RP with 5-OHBza and 5-OMeBza respectively. The retention of regioselectivity despite the presence of a methyl group indicates that the latter three enzymes might employ, yet unknown, mechanisms for substrate recognition which are similar to the bacterial CobT3 homologs described in chapter 2.

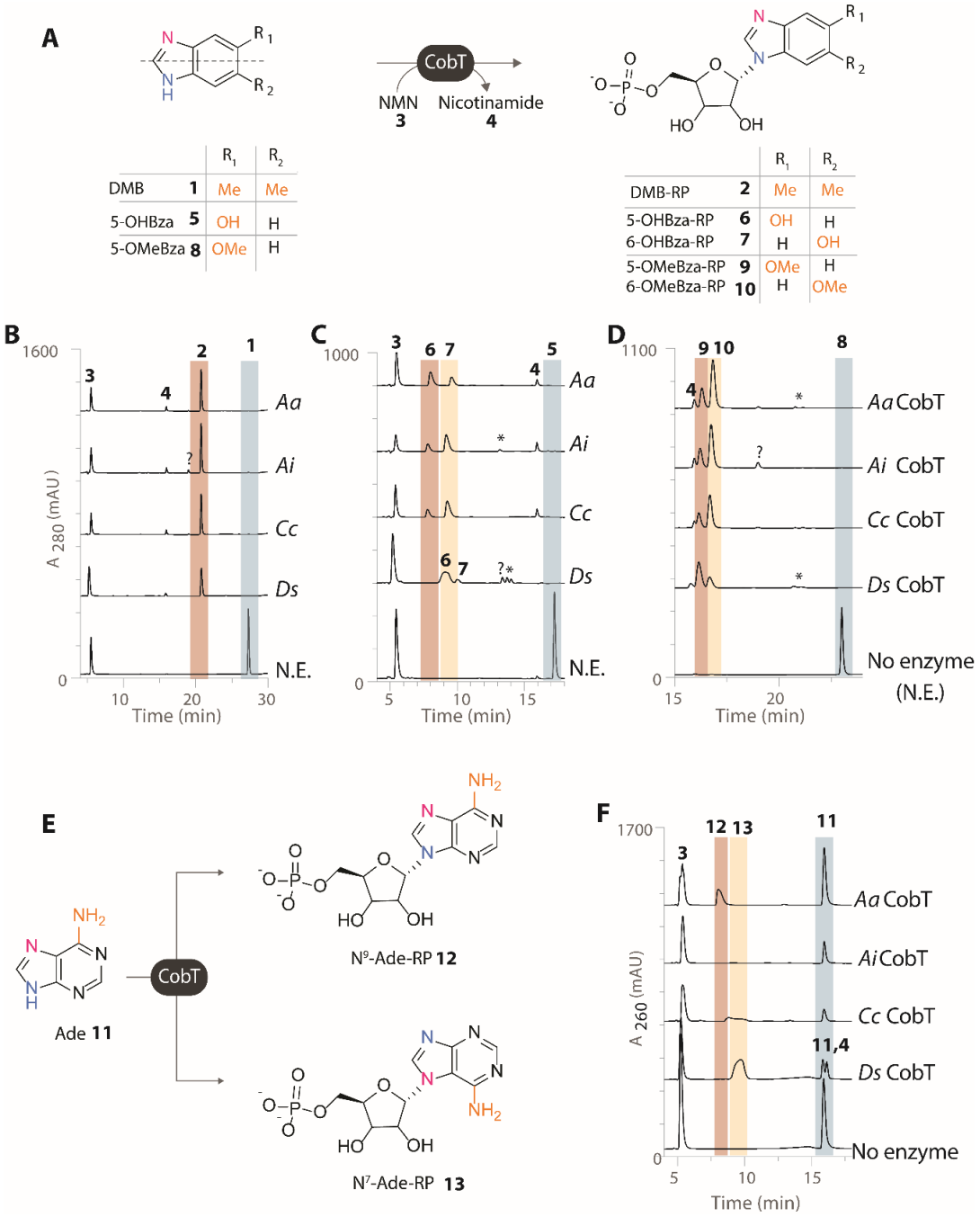


Figure 3.4 Reconstitution of eukaryotic CobT homologs with DMB. A) Reaction scheme for CobT reaction with DMB **1**, 5-OHBza **5**, and 5-OMeBza **8. B)** Reconstitution of phosphoribosyltransferase activity of eukaryotic CobT homologs shows *Aa*CobT, *Ai*CobT, *Cc*CobT, and *Ds*CobT efficiently activate DMB **1** (peak at 25 min) to produce DMB-RP **2** (peak at 20 min). **C)** The eukaryotic CobT homologs activate 5-OHBza **5** to produce 5-OHBza-RP **6** and 6-OHBza-RP **7** in varying ratios of regioselectivity. The *Aa*CobT and *Ds*CobT show regioselective formation of 5-OHBza-RP **6** whereas the *Ai*CobT and *Cc*CobT show regioselective formation of 6-OHBza-RP **7**. **D)** The reactions with 5-OMeBza **8** also yield both expected products where 6-OMeBza-RP **10** is preferred product for *Aa*CobT, *Ai*CobT, and *Cc*CobT where as the *Ds*CobT regioselective formation of 5-OMeBza-RP **9**. The asterisk (*) marked peaks correspond to the riboside derivatives that is the dephosphorylated products formed from spontaneous degradation of the riboside phosphate products. The peaks of unidentified nature that appear to be derived from the enzyme are marked by a question (?) mark. **E)** Activation of adenine by CobT can lead to two isomeric products N^9-Ade-RP **12** and N^7-Ade-RP **13. F)** The tested eukaryotic CobT homologs show stark differences in activation of adenine such that *Aa*CobT and *Ds*CobT show efficient conversions, *Cc*CobT show poor activity, and *Ai*CobT does not activate adenine. The identity of all product peaks were confirmed using LCMS analyses.

Next, we tested the activity of eukaryotic CobTs for activity with adenine (Figure 3.4E). Previous studies have established that most bacterial CobTs typically show poor activity with adenine (Hazra et al. 2013). The HPLC chromatogram with CobT reconstitution with adenine shows substrate peak at 16 minutes, and the two possible products N^9-adenine-RP, and N^7-adenine-RP are expected to elute at 8 min and 10 min, respectively. Similar to bacterial CobT homologs, the *Aa*CobT shows poor activity with adenine conversion with adenine (Figure 3.4F). We observe two peaks for reaction between *Cc*CobT and adenine, and further analysis to ascertain the molecular identity of each the products is underway. On the contrary, the *Ai*CobT and *DsC*obT homologs show a higher activity with adenine of 15% and 82 % conversion with an apparent preference to produce N^9-adenine-RP (N^9-Ade-RP) and N^7-adenine-RP (N^7-Ade-RP), respectively. We also detect phosphoribosyltransferase activity with all four homologs with benzimidazole, 5-methylbenzimidazole, which are naturally occurring cobamide lower ligands and with 5-chlorobenzimidazole which can be attached as lower ligands via guided biosynthesis. We are currently examining the scope of activation of a range of purinyl substrates by these homologs which might expand our understanding the physiological function of the eukaryotic CobTs.

3.4.4 A comparative genomics approach to grasp B_{12} acquisition, assembly, and utilization in eukaryotes

To grasp the physiological role of the eukaryotic CobT homologs in the context of B_{12} metabolism, we questioned whether the organisms that harbor a cobT gene require B_{12} as a cofactor in the first place. Despite widespread use of B_{12} in an array of microbial pathways, the eukaryotes are known to use B_{12} only for methionine synthesis and isomerization of methylmalonyl-CoA, a catabolic product of branched chain fatty acids and amino acids (Banerjee and Ragsdale 2003; Rodionov et al. 2003; Y. Zhang et al. 2009). We probed the translated genomes that encode CobT for B_{12}-dependent methionine synthase (MetH), methylmalonyl-CoA mutase (MCM), and the enzymes required for installing the methyl and the deoxyadenosyl upper ligands (MeaB, CblABCDE). We also tested the presence of B_{12}-dependent ribosnucleotide reductase (RNR II), which catalyzes formation of deoxyribonucleoties from ribonucleotide in several bacteria (Rodionov et al. 2003; Shelton et al. 2018). To do so, we used hidden Markov models (HMM) available on PFAM database as queries to search the reference genomes using HMMsearch available on EBI. We find a widespread occurrence of MetH and MCM as opposed to sparsely present RNR-II, and most organisms that encode a B_{12}-dependent enzyme harbor the enzymes required for B_{12} trafficking and upper ligand metabolism (Figure 3.5, columns 1-9). We then probed the translated genomes for occurrence of B12 uptake and transport pathway, and we looked for bacterial homologs for

BtuB, BtuC, BtuD, BtuE, BtuF, and BtuR proteins (Figure 3.5, columns 12-16). Only a handful of organisms show positive hits for the transport machinery, and since the *btu* operon proteins work in a cascade for B_{12} transport, the organisms need to encode for the complete machinery to successfully import B_{12} from its environment.

Finally, organisms use various strategies to obtain B_{12} for cellular metabolism (Figure 3.6A). A small subset of bacteria and archaea such as *Sinorhizobium meliloti*, *Eubacterium limosum* conduct *de novo* biosynthesis pathway (Figure 3.6A-i). Most B_{12}-dependent organisms including microbes, algae, and animals, including humans, rely on importing the entire B_{12} as a nutrient. Among microbes, some bacteria such as *Escherichia coli* lacks the complete pathway but encodes the CobUTSC genes required for B_{12} assembly. Thus *E. coli* can import cobinamide or corrin ring, and uptake lower ligands available in the environment, and assemble a cobamide- this mode of B_{12} biosynthesis is called the salvage pathway (Figure 3.6A-ii). Some exceptional organism such as *Listeria innocua* salvages only activated lower ligand through ribazole salvaging pathway and then use the CobU and CobS enzymes to assemble B_{12} (Figure 3.6A-iii). Lastly, some organisms such as *Dehalococcoides mccartyii* use a much-complicated method called the remodeling pathway, wherein the organism uptakes any available cobamide from their environment, then removes it the nucleotide loop to obtain a cobinamide backbone which undergoes B_{12} assembly with a new lower ligand (Figure 3.6A-iv). The lower ligand in this case can either be made by the organisms via de novo pathway or imported from the environment. Thus, the nucleotide assembly pathway catalyzed by CobU, CobT, CobS, and CobC are common to all the modes of B_{12} assembly (Figure 3.6A).

Based on this information, we test the present of CobU, CobS, and CobC in the eukaryotic organisms that harbor a CobT. Since, CobC is a phosphatase with an unusually high sequence similarity with histidine phosphatase, a reliable HMMsearch for this class could not be conducted. We also looked for CbiZ and CobD which are known to be catalyze excision of nucleotide loop, and rebuilding of aminopropanol tail in the remoedlling pathway. We find that CobS is present in several organism indicating the possibility of cobamide assembly step (Figure 5, column 10). However, the infrequent occurrence of CobU, CbiZ, and CobD raises questions on how the organisms would obtain an activated corrin ring (Figure 3.5, columns 11, 17, 18). Recently, the evidence for CbiZ-independent remodeling pathways have surfaced. A newly discovered enzyme CbiR cleaves the phosphoribosyl bond to initiate remodeling in *Akkermansia muciniphila* and the CobS enzyme appears to catalyze a direct remodeling in *Vibrio cholera* (Mok et al. 2020; Ma and Beld 2021). While the former pathway is predicted to utilize CobU, the direct remodeling by CobS scheme appear to the bypass the CobU step by a yet unknown mechanism. On similar lines, the B_{12}-dependent strain of marine alga *Chlamydomonas reinhardtii* which lacks a CbiZ but encodes a CobS might be involved in direct

cobamide remodeling. Further analyses of CobS-dependent remodeling mechanism and the adaptations in CobS sequences that renders an additional hydrolase activity, will help in reliably predicting cobamide remodeling in prokaryotes as well as eukaryotes.

Figure 3.5 A comparative genomics approach to understand the scope of B12 utilization, import, remodeling, and assembly in eukaryotes harboring a CobT. RNR II, MetH, MCM (column 1-3) are genes that require B_{12} as a cofactor and MeaB, CblABCDE (column 4-9) are the associated enzymes required for installing catalytic upper ligands. The CobS and CobU (column 10-11) are required alongwith CobT to complete the cobamide assembly. The BtuBCDFR (lane 12-16) are proteins involved B_{12} transport, and CbiZ and CobD (lane 17-18) participate in cobamide remodeling. Overall, the comparative genomics provide the basis for finding the physiological role of CobT in the eukaryotic organisms in the context of B_{12} assembly, and related metabolism. The grey boxes indicate asbence of the gene of interest, purple box

indicates detection of a hit below the reliable cut-off for HMM, and the yellow box show the presence of the gene of interest in the corresponding organism.

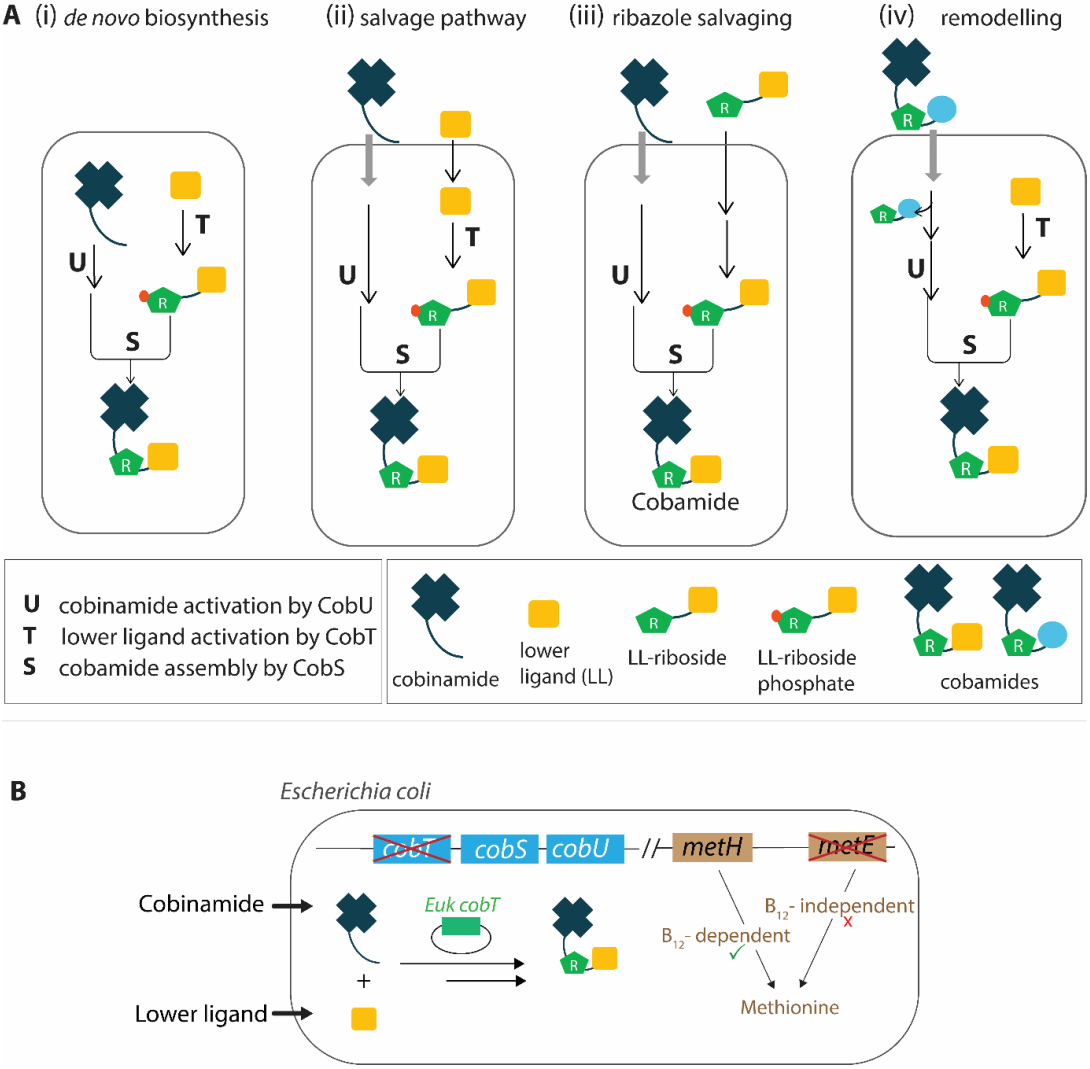


Figure 3.6 Role of CobT in different modes of cobamide biosynthesis and proposed methods for validating *cobT* function. A) Microorganisms use various modes to produce B_{12}. i) the *de novo* biosynthesis pathway – organisms can produce a complete cobamide using the ~30 steps pathway. The nature of lower ligand can vary acorss microbes. ii) the salvage pathway- certain organisms lack genes for biosynthesis of the corrin ring, however the corrin ring or cobinamide can be imported from the environment. The lower ligand can either be imported from the environment, or produced within the organism. iii) the ribazole salvaging pathway- a variation of the salvage pathway where the organism imports DMB-riboside phosphate (called α-ribazole), which is phosphorylated and then fluxed into the cobamide assembly. iv) the remodeling pathway- certain organisms can import various cobamides available in the environment and remodel the strcuture replace the lower ligand. The various modes of cobamide biosynthesis converge at the nucleotide assembly pathway which involves three steps- activation of corrin ring by CobU (labelled U), activation of lower ligand by CobT (lablled T), and cobamide assembly by CobS (labelled S) and removal of the 5' phosphate group by CobC (not shown here for simplicity). **B)** Experimental design for testing the activity of eukaryotic CobT homologs in an engineered strain *E.coliΔmetEΔcobT*. The strain lacks the B_{12}-independent methionine synthase and hence when grown under minimal media, the strain requires B_{12} or any other complete cobamide. *E.coli* can salvage imported cobinamide and lower ligand to assemble a cobamide, which in turn can support B_{12}-dependent methionine synthease and thus supports the growth of bacteria. In absence of the genomic *cobT*, the strain relies on the heterologous expression of the eukaryotic CobT homologs which are expressed under a constitutive promoter to produce cobamides. The screen can help verify whether the eukaryotic CobT homologs can support B_{12} assembly within a cell as well as probe the physiological implications of different substrate preferences and regioselectivity commonly observed among CobT homologs.

3.5 Discussion

In this study, we used the insights gained form our studies on bacterial CobT homologs to discover and characterize CobT homologs in eukaryotic genomes. We filter the sequences based on sequence conservation in the known binding sites for substrate, phosphoribosyl donor, and the catalytic residues (Figure 3.2). The *in vitro* assays prove that the eukaryotic CobT homologs from *Aphanomyces astaci* (*Aa*CobT), *Aphanomyces invadans* (*Ai*CobT), *Chytriomyces confervae* (*Cc*CobT), and *Dunaliella salina* (*Ds*CobT). We find that all the four tested homologs activate DMB and a range of other benzimidazole derivates, however only *Aa*CobT and *Ds*CobT can activate adenine (Figure 3.4A, 3.4B). Moreover, the homologs show differences in regioselectivity in activation of 5-OHBza and 5-OMeBza. Notably, the *Aa*CobT and *Ai*CobT share 75% sequence identity (Figure 3.2), and yet show difference in the enzyme activity in terms of range of substrate as well as changes in regioselectivity with a subtle change in substrate structure (Figure 3.4C, 3. 4D). Thus, *Aa*CobT and *Ai*CobT homologs are reasonable candidates to further explore molecular mechanisms behind substrate recognition and regioselectivity among CobT homologs. Thus, on the lines with our comparative study with bacterial CobTs encoded in different gene neighborhood which lays the foundation for investigating the residues responsible for differences in regioselectivity, eukaryotic CobT homologs described here can also be further analyzed to these nuances of CobT activity. The ability to pin the patterns in sequences that enable the enzyme to recognize a purinyl substrate *vs* benzimidazolyl substrate will be help us detect B_{12} producers apart from pB_{12} producers among metagenomic data such as from human gut or fermented food products.

Additionally, the activation of adenine by *Aa*CobT and *Ds*CobT hint towards possibility of production of pseudocobalamin (pB_{12}) which is the cobamide with adenine as the lower ligand and is produced by a large proportions of microbes (Degnan, Taga, and Goodman 2014; R. H. Allen and Stabler 2008; Katherine Emma Helliwell et al. 2016). Previous experiments have shown that addition of DMB in even low concentrations can direct the organism to produce B_{12} instead, and this observation has been attributed to general preference of CobT for DMB (Anderson et al. 2008; Hazra et al. 2013). Though bacteria can use various cobamides including pB_{12}, eukaryotes are known to specifically require B_{12}, that is the cobamide with DMB as the lower ligand (Sokolovskaya et al. 2021; 2019; Helliwell et al. 2016; Cooper et al. 2019; Baum et al. 2020). For instance, pB_{12} does not support methionine synthesis in eukaryotic organisms and is rather detrimental to the growth of the organism (Helliwell et al. 2016). Hence, the substrate preferences of the eukaryotic CobT homologs can be instrumental in determining survival of the organism.

To verify the physiological significance of CobT in eukaryotes, experiments with native organisms encoding the *cobT* gene needs to be conducted. Among the sources of CobT

homologs tested in this study, there is no literature evidence for B_{12}-related metabolism in parasitic fungal species *Aphanomyces astaci* and *Aphanomyces invadans*, and soil fungus *Chytriomyces confervae*. Interestingly, *Dunaliella salina*, is a halotolerant microalga which is known to contain B_{12} (Kumudha et al., 2016). Several other algae such as *Chlamydomonas reinhardtii, Thalassiosira oceanica*, and *Ostreococcus tauri* that have been shown to acquire and assimilate B_{12} via symbiotic relationship with B_{12}-producing bacteria also appear to contain a CobT and CobS gene in our comparative genomics study (Cooper et al. 2019; Sokolovskaya, Shelton, and Taga 2020; Helliwell et al. 2016). However, limited methods to genetically manipulate and grow axenic cultures of such eukaryotic organism pose challenges to validate gene function. As a first step towards *in vivo* characterization of eukaryotic CobT homologs, we are currently investigating whether these homologs retain their activity, substrate preference and regioselectivity under physiological conditions using an engineered *E.coli* model system. We have cloned the *Aa*CobT, *Ai*CobT, *Cc*CobT, and *Ds*CobT in the pETmini vector or a modified pCA24N vectors (Mok et al. 2020, REF) under a constitutive promoter (Figure 3.6B). The *E.coliΔmetEΔcobT* strain is a B_{12}-dependent methionine auxotroph which lacks a *cobT* gene in the genome, however the native *cobU*, *cobS* and *cobC* are present which can participate in the cobamide assembly(Hazra et al. 2015). Hence, when grown in minimal media in presence of cobinamide and lower ligands such as DMB or adenine, the strain can survive only if the CobT expressed from the plasmid can activate the lower ligand. The screen will help provide evidence if the observed *in vitro* activity of eukaryotic CobT homologs holds true under physiological conditions.

Apart from algae, we also find CobT in protists and fungi, several of which show genes for B_{12} dependent enzymes, B_{12} transport, and CobS for B_{12} assembly. However, the B_{12} utilization and metabolism in fungi is only beginning to be explored, and hence extensive experimentation with species that were predicted to be B_{12}-dependent (Orłowska, Steczkiewicz, and Muszewska 2021) and the organisms we found to harbor genes for B_{12} assembly and uptake can inform more details on B_{12} metabolism among fungi.

Notably, DMB and other benzimidazoles that act as cobamide lower ligands are a niche class of biomolecules known, to date, to be produced only by a small fraction of cobamide producing bacteria and archaea. A previous study has proven presence of benzimidazoles in several environmental samples indicating the abundance of these molecules and the projected influence of benzimidazoles on natural microflora (Crofts et. al., 2014). Thus, if the subset of eukaryotic microorganisms reported here express the encoded CobT, they can feasibly produce α-ribotide derivatives. The fate of the α-ribotide products is puzzling and presents multiple possibilities.

The downstream nucleotide assembly pathway requires CobU and CobS for cobamide assembly as well as the entire machinery for transport of cobamide precursors. The increased genetic load and the metabolic expenditure might have been a limiting factor which may explain why far less eukaryotic genomes show the lower ligand attachment genes. Thus, it'd be an interesting approach to probe whether the eukaryotic microorganisms use CobT to produce activated lower ligands as molecules of communication or exchange with other organisms in their communities. Interestingly, benzimidazole-based molecules have also been shown to potently inhibit bacterial cytoskeletal protein (Hurley et. al., 2016). Thus, can ribotide or riboside derivatives of the benzimidazoles can participate as agents of interspecies competition is an interesting avenue? From an enzyme function perspective, CobT is a unique phosphoribosyltransferase which produces α-glycosides which in turn are rare biomolecules. We observe several instances of *cobT* genes in non-B_{12} related bacterial and archaeal genomic contexts in chapter-2 and the function of such homologs are yet unknown. Thus, it is compelling to probe the function of CobT and α-riboside derivatives in biochemical pathways other than B_{12} biosynthesis in eukaryotic as well as prokaryotic systems.

Overall, identification and validation of the eukaryotic CobT homologs partly provides an explanation for previous studies that show cobamide remodeling in unicellular algae. We find that cobamide synthase (CobS) is also widespread in algae that encode a CobT, and further verification of CobS activity will help us understand the possibility and the mechanism of cobamide remodeling and/or assembly in such organisms.

References:

Allen, Robert H., and Sally P. Stabler. 2008. "Identification and Quantitation of Cobalamin and Cobalamin Analogues in Human Feces." American Journal of Clinical Nutrition 87 (5): 1324–35.

Anderson, Peter J., Jozsef Lango, Colleen Carkeet, Audrey Britten, Bernhard Kräutler, Bruce D. Hammock, and John R. Roth. 2008. "One Pathway Can Incorporate Either Adenine or Dimethylbenzimidazole as an α-Axial Ligand of B12 Cofactors in Salmonella Enterica." Journal of Bacteriology 190 (4): 1160–71. https://doi.org/10.1128/JB.01386-07.

Banerjee, Ruma, and Stephen W Ragsdale. 2003. "The Many Faces of Vitamin B12 : Catalysis by Cobalamin-Dependent Enzymes." Annual Reviews Biochemistry 72: 209–47. https://doi.org/10.1146/annurev.biochem.72.121801.161828.

Baum, Christoph, Riya C. Menezes, Aleš Svatoš, and Torsten Schubert. 2020. "Cobamide Remodeling in the Freshwater Microalga Chlamydomonas Reinhardtii." FEMS Microbiology Letters 367 (20): 1–6. https://doi.org/10.1093/femsle/fnaa171.

Cooper, Matthew B., Elena Kazamia, Katherine E. Helliwell, Ulrich Johan Kudahl, Andrew Sayer, Glen L. Wheeler, and Alison G. Smith. 2019. "Cross-Exchange of B-Vitamins Underpins a Mutualistic Interaction between Ostreococcus Tauri and Dinoroseobacter Shibae." ISME Journal 13 (2): 334–45. https://doi.org/10.1038/s41396-018-0274-y.

Crofts, Terence S., Yujie Men, Lisa Alvarez-Cohen, and Michiko E. Taga. 2014. “A Bioassay for the Detection of Benzimidazoles Reveals Their Presence in a Range of Environmental Samples.” Frontiers in Microbiology 5 (NOV): 1–12. https://doi.org/10.3389/fmicb.2014.00592.

Croft, Martin T., Andrew D. Lawrence, Evelyne Raux-Deery, Martin J. Warren, and Alison G. Smith. 2005. “Algae Acquire Vitamin B12 through a Symbiotic Relationship with Bacteria.” Nature 438 (7064): 90–93. https://doi.org/10.1038/nature04056.

Degnan, Patrick H., Michiko E. Taga, and Andrew L. Goodman. 2014. “Vitamin B12as a Modulator of Gut Microbial Ecology.” Cell Metabolism 20 (5): 769–78. https://doi.org/10.1016/j.cmet.2014.10.002.

Edgar, Robert C. 2004. “MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput.” Nucleic Acids Research 32 (5): 1792–97. https://doi.org/10.1093/nar/gkh340.

Finn, Robert D, Alex Bateman, Jody Clements, Penelope Coggill, Y Eberhardt, Sean R Eddy, Andreas Heger, et al. 2014. “Pfam : The Protein Families Database” 42 (November 2013): 222–30. https://doi.org/10.1093/nar/gkt1223.

Finn, Robert D, Jody Clements, and Sean R Eddy. 2011. “HMMER Web Server : Interactive Sequence Similarity Searching.” Nucleic Acids Research 39 (May): W29–37. https://doi.org/10.1093/nar/gkr367.

Gerlt, John A. 2017. “Genomic Enzymology: Web Tools for Leveraging Protein Family Sequence-Function Space and Genome Context to Discover Novel Functions.” Biochemistry 56 (33): 4293–4308. https://doi.org/10.1021/acs.biochem.7b00614.

Hall, T.A. 1999. “BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT.” Nucleic Acids Symposium 41: 95–98. https://doi.org/citeulike-article-id:691774.

Hazra, Amrita B., Andrew W. Han, Angad P. Mehta, Kenny C. Mok, Vadim Osadchiy, Tadhg P. Begley, and Michiko E. Taga. 2015. “Anaerobic Biosynthesis of the Lower Ligand of Vitamin B12.” Proceedings of the National Academy of Sciences of the United States of America 112 (34): 10792–97. https://doi.org/10.1073/pnas.1509132112.

Hazra, Amrita B., Jennifer L.A. Tran, Terence S. Crofts, and Michiko E. Taga. 2013. “Analysis of Substrate Specificity in CobT Homologs Reveals Widespread Preference for DMB, the Lower Axial Ligand of Vitamin B12.” Chemistry and Biology 20 (10): 1275–85. https://doi.org/10.1016/j.chembiol.2013.08.007.

Helliwell, Katherine E., Glen L. Wheeler, and Alison G. Smith. 2013. “Widespread Decay of Vitamin-Related Pathways: Coincidence or Consequence?” Trends in Genetics 29 (8): 469–78. https://doi.org/10.1016/j.tig.2013.03.003.

Helliwell, Katherine Emma, Andrew David Lawrence, Andre Holzer, Ulrich Johan Kudahl, Severin Sasso, Bernhard Kräutler, David John Scanlan, Martin James Warren, and Alison Gail Smith. 2016. “Cyanobacteria and Eukaryotic Algae Use Different Chemical Variants of Vitamin B12.” Current Biology 26 (8): 999–1008. https://doi.org/10.1016/j.cub.2016.02.041.

Hurley, Katherine A., Thiago M.A. Santos, Gabriella M. Nepomuceno, Valerie Huynh, Jared T. Shaw, and Douglas B. Weibel. 2016. “Targeting the Bacterial Division Protein FtsZ.” Journal of Medicinal Chemistry 59 (15): 6975–98. https://doi.org/10.1021/acs.jmedchem.5b01098.

Kumudha, A., & Sarada, R. (2016). Characterization of vitamin B12 in Dunaliella salina. *Journal of food science and technology*, *53*(1), 888–894. https://doi.org/10.1007/s13197-015-2005-y

Ma, Amy T, and Joris Beld. 2021. “Direct Cobamide Remodeling via Additional Function of Cobamide Biosynthesis Protein CobS from Vibrio Cholerae.” MBio 203 (15): e00172-21.

Mok, Kenny C., Olga M. Sokolovskaya, Alexa M. Nicolas, Zachary F. Hallberg, Adam Deutschbauer, Hans K. Carlson, and Michiko E. Taga. 2020. “Identification of a Novel Cobamide Remodeling Enzyme in the Beneficial Human Gut Bacterium Akkermansia Muciniphila.” MBio 11 (6): 1–18.

Orłowska, Małgorzata, Kamil Steczkiewicz, and Anna Muszewska. 2021. “Utilization of Cobalamin Is Ubiquitous in Early-Branching Fungal Phyla.” Genome Biology and Evolution 13 (4): 1–12. https://doi.org/10.1093/gbe/evab043.

Potter, Simon C, L. Aurelien, Sean R Eddy, Youngmi Park, Rodrigo Lopez, and Robert D Finn. 2018. “HMMER Web Server : 2018 Update.” Nucleic Acids Research 46 (June): W200–204. https://doi.org/10.1093/nar/gky448.

Procter, James B, G Mungo Carstairs, Ben Soares, Kira Moura, T Charles Ofoegbu, Daniel Barton, Lauren Lui, et al. 2021. “Alignment of Biological Sequences with Jalview James.” In Multiple Sequence Alignment: Methods and Protocols, Methods in Molecular Biology, 2231:203–23.

Rodionov, Dmitry A., Alexey G. Vitreschak, Andrey A. Mironov, and Mikhail S. Gelfand. 2003. “Comparative Genomics of the Vitamin B12 Metabolism and Regulation in Prokaryotes.” Journal of Biological Chemistry 278 (42): 41148–59. https://doi.org/10.1074/jbc.M305837200.

Roth, J R, J G Lawrence, and T A Bobik. 1996. “Cobalamin (Coenzyme B12): Synthesis and Biological Significance.” Annual Review of Microbiology 50: 137–81.

Scherer, Paul, Vera Höllriegel, Christine Krug, Michael Bokel, and Paul Renz. 1984. “On the Biosynthesis of 5-Hydroxybenzimidazolylcobamide (Vitamin B12-Factor III) in Methanosarcina Barkeri.” Archives of Microbiology 138 (4): 354–59. https://doi.org/10.1007/BF00410903.

Seth, Erica C., and Michiko E. Taga. 2014. “Nutrient Cross-Feeding in the Microbial World.” Frontiers in Microbiology 5 (JULY): 1–6. https://doi.org/10.3389/fmicb.2014.00350.

Shelton, Amanda N., Erica C. Seth, Kenny C. Mok, Andrew W. Han, Samantha N. Jackson, David R. Haft, and Michiko E. Taga. 2018. “Uneven Distribution of Cobamide Biosynthesis and Dependence in Bacteria Predicted by Comparative Genomics.” ISME Journal 13 (3): 789–804. https://doi.org/10.1038/s41396-018-0304-9.

Sokolovskaya, Olga M., Kenny C. Mok, Jong Duk Park, Jennifer L.A. Tran, Kathryn A. Quanstrom, and Michiko E. Taga. 2019. “Cofactor Selectivity in Methylmalonyl Coenzyme a Mutase, a Model Cobamide-Dependent Enzyme.” MBio 10 (5): 27–32. https://doi.org/10.1128/mBio.01303-19.

Sokolovskaya, Olga M., Tanja Plessl, Henry Bailey, Sabrina Mackinnon, Matthias R. Baumgartner, Wyatt W. Yue, D. Sean Froese, and Michiko E. Taga. 2021. “Naturally Occurring Cobalamin (B12) Analogs Can Function as Cofactors for Human Methylmalonyl-CoA Mutase.” Biochimie 183: 35–43. https://doi.org/10.1016/j.biochi.2020.06.014.

Sokolovskaya, Olga M., Amanda N. Shelton, and Michiko E. Taga. 2020. “Sharing Vitamins: Cobamides Unveil Microbial Interactions.” Science 369 (6499). https://doi.org/10.1126/science.aba0165.

Woodson, Jesse D., and Jorge C. Escalante-Semerena. 2004. “CbiZ, an Amidohydrolase Enzyme Required for Salvaging the Coenzyme B 12 Precursor Cobinamide in Archaea.” Proceedings of the National Academy of Sciences of the United States of America 101 (10): 3591–96. https://doi.org/10.1073/pnas.0305939101.

Zhang, Yan, Dmitry A. Rodionov, Mikhail S. Gelfand, and Vadim N. Gladyshev. 2009. “Comparative Genomic Analyses of Nickel, Cobalt and Vitamin B12 Utilization.” BMC Genomics 10 (78). https://doi.org/10.1186/1471-2164-10-78.

Chapter 4

Reconstitution of enzyme activity of *Mt*BzaC and understanding its role in the *bza* operon pathway

4.1 Introduction

Cobamide diversity among microorganisms arise from variations nucleotide loop wherein the variations are primarily occurring on the lower ligand. Several anaerobic bacteria, specially the acetogens and methanogens, have been extensively explored for the nature of cobamides made and use in the respective organisms(Erhard Stupperich 1988; Kräutler, Kohler, and Stupperich 1988; P Renz 1993; Sokolovskaya, Shelton, and Taga 2020). For example, *Moorella thermoacetica* produces 5-methoxybenzimidazolylcobamide ([5-OMeBza]Cba) which acts as a cofactor for methyl trafficking in methanogenesis in the organism. An acetogen *Acetobacterium woodii* and the gut microbe *Eubacterium limosum* make B_{12} which as 5,6-dimethylbenzimidazole (DMB). Other anaerobes such as *Desulfobulbous propionicus* and *Methanobacterium thermoautotrophicum* make 5-methylbenzimidazolylcobamide ([5-MeBza]Cba), and 5-hydroxybenzimidazolylcobamide ([5-OHBza]Cba) respectively. Most of the variations observed in lower ligand structure typically arise from differences in the number and position of the methyl group. The discovery of the *bza* operon explained origin of four such benzimidazole derivatives that exists as natural lower ligands of cobamides as well as elucidated the genetic reasons leading to differences in methylations on the benzimidazole scaffold.

The *Eubacterium limosum bza* operon was found to have *bzaA-bzaB-cobT-bzaC-bzaD-bzaE* yields 5-hydroxybenzimidazole (5-OHBza **1**), 5-methoxybenzimidazole (5-OMeBza **2**), 5-methoxy-6-methylbenzimidazole (5-OMe-6MeBza) as intermediates and finally 5,6-dimethylbenzimidazole (DMB) as the ultimate lower ligand (Chapter 1, Figure 1.6A) (Hazra et al. 2015). On the other hand, the *Moorella thermoacetica bza* operon has *bzaA-bzaB-cobT-bzaC* genes which result in 5-OHBza as an intermediate and 5-OMeBza as the final lower ligand of the cobamide (Chapter 1, Figure 1.6A)(Hazra et al. 2015). The *bzaC-bzaD-bzaE* enzymes were shown to encode for three methyltransferases that were predicted to catalyze three sequential but chemically distinct methylations (Chapter 1, Figure 1C). According to the pathway predicted by Hazra et. al., (2015), the first pathway intermediate 5-hydroxybenzimidazole (5-OHBza) undergoes an O-methylation catalyzed by gene product of *bzaC* (Figure 4.1A). The predicted domain architecture of BzaC indicates that the enzyme is typical Class-I S-adenosylmethionine (SAM) dependent methyltransferases (Figure 4.1B).

Methyltransferases are ubiquitously found in various biochemical processes and have diverse roles and mechanisms. The SAM Class-I methyltransferases are the most abundant methyltransferases and accept a range of substrates such as DNA, RNA, proteins, and small metabolites such as neurotransmitters, flavonoids, glycine, arginine etc(Martin, Mcmillan, and S-adenosylmethionine-dependent 2002). Since the predicted substrate of BzaC, 5-OHBza, has an aromatic hydroxyl group which is similar to those in catechols, we hypothesized that BzaC to show activity and mechanism similar to catechol O-methyltransferases, which are extensively_studied in literature(Farnberger et al. 2018; Siegrist et al. 2017). A common mechanism used by O-methyltransferases involve a nucleophilic attack by negatively charged oxygen of the substrate onto the methyl group on the positively charged sulfur of S-adenosylmethionine (SAM **3**) (Figure 4.1C). Notably, upon comparing the predicted domain architectures for BzaC homologs from *Moorella thermoacetica* (*MtBzaC*) and *Eubacterium limosum* (*ElBzaC*), we find that the two differ by presence of an additional domain of unknown function 2284 (DUF2284). In this chapter, we conduct primary characterization of *Mt*BzaC, validate its function as O-methyltransferase, and find its preferred substrate. We combine our observations from characterization of *Moorella thermoacetica CobT* (*MtCobT*, described in chapter -2) and *Mt*BzaC. The results allow us to put forth a revised pathway for the *bza* operon wherein the activation of the first intermediate, 5-OHBza **1** preceded the methylation steps, and thus ensures a regiospecific attachment of the lower ligand.

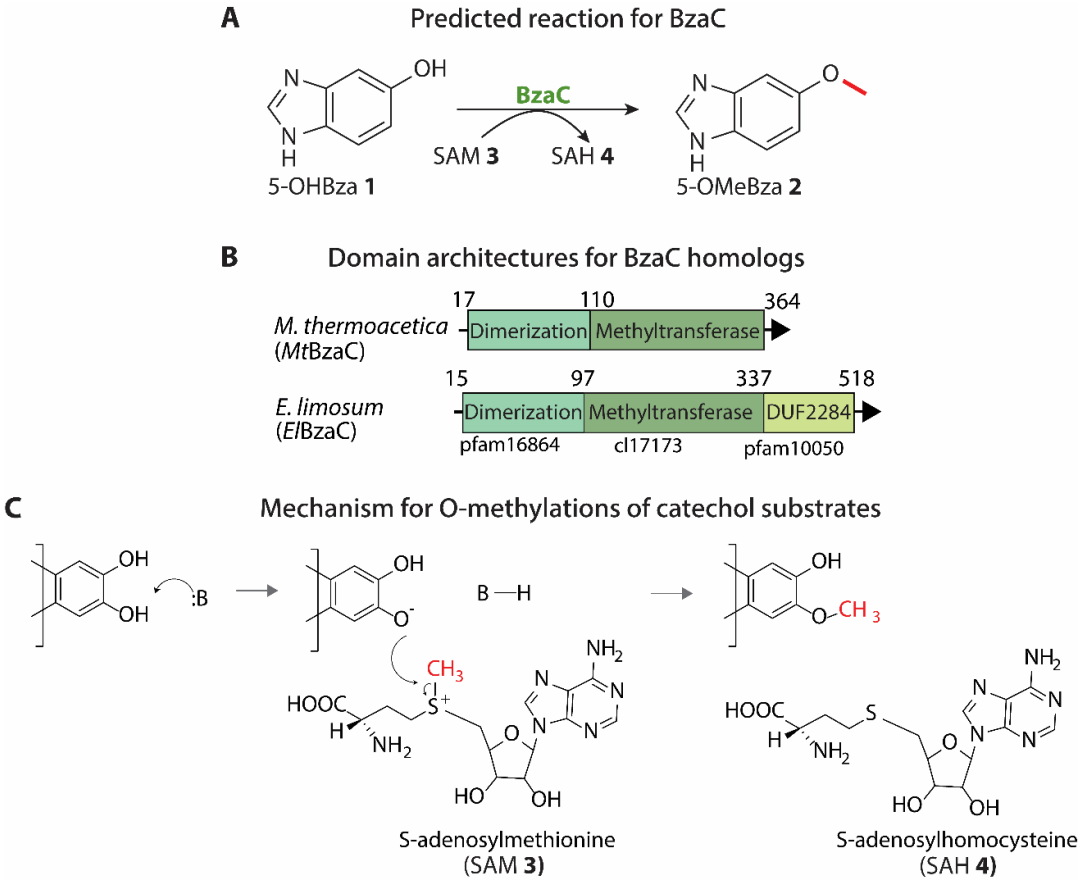


Figure 4.1. Predicted function for BzaC enzymes A) The BzaC is predicted to catalyze the methylation of 5-hydroxybenzimidazole (5-OHBza **1**) using S-adenosylmethionine (SAM **3**) as the methyl donor resulting in 5-methoxybenzimidazole (5-OMeBza **2**) using S-adenosylhomocystein (SAH **4**) as products. **B)** The predicted domain architectures for *Moorella thermoacetica* BzaC (*Mt*BzaC) and *Eubacterium limosum* BzaC (*El*BzaC) show presence of a N-terminal dimerization domain (pfam 16864), a Class-I SAM-dependent methyltransferase (cl 17173). The *El*BzaC additionally contain a domain of unknown function (DUF) 2284 (pf10050). **C)** The general reaction scheme for O-methylation by Catechol-O-methyltransferases which utilize SAM as the methyl donor similar to the predicted function of BzaC enzymes.

4.2 Methods:

4.2.1 Phylogeny and comparative genomics analysis

The candidate BzaC homologs were identified using BLASTp (https://blast.ncbi.nlm.nih.gov/Blast.cgi) (S. Altschul et al. 1990) search by using BzaC sequences from *Eubacterium limosum*, *Moorella thermoacetica*, *Geobacter lovelyi*, *Desulfofundulus kuznetsovii*, *Thermicola potens*, and *Syntrophaceticus schinkii* as queries against the non-reduntant protein sequences database at NCBI GenBank. To identify taxonomically diverse BzaC homologs, BLASTp searches were conducted against each bacterial phylum and then each class within the phyla which returned >100 sequences. Unique sequences were retrieved using Batch entrez (https://www.ncbi.nlm.nih.gov/sites/batchentrez). The resulting list of 833 sequences comprising of O-methyltransferases and hypothetical proteins that share sequence homology with known BzaC proteins. From the dataset, the BzaC

sequences were identified based on scores from HMMsearch HMM models(Shelton et al. 2018), and gene neighborhood search in the GenBank(Benson et al. 2005). Since most of the *bza* operon gene products were annotated as hypothetical proteins at the time of this study, the identity of the neighboring genes was verified using Conserved domain search(Marchler-Bauer and Bryant 2004). The O-methyltransferases that were encoded downstream to cobalamin riboswitch or in a neighborhood with *thiC* homologs (which correspond to *bzaAB* or *bzaF*), phosphoribosyltransferase (*cobT*), and B_{12} radical-SAM enzymes (*bzaD* and/or *bzaE*) were considered as BzaC homologs. The resulting set of 90 unique BzaC-like protein sequences was aligned using MUSCLE (https://www.ebi.ac.uk/Tools/msa/muscle/) (Edgar 2004) and multiple sequence alignment was visualized, edited to remove large gaps, and then converted to phylip4.0 format using BioEdit(Hall 1999). The phylip4.0 file was used to build a maximum likelihood tree using rapid bootstrapping RaXML-HPC2(Stamatakis 2014) on XSEDE (Towns et al. 2014) on the CIPRES Science gateway server (http://www.phylo.org) (Miller, Pfeiffer, and Schwartz 2010). The results were extracted as a Newick tree which was visualized and annotated on the iToL tool (http://itol.embl.de) (Letunic and Bork 2016). The gene neighborhood-based search allows us to confidently predict BzaC homologs that may accept 5-OHBza as the substrate from the superfamily of thousands of Class-I SAM-dependent O-methyltransferases. The patterns and consensus conserved in this dataset of 93 sequences can be used to build HMM models which can enable further detection of candidate 5-OHBza methyltransferases.

4.2.2 Molecular cloning and construction of plasmids

Genes for *bzaC* from *M.thermoacetica* was amplified from respective *bza* operons from the plasmids pKM077 and pKM076 (Hazra et al. 2015), respectively(Cheng et al. 2007). The restriction-free (RF) cloning protocol as described in section 2.3.3 was used to clone the gene into pET28a that provides a N-terminal 6x-His tag(Van Den Ent and Löwe 2006). The resultant plasmids pYMH002 was used to create a mutant with dimerization domain deleted, *Mt*BzaCΔDiD. To determine the boundaries of the dimerization domain, we modelled the structure of *Mt*BzaC using iTasser (Yang and Zhang 2016) and PHYRE2 (Kelly et al. 2015). Five models with highest scores indicated that the dimerization domain ranges from Gly37 to Leu115/Val116 (numbered as in the tagged protein), where latter residue is predicted to be a part of the loop connecting to the next domain. Using restriction-free cloning, the DNA sequence corresponding to G37 to Leu115 was removed from pYMH002 resulting in deletion of the dimerization domain (pYMH014). In this study, we also used pET28a pMSH001 and pDH004 which contain *M.thermoacetica cobT* and *Escherichia coli cobT* in pET28a, respectively. Additionally, we cloned *E. coli metK* and *E. coli mtn* in pET15b and pET28a

respectively (pMSH002, pYMH015), which allow overexpression of SAM synthetase and 5-methyladenosine nucleosidase (MTAN).

4.2.3 Overexpression and purificaton of the recombinant proteins

The *E. coli* BL21 (DE3) cells harboring the plasmid of interest grown in LB broth containing 50 μg/ml kanamycin and incubated at 37˚C till $O.D._{600}$ of 0.6. Protein overexpression was induced by 0.25 to 0.5 mM IPTG and the culture was incubated for 15 h at 25 ˚C or 20 ˚C depending on optimized expression conditions for each construct. Cells were harvested by centrifugation at 4000 rpm at 4 ˚C and stored at −80 °C.

Purification of *Mt*BzaC, *Ec*MetK, and *Ec*MTAN: Cell pellets were resuspended in lysis buffer (50 mM Tris-Cl pH 8.0, 300 mM NaCl, 10mM imidazole, 10 mM $MgCl_2$, 0.025% β-mercaptoethanol, and 100 μM PMSF) followed by lysis by sonication at 60 % amplitude for 15 minutes of 1 sec on, 3 sec off cycles. The lysate centrifuged at 18000 rpm at 4°C for 20 minutes. The clarified lysate was loaded onto a Ni-NTA column pre-equilibrated with lysis buffer. The column was washed with buffers containing 10 mM and 50 mM imidazole, the protein was by eluted using 50 mM Tris-Cl pH8.0, 300mM NaCl, 10 mM $MgCl_2$, 250 mM imidazole, and 0.025% β-mercaptoethanol. The eluent protein was desalted in 50 mM Tris-Cl pH 8.0, 150 mM NaCl, 5 mM $MgCl_2$, and 0.025% β-mercaptoethanol.

For assays with *Mt*BzaC, the desalting step was conducted in an anaerobic glove bag as required using desalting buffer which was previously buffer exchanged to remove dissolved oxygen. The concentration of the desalted protein was measured using absorbance at 280 nm and the predicted molar extinction coefficient 38765 $M^{-1}cm^{-1}$. Purified *Mt*BzaC lost activity upon freezing and hence, the desalted protein was freshly used for setting up the reactions. The concentration of purified *Ec*MetK and *Ec*MTAN was calculated using Bradford assay and the protein was stored in -80°C as 15% glycerol stocks (Ernst and Zor 2010). The *Mt*CobT and *Ec*CobT was purified using methods described in section 2.3.3.

4.2.4 Synthesis of the 5-OHBza substrate

5-OHBza **1** was synthesized from 5-OMeBza **2** as described previously(Crofts et al. 2013) with few modifications (Figure 4.4A). Briefly, 600 mg of 5-OMeBza **2** and 27 mL of 48% HBr at 105°C were reacted under reflux with constant stirring and inert atmosphere for 8 hours. We monitored the reaction progress was monitored using thin layer chromatography using solvent system with chloroform: methanol: trimethylamine:: 18:2:0.1. Once the 100% conversion was achieved, the reaction mixture was placed on ice and neutralized with NaOH with intermittent stirring. Further, water was evaporated from the reaction mix by rotatory

evaporation under vaccuum. 5-OHBza **1** was separated from NaBr through multiple extractions in small volumes of acetonitrile and was later filtered. Filtrate was used to make a slurry with silica and subjected to normal phase column chromatography. The desired compound was eluted in 5%: 10%:: CH_3OH: $CHCl_3$. Colored impurities were removed upon treatment with activated charcoal. Purity of the final compound obtained was checked through ^1H NMR. The final yield was calculated to be 30%.

4.2.5 Size exclusion chromatography

Size exclusion chromatography was performed on analytical GE Sephadex-200 column calibrated from 600 kDa to 29 kDa (Sandanaraj et al. 2018). The column was equilibrated with 50 mM Tris-Cl pH 8.0 containing 150 mM NaCl and 0.025% β-mercaptoethanol. Protein eluted from Ni-NTA chromatography was directly loaded onto the equilibrated column and the 30 ml of the buffer was run through the column at 0.25 ml/min. The chromatographs were recorded at 280 nm to monitor protein elution profile.

4.2.6 Intrinsic fluorescence assays

To study protein-cofactor binding, we used the phenomenon of fluorescence quenching of the protein. Purified *Mt*BzaC show fluorescence emission maxima at 328 nm upon an excitation at 280 nm. We calculated the K_d value for SAM or SAH binding to the BzaC as follows: In a pre-chilled 96-well fluorescence plate, 100 µL of 10µM protein was added in 9 wells. Then, pre-chilled 2x stock solution of the ligand was added to the protein. The plate was incubated in the plate-reader at 25°C for 10 minutes and then a fluorescence emission was recorded in the range of 300 nm to 600 nm. The fold change in fluorescence at 328 nm was plotted against the final concentration of ligand in each well. The plot was fit to following equation to find the K_d value: $F = F_0 + \frac{\Delta F\,[S]}{[S]+K_d}$. Where, F = fluorescence for Protein & ligand mix, F_0 = fluorescence for Protein only, $\Delta F = F_{max} - F$, [S] = concentration of ligand, K_d = dissociation constant.

4.2.7 Enzymatic Reactions

For all endpoint enzyme assays with *Mt*BzaC, the purified protein was desalted and reactions were setup under either ambient lab conditions or anoxic conditions in specified experiments. A typical reaction was set up in 50 mM Tris-Cl pH 8.0 with 46µM to 50µM of protein, 500 µM of substrate and 1000 µM of SAM **3**, 10mM $MgCl_2$, and 0.025% β- mercaptoethanol. The reaction was incubated for 48 hours at 25°C or at temperatures specified in the experiments, and thereafter the reaction was quenched by addition of 4% formic acid. The mix was

centrifuged at 14000 rpm at 4°C to remove precipitated protein(s), and the supernatant was used for HPLC and LC-MS analyses.

4.2.8 Reaction monitoring through reverse phase HPLC and LCMS

The reaction turnover was monitored using the reverse phase HPLC on Agilent 1260 infinity II UHPLC system paired with a UV-Vis DAD detector. An Agilent Zorbax Eclipse Plus- C18 5µ 150mm x 4.60 mm column was used at 0.5mL/min at 25°C and analyzed on HPLC using method-I.

Method-I: Mobile phases used were Solvent A: ammonium acetate, 10mM, pH 6.5 and solvent B: methanol. The gradient was run as follows: 100% solvent A from 0 to 2 minutes, gradient from 100% solvent A to 70% solvent B from 2 to 25 minutes and a gradient to 100% solvent B from 25 to 28 minutes and then column was re-equilibrated in 100% solvent A from 28 to 30 minutes. The chromatograms were recorded at 250 nm, 260nm, and 280 nm. LCMS analyses were performed on Exion-LC series UHPLC attached to a ScieX X500R QTOF Mass spectrometer system. The HPLC Method I was coupled with the MS method-1 which is as follows: The TOF-MS analyses were conducted in +ESI mode with +4500 V ion spray voltage, de-clustering potential of 80V, and low collision energy of 5V at 400 °C. TOF-MS was followed by targeted MRM-HR analysis using following precursor ion mass and fragment ion mass for the analyte of interest: SAM **3** (399.1, 399.1445), 5-OMeBza **2** (149.0, 149.0709), 5-OHBza **1** (135.0, 135.0553), SAH **4** (385.1, 385.1289), MTA (298.1, 298.0968), 5-OHBza-RP **5** and 6-OHBza-RP **6** (347.0693, 135.0553), 5-OHBza-R **7** and 6-OHBza-R **8** (267.0975, 135.0553), 5-OMeBza-RP (361.0795, 149.0709), and 5-OMeBza-R **9** and 6-OMeBza-R **10** (281.1132, 149.0709). Data was processed on ScieX-OS to obtain extracted ion chromatogram and mass spectra from the total ion chromatogram.

4.2.9 Purification of reaction product and NMR analysis

To scale-up the reaction for NMR of product peak, a 25 mL of reaction using 94.2 µM *Mt*BzaC with 1 mM 5-OHBza **1** and 2 mM SAM **3** as substrates for 48 hours at 25°C. The product peak was purified through reverse phase HPLC on a Phenomenex Luna 5u C18 (2) 100Å 250mm x 4.60 mm column using LC method I. The collected product was dried on a under vacuum, dissolved into 400 µl D_2O. The ^1H NMR was recorded at 400 MHz for 32 scans. As a standard, 4 mg of 5-OMeBza **2** standard was dissolved in 500 µl D_2O and the ^1H NMR at 400 MHz was analyzed. The NMR spectra were analyzed using MNOVA 7.2.1.

4.2.10 Synthesis of the phosphoribosylated and ribosylated derivatives of 5-OHBza

The phosphoribosylated and ribosylated derivatives of 5-OHBza **1** were synthesized enzymatically using CobT (Figure 4.7). To synthesize 5-OHBza-RP **5**, 1mM of 5-OHBza **1** was incubated with 10 µM of purified *Mt*CobT enzyme and 2 mM of nicotinamide mononucleotide (NMN) at 25°C for 48 hours in 50 mM Tris-Cl pH 8.0 and 10 mM $MgCl_2$. To synthesize 6-OHBza-RP **6**, 1mM of 5-OHBza **1** was incubated with 10 µM of purified *Ec*CobT enzyme and 2 mM of NMN at 25°C for 48 hours in 50 mM Tris-Cl pH 8.0 and 10 mM $MgCl_2$. The products were individually purified using reverse phase HPLC using Phenomenex Luna 5µ C18 (2) 100Å using Method-II that is as follows: Mobile phases that were used: Milli-Q H_2O (solvent A), methanol (solvent B), and ammonium acetate, 10 mM, pH 6.5 (solvent C). 0-3 min, 10% A and 90% C; 3-6 min 15% A and 85% C; 6-10 min, 50% A and 50% C; 10-13 min, 50% B and 50% C; 13-16 min, 100% solvent B; 16-20 min, 10% A and 90% C. The chromatograms were recorded at 250 nm, 260nm, and 280 nm. The peaks at retention times corresponding to 5-OHBza-RP **5** (Figure 4.5, peak at 7 minutes) or 6-OHBza-RP **6** (Figure 4.5, peak at 11 minutes) were collected, lyophilized to 100 µl and verified using UV-Vis spectra and the mass spectra. From 15 mL reactions, a total of 200µl each of 5.5 mM of 5-OHBza-RP **5** and 7.1 mM of 6-OHBza-RP **6** were obtained. To obtain the ribosyl derivatives, 100µL of the 5-OHBza-RP **5** and 6-OHBza-RP **6** were treated with alkaline phosphatase for 4 hours at 37°C which yielded 100% conversion to 5-OHBza-R **8** and 6-OHBza-R **9**, respectively.

4.2.11 Quantitation of products formed in the reaction

A standard curve for 5-OMeBza **2** was prepared in the range of 50 picomoles to 500 picomoles. 100µl of each dilution was run on reverse phase HPLC method-I and area under curve (AUC) with UV absorbance at 280nm was calculated. For fluorescence detection, only 10µl of each dilution was chromatogram was recorded with excitation at 280 nm and emission at 312 nm. The standard curves were plotted against AUC obtained for each dilution against picomoles of 5-OMeBza **2**. Typically, 100µL of reaction samples were analyzed with HPLC method-I and AUC of the peak corresponding to 5-OMeBza **4** was interpolated from the standard curves.

A standard curve for 5-OMeBza-RP was obtained as per the following method: A 1mL reaction with 500µM of 5-OMeBza **2**, 1mM NMN, and 10µM *Mt*CobT was set up. An 200µL aliquot of the reaction mix was immediately acid quenched as zero-minute sample. The zero-minute sample was diluted ten times and analyzed on HPLC method-II with increasing injection volumes and analyzed with fluorescence with excitation at 250nm and emission at 312 nm. The dilutions were also analyzed with excitation at 280nm and emission at 312 nm. The values obtained for AUC of the 5-OMeBza **4** peak was plotted against picomoles at 250 nm and 280

nm excitation wavelengths. Further, the remaining reaction was incubated for 3 hours at 25°C and then analyzed on HPLC similar to zero-minute sample. The AUC for residual 5-OMeBza **2** and the product 5-OMeBza-RP were calculated. The concentration of residual 5-OMeBza **2** was calculated using its standard curve. For each run, the concentration of the product 5-OMeBza-RP was calculated as $[5OMeBza]_{initial}$ - $[5OMeBza]_{3hours.}$ A standard curve was plotted with AUC of 5-OMeBza-RP peak against picomoles of 5-OMeBza-RP formed (concentration x volume of each injection). Since 5-OMeBza-R **9** and 5-OMeBza-RP have similar absorbance spectra, the standard curve obtained with 5-OMeBza-RP was used to interpolate the concentration of 5-OMeBza-R **9** formed in the assays. Since, 6-OMeBza-R **10** and 6-OMeBza-RP have absorbance spectra and fluorescence properties similar to base 5-OMeBza **2**, the standard curve for 5-OMeBza **2** was used to calculate their concentrations.

4.3 Results

4.3.1 Two major domain architectures of BzaC methyltransferase

To initiate the characterization of BzaC, we conducted a bioinformatics search to understand how it compares to previously studied methyltransferases. A BLASTp (Basic alignment search tool-protein) for *M. thermoacetcia* BzaC (*Mt*BzaC) and *E. limosum* BzaC (*El*BzaC) against the non-redundant UniProtKB/ SwissProt database shows that *Mt*BzaC and *El*BzaC share 18%-27% identity with previously studied methyltransferases, 3-hydroxy-5-methyl-1-naphthoate O-methyltransferase (AziB2), acetylserotonin O-methyltransferase (ASMT), 4-amino-4-de-(dimethylamino)-anhydrotetracycline-N,N-dimethyl-methyltransferase (OxyT) (S. Altschul et al. 1990; Ding et al. 2010; Botros et al. 2013; W. Zhang et al. 2008).

To further obtain insights about the biochemical nature of BzaC homologs, we examined the domain architecture of BzaC using the Conserved Domain (CD) search(Marchler-Bauer and Bryant 2004). The *Mt*BzaC and *El*BzaC sequences contain a dimerization domain belonging to Pfam16864 at the N-terminus, which is followed by a methyltransferase domain, belonging to the superfamily cl17173 (Figure 4.1B). The dimerization domain of Pfam16864, called dimerization2, is also found in other methyltransferase subfamilies (Finn et al. 2014; K. Zhang et al. 2017; Botros et al. 2013). The SAM-dependent class-I methyltransferases can methylate a wide range of substrates including DNA, protein, and small molecules such as catechols, ubiquinone, and flavones. All members of this class employ S-adenosylmethionine (SAM **3**) as a methyl-donor and yield S-adenosylhomocysteine (SAH **4**) as the by-product (Petrossian and Clarke 2009; Schubert, Blumenthal, and Cheng 2003; Struck et al. 2012). A sequence alignment of BzaC show the

presence of conserved GXGXG consensus which is characteristic of SAM binding sites (Kozbial and Mushegian 2005) (Figure 4.2A).

Additionally, the *El*BzaC homolog contains a 145-amino acid domain of unknown function called DUF2284 belonging to Pfam10050 at the C-terminal (Figure 4.1B). A recent *in silico* survey for occurrence and distribution of ~3400 domains of unknown functions reported that DUF2284 is present in several bacterial and archaeal species (Goodacre, Gerloff, and Uetz 2014). We also found that along with being associated with BzaCs, DUF2284 is also present as separate protein encoded within other genomic contexts. A sequence alignment of the DUF2284 domain of BzaC sequences shows two conserved cysteine rich consensus of CX_3CX_7C and $CX_2CX_2CX_5C$, which resemble the consensus found for Fe-S cluster binding motifs (Figure 4.2A, marked with black asterisk) (Broderick et al. 2014; Johnson et al. 2005; Drennan, Bridwell-Rabb, and Grell 2018). Overall, the analysis of the protein sequences shows that there exist two types of BzaC enzymes distinguished by the presence or absence of the DUF2284 domain (Figure 4.2A). We constructed a phylogeny tree with dimerization and methyltransferase domains of BzaC homologs which show sequences are taxonomically conserved (Figure 4.2B). The overall distribution of the BzaC homologs shows that the ones that lack DUF2284 are scattered among the clades of BzaC sequences that possess DUF2284. This suggests that the dimerization and methyltransferase domains are conserved across BzaC homologs from various taxonomies and the presence of the DUF2284 is independent of these two domains. We therefore predict that the role of DUF2284 is likely beyond the methyltransferase reaction of BzaC enzymes.

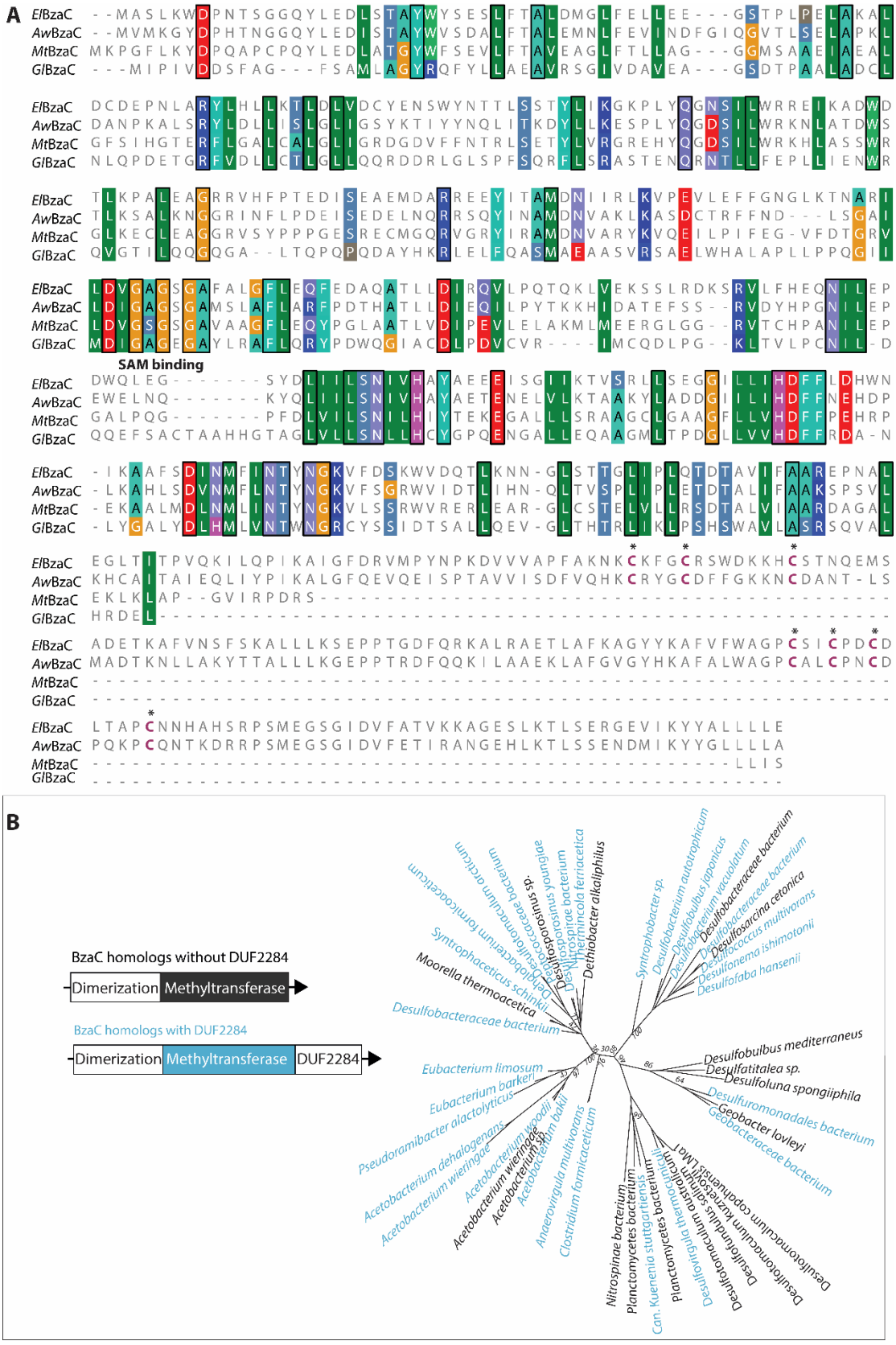


Figure 4.2 Bioinformatics analysis of BzaC sequences. A) A multiple sequence alignment with BzaC homologs from the *bza* operon of cobamide producing anaerobes *Eubacterium limosum* (*El*BzaC), *Acetobacterium woodii* (*Aw*BzaC), *Moorella thermoacetica* (*Mt*BzaC), and *Geobacter lovelyii* (*Gl*BzaC). The conserved GXGXG site indicates presence of SAM binding site typical of Class-I methyltransferases. The *El*BzaC and *Aw*BzaC, show an extra C-terminal domain of unknown function (DUF2284) which has

two conserved cysteine repeats, which resemble iron-sulfur cluster binding sites. **B)** A phylogeny tree constructed from sequences corresponding to the dimerization and methyltransferase domains of BzaC homologs. The presence of DUF2284 in the full-length sequence of each homolog is indicated in the blue color, and the sequences that lack DUF2284 are shown in black font. The tree indicates that methyltransferase domain of the BzaC homolog is tightly conserved across homologs irrespective of the presence or absence of the DUF2284.

4.3.2 Biochemical analysis of the dimerization domain and methyltransferase domain of MtBzaC

We initiated the characterization of BzaC homologs with *Mt*BzaC, which lacks the additional DUF2284. We cloned and overexpressed *Mt*BzaC and a dimerization domain deletion construct *Mt*BzaCΔDiD which purify at expected masses of 41.9 kDa and 30.8 kDa respectively (Figure 4.3A). To validate the function of the dimerization domain of the *Mt*BzaC, we conducted size exclusion chromatography with the purified enzymes. The *Mt*BzaC showed a peak corresponding to the dimer size and some higher oligomers of the protein (Figure 4.3B, solid trace). The *Mt*BzaCΔDiD showed a major peak corresponding to the expected size of its monomer and fewer peaks for aggregates (Figure 4.3B, dashed trace). This confirms that the dimerization domain of the *Mt*BzaC is involved in oligomerization of the enzyme under *in vitro* conditions.

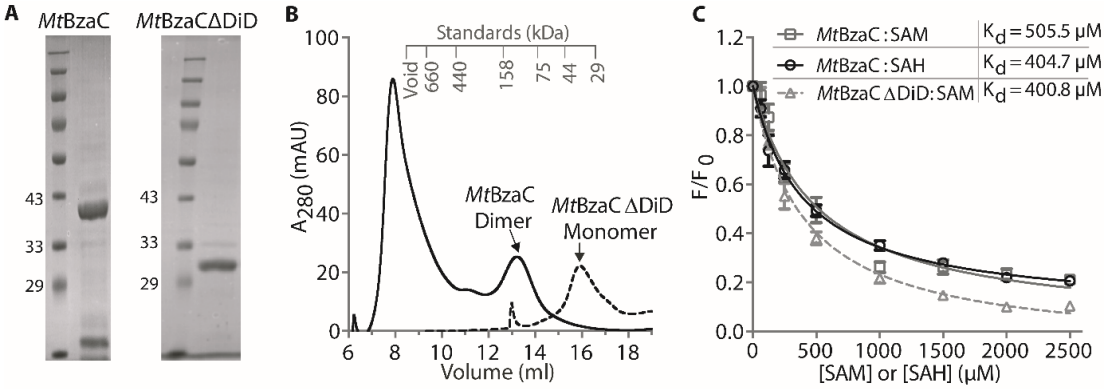


Figure 4.3 Biochemical characterization of *Mt*BzaC. A) SDS-PAGE for purified *Mt*BzaC and the dimerization deleted mutant *Mt*BzaCΔDiD show proteins eluted at expected masses of 41.9 kDa and 30.8 kDa respectively. **B)** The size exclusion chromatography validates the role of dimerization domain into *Mt*BzaC dimerization and oligomerization. The full length *Mt*BzaC construct shows a peak corresponding to the size of the dimer as well as oligomers of higher masses (solid line). The *Mt*BzaCΔDiD, however, predominantly elute at mass corresponding to the monomer (dashed line). **C)** The intrinsic fluorescence of *Mt*BzaC decreases upon binding with the SAM, thus indicating that the predicted Class-I methyltransferase domain is indeed functional. The decay in fluorescence with SAM and SAH return dissociation constant (K_d) values of 505.5 μM and 404.7 μM. The *Mt*BzaCΔDiD construct also binds to SAM at a K_d 400.8 μM which indicates that the deletion of dimerization domain does not hinder the SAM binding ability of the methyltransferase domain.

Next, we tested the ability of the predicted methyltransferase domain of *Mt*BzaC to bind the cofactor SAM. To do so, we probed the quenching in the intrinsic fluorescence of the protein upon ligand binding (Warner and Copley 2007). *Mt*BzaC contains 4 tryptophan and 11 tyrosine residues and shows intrinsic fluorescence with excitation maxima at 280 nm and

emission maxima at 328 nm. When SAM was added to the enzyme in increasing concentrations, we observed a steady decrease in the protein fluorescence which saturated at a concentration ~1.5 mM SAM. A plot of normalized fluorescence intensity (F/F_o) at 328 nm against concentration of SAM was fit to the equation-1 (see method 4.2.6) to obtain the dissociation constant (K_d) of SAM for *Mt*BzaC as 505.5 µM (Figure 4.3C, square). S-adenosylhomocysteine (SAH), which is a known inhibitor for Class-I methyltransferases also shows a similar trend with a K_d value of 404.7 µM suggesting that SAH may be an inhibitor for *Mt*BzaC (Figure 4.3C, circle) (J. Zhang and Zheng 2016) . The *Mt*BzaCΔDiD construct yielded a K_d value of 400.8µM with SAM, indicating that deletion of the dimerization domain does not impair the affinity of the methyltransferase domain towards the co-factor (Figure 4.3C, triangle).

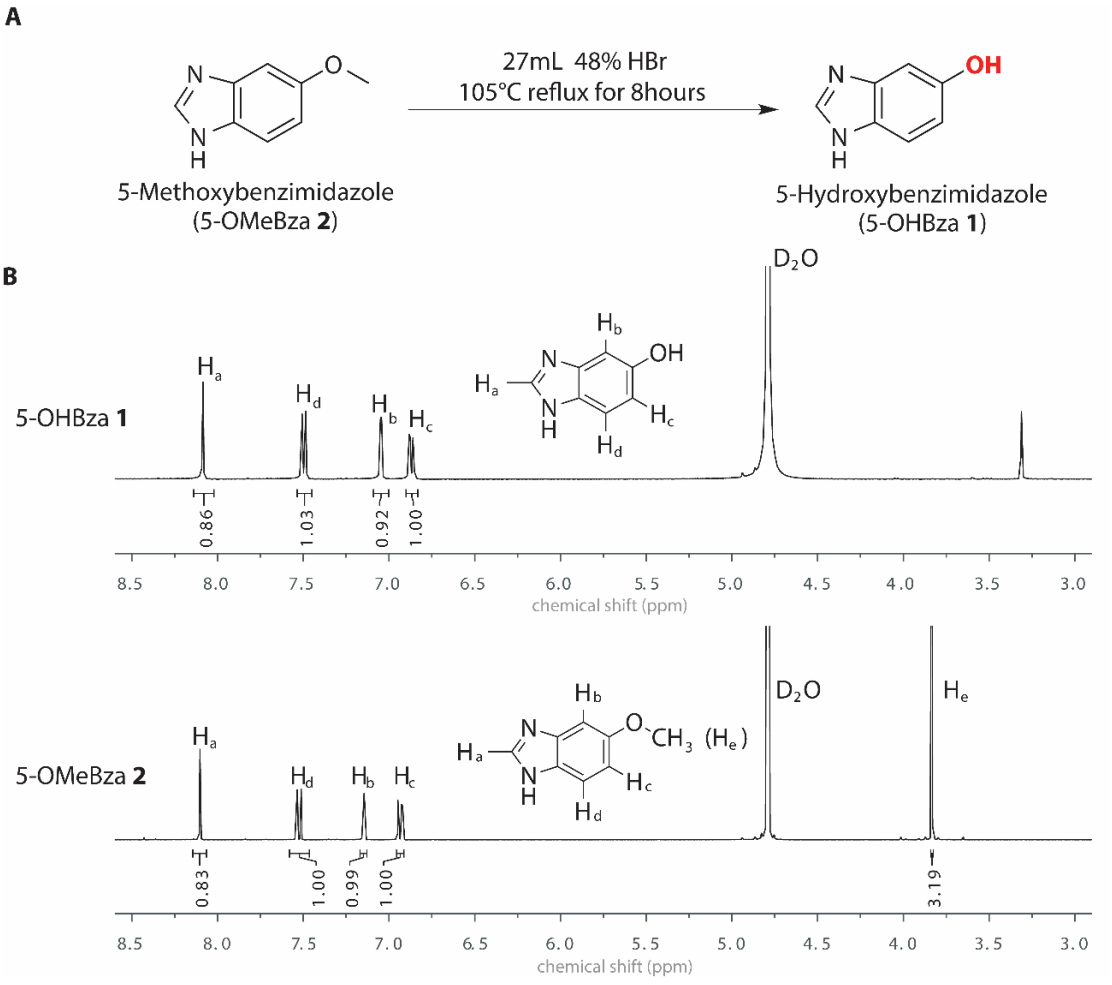


Figure 4.4 Chemical synthesis of the substrate 5-OHBza. A) The synthesis scheme for demethylation of 5-OMeBza **2** using hydrobromic acid (HBr) to produce 5-OHBza **1** (adapted from previous literature (Renz et al. 1993; Crofts et al. 2013)). **B)** The ^{1}H-NMR for the synthesized 5-OHBza **1** (top) and standard 5-OMeBza **2** (bottom). The peak corresponding to the methoxy group (H_e) is absence in NMR for 5-OHBza which confirms the purity of the 5-OHBza **1** produced.

4.3.3 In vitro reconstitution of SAM: 5-hydroxybenzimidazole methyltransferase activity of *Mt*BzaC

The *Mt*BzaC was predicted to catalyze methylation of 5-OHBza **1** using SAM **3** as the methyl donor. We chemically synthesized 5-OHBza **1** from commercially available 5-OMeBza **2** as described in the methods (Figure 4.4A). Since the starting material for synthesis, 5-OMeBza is also the final expected product in the enzymatic reaction, we confirmed the complete removal of 5-OMeBza **2** from the batch of 5-OHBza synthesized through ^{1}H-NMR (Figure 4.4B).

The gene product of *MtbzaC* has been previously demonstrated to show [5-OMeBza]Cba formation when expressed in *E. coli* under aerobic conditions as well (Chapter 1, Figure 1.5B)(Hazra et al. 2015). Thus, we attempted to reconstitute the *in vitro* activity of *Mt*BzaC under aerobic conditions with the substrates 5-OHBza **1** and SAM **3** (Figure 4.5A). The end-point reaction with *Mt*BzaC for 72 hours at 25°C showed a new peak at 19.8 minutes on the HPLC (data not shown) and 16.5 minutes on LCMS, with retention time and mass identical to that of a standard of 5-OMeBza **2**, the expected product (Figure 4.5B, 4.5C). The UV-Vis (Figure 4.5D) and the NMR spectra (Figure 4.5E) of the product confirms its identity as 5-OMeBza **2**, thus verifying that *Mt*BzaC is a functional methyltransferase. Moreover, the reconstitution with *Mt*BzaCΔDiD does not show product which indicates that the dimerized or oligomerized form of *Mt*BzaC is essential for the methyltransferase activity (data not shown). Notably, the quantity of 5-OMeBza **2** formed by 50 µM *Mt*BzaC from 500 µM 5-OHBza **1** and 1000 µM of SAM **3** was only 13 -17 µM (3.1 to 4.4 % conversion). To examine the reasons underlying the low turnover, we reconstituted the reaction under several biochemical conditions. We find that *Mt*BzaC does not require metal ion for activity, and the turnover remains comparable even with excess EDTA (Figure 4.6A). Addition of $FeCl_3$ caused enzyme to precipitate which results in very poor turnover. Further, the presence of reducing agent dithiothreitol (DTT) and the crowding agent BSA did not enhance the enzyme activity (Figure 4.6B, 4.6C). Moreover, desalting and reconstitution of *Mt*BzaC inside an anaerobic glove bag improved the yield by only 1.34-folds which implies that *Mt*BzaC may not be affected by oxygen (Figure 4.6D).

Interestingly, a linear increase in the enzyme concentration resulted in a linear increase in the concentration of product formed (Figure 4.6E). This implies that *Mt*BzaC itself may be the limiting factor in the reaction because of irreversible inhibition by a reaction by-product, or absence of other factors required in the reaction. The HPLC traces of purified enzyme and commercially bought SAM **3** shows degradation products, SAH **4**, and 5-methylthioadenosine (MTA) which are known inhibitors of class-I methyltransferases (Figure 4.5B)(J. Zhang and Zheng 2016; Siegrist et al. 2015). Thus, we co-incubated *Mt*BzaC with SAM synthetase (*Ec*MetK) to produce SAM **3** *in situ*, and we observe a 1.44 ± 0.10 folds increase in 5-OMeBza

yield (Figure 4.6F). Next, we introduced the enzyme methylthioadenosine nucleosidase (*Ec*MTAN) in the reaction in order to hydrolyse the potential inhibitors MTA and SAH **4** (Siegrist et al. 2015; Cornell et al. 1996). Co-incubation with *Ec*MTAN increased the *Mt*BzaC activity by 5.78 ± 0.49 folds with commercial SAM **3** and 4.75 +0.69 folds with the *in situ* SAM **3** synthesis (Figure 4.6F). We also observe an increase in the optimum reaction temperature from 25°C to 37°C (Figure 4.6G). Thus, we conclude that SAH **3** and MTA which formed due to spontaneous degradation of SAM **3** inhibit the methyltransferase activity of *Mt*BzaC.

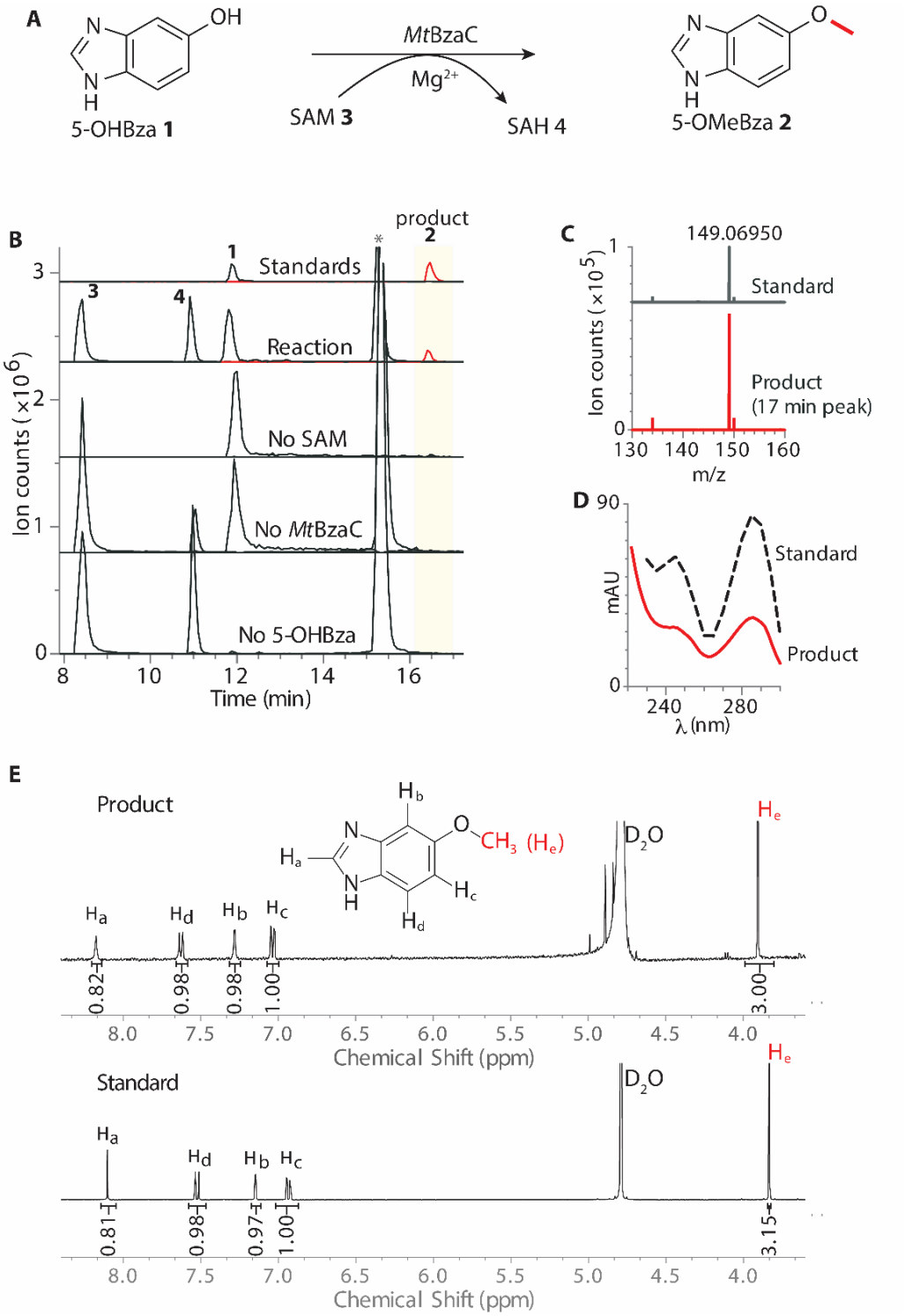


Figure 4.5 ***Mt*BzaC is an O-methyltransferase. A)** The *Mt*BzaC activity was reconstituted under in vitro conditions with 5-OHBza **1** as substrate, SAM **3** as the methyl donor. Most O-methyltransferases are known to require a divalent metal ion such as Mg^{2+}. 5-OMeBza **2** and SAH **3** the expected products of the reaction. **B)** Extracted ion chromatogram for reconstitution of *Mt*BzaC with 5-OHBza **1** and SAM **3** shows a product peak at 16.5 min (red trace in reaction) which matches the retention time of the 5-OMeBza **2** standard. **C)** The mass spectrum of the product peak is identical to the mass spectrum of the 5-OMeBza **2. D)** The product peak shows UV-Vis spectrum characteristic of 5-OMeBza **2. E)** The identity of the product was confirmed as 5-OMeBza **2** through ^{1}H-NMR. The experiment confirms that *Mt*BzaC is a functional O-methyltransferase that transfers a methyl group from SAM to 5-OHBza **1** and results in 5-OMeBza **2** as the product. The peak marked with asterisk (*) corresponds to 5-methylthioadenosine (MTA), which formed due to spontaneous degradation of SAM.

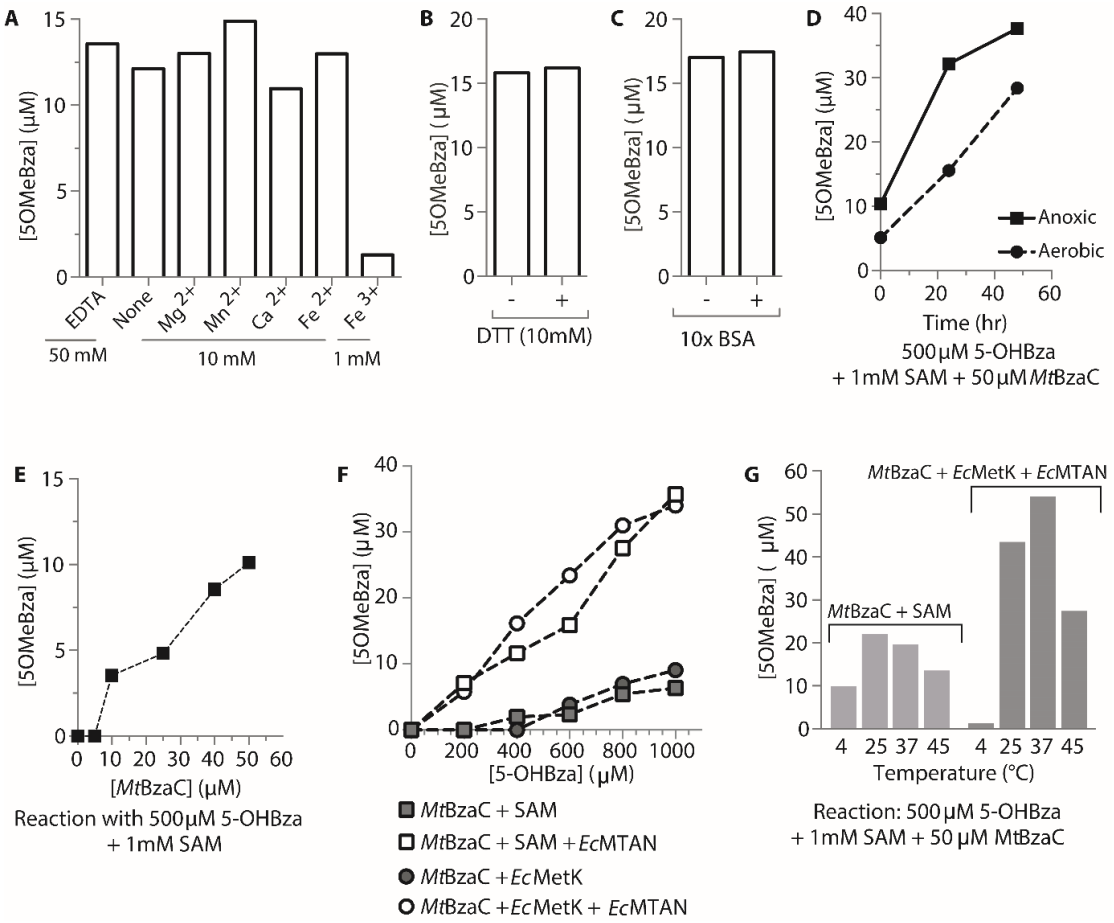


Figure 4.6 Optimization of MtBzaC activity. A) The 5-OMeBza produced from *Mt*BzaC reconstitution in presence of various metal ions shows that *Mt*BzaC does not require a metal ion for activity. **B)** Addition of dithiothreitol (DTT) does not enhance the reaction turnover. **C)** Crowding agent bovine serum albumin (BSA) does not increase the *Mt*BzaC activity. **D)** A comparison between the *Mt*BzaC activity reconstituted under aerobic and anoxic conditions over 48 hours shows slightly better turnover under anoxic conditions. **E)** The reaction turnover linearly increases with increase in enzyme concentration indicating near to single turnover conditions by *Mt*BzaC under the tested conditions. **F)** The *Mt*BzaC activity is slightly improved in presence in situ SAM synthesis by *E.coli* SAM synthetase (*Ec*MetK). Addition of *E.coli* 5-methyladenosyl nucleosidase (*Ec*MTAN) relieves the enzyme inhibition imposed by SAH, the byproduct of the reaction, and MTA, which accumulates in reaction due to SAM degradation. **G)** The *Mt*BzaC reaction with commercial SAM shows maximum activity at 25°C, which can be accredited to higher rate of SAM degradation at higher tempteratures. The *Mt*BzaC activity improved at 37°C when coupled with *Ec*MetK and *Ec*MTAN.

4.3.4 Investigating the repertoire of possible substrates for BzaC

In *Moorella thermoacetica*, the *bza* operon encodes for the *bzaA-bzaB-cobT-bzaC* genes. The order of reactions catalyzed that follows the 5-OHBza **1** biosynthesis of BzaAB enzymes can proceed in three ways:

i) *Mt*BzaC can methylate 5-OHBza **1** to make 5-OMeBza **2** which is then phosphoribosylated by *Mt*CobT and subsequently attached to the corrin ring to produce a cobamide

ii) *Mt*CobT, next gene in the pathway can activates the 5-OHBza **1** form 5-OHBza-RP **5** which can be methylated by *Mt*BzaC to form 5-OMeBza-ribotide (5-OMeBza-RP)

iii) *Mt*CobT activates 5-OHBza **3** to make 5-OHBza-RP **5** which then undergoes cobamide assembly to [5-OHBza]Cba which is lastly methylated by *Mt*BzaC to yield [5-OMeBza]Cba.

The net low activity of *Mt*BzaC with 5-OHBza as the substrate indicates that 5-OHBza is possibly not the physiological substrate of the enzyme. Since, the *Mt*BzaC sequence does not show any known cobamide-binding site, we tested if MtBzaC preferentially methylated the CobT product. Notably, the *Mt*CobT enzyme show regiospecific activation of 5-OHBza to yield 5-OHBza-RP **5** as the sole product. However, when reacted with 5-OMeBza, *Mt*CobT produces both 5-OMeBza-RP and 6-OMeBza-RP isomers.

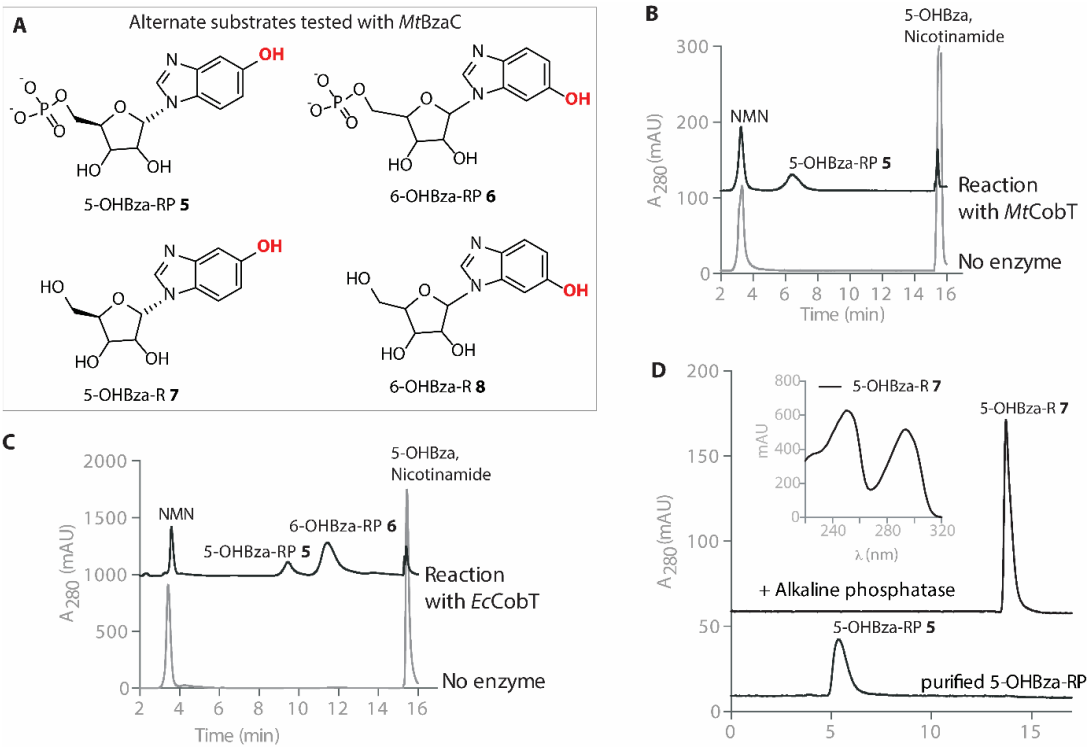


Figure 4.7 CobT products as the plausible substrates for *Mt*BzaC. A) The phosphoribosylated derivatives (or ribotides) of 5-OHBza namely, 5-OHBza-RP **5** and 6-OHBza-RP **6** are isomeric products formed by a CobT enzyme upon reaction with 5-OHBza **1.** The ribosylated derivatives namely, 5-OHBza-R **7** and 6-OHBza-R **8** can result from spontaneous dephosphoribosylation of the ribotide. **B)** HPLC chromatogram for purification of 5-OHBza-RP from a reaction with *Mt*CobT and 5-OHBza **1**. The peak at 7 mins corresponding to 5-OHBza-RP **5** was collected. **C)** HPLC chromatogram for purification of 6-OHBza-RP from a reaction with *Ec*CobT and 5-OHBza **1**. The peak at 11 mins corresponding to 6-OHBza-RP **6** was collected. **D)** The 5-OHBza-R, the riboside derivative was obtained by treating the purified ribotide 5-OHBza-RP **5** with alkaline phosphatase. The inset shows the UV-Vis spectra of 5-OHBza-R substrate. The same method was used to obtain 6-OHBza-R **8** from 6-OHBza-RP **6**.

To test if 5-OHBza **1** activation by *Mt*CobT occurs prior to methylation by *Mt*BzaC, we tested *Mt*BzaC activity with 5-OHBza-RP **6** and 6-OHBza-RP **7** (Figure 4.7A). We enzymatically synthesized 5-OHBza-RP **6** and 6-OHBza-RP **7** using 5-OHBza **1** and *Mt*CobT and *Ec*CobT respectively (as described in methods, Figure 4.7B, 4.7C). The methyltransferase activity of *Mt*BzaC was individually tested with 5-OHBza-RP **6** and 6-OHBza-RP **7** as substrates and SAM **3** as methyl donor under reaction conditions optimized with 5-OHBza. The LCMS analysis of these reactions showed that *Mt*BzaC does not methylate either of the two phosphoribosylated substrates. Instead, the LCMS chromatogram of the end-point reaction of *Mt*BzaC with 5-OHBza-RP **6** for 48 hours repeatedly (n=3) showed new peaks corresponding to mass for 5-hydroxybenzimidazole riboside (5-OHBza-R **7**) and 5-methoxybenzimidazole riboside (5-OMeBza-R **9**) at 11.2 min and 14.6 min respectively. Under *in vitro* conditions, the riboside derivatives 5-OHBza-R **7** and 5-OMeBza-R **9** can result from spontaneous dephosphorylation of 5-OHBza-RP **5** and 5-OMeBza-RP respectively. Thus, the formation of 5-OMeBza-R **9** could occur either if 5-OHBza-R **7** was methylated by *Mt*BzaC or if 5-OMeBza-RP formed in the reaction was spontaneously degrading to 5-OMeBza-R **9**. To investigate this observation, we enzymatically synthesized 5-OHBza-R **7** (as described in methods, Figure 4.7D) and 6-hydroxybenzimidazole riboside (6-OHBza-R **8**) and tested the methyltransferase activity of *Mt*BzaC with both the riboside isomers (Figure 4.8A). Surprisingly, we found that *Mt*BzaC methylates the two isomers to form 5-OMeBza-R **7** and 6-OMeBza-R **8** under both aerobic and anoxic conditions (Figure 4.8B). The *Mt*BzaC activity was 5.05 times higher with 5-OHBza-R **7** and 1.22 times with 6-OHBza-R **9** than the activity with 5-OHBza **1**, which indicates that inherent regiospecificity in methylation by BzaC (Figure 4.8B). LCMS analysis of individual reactions confirmed the product peaks as 5-OMeBza-R 17 and 6-OMeBza-R 19 (Figure 4.8C, 4.8D). In summary, of all the substrates presented to *Mt*BzaC under *in vitro* conditions, 5-OHBza-R **7** is preferentially methylated (Figure 4.8B). Thus, we conclude that 5-OHBza-R **7** is the likely physiological substrate of *Mt*BzaC and that activation of the lower ligand precedes its methylation in benzimidazole biosynthesis pathway.

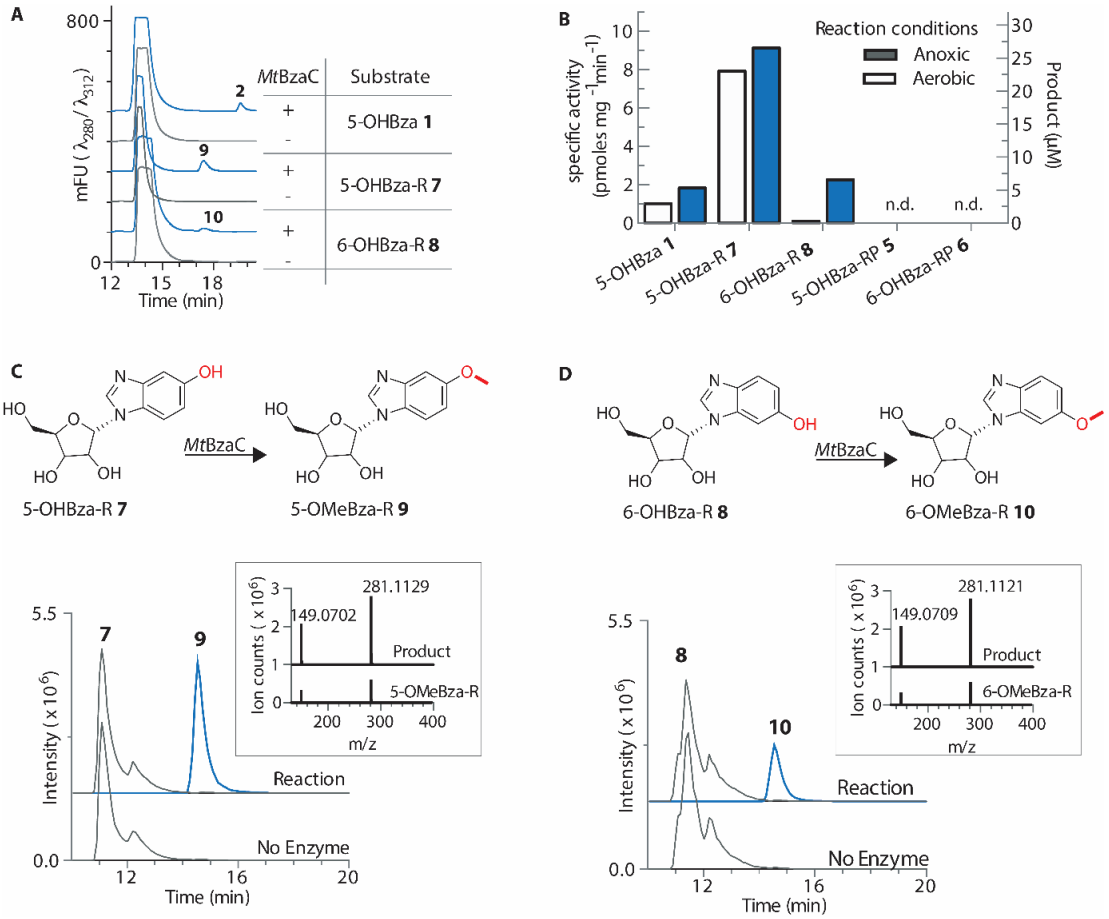


Figure 4.8 *Mt*BzaC preferentially methylates activated 5-OHBza. **A)** HPLC- fluorescence chromatogram for *Mt*BzaC reaction with 5-OHBza **1** and its riboside derivatives 5-OHBza-R **7** and 6-OHBza-R **8**. All three substrates are methylated by *Mt*BzaC to form 5-OMeBza **2**, 5-OMeBza-R **9** and 6-OMeBza-R **10,** respectively**. B)** Specific activity of *Mt*BzaC with 5-OHBza **1**, 5-OHBza-R **7**, 6-OHBza-R **10**, 5-OHBza-RP **5** and 6-OHBza-RP **6** as the substrates. Relatively, 5-OHBza-R **7** is the most preferred substrate whereas no detectable amount of riboside phosphate products i.e. 5-OMeBza-RP **5** and 6-OMeBza-RP **6** were found. **C)** LCMS EIC of the MtBzaC and 5-OHBza-R **7** reaction. The EIC corresponding to 5-OMeBza-R **9** (blue trace) shows a peak at 14.5 min in the reaction. The mass spectrum of the product peak confirms its identity as 5-OMeBza-R **9** with a fragment of m/z 149.0702 corresponding to 5-OMeBza **2** base as shown in the inset. **D)** LCMS EIC of the *Mt*BzaC and 6-OHBza-R **6** reaction. The EIC corresponding to 6-OMeBza-R **10** (black trace) shows a peak at 14.5 min in the reaction. The mass spectrum of the product peak confirms its identity as 6-OMeBza-R **10** with a fragment of m/z 149.0702 corresponding to 5-OMeBza **2** base as shown in the inset.

4.4 Discussion

Previously, the heterologous expression of *bza* operon from *Eubacterium limosum* and *Moorella thermoacetica* informed that the *bzaC*, *bzaD*, and *bzaE* genes products catalyze three sequential and chemically distinct methylations (Hazra et al. 2015). The *bza* operon was predicted as follows: gene products of *bza*A-*bza*B (or its single gene homolog *bza*F) catalyzes rearrangement of 5-OHBza **1** which undergoes three methylations by *bzaC*, *bzaD*, and *bzaE* genes products yielding 5-OMeBza **2,** 5OMe-6MeBza as intermediates and DMB as the final product. Since, CobT homologs from several bacteria were shown to preferentially activate DMB, it was proposed that bza operon CobT would similarly activate DMB for cobalamin assembly(Hazra et al. 2015). The BzaF enzyme involved in biosynthesis of 5-OHBza was

characterized *in vitro* prior to our explorations with BzaC-BzaD-BzaE methyltransferases (Hazra et al. 2015; Mehta et al. 2015; Gagnon et al. 2018a). In this study, we combine our observations from characterization of next two gene products CobT and BzaC using the *M. thermoacetica bza* operon, which contains the *bza*A-*bza*B-*cob*T-*bza*C genes and produces [5-OMeBza]Cba as the native cobamide. We find that contrary to the previously proposed pathway, the activation of the 5-OHBza by CobT precedes its methylation by BzaC, that is, 5-OHBza is first phosphoribosylated, then dephosphorylated to form 5-OHBza-R which is finally methylated to yield 5-OMeBza-R (Figure 4.9).

In this chapter, we validated the function of predicted dimerization and methyltransferase domains of *Mt*BzaC. We show, for the first time, that BzaC is a functional methyltransferase which utilizes SAM to methylate 5-OHBza to form 5-OMeBza (Figure 4.5). The SAM-dependent methylation observed *in vitro* agrees with previous studies in *M. thermoacetica* using ^{14}C-labelled methionine(Wurm, Renz, and Heckmann 1980b). However, the low net activity of the reaction under varying optimization methods indicated that some key element in the reaction was missing (Figure 4.6). Finally, when we explored the possibility of CobT acting prior to BzaC in the pathway. We tested *Mt*BzaC with 5-OHBza-derived ribotide and riboside as alternate substrates and find that *Mt*BzaC shows highest activity with 5-OHBza-R, the riboside derivative (Figure 4.8B, 4.8C). Hence, we characterize *Mt*BzaC as a SAM-dependent 5-OHBza-R methyltransferase (HBIR-OMT).

How do the observations fit the overall cobamide assembly pathway?

The last steps in cobalamin biosynthesis pathway, as studied in a facultative anaerobe *Salmonella enterica*, involves the CobS and CobC enzymes. Once DMB is activated by CobT to form DMB-RP, the pathway can proceed in either of the two ways- a) the phosphatase enzyme CobC removes the 5'-phosphate from DMB-RP to make the DMB-riboside intermediate which is then attached to the corrin ring by CobS(Maggio-Hall and Escalante-Semerena 1999) (Figure 4.9A, route 1), b) the DMB-RP is attached to the activated corrin ring by the enzyme CobS, and then the phosphatase CobC can remove the 5-phosphate group to yield cobalamin (Maggio-Hall, Claas, and Escalante-Semerena 2004) (Figure 4.9A, route 2).

If the observed the substrate preference for *Mt*BzaC holds true under physiological conditions, the activation of 5-OHBza by should be followed by a dephosphoribosylation to produce 5-OHBza-R for methylation by *Mt*BzaC. The dephosphoribosylation can either be catalyzed by CobC encoded in the *Moorella thermoacetica* genome or can also occur non-enzymatically (Mattes and Escalante-Semerena 2017). To enable either of the reactions, the *Moorella thermoacetica* genome encodes for CobS and CobC genes within the locus encoding several corrin ring biosynthesis and activation genes (Figure 4.9B). To date, none of the homologs of cobamide assembly pathway enzymes have been studied from obligate anaerobes.

Hence, understanding substrate promiscuity and regioselectivity of CobS and CobC enzymes from organisms that contain a *bza* operon, thus make more than one type of benzimidazole derivatives, will lead to a better understanding of the last steps in the pathway.

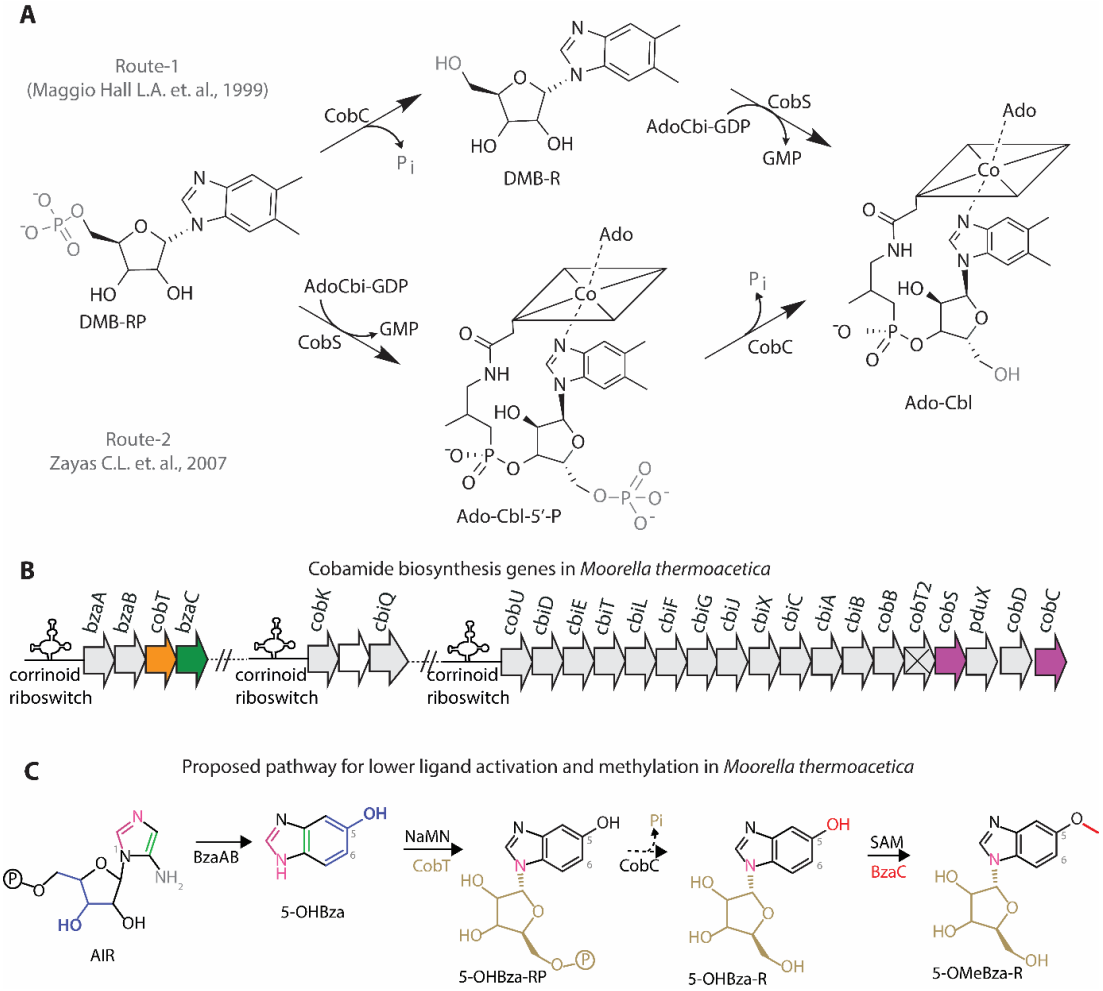


Figure 4.9 The activation of lower ligand precedes methylation in the *bza* operon pathway. A) Fate of activated lower ligand DMB-riboside phosphate (DMB-RP) in B_{12} biosynthesis can follow either of the two routes. Route-1: The DMB-RP undergoes a dephosphorylation catalyzed by phosphatase enzyme CobC, and the DMB-riboside (DMB-R) is attached to the corrin ring by the enzyme CobS to produce cobalamin. Route-2: The DMB-RP can be attached directly to the corrin ring by CobS to form cobalamin-5'-phosphate and then CobC acts last in the pathway to produce cobalamin. **B)** The *Moorella thermoacetica* shows a beautiful arrangement of all the genes involved in biosynthesis of [5-OMeBza]Cba, its native cobamide. The pathway is regulated under riboswitches classified as corrinoid riboswitches (also known as cobalamin riboswitches, RFam1482). The *cobT*, *bzaC* genes in the *bza* operon and the *cobS* and *cobC* genes in the *cob/cbi* operon are shown in color. The *cobT2* gene in the *cob* operon shows a deletion of the sequence corresponding to the active site (hence marked with 'x'), and thus *Moorella thermoacetica* has only one functional gene for CobT which is localised in the bza operon (gene in yellow). **C)** Based on our observed activities for *Moorella thermoacetica* bza operon CobT and BzaC, we propose the following revision in the pathway: The BzaAB genes rearrange the purine biosynthesis intermediate, 5-aminoimidazole ribotide (AIR) to make 5-OHBza which is then activated by CobT in a regiospecific manner that make 5-OHBza-RP. Next, we propose that the 5'-phosphate is removed by the phosphatase enzyme CobC to yield 5-OHBza-R. Lastly, BzaC methylates 5-OHBza-R and produces 5-OMeBza-R which can then be attached to the activated corrin ring by the enzyme CobS to produce [5-OMeBza]Cba as the final cobamide. Thus, we propose that the activation of first pathway intermediate, 5-OHBza, by CobT precedes the methylations in the bza operon pathway.

Moreover, in chapter-2, we find that CobT homologs encoded in the *bza* operon of cobamide producers *Moorella thermoacetica, Eubacterium limosum, Eubacterium barkeri* activate 5-OHBza in a regiospecific manner, thus producing 5-OHBza-RP as the only product. However, the *bza* operon CobT does not show regiospecificity with 5-OMeBza, and thus produces both 5-OMeBza-RP and 6-OMeBza-RP. If the activation of lower ligand by *Mt*CobT were to occur after the methylation of 5-OHBza to form 5-OMeBza, the pathway would likely yield both [5-OMeBza]Cba and [6-OMeBza]Cba isomers as *Mt*CobT lacks regiospecificity with 5-OMeBza. Instead, based on our findings *Mt*CobT phosphoribosylates 5-OHBza, the first reaction intermediate, to produce a single phosphoribosylated isomer 5-OHBza-RP which is later dephosphorylated and methylated by BzaC. Thus, the pathway would result in [5-OMeBza]Cba exclusively. Our findings are corroborated by previous literature that shows *M. thermacetica* naturally produces only [5-OMeBza]Cba (Wurm, Weyhenmeyer, and Renz 1975; Wurm, Renz, and Heckmann 1980b). Furthermore, the heterologous expression of the *M. thermacetica bza* operon in *E. coli* showed the formation of [5-OMeBza]Cba exclusively as compared to an *E. coli* control with added 5-OMeBza where both [5-OMeBza]Cba and [6-OMeBza]Cba isomers were formed (Hazra et al. 2015). Finally, we observe that despite 6-OHBza-R also being an activated form of 5-OHBza, *Mt*BzaC shows poor activity with this isomer as compared to 5-OHBza-R. All of this taken together strongly puts forth that the activation of the first intermediate in the pathway precedes the methylations in the *bza* operon pathway (Figure 4.9C).

Finally, our studies shed light on a long-standing unexplained observation in anaerobic DMB biosynthesis (Lamm, Hecknmann, and Renz 1982; Paul Renz et al. 1993). Extensive labelling studies conducted to understand how DMB is synthesized in anaerobic organisms had shown that the two N's in DMB are derived from glycine and glutamine (Lamm et al. 1980; Vogt and Renz 1988). Intriguingly, even though DMB is a symmetric molecule, the nucleotide loop is attached specifically through the nitrogen atom derived from glutamine (Lamm, Hecknmann, and Renz 1982; Vogt and Renz 1988). This puzzling result was justified by a hypothesis that an asymmetric intermediate from benzimidazole biosynthesis must undergo regiospecific phosphoribosylation and the resulting intermediate would be a substrate for subsequent methylations (Schulze, Vogler, and Renz 1998; Paul Renz et al. 1993; Hazra et al. 2015). Since the *Moorella thermoacetica bza* operon produces two asymmetric benzimidazoles, the activities of the *Mt*CobT and *Mt*BzaC provided significant insights into this long-standing puzzle. The co-occurrence of the *cobT* gene with the *bzaA-bzaB* or *bzaF* genes and our finding that *Mt*CobT produces one regiospecific product which is then methylated by *Mt*BzaC provides evidence for the convergence of biosynthesis of the benzimidazole lower ligand with its final incorporation into the cobamide.

References:

Altschul, SF, W Gish, W Miller, EW Myers, and DJ Lipman. 1990. “Basic Local Alignment Search Tool.” J. Mol. Biol 215: 403–10.

Benson, Dennis A., Ilene Karsch-Mizrachi, David J. Lipman, James Ostell, and David L. Wheeler. 2005. “GenBank.” Nucleic Acids Research 33 (DATABASE ISS.): 34–38. https://doi.org/10.1093/nar/gki063.

Botros, Hany Goubran, Pierre Legrand, Cecile Pagan, Vincent Bondet, Patrick Weber, Mariem Ben-Abdallah, Nathalie Lemière, et al. 2013. “Crystal Structure and Functional Mapping of Human ASMT, the Last Enzyme of the Melatonin Synthesis Pathway.” Journal of Pineal Research 54 (1): 46–57. https://doi.org/10.1111/j.1600-079X.2012.01020.x.

Broderick, Joan B., Benjamin R. Duffus, Kaitlin S. Duschene, and Eric M. Shepard. 2014. “Radical S-Adenosylmethionine Enzymes.” Chemical Reviews 114 (8): 4229–4317. https://doi.org/10.1021/cr4004709.

Cheng, Jiujun, Christopher D Sibley, Rahat Zaheer, and Turlough M Finan. 2007. “A Sinorhizobium Meliloti MinE Mutant Has an Altered Morphology and Exhibits Defects in Legume Symbiosis.” Microbiology 153: 375–87. https://doi.org/10.1099/mic.0.2006/001362-0.

Cornell, Kenneth A, William E Swarts, Ronald D Barry, and Michael K Riscoe. 1996. “Characterization of Recombinant Eschericha Coli 5’-Methyladenosine/S-Adenosylhomocysteine Nucleosidase Analysis of Enzymatic Activity and Substrate Specificity.” Biochemical and Biophysical Research Communications 732: 724–32.

Crofts, Terence S., Erica C. Seth, Amrita B. Hazra, and Michiko E. Taga. 2013. “Cobamide Structure Depends on Both Lower Ligand Availability and CobT Substrate Specificity.” Chemistry and Biology 20 (10): 1265–74. https://doi.org/10.1016/j.chembiol.2013.08.006.

Ding, Wei, Wei Deng, Mancheng Tang, Qi Zhang, Gongli Tang, Yurong Bi, and Wen Liu. 2010. “Biosynthesis of 3-Methoxy-5-Methyl Naphthoic Acid and Its Incorporation into the Antitumor Antibiotic Azinomycin B.” Molecular BioSystems 6 (6): 1071–81. https://doi.org/10.1039/b926358f.

Drennan, Catherine L., Jennifer Bridwell-Rabb, and Tsehai A.J. Grell. 2018. “A Rich Man, Poor Man Story of S-Adenosylmethionie and Cobalamin Revisited.” Annual Review of Biochemistry 87: 555–84.

Edgar, Robert C. 2004. “MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput.” Nucleic Acids Research 32 (5): 1792–97. https://doi.org/10.1093/nar/gkh340.

Ent, Fusinita Van Den, and Jan Löwe. 2006. “RF Cloning: A Restriction-Free Method for Inserting Target Genes into Plasmids.” Journal of Biochemical and Biophysical Methods 67 (1): 67–74. https://doi.org/10.1016/j.jbbm.2005.12.008.

Ernst, Orna, and Tsaffrir Zor. 2010. “Linearization of the Bradford Protein Assay.” Journal of Visualized Experiments, no. 38: e1918. https://doi.org/10.3791/1918.

Farnberger, Judith E., Nina Richter, Katharina Hiebler, Sarah Bierbaumer, Mathias Pickl, Wolfgang Skibar, Ferdinand Zepeck, and Wolfgang Kroutil. 2018. “Biocatalytic Methylation and Demethylation via a Shuttle Catalysis Concept Involving Corrinoid Proteins.” Communications Chemistry 1 (1): 1–8. https://doi.org/10.1038/s42004-018-0083-2.

Finn, Robert D, Alex Bateman, Jody Clements, Penelope Coggill, Y Eberhardt, Sean R Eddy, Andreas Heger, et al. 2014. “Pfam : The Protein Families Database” 42 (November 2013): 222–30. https://doi.org/10.1093/nar/gkt1223.

Gagnon, Derek M., Troy A. Stich, Angad P. Mehta, Sameh H. Abdelwahed, Tadhg P. Begley, and R. David Britt. 2018. “An Aminoimidazole Radical Intermediate in the Anaerobic Biosynthesis of the 5,6-Dimethylbenzimidazole Ligand to Vitamin B12.” Research-article.

Journal of the American Chemical Society 140 (40): 12798–807. https://doi.org/10.1021/jacs.8b05686.

Goodacre, Norman F, Dietlind L Gerloff, and Peter Uetz. 2014. “Protein Domains of Unknown Function Are Essential in Bacteria.” MBio 5 (1): e00744-13. https://doi.org/10.1128/mBio.00744-13.Editor.

Hall, T.A. 1999. “BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT.” Nucleic Acids Symposium 41: 95–98. https://doi.org/citeulike-article-id:691774.

Hazra, Amrita B., Andrew W. Han, Angad P. Mehta, Kenny C. Mok, Vadim Osadchiy, Tadhg P. Begley, and Michiko E. Taga. 2015. “Anaerobic Biosynthesis of the Lower Ligand of Vitamin B12.” Proceedings of the National Academy of Sciences of the United States of America 112 (34): 10792–97. https://doi.org/10.1073/pnas.1509132112.

Johnson, Deborah C., Dennis R. Dean, Archer D. Smith, and Michael K. Johnson. 2005. “Structure, Function, and Formation of Biological Iron-Sulfur Clusters.” Annual Review of Biochemistry 74 (1): 247–81. https://doi.org/10.1146/annurev.biochem.74.082803.133518.

Kelly, Lawrence A. L.A., Stefans Mezulis, Christopher M. C.M. Yates, M.N. Mark N Wass, Michael J. E. Strenberg, M. Stefans, Christopher M. C.M. Yates, M.N. Mark N Wass, and Michael J E Sternberg. 2015. “The Phyre2 Web Portal for Protein Modelling, Prediction and Analysis.” Nature Protocols 10 (6): 845–58. https://doi.org/10.1038/nprot.2015-053.

Kozbial, Piotr Z., and Arcady R. Mushegian. 2005. “Natural History of S-Adenosylmethionine-Binding Proteins.” BMC Structural Biology 5 (19). https://doi.org/10.1186/1472-6807-5-19.

Kräutler, Bernhard, Hans-Peter E. Kohler, and Erhard Stupperich. 1988. “5′-Methylbenzimidazolyl-cobamides Are the Corrinoids from Some Sulfate-reducing and Sulfur-metabolizing Bacteria.” European Journal of Biochemistry 176 (2): 461–69. https://doi.org/10.1111/j.1432-1033.1988.tb14303.x.

Lamm, Lydia, Gernot Hecknmann, and Paul Renz. 1982. “Biosynthesis of Vitamin B12 in Anaerobic Bacteria. Mode of Incorporation of Glycine into the 5,6-Dimethylbenzimidazole Moiety in Eubacterium Limosum.” European Journal of Biochemistry 122 (3): 569–71. https://doi.org/10.1111/j.1432-1033.1982.tb06476.x.

Lamm, Lydia, Joachim A. Hörig, Paul Renz, and Gernot Heckmann. 1980. “Biosynthesis of Vitamin B12. Experiments with the Anaerobe Eubacterium Limosum and Some Labelled Substrates.” European Journal of Biochemistry 109 (1): 115–18. https://doi.org/10.1111/j.1432-1033.1980.tb04775.x.

Letunic, Ivica, and Peer Bork. 2016. “Interactive Tree of Life (ITOL) v3: An Online Tool for the Display and Annotation of Phylogenetic and Other Trees.” Nucleic Acids Research 44 (W1): W242–45. https://doi.org/10.1093/nar/gkw290.

Maggio-Hall, Lori A., Kathy R. Claas, and Jorge C. Escalante-Semerena. 2004. “The Last Step in Coenzyme B12 Synthesis Is Localized to the Cell Membrane in Bacteria and Archaea.” Microbiology 150 (5): 1385–95. https://doi.org/10.1099/mic.0.26952-0.

Maggio-Hall, Lori A., and Jorge C. Escalante-Semerena. 1999. “In Vitro Synthesis of the Nucleotide Loop of Cobalamin by Salmonella Typhimurium Enzymes.” Proceedings of the National Academy of Sciences 96 (21): 11798–803. https://doi.org/10.1073/pnas.96.21.11798.

Marchler-Bauer, Aron, and Stephen H. Bryant. 2004. “CD-Search: Protein Domain Annotations on the Fly.” Nucleic Acids Research 32 (Web server issue): W327–31. https://doi.org/10.1093/nar/gkh454.

Martin, Jennifer L, Fiona M Mcmillan, and The S-adenosylmethionine-dependent. 2002. “SAM (Dependent) I AM : The S-Adenosylmethionine-Dependent Methyltransferase Fold.” Current Opinion in Structural Biology 12: 783–93.

Mattes, Theodoric A., and Jorge C. Escalante-Semerena. 2017. “Salmonella Enterica Synthesizes 5,6-Dimethylbenzimidazolyl- (DMB)-α-Riboside. Why Some Firmicutes Do Not Require the Canonical DMB Activation System to Synthesize Adenosylcobalamin.” Molecular Microbiology 103 (2): 269–81.

Mehta, Angad P., Sameh H. Abdelwahed, Michael K. Fenwick, Amrita B. Hazra, Michiko E. Taga, Yang Zhang, Steven E. Ealick, and Tadhg P. Begley. 2015. “Anaerobic 5-Hydroxybenzimidazole Formation from Aminoimidazole Ribotide: An Unanticipated Intersection of Thiamin and Vitamin B12 Biosynthesis.” Journal of the American Chemical Society 137 (33): 10444–47. https://doi.org/10.1021/jacs.5b03576.

Miller, Mark A., Wayne Pfeiffer, and Terri Schwartz. 2010. “Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees.” 2010 Gateway Computing Environments Workshop, GCE 2010.

Petrossian, Tanya, and Steven Clarke. 2009. “Bioinformatic Identification of Novel Methyltransferases.” Epigenomics 1 (1): 163–75. https://doi.org/10.2217/epi.09.3.

Renz, P. 1993. “Biosynthesis of Vitamin B12 in Anaerobic Bacteria” 1121: 1117–21.

Renz, Paul, Birgit Endres, Birgit Kurz, and Jörg Marquart. 1993. “Biosynthesis of Vitamin B12 in Anaerobic Bacteria. Transformation of 5-Hydroxybenzimidazole and 5-Hydroxy-6-Methylbenzimidazole into 5,6-Dimethylbenzimidazole in Eubacterium Limosum.” European Journal of Biochemistry 217 (3): 1117–21. https://doi.org/10.1111/j.1432-1033.1993.tb18344.x.

Sandanaraj, Britto S., Mullapudi Mohan Reddy, Pavankumar Janardhan Bhandari, Sugam Kumar, and Vinod K. Aswal. 2018. “Rational Design of Supramolecular Dynamic Protein Assemblies by Using a Micelle-Assisted Activity-Based Protein-Labeling Technology.” Chem. Eur. J. 24 (60): 16085–96. https://doi.org/10.1002/chem.201802824.

Schubert, Heidi L, Robert M Blumenthal, and Xiaodong Cheng. 2003. “Many Paths to Methyltransfer : A Chronicle of Convergence.” Trends 28 (6): 329–35. https://doi.org/10.1016/S0968-0004(03)00090-2.

Schulze, Bettina, Bernhard Vogler, and Paul Renz. 1998. “Biosynthesis of Vitamin B12 in Anaerobic Bacteria - Experiments with Eubacterium Limosum on the Transformation of 5-Hydroxy-6-Methyl- Benzimidazole, Its Nucleoside, Its Cobamide, and of 5-Hydroxybenzimidazolylcobamide in Vitamin B12.” European Journal of Biochemistry 254 (3): 620–25. https://doi.org/10.1046/j.1432-1327.1998.2540620.x.

Shelton, Amanda N., Erica C. Seth, Kenny C. Mok, Andrew W. Han, Samantha N. Jackson, David R. Haft, and Michiko E. Taga. 2018. “Uneven Distribution of Cobamide Biosynthesis and Dependence in Bacteria Predicted by Comparative Genomics.” ISME Journal 13 (3): 789–804. https://doi.org/10.1038/s41396-018-0304-9.

Siegrist, Jutta, Simon Aschwanden, Silja Mordhorst, Linda Thöny-Meyer, Michael Richter, and Jennifer N. Andexer. 2015. “Regiocomplementary O-Methylation of Catechols by Using Three-Enzyme Cascades.” ChemBioChem 16 (18): 2576–79. https://doi.org/10.1002/cbic.201500410.

Siegrist, Jutta, Julia Netzer, Silja Mordhorst, Lukas Karst, Stefan Gerhardt, Oliver Einsle, Michael Richter, and Jennifer N. Andexer. 2017. “Functional and Structural Characterisation of a Bacterial O-Methyltransferase and Factors Determining Regioselectivity.” FEBS Letters 591 (2): 312–21. https://doi.org/10.1002/1873-3468.12530.

Sokolovskaya, Olga M., Amanda N. Shelton, and Michiko E. Taga. 2020. “Sharing Vitamins: Cobamides Unveil Microbial Interactions.” Science 369 (6499). https://doi.org/10.1126/science.aba0165.

Stamatakis, Alexandros. 2014. “RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies.” Bioinformatics 30 (9): 1312–13. https://doi.org/10.1093/bioinformatics/btu033.

Struck, Anna Winona, Mark L. Thompson, Lu Shin Wong, and Jason Micklefield. 2012. “S-Adenosyl-Methionine-Dependent Methyltransferases: Highly Versatile Enzymes in Biocatalysis, Biosynthesis and Other Biotechnological Applications.” ChemBioChem 13 (18): 2642–55. https://doi.org/10.1002/cbic.201200556.

Stupperich, Erhard. 1988. “Diversity of Corrinoids in Acetogenic Bacteria.” European Journal of … 464: 459–64. http://onlinelibrary.wiley.com/doi/10.1111/j.1432-1033.1988.tb13910.x/full.

Towns, John, Tim Cockerill, Maytal Dahan, Ian Foster, Kelly Gaither, Andrew Grimshaw, Victor Hazlewood, et al. 2014. “XSEDE: Accelerating Scientific Discovery.” Computing in Science and Engineering 16 (October): 62–74. https://doi.org/10.1109/MCSE.2014.80.

Vogt, Joachim R.A., and Paul Renz. 1988. “Biosynthesis of Vitamin B12 in Anaerobic Bacteria. Experiments with Eubacterium Limosum on the Origin of the Amide Groups of the Corrin Ring and of N-3 of the 5,6-dimethylbenzimidazole Part.” European Journal of Biochemistry 171 (3): 655–59. https://doi.org/10.1111/j.1432-1033.1988.tb13836.x.

Warner, Joseph R., and Shelley D. Copley. 2007. “Pre-Steady-State Kinetic Studies of the Reductive Dehalogenation Catalyzed by Tetrachlorohydroquinone Dehalogenase.” Biochemistry 46 (45): 13211–22. https://doi.org/10.1021/bi701069n.

Wurm, Rainer, Paul Renz, and Gernot Heckmann. 1980. “The Biosynthesis of the Vitamin B12 Analog 5-Methoxybenzimidazolylcobamide in Clostridium Thermoaceticum.” FEMS Microbiology Letters 7: 11–13.

Wurm, Rainer, Roland Weyhenmeyer, and Paul Renz. 1975. “On the Biosynthesis of 5-Methoxybenzimidazole: Precursor-Function of 5-Hydroxybenzimidazole, Benzimidazole and Riboflavin.” European Journal of Biochemistry 56 (2): 427–32. https://doi.org/10.1111/j.1432-1033.1975.tb02249.x.

Yang, Jianyi, and Yang Zhang. 2016. “Protein Structure and Function Prediction Using I-TASSER.” Current Protocol Bioinformatics 52 (5): 5.8.1-5.8.15. https://doi.org/10.1002/0471250953.bi0508s52.Protein.

Zhang, Jing, and Yujun George Zheng. 2016. “SAM/SAH Analogs as Versatile Tools for SAM-Dependent Methyltransferases.” ACS Chemical Biology 11 (3): 583–97. https://doi.org/10.1021/acschembio.5b00812.

Zhang, Kai, Zhiqiang Ruan, Jia Li, Chao Bian, Xinxin You, Steven L. Coon, and Qiong Shi. 2017. “A Comparative Genomic and Transcriptomic Survey Provides Novel Insights into N-Acetylserotonin Methyltransferase (ASMT) in Fish.” Molecules 22 (10): 1–15. https://doi.org/10.3390/molecules22101653.

Zhang, Wenjun, Kenji Watanabe, Xiaolu Cai, Michael E. Jung, Yi Tang, and Jixun Zhan. 2008. “Identifying the Minimal Enzymes Required for Anhydrotetracycline Biosynthesis.” Journal of the American Chemical Society 130 (19): 6068–69. https://doi.org/10.1021/ja800951e.

Chapter 5

Investigations to decipher the function of DUF2284 and the last two methyltransferases in *bza* operon

5.1 Introduction

The *Eubacterium limosum bza* operon contains *bzaA-bzaB-cobT-bzaC-bzaD-bzaE* genes and produce wherein the *bzaC, bzaD, bzaE* gene products are predicted to catalyze three distinct methylation reactions leading to DMB as the final product. There exist two major domain variants of the BzaC protein, and we define BzaC as a Class-I methyltransferase found in obligate anaerobes localized downstream of a corrinoid/B_{12} riboswitch or in the neighborhood of ThiC paralogs *bzaA-bzaB* or *bzaF*. The *Moorella thermoacetica* (*Mt*BzaC), as described in chapter 4, contains a dimerization domain and Class-I SAM-dependent methyltransferase domain. The *bzaC* gene product from *Eubacterium limosum* (*El*BzaC) also contain an N-terminal dimerization and a central Class-I SAM-dependent methyltransferase domain which share high sequence similarity and phylogeny with *Mt*BzaC (Chapter 4, figure 4.2). Interestingly, we observe that *El*BzaC contains an additional C-terminal domain which is annotated as domain of unknown function 2284 (DUF2284).

The DUF2284 is classified as the protein family (pf10050) and is a yet uncharacterized protein domain with little to no literature evidence. As per the InterPro database, the pf10050 has >2200 bacterial and >750 archaeal predicted to carry the DUF2284 domain. From an overview of patterns of predicted domain architectures for proteins in the pf10050 family, the DUF2284 domain abundantly exists as single domain protein, however several sequences show DUF2284 fused with a methyltransferase, adenylate cyclase, or racemase domains. Certain interesting entries show two sequentially placed DUF2284 in the predicted protein domain architecture. The DUF2284 is one of over thousands of orphan protein domains whose function, biochemical properties, and significance in life processes remain unknown. As of 2013, the DUFs were calculated to account for 20% of all known protein sequences indicating that we might not yet know the properties of several gene products involved in essential pathways. The recent boom in data from genomic and metagenomic sequencing has accelerated research into such novel protein domains. Previous efforts into de-orphaning various domains of unknown function have require systematic and combinatorial approaches with methods in bioinformatics, genetics, protein chemistry, and structural biology. In the past decade, streamlined genomic enzymology workflows built by Enzyme Function Initiative allows large-scale analyses of protein sequences and gene neighborhoods, and thus, these methods help with help to decipher function novel proteins such as DUFs as well find new functions within known protein families.

Further, the last two methyltransferases in the pathway, the gene product of *bzaD* and *bzaE* are predicted to contain similar domain architecture with an N-terminal B_{12}-binding domain and a C-terminal radical-SAM/ Fe-S cluster binding sites. The presence of the B_{12} as a cofactor in the enzymes involved in its own biosynthesis is intriguing. Recently, a new family of radical-SAM enzymes caught attention for utilizing B_{12}, SAM, and Fe-S cluster. Several of these enzymes are now beginning to be characterized through in vitro reconstitution and structural analyses. The mechanisms for each of these enzymes are predicted to be distinct from each other and are currently being investigated by various research groups. Enzymes of B_{12}/radical-SAM family are shown to catalyze a range of chemical transformations including formation of C-C, C-P, C-N, and C-S bonds which result in intramolecular rearrangements, thioether bond formation as well as radical and non-radical mediated methylations (Zhong et al. 2021; Knox et al. 2021; 2022; Wang et al. 2018; J. Y. Kim et al. 2008; Werner et al. 2011; Sinner et al. 2022). To date, structures exist for only four B_{12}-RSAM enzymes, namely OxsB, TsrM, TokK, and Mmp10, of which the latter three are methyltransferases (Bridwell-Rabb et al. 2017; Knox et al. 2021; 2022; Fyfe et al. 2022).

Some notable and historically important examples of B_{12}/radical-SAM include Fom3, Genk, OxsB, and TsrM. Fom3 is a B_{12} and radical-SAM-dependent methyltransferase was shown to use B_{12} as the direct methyl donor and SAM and Fe-S cluster were proposed to assist as radical donors in the methylation mechanism(K. D. Allen and Wang 2014). GenK is another enzyme with similar binding sites that uses two molecules of SAM, one that act as radical donor and second donates the methyl group. In its proposed mechanism, B_{12} as an intermediate carrier of the methyl group from SAM to the substrate(H. J. Kim et al. 2013). OxsB, is a B_{12}/radical-SAM enzyme which is not a methyltransferase, and catalyzed a intramolecular rearrangement of a five-membered ribose ring to a four-membered ring. TsrM- a relatively more explored B_{12}/radical-SAM enzyme which is proposed to catalyze methylation through a non-radical (called polar methylation) and requires two molecules of SAM per turnover(Pierre et al. 2012). All of the characterized B_{12}/radical-SAM proteins contain a [4Fe-4S] cluster within the SAM-binding TIM barrel or modified TIM barrel like architecture (Bridwell-Rabb, Li, and Drennan 2022). The [4Fe-4S] cluster involve the consensus CX_3CX_2C sequence, except for OxsB which contains a CX_4CX_2C sequence. The hallmarks of SAM binding site include a glycine rich sequence such as GGE and GXIXGXXE motifs. Notably, the B_{12} binding domains in the B_{12}/radical-SAM enzymes show interesting features in the region near to the nucleotide loop. The B_{12} binding sites in the methionine synthase (MetH) and other methyltransferases that do not use a radical-SAM mechanism typically involves a coordination between the cobalt ion and an active site histidine residue resulting in a 'base-off, His-on' conformation of the B_{12} cofactor. However, the B_{12}/radical-SAM enzymes lack the corresponding His residue in the structural

loop near the DMB base, and result in a 'base-off, His-off' conformation. In these enzymes, the cobalt center is protected from water via physical obstruction by residues such as arginine, tyrosine, and leucine in TsrM, TokK, and Mmp10, respectively.

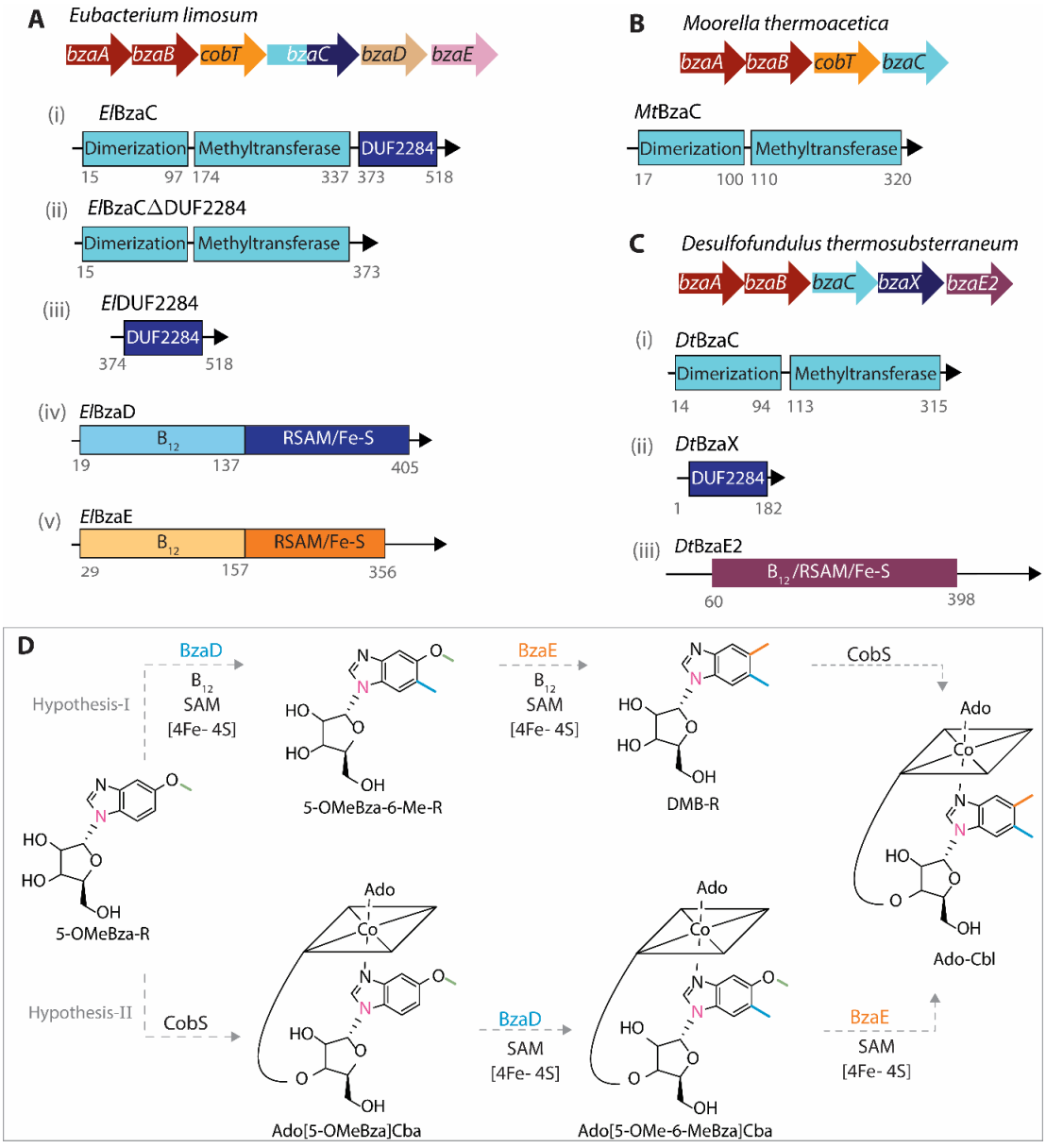


Figure 5.1 Predicted domain architecture of BzaC homologs reveal a domain of unknown function called DUF2284. A) The *bzaC* gene product from the *bza* operon of *Eubacterium limosum* (*El*BzaC) codes for a protein with 518 amino acid residues which contain a dimerization domain, a methyltransferase domain, and DUF2284. The DUF2284 in spans 173 amino acid residues in *El*BzaC and is predicted to be a metal binding domain. The constructs of *El*BzaC were used in the study- (i) *El*BzaC full length protein, (ii) *El*BzaCΔDUF2284 wherein the sequence corresponding to the DUF2284 was deleted, (iii) *El*DUF2284 which contains only the sequence corresponding to the DUF2284 of *El*BzaC protein. (iv, v) The BzaD and BzaE are the last two enzymes in the *bza* operon pathway that finally produce DMB as the lower ligand. The predicted domain architecture includes an N-terminal B_{12} binding domain and a C-terminal radical-SAM/[4Fe-4S] binding domain which resembles the architecture of the newly discovered class of B_{12}-RSAM proteins. **B)** The gene product of *bzaC* from *Moorella thermoacetica* (*Mt*BzaC) codes for a shorter variant of BzaC which contains a dimerization and a methyltransferase domain. **C)** We discovered a new variant of *bza* operon from the organism *Desulfofundulus thermosubterraneum* which contains *bzaA-bzaB-bzaC-bzaX-bzaE2* genes. (i) The *D. thermosubterraneum bzaC* gene product (*Dt*BzaC) is a *Mt*BzaC-like

homolog that is composed of a dimerization and a methyltransferase domain. (ii) The *D. thermosubterraneum bzaC* gene is followed by a gene that encodes a 182 amino acid residue long DUF2284, and we name the gene as *bzaX*. (iii)The *bzaE2* is a homolog of the *bzaE* gene which contains also a B_{12}/ radical-SAM/[4Fe-4S] binding domain and share sequence similarity with BzaE. **D)** The requirement of B_{12}-binding domain in enzymes involved in its biosynthesis is seemingly a chicken-and-egg problem. We propose two hypotheses for the function for the predicted B_{12}-binding sites in the enzymes BzaD and BzaE (Hazra et al. 2015). According to our first hypothesis, the B_{12} -binding sites indeed serve as cofactor-binding sites and hence, the BzaD and BzaE conduct methylations at C5 and C6 using B_{12}-RSAM chemistry and the riboside derivatives 5-OMeBza-R and 5-OMe-6MeBza-R serve as pathway intermediates to yield DMB-R which finally undergoes cobamide assembly. According to our second hypothesis, the cobamide assembly precedes the last methylations, and the predicted B_{12}-binding sites in BzaD and BzaE accommodate the cobamide substrates Ado[5-OMeBza]Cba and Ado[5-OMe-6-MeBza]Cba, respectively. The methylations are then expected to be catalyzed by SAM in association with [4Fe-4S] clusters.

The cobalamin-dependent radical-SAM methyltransferases that are characterized and mechanistically explored employ distinct cofactor utilization and reaction mechanisms(Bridwell-Rabb, Li, and Drennan 2022)(Knox et al. 2021). Hence, the reconstitution and mechanistic analysis for BzaD and/or BzaE based solely on predicted binding sites is challenging. Moreover, the reaction to replace a methoxy group by a methyl group proposed to be catalyzed by BzaE has no precedence in biochemistry literature. Moreover, since BzaD and BzaE are involved in biosynthesis of B_{12}, it would not be surprising if the B_{12}-binding domain turns out to be a substrate binding domain for cobamides containing DMB precursors.

Based on our findings with CobT and BzaC enzymes, we propose two hypotheses for the function of BzaD and BzaE enzymes. In the first scenario, BzaD and BzaE act as B_{12}-RSAM enzymes which successively methylate 5-methoxybenzimdazole-riboside (5-OMeBza-R), the product of BzaC, to produce DMB-riboside, which is then utilized by CobS to form cobalamin. In the second scenario, the B_{12}-binding domain in BzaD and BzaE are substrate-binding sites for 5-methoxybenzimidazolylcobamide and 5-methoxy-6-methylbenzimidazolylcobamide, respectively (Figure 5.1). The two hypotheses need to be tested, by testing the reconstitution with either cofactor forms of B_{12} or substrate cobamides, and based our previous experience with enzyme characterizations, only one of the conditions is likely to yield a successful turnover. Hence, the underlying work for reconstitution of BzaD and BzaE will also help with reviewing the current knowledge about the anaerobic biosynthesis of vitamin B_{12}.

In this chapter, we address the characterization of the novel domain of unknown function, DUF2284 present in the *bza* operon, and initiate the primary experimental analyses for BzaD and BzaE enzymes. To do so, we first purified the *El*BzaC and *Eubacterium barkeri* BzaC *(Eb*BzaC) homolog and study biochemical properties imparted by the C-terminal DUF2284. We first probed the function of dimerization domain and SAM-binding domain using methods that we employed to characterize *Mt*BzaC (Chapter 4, section 4.3.2). We then

undertake a bioinformatic approach to study the conservation patterns in DUF2284 sequences, which indicate presence of cysteine repeats that resemble iron-sulfur cluster binding site. We used this information to validate the presence of iron-sulfur cluster in the purified protein. We then employ comparative genomics-based approach to study the genomic context of DUF2284 in B_{12} biosynthesis pathway, and as a result we found 26 new variants of *bza* operon which remain to be characterized. We also find two new genes within the *bza* operon of certain anaerobes, one the that encode for a single domain DUF2284 (hereafter named *bzaX*), and two a B_{12}/radical-SAM like homolog which shows sequence similarity *bzaE* (hereafter named *bzaE2*). For example, the organism *Desulfofundulus thermosubterraneum* codes for a BzaC homolog that is similar to *Mt*BzaC and contains dimerization and methyltransferase domain. The *bzaC* gene in *D. thermosubterraneum* is followed by a gene that encodes for a DUF2284 protein which we name as '*bzaX'*. Usually, organisms typically contain both *bzaD* and *bzaE* genes, and previous work has shown that *bzaE* does not function in absence of *bzaD* (Hazra et al. 2015). However, in organisms such as *D. thermosubterraneum,* we find only one gene that codes for a B_{12}/radical-SAM protein which shows slight resemblance to BzaE, and we name the gene '*bzaE2'*.

The primary bioinformatics and biochemical characterization of BzaC homologs that contain DUF2284 allow us to establish it as a as Fe-S cluster binding domain. The comparative genomics show a strong cooccurrence between DUF2284 in the *bza* operon with *bzaD* and *bzaE/ bzaE2* genes, which opens an avenue for to explore any interactions between the DUF2284, BzaD, and BzaE that might be essential for the biosynthesis of DMB by the *bza* operon pathway. We then initiated the primary characterization of BzaD and BzaE with optimizing the synthesis of required substrates, purification of proteins, and reconstitution of the Fe-S clusters within the proteins. Additionally, our comparative genomics experiment uncover a wide range of *bza* operon that exists in anaerobic bacteria. We also explore the occurrence of DUF2284 beyond the *bza* operon. In summary, the work presented here forms the foundation for understanding the function of DUF2284 and mechanistic analysis of the last two methyltransferases of the *bza* operon- which largely are the final missing pieces of information in the anaerobic biosynthesis of B_{12}.

5.2 Methods

5.2.1 Molecular cloning, overexpression, and purification

The *bzaC* genes from *Eubacterium limosum* (*ElbzaC*) and *Eubacterium barkeri* (*EbbzaC*) were amplified from pKM077 and *E. barkeri* genomic DNA, respectively. The genes were inserted into the pET28a vector using restriction free (RF) cloning as described in chapter 2. The resulting in plasmids were called pYMH001 and pYMH003, respectively. We

additionally created two constructs from the full length *ElbzaC.* First, *El*BzaCΔDUF2284- we used RF cloning to exclude the sequence corresponding DUF2284 domain from pYMH001 to give pYMH004. As a result, the *El*BzaC was truncated at Tyr373 which marks the beginning of the DUF2284. Second, the sequence corresponding to the DUF2284 domain starting from Asn374 in *El*BzaC was subcloned into pET28a that allowed overexpression of N-terminal 6xHis tagged *El*DUF2284 (pYMH007). We also created a *Mt*BzaC+DUF2284 where we inserted the DUF2284 sequence at the C-terminal of the *Mt*BzaC protein, however, this construct failed to produce a stable protein.

The *E.limosum bzaD* (*ElbzaD*) and *bzaE* (*ElbzaE*) were amplified from plasmids pKM077(Hazra et al. 2015) which contain the corresponding *bza* operons in the pTH1227 vector. The *E.barkeri bzaD* (*EbbzaD*) and *bzaE* (*EbbzaE*) were amplified from the genomic DNA of the organism.

We additionally cloned the *DtbzaC, DtbzaX, and DtbzaE2* genes from *Desulfofundulus thermosubterraneum* were amplified from the organism's genomic DNA and subsequently cloned into pET28a using RF cloning method (pYMH020, pYMH021).

The *bzaC* genes were overexpressed in *E.coli* BL21(DE3) and purified using Ni-NTA affinity chromatography as described in 4.2.3. For experiments with Fe-S clusters proteins, the plasmid of interest was co-transformed in *E. coli* BL21(DE3) cells along with pDB1282 (Seravalli et al., n.d.), a plasmid that harbors *isc* genes from *Acetobactor vinelandii* for Fe-S cluster biogenesis. Cells were grown in LB broth with 50 µg/ml kanamycin, 100 µg/ml ampicillin and to assist the Fe-S cluster biogenesis process, cells were fed with 100 µM ferrous ammonium sulphate and 200 mg/l cysteine. After an $O.D._{600}$ = 0.4 was reached, the genes of pDB1282 system were induced with 0.2% L-arabinose and the protein expression by pET28a was induced by 0.5 mM IPTG. The induced cultures were incubated at 20 °C for 16 hours shaking at 100 rpm. Cells were harvested by centrifugation and pellets were stored at −80 °C.

5.2.2 UV-Vis spectroscopy to probe the presence of Fe-S cluster

To test for the presence of Fe-S cluster using UV-Vis spectrometry, the proteins were freshly purified under anoxic conditions (95% N_2: 5% H_2; O_2 < 70ppm) using Ni-NTA affinity chromatography and desalted into 50 mM Tris-Cl pH 8.0 and 150 mM NaCl using 10 kDa cut-off desalting columns (Bio-rad laboratories). To avoid exposure of air during the spectroscopic analysis, 600 µL of buffer or protein was transferred to a Hellma cuvette (Suprasil® quartz 200-2,500 nm; 10 mm) and sealed using airtight septa and paraffin film, and when required, the cuvettes were opened only inside the anaerobic chamber in between runs.

A UV-Vis spectrum from 200 nm to 600 nm was recorded on the Shimadzu UV-2600 instrument. The desalting buffer was used for baseline correction, following the measurement

of the protein spectrum. To test for the presence of a functional [Fe-S], dithionite was titrated in micromolar concentrations using a Hamilton- GasTight 25 μL syringe and the UV-Vis spectra was recorded after each addition. The change in the peak intensity and shape at 410-420 nm were monitored to estimate the saturation of titration. In case, the protein precipitated during the titration, dithionite was not added any further.

5.2.3 Elemental analysis for detection of iron bound to the protein

The 6xHis-tagged proteins were purified under anoxic conditions through Ni-NTA affinity chromatography as described above. During elution, fractions of 0.5 mL each were collected and protein concentrations were estimated using Bradford's reagent (Bradford 1976) and A280 (Aitken and Learmonth 1996). Most concentrated fractions were pooled to get 3 mL of protein that was subsequently desalted into 50 mM Tris-Cl pH 8.0 containing 150 mM NaCl. The 3 mL was mixed with 3mL of 4% of HNO_3 to denature the protein and centrifuged at 7000 rpm for 20 mins at 4-20°C. The supernatant was collected in fresh tubes which were used for ICP-MS analysis after a 10x dilution. ICP-MS analysis was conducted on Element, XR (Thermo Fischer Scientific). The instrument was first tuned and calibrated using commercially available standard metal solutions prior to each analysis.

5.2.4 Biochemical characterization of the dimerization and methyltransferase domain

The purified *El*BzaC was subjected to gel filtration chromatography and SAM-binding assay using intrinsic fluorescence of the protein as described in section 4.2.5 and 4.2.6. Due to presence of a Fe-S cluster in the purified protein, we conducted the protocols inside an anoxic glove bag maintained at <70 ppm oxygen using 95% N_2: 5% H_2.

5.2.5 Bioinformatics analysis and comparative genomics

For analyzing the DUF2284 sequences present as a part of BzaC homologs, we used the dataset curated using extensive BLASTp searches against diverse bacterial phyla as described in section 4.2.1. We manually tested the gene neighborhood of each *bzaC* gene, which led to discovery of *bzaX* genes that encode for individual DUF2284 proteins. All sequence alignments were done using MUSCLE2.0 and prepared for visualization using BioEdit (Edgar 2004; Hall 1999).

The dataset for phylogeny analysis of DUF2284 was curated using BLASTp searches against bacterial and archaeal taxa. We selected approximately 5 random sequences from 100 results per BLASTp search. The sequences were aligned in MUSCLE2.0, converted to PHYLIP4.0 format, and a maximum likelihood tree was calculated using JTT model using RaXML-HPC2 (Stamatakis 2014) on XSEDE (Towns et al. 2014) on the CIPRES Science gateway server (http://www.phylo.org)(Miller, Pfeiffer, and Schwartz 2010). The Newick tree

was visualized and annotated on iToL (http://itol.embl.de)(Letunic and Bork 2016). We generated multiple sequence alignments for each clade observed in the tree, and we manually determined conservation patterns specially in the cysteine rich regions unique to each clade.

The dataset for BzaC phylogeny tree includes all BzaC sequences available in NCBI GenBank (as of July 2018), which emphasizes the rarity of *bza* operon genes in nature. We cross-checked our databases by searching for *bzaC*, *bzaD*, and *bzaE* genes in organisms previously shown to contain a *bzaF* or *bzaA* and *bzaB*. We manually recorded the composition of *bza* operon in each of the resulting 93 genomes. We validated the authenticity of the predicted BzaA, BzaB, BzaC, BzaD, BzaE genes using HMM models developed by Shelton et. al., (2018) (Shelton et al. 2018) using HMMer3.0 (http://hmmer.janelia.org/).

For quantitating the comparative genomics data, the dataset was arranged on the basis of presence or absence of *bzaAB*, *bzaF*, *cobT*, *bzaD*, *bzaE* and the type of *bzaC* (i.e. whether the BzaC sequence has a DUF2284 or not) in the corresponding operons. In certain cases, genomic DNA sequence contig containing the *bza* operon ended in the middle of the operon. Those entries of BzaC sequences were excluded from the quantitation but are listed in the compiled results. For certain interesting operon variants, we searched and stitched the contigs, but using the translated BLAST (tBLASTn) with sequences for BzaA, BzaB, BzaC, BzaD, BzaD, BzaE, and CobT sequences from taxonomically closet available organism. Finally, the co-occurrence of the DUF2284 with BzaAB / F, CobT and BzaD/E were tabulated as contingency tables and analyzed using Fisher's exact test (two tailed) with 95% confidence interval on GraphPad Prism 6.

5.2.6 Purification of cobamides

The traditional methods to purify cobamides from microbial cultures require cyanylation step using postassium cyanide, which posed technical challenges owing to the restricted availability in India and health hazard posed by the cyanide. We thus, used a recently developed cyanide-free method for cobamide purification as described in literature (Deery, Lawrence, and Warren 2022), with a few optimizations as follows:

a. Production of cobamides:

We used the *E.coli* ED674 strain, an engineered strain which contain all the B_{12} biosynthesis genes from *Rhodobacter capsulatus* integrated into the genome of *E.coli*. Additionally, the strain contains a T7-RNA polymerase system for overexpression of genes under IPTG inducible system. The strain does not contain genes *bluB*, or *bza* operon required for lower ligand biosynthesis, and hence produces pseudocobalamin (pB_{12}, which contains adenine as the lower ligand) as the native cobamide. We hence use guided biosynthesis based method for producing 5-hydoxybenzimidazolylcobamide [5-OHBza]Cba, 5-

methoxybenzimidazolylcobamide [5-OMeBza]Cba, and B_{12}. We also introduced the vectors containing the *bza* operon from *E. limosum* and *M. thermoacetica* as well as pET28a vectors containing *ElbzaC, EbbzaC, MtbzaC, DtbzaC, MtcobT, EbcobT, EbbzaD, EbbzaE, ElbzaD,* and *DtbzaE2* individually into the ED674 strain.

E.coli ED674 or the derived strains were grown in media containing: M9 minimal media (in 1 liter- 6 g Na_2HPO4, 3 g KH_2PO4, 1 g NH_4Cl, and 0.5 g NaCl), 1 g yeast extract, 0.5% glycerol, 0.1 mM $CaCl_2$, 2 mM $MgCl_2$, 50 mg/ l $CoCl_2$, $6H_2O$, 0.05µM $MnCl_2$. An overnight primary culture was started with ED674 in LB + 0.2 % glucose + 5mg/ l $CoCl_2$ at 28°C.

The secondary culture inoculated with 1% primary culture and was incubated at 28°C for 6 hours at 150 rpm shaking. When the $O.D._{600}$ reached 1.0, we added IPTG to a final of 0.4mM. As per requirement of the experiment, we added 25mg/L of desired lower ligand.

The culture was incubated for 24 hours at 28°C at 150 rpm. Upon successful production of cobamides, the cultures turn orange to pink colour over time. The cells were harvested by centrifugation at 3000 rpm (on swinging) or 6000 rpm (on fixed rotor) at 4°C for 10-20 mins. The pellet was stored at -20 °C or immediately use for cobamide purification.

b. Preparation of resin with immobilized BtuF of cobamides:

The strain BL21 star (DE3)(pLysS)(pET14b-*E. coli btuF*) was grown in 1L of LB broth with 100 µg/ml Ampilcillim, 34 µg/ml Chloramphenicol. The protein overexpression was induced with 400 µM IPTG at $O.D._{600}$ 1.0. The culture was incubated at 18-20 °C with shaking at 180 rpm for overnight. The cells were harvested with centrifugation at 4°C, and pellet was stored at -80°C. The Ni-NTA resin was charged with 50mM NiSO4 as per standard protocol and was then equilibrate the charges resin with 20mM Tris-Cl pH 7.5, 500mM NaCl, 5mM Imidazole buffer. The cell pellet from 1L culture was resuspended in 25 mL of 20mM Tris-Cl pH 7.5, 500mM NaCl, 5mM Imidazole buffer. The cell suspension was sonicated at 30 s on/ 30 s off cycle at 55% amplitude for 4 min of on time. The lysate was centrifuged at 18000 rpm at 4°C for 20 mins, and supernatant was loaded on the resin. Next, three washes were given using 25 mL BtuF-Resuspension buffer (20mM Tris-Cl pH 7.5, 500mM NaCl, 5mM Imidazole), 10 mL of BtuF-wash buffer (20mM Tris-Cl pH 7.5, 500mM NaCl, 60mM Imidazole), and 10 mL of BtuF- storage buffer (20mM Tris-Cl pH 7.5, 100mM NaCl, 5mM imidazole). Lastly, 10 mL of BtuF- storage buffer was added and the resuspended resin was removed from column, and then transfered to a glass bottle (20-25 mL). The immobilized BtuF can be stored at 4°C for a month, condition, that is enough buffer and the resin does not dry out during storage.

c. Purification of cobamides using immobilized BtuF

The cell pellets from 25 mL of ED674 culture were resuspend in 5 mL of low salt binding buffer (20mM Tris-Cl pH 7.5, 100mM NaCl, 5mM Imidazole). The cell suspensions were boiled at 90 °C in a water bath for 10-15 minutes and then centrifuge at 18000 rpm for 20 mins. The supernatant was collected in a fresh 15 mL tube. Next, the 0.6 mL of BtuF-resin was transferred into an empty EconoPac column and washed with Low salt binding buffer. The supernatant was loaded on the 0.6 mL of charged and equilibrated BtuF-resin, and flow-through was discarded. The column was washed with 2 mL low salt wash buffer (20mM Tris-Cl pH 7.5, 100mM NaCl, 60mM Imidazole) and then with 2 mL low salt buffer (20mM Tris-Cl pH 7.5, 100mM NaCl) to remove all imidazole. Finally, 2 mL of Urea buffer (20mM Tris-Cl pH 7.5, 100mM NaCl, 8M Urea) was added and the pink/orange fractions were collected. [Upon addition of Urea, the BtuF denatures and thus cobamides are released.] To revive the resin, the BtuF can be stripped along with Ni using EDTA chelation. The column has to be washed with plenty of water to remove traces of Urea and EDTA.

To remove urea from the eluted cobamides, Bond Elut C-18 columns were used. The C-18 resin was conditioned with 1 mL of MilliQ H_2O and then the sample containing cobamides were loaded. Next, the sample was washed with 5 mL of MilliQ H_2O. Here, the cobamides form a pink band on the resin. To elute the cobamides, 2-5 ml HPLC grade Methanol was added slowly. The methanol was removed from the samples on centrivap at 45°C at 6000 rpm. Finally, the cobamides were resuspended in MilliQ and quality check was performed on HPLC and LCMS analysis using LC method described previously (Hazra et al. 2015; Crofts et al. 2014).

5.3 Results

5.3.1 Analysis of the DUF2284 sequence

To get insights into the possible biochemical properties about DUF2284, we closely inspected the sequences of DUF2284 domains present as BzaC protein sequences as well as those encoded by the *bzaX* in the *bza* operon. A sequence alignment with DUF2284 sequences from *bza* operon show a high conservation among the homologous domain sequences with two cysteine rich repeats with a $CX_3C_{7\text{-}8}C$ and $CX_2CX_2CX_4C$ consensus sequences situated about 58-60 amino acid residues apart (Figure 5.2). In literature, proteins that bind to iron-sulfur (Fe-S) clusters are known to contain such cysteine rich repeats, wherein the sulfur in the cysteine side chain adhere to the inorganic iron. For example, radical-SAM enzymes such as lysine aminomutase, pyruvate formate lyase, and biotin synthase contain a CX_3CX_2C consensus and binds to a [4Fe-4S] clusters. Certain [2Fe-2S] cluster binding proteins such as some ferredoxins contain $CX_2CX_5CX_{35}C$ conserved sequence. Thus, based on several evidences in literature for the role of conserved cysteines in binding to Fe-S clusters, we predict that DUF2284 is a candidate Fe-S cluster binding domain. Furthermore, upon Ni-NTA affinity chromatography,

the *El*BzaC protein always elute as dark colored fractions which are typical of iron binding proteins. However, the *Mt*BzaC homolog which lacks the DUF2284 domain purified as a colorless protein. Hence, we further ventured into biochemically testing if the DUF2284 domain is an Fe-S cluster binding domain.

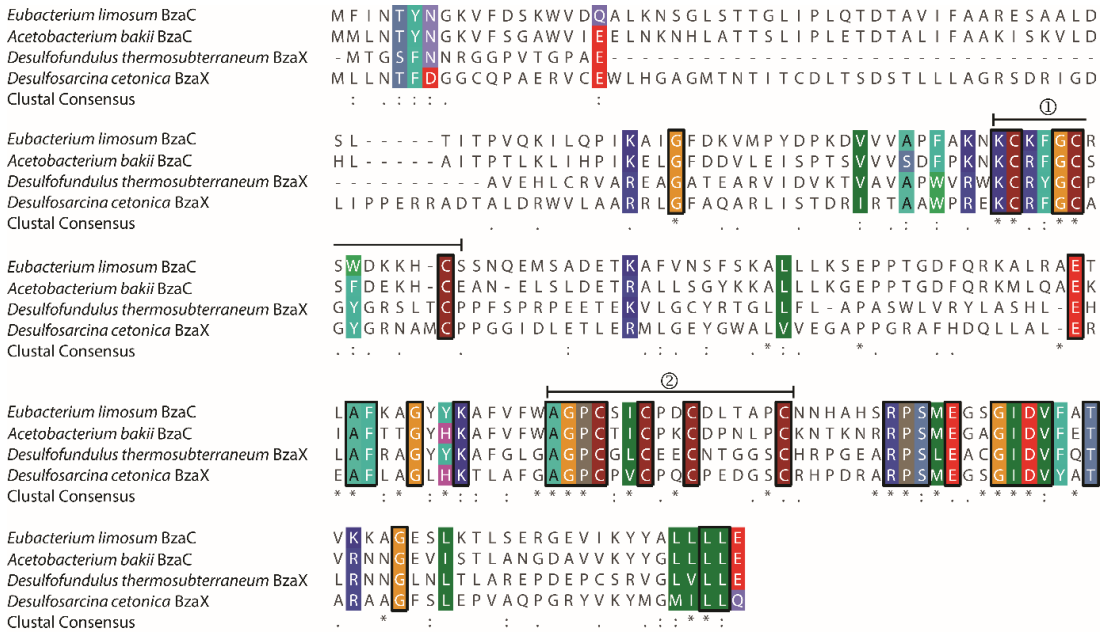


Figure 5.2 Sequence alignment with DUF2284 domain present in *bza* operon. Multiple sequence alignment with DUF2284 which is present as the C-terminal domain of BzaC and as separate protein encoded by the gene *bzaX*. Overall, the DUF2284 sequences show high sequence conservation, and the most striking feature of DUF2284 sequence is the presence of two cysteine rich conserved repeats- $CX_3C_{7-8}C$ (marked as ①) and $CX_2CX_2CX_4C$ (marked as ②).

5.3.2 Characterization of *El*BzaC reveals DUF2284 as an iron-sulfur cluster domain

We undertook the characterization of the activity of *El*BzaC and explored the biochemical nature of DUF2284. We expressed and purified *El*BzaC under aerobic, semi-anaerobic and completely anaerobic conditions. The protein purified with a slightly brown color and exhibited a peak at 420 nm in the UV-Vis spectrum under all three conditions which are regarded as characteristics of an iron-sulfur (Fe-S) cluster containing proteins (Figure 5.3A, red solid trace) (Chatterjee et al. 2008). However, attempts to reconstitute the methyltransferase activity of the *El*BzaC did not show any 5-OMeBza formation. The major difference between the *El*BzaC with the *Mt*BzaC homologs is the presence of the DUF2284 domain. Since experiments with *Mt*BzaC show that the first two domains are necessary and sufficient for the methylation of 5-OHBza, we deleted the DUF2284 domain from *El*BzaC. The *El*BzaC was truncated at Tyr373, based on the domain boundaries predicted by Conserved domain database (CDD). The *El*BzaCΔDUF2284 construct was obtained as a soluble and stable protein of expected molecular weight in SDS-PAGE and MALDI-TOF analysis.

Interestingly, purified *El*BzaCΔDUF2284 is colorless as compared to *El*BzaC. To verify the presence of Fe-S clusters, we co-expressed both proteins with the pDB1282 plasmid harboring the *isc* operon from *Acetobacter vinelandii* that facilitates the reconstitution of Fe-S clusters (Seravalli et al., 2008.). Under anoxic conditions, *El*BzaC purified as dark brown colored protein and showed an enhanced peak at the 420 nm wavelength in the UV-Vis spectrum as compared to the *El*BzaC expressed without pDB1282 (Figure 5.3A, red solid trace). The 420 nm peak reduces upon addition of dithionite signifying presence of functional Fe-S cluster associated with *El*BzaC (Figure 5.3A, red dashed trace) (Chatterjee et al. 2008). Next, we purified another homolog of BzaC from *Eubacterium barkeri* (*Eb*BzaC), which also purifies with a dark color and shows the characteristic peak at 420 nm under UV-vis spectrum which reduces upon addition of dithionite (data not shown).

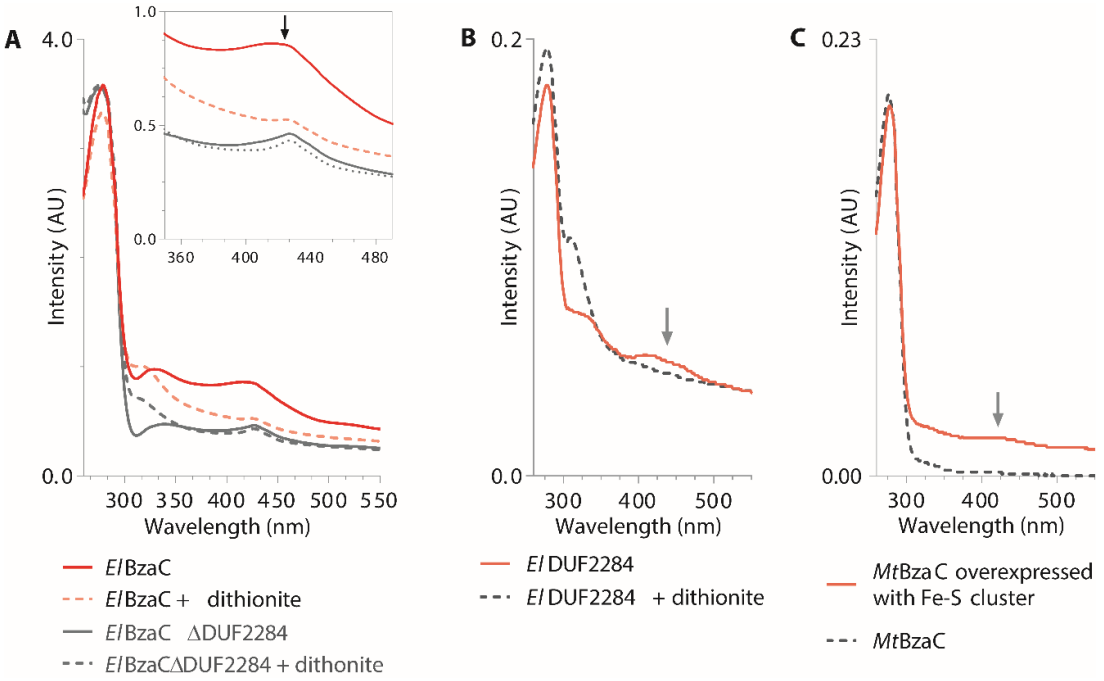


Figure 5.3 DUF2284 is a candidate Fe-S cluster domain. A) UV-Vis absorption spectrum of *El*BzaC overexpressed along with pDB1282 (plasmid coding Fe-S cluster biogenesis genes) for facilitation of Fe-S cluster reconstitution shows a peak at 420 nm (red solid trace) which is subsequently reduced upon addition of dithionite (red dashed trace). The *El*BzaCΔDUF2284 construct does not show a significant peak at 420 nm (grey solid trace) or any change with dithionite implying a lack of Fe-S cluster (grey dashed trace). **B)** UV-Vis spectrum of purified DUF2284 fragment from *El*BzaC, overexpressed with pDB1282 also shows a peak at 420nm that reduces upon addition of dithionite. **C)** UV-Vis spectrum of purified *Mt*BzaC does not show a similar UV-Vis spectrum despite co-expression with pDB1282 which produces the Fe-S cluster in the cell.

In contrast, no significant peak at 420 nm was seen in the UV-Vis spectrum of *El*BzaCΔDUF2284 protein overexpressed under similar conditions as *El*BzaC (Figure 5.3A, grey solid and dashed traces). Next, the sequence corresponding to the DUF2284 domain of *El*BzaC (*El*DUF2284) was cloned, overexpressed, and purified. The *El*DUF2284 construct exhibited low solubility when expressed on its own, which improved when co-expressed with

pDB1282. The UV-Vis spectrum of the purified *El*DUF2284 shows a peak at 420 nm similar to as observed for *El*BzaC which reduces in response to dithionite addition (Figure 5.3B, solid and dashed traces). Further, when *Mt*BzaC was expressed with or without pDB1282, it failed to bind any Fe-S cluster (Figure 5.3C). Hence, we conclude that the *El*BzaC is an Fe-S cluster binding protein and the cluster is bound within the DUF2284 domain. A primary elemental analysis using inductively coupled plasma -mass spectrometry (ICP-MS) show that *El*BzaC constructs purify with 3.6 iron/ protein molecule, whereas ~0.5 iron bind to the *El*BzaCΔDUF2284 protein. Further replicates of the elemental analysis remain to be conducted. The advanced spectroscopic analysis to determine the nature of Fe-S cluster bound to the DUF2284 remains to be done.

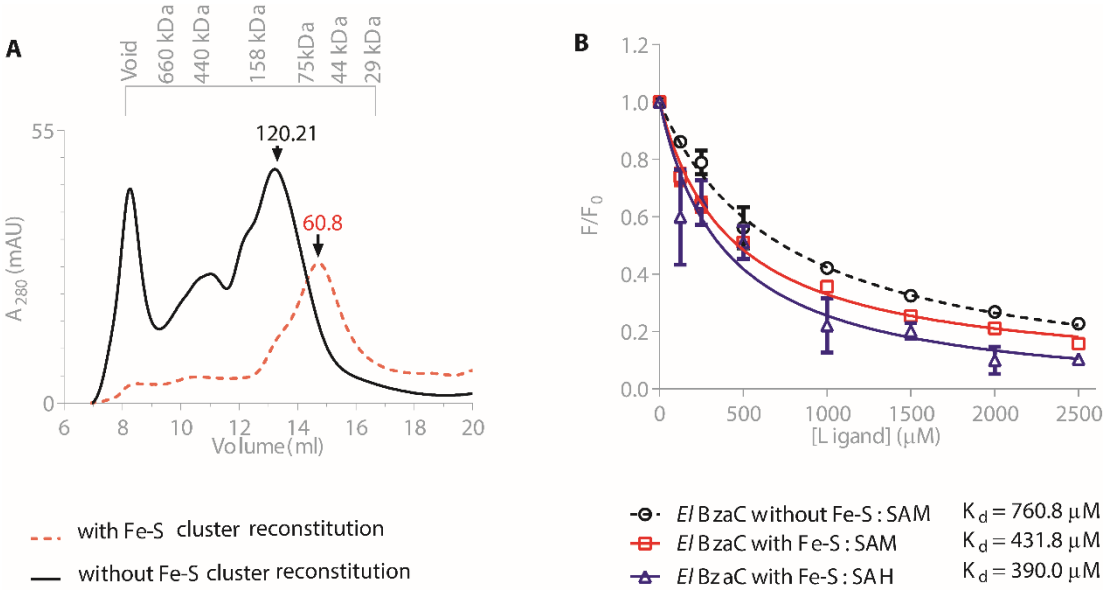


Figure 5.4 Biochemical characterization of *El*BzaC. A) Size exclusion chromatography to evaluate the role of dimerization domain and the effect of iron-sulfur cluster reconstitution on oligomerization of *El*BzaC. **B)** Intrinsic fluorescence based binding assay for *El*BzaC shows a dissociation constant value of $K_{d, SAM}$ = 760.8 ± 109.2 µM for SAM. For *El*BzaC with an *in vivo* reconstituted Fe-S cluster, the values are $K_{d, SAM}$ = 431.8 ± 38.41 µM and $K_{d, SAH}$ = 390.0 ± 117.7 µM for SAM and SAH, respectively.

Next, we conducted size exclusion chromatography and intrinsic fluorescence based binding assays with *El*BzaC and tested the effect of Fe-S cluster reconstitution on the enzyme. Size exclusion chromatography shows that *El*BzaC without a reconstituted Fe-S cluster forms dimers and higher order oligomers, however the fraction of monomer increases when the Fe-S cluster is reconstituted (Figure 5.4A). The binding affinity of *El*BzaC for SAM increases by 1.7-fold in the presence of Fe-S cluster - the K_d value for Fe-S reconstituted protein is 431.8 µM as opposed to K_d value of 760.8 µM without the aided cluster formation. Similar to *Mt*BzaC, *El*BzaC also binds to SAH as efficiently as it does to SAM, with K_d value of 390.0 µM (Figure 5.4B). When tested for *in vitro* reconstitution, *El*BzaCΔDUF2284 and *El*BzaC

purified with and without reconstituted Fe-S clusters failed to yield any product (data not shown).

5.3.3 Mapping the presence of DUF2284 to the phylogeny analysis of BzaC homologs

The presence of a DUF2284 domain in *El*BzaC and its absence in *Mt*BzaC may suggest that the two homologs of BzaC engage in methylation via different reaction mechanisms. Hence, we proceeded with phylogeny and comparative genomic analyses of putative BzaC homolog sequences available in the Genbank database.

We compiled a set of methyltransferase sequences similar to either *Mt*BzaC or *El*BzaC using NCBI-BLASTp search. A search for sequences similar to BzaC among diverse bacterial phyla revealed that the occurrence of BzaC is limited to members of Clostridiales of the Firmicutes phylum and a few classes of the Deltaproteobacteria phylum (data not shown). A list of 833 unique methyltransferases was obtained and subjected to HMM-search using the recently defined BzaC HMM models (Shelton et al. 2018). We find that only 71 sequences score higher than the prescribed cut-off. We also manually searched for the gene neighborhood of all the sequences to identify other putative BzaC homologs. We define BzaC as a class-I methyltransferases coded by a gene present in proximity of either cobalamin riboswitch, *bza* genes or *cob* genes (namely *bzaA*, *bzaB*, *bzaF*, *bzaD*, *bzaE* and *cobT*, *cob*S, *cobU*). Finally, a dataset comprising of 90 sequences was curated. We observed three possible arrangements of BzaC and DUF2284 in the *bza* operon (Figure 5.1, Figure 5.5)

1) the presence DUF2284 coding region with the *bzaC* gene which results in *El*BzaC-like homologs

2) Absence of DUF2284 presence of coding region with the *bzaC* gene which results in *Mt*BzaC-like homologs

3) Presence of DUF2284 coding gene adjacent to the bzaC gene, thus resulting in *Dt*BzaC and *Dt*BzaX like combination

We then subjected the protein sequences to a phylogenetic analysis using the maximum likelihood method (Stamatakis 2014). To eliminate the possibility of the DUF2284 sequence biasing the phylogenetic distribution because when present it constitutes ~1/3 of the sequences, the analysis was conducted by aligning only the sequences corresponding to the dimerization and methyltransferase domains. The phylogenetic tree (Figure 5.5) shows that the BzaC sequences of related organisms cluster together, suggesting that the first two domains of BzaCs are taxonomically conserved. We then mapped the domain architecture of each homolog on the resultant phylogeny tree to examine the clustering pattern of BzaC sequences. The overall distribution of the BzaC homologs shows small clusters of *Mt*BzaC-like sequences (Figure 5.5, grey font) scattered among the clades of *El*BzaC-like sequences (Figure 5.5, black font). This

suggests that the methyltransferase domain is conserved among these clades and the incidence of DUF2284 domain is not reflected in the sequences of the first two domains of BzaC. Hence, the effect of DUF2284 on the activity of the two types of BzaCs might be beyond the methyltransferase mechanism.

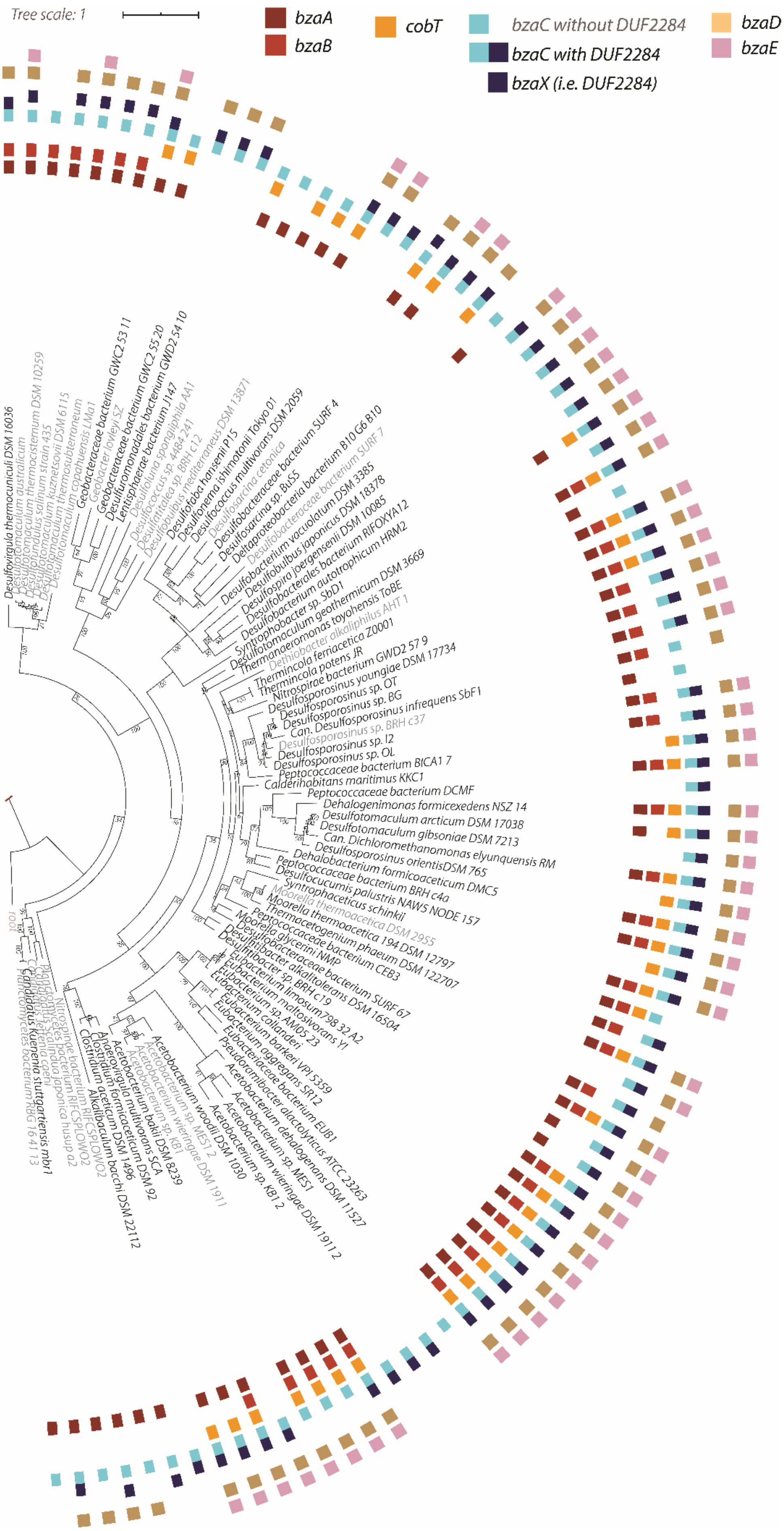

Tree scale: 1
bzaA
bzaB
cobT
bzaC without DUF2284
bzaC with DUF2284
bzaX (i.e. DUF2284)
bzaD
bzaE

Figure 5.5 Phylogeny and comparative genomics analysis of BzaC. A phylogeny tree of BzaC homologs with the corresponding composition of the *bza* operon. A colored square box indicates the presence of the respective gene in the neighborhood *bzaC* genes on the tree. The *bzaA* and *bzaB* are represented by two red squares, while a single red box signifies presence of bzaF gene instead. We observe three domain architectures for *bzaC* and DUF2284 domain. The genes encoding *Mt*BzaC-like homologs which contain a dimerization domain and methyltransferase domain are shown in light blue box. The genes for *El*BzaC-like homologs which contain a dimerization domain, a methyltransferase domain, and a DUF2284 domain are shown as fusion of light and dark blue boxes to signify presence of additional DUF2284 at C-terminal. The genes encoding for a single domain DUF2284 are named *bzaX* and are represented with dark blue boxes. The presence of either or both putative radical-SAM methyltransferases *bzaD* and *bzaE* in the corresponding gene neighborhood (brown and pink squares) are also mapped on the tree. We notice that majority of operons that encode a DUF2284 domain, either as a part of *bzaC* or as *bzaC*, also encode for either or both of *bzaD* and *bzaE* genes. A quantitation of this data to understand the co-occurrence of *bza* operon genes with DUF2284 is shown in Figure 5 6.

5.3.4 A comparative genomics approach to understand the DUF2284 in context of the *bza* operon

To further explore the role of the DUF2284 in context of the *bza* operon, the putative *bza* genes present in the vicinity of *bzaC* were examined in a comparative genomic analysis (Figure 5.5, Table 1). The curated dataset includes 19 BzaC sequences corresponding to *Mt*BzaC-like domain architecture that is, the DUF2284 sequence was absent. We found 63 sequences with *El*BzaC-like domain architecture that is, the DUF2284 sequence was present within the protein. Lastly, only 8 sequences similar to *Dt*BzaC-like architecture where a separate gene *bzaX* coded for DUF2284 (Figure 5.1, Figure 5.6). Inspection of the gene neighborhood of the *bza* operon unveiled a large diversity of the combinations of genes of the *bza* operon in various organisms.

We observe that among the group with *Mt*BzaC-like entries, 63% (12 out of 19) are not followed by *bzaD* and/or *bzaE* (Figure 5.6, 5.7, Table 1) whereas, in the group with *El*BzaC-like entries, 76.1% (48 out of 63) possess *bzaD* and *bzaE*, 15.8% (10 out of 63) have either *bzaD* or *bzaE*, and only 7.9% (5 out of 63) lack the radical SAM coding genes. Similarly, among the 8 organisms constituting the group with *Dt*BzaC-*Dt*BzaX- like combination, 25% (2 out of 8) have both *bzaD* and *bzaE* and 75% (6 out of 8) have *bzaE*. In summary, in all the *bza* operon variants possessing DUF2284 (Figure 6, 'DUF2284 present' category) we see that, 92.9% (66 out of 71) of the operons that encode DUF2284 also contain *bzaD* and *bzaE*. On the contrary, 85% (12 out of 19) of the operons lacking DUF2284 (Figure 6, DUF2284 absent category) also lack *bzaD* and *bzaE* as well (See Table 1 for details of individual *bza* operon variants). The co-occurrence of DUF2284 with BzaD and BzaE in the *bza* operon is statistically significant (Fishers's exact test, two-tailed, $p < 0.0001$, relative risk = 2.523). In contrast, DUF2284 exhibit less significant co-occurrence with the first gene(s) of the biosynthesis pathway, *bzaAB/F* (Fishers's exact test, two-tailed, $p = 0.0050$, relative risk = 0.6901) and the

co-occurrence with the phosphoribosyltransferase gene *cobT* was found to be non-significant (Fishers's exact test, two-tailed, p = 0.6112, relative risk = 1.16) (Figure 5.6).

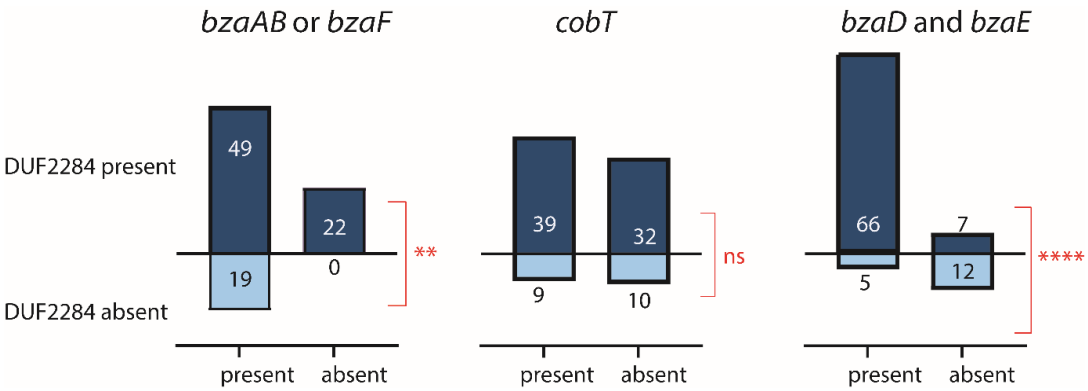


Figure 5.6 Statistical analysis for verifying the pattern of co-occurrence of DUF2284 with *bza* genes. The significance of co-occurrence of *bza* genes with the DUF2284 were analyzed using a Fisher's exact test. Operon showing DUF2284 either as a part of *bzaC* gene product or encoded as *bzaX* were clubbed under the DUF2284 present category. The rest of the operon that lacked DUF2284 were recorded as DUF2284 absent category. In three independent experiments, we sub-categorized the datasets based on presence of genes coding for 1) the first enzyme of the pathway, HBI-synthase (i.e. gene product of either *bzaA-bzaB* or *bzaF*); 2) phosphoribosyltransferase for activation of the lower ligands (i.e. *cobT*); 3) the last two radical-SAM methyltransferases (i.e. either one or both of *bzaD* or *bzaE*). The cases where *bzaAB*, *bzaF* or *cobT* were found at genomic locations distant from the *bza* operon were also counted in the absent category. The p-value for the co-occurrence of DUF2284 with *bzaAB*/F, *cobT*, *bzaD* and/or *bzaE* were 0.005, 0.61, <0.0001 which strongly suggests non-random co-localization of the DUF2284 with the BzaD and BzaE coding genes.

5.3.5 Primary characterization of the last two methyltransferases

Previously, the heterologous expression of the complete *bza* operon from *Eubacterium limosum*, that is the *bzaA-bzaB-cobT-bzaC-bzaD-bzaE* resulted in formation of B_{12} as the final product(Hazra et al. 2015). The expression of *bzaA-bzaB-cobT-bzaC-bzaD* construct results in formation of 5-methoxy-6-methylbenzimidazolylcobamide [5-OMe-6MeBza]Cba as the product, and the expression of *bzaE* without the *bzaD* gene did not result in B_{12}. Furthermore, the expression of *bzaA-bzaB-cobT-bzaC* construct only resulted in small amount of 5-methoxybenzimidazolylcobamide [5-OMeBza]Cba. Importantly, while *bzaC* gene product appears to function even in the presence of oxygen, the activity of *bzaD* and *bzaE* strictly required anaerobic growth conditions.

Thus, previous observations agree with the oxygen-sensitivity owing to predicted presence of Fe-S clusters and indicate that presence of *bzaD* is imperative for the function of *bzaE*. When we tested the activity of full length *El*BzaC with 5-OHBza, we did not observe any methylated product formation. We are currently testing the activity *El*BzaC with 5-OHBza-riboside (5-OHBza-R), the preferred substrate of *Mt*BzaC homolog. Further, to test whether DUF2284 is essential for activity of BzaC homologs from the operons that contain a DUF2284, we overexpressed *Dt*BzaC in ED674 strain, which produces the cobamide backbone. We fed

5-OHBza to the ED674 strain harboring a *DtbzaC* gene on a pET28a vector, and we compare the cobamide produced by the strain against ED674 contain an empty pET28a vector. We observe, that ED674 strain expressing a *Dt*BzaC fed with 5-OHBza efficiently produced [5-OMeBza]Cba whereas the strain without *Dt*BzaC produced only 5-hydoxybenzimidazolylcobamide [5-OHBza]Cba. Thus, providing another proof that DUF2284 is not required for the first methylation in the pathway (data not shown). We then tested the activity of *El*BzaCΔDUF2284 under *in vitro* and in ED674 strain, which failed to produce a methylated product (data not shown). Based on our comparative genomics assay, the high co-occurrence of DUF2284 with BzaD and BzaE coding genes might signify that some sort of interaction between the DUF2284, BzaD, and/ or BzaE might be required for the activity of methyltransferases in organisms that contain the *bzaA-bzaB-cobT-bzaC-bzaD-bzaE* construct of the *bza* operon.

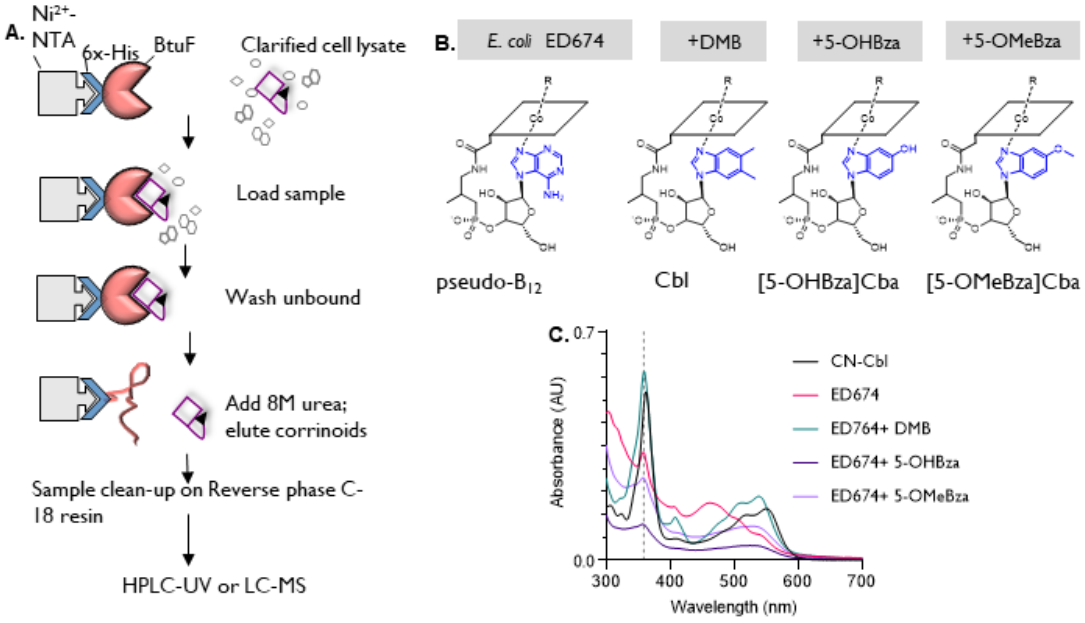


Figure 5.7 Affinity purification of B_{12} and production of alternate cobamides. A) Schematic for affinity purification of cobamides using immobilized BtuF which is a B12 transporter. **B)** The strain E. coli ED674 produces pseudo-B_{12} as native cobamide and hence various benzimidazole derivatives can be added to the culture to enable biosynthesis of cobamides with 5-OHBza or 5-OMeBza as lower ligands. When DMB is added, cobalamin is obtained as the resulting cobamide. **C)** UV-Vis spectra of commercially available cyanocobalamin (CN-Cbl, black trace), pseudocobalamin (ED674-pink trace), cobalamin formed with DMB added (green trace), [5-OHBza]Cba produced with 5-OHBza added to culture (dark purple trace), and [5-OMeBza]Cba produced with 5-OMeBza added to culture (light purple trace).

We then aimed to purify, test the biochemical properties of BzaD and BzaE, and attempt the reconstitution of the enzymes as per the two hypotheses shown in Figure 5.1. To do so, the first step was to obtain the required substrates. To produce 5-OMeBza-riboside we used methods using CobT and alkaline phosphatase similar to that used for production of 5-OHBza-riboside in Chapter 4. However, the chemical synthesis of 5-OMe-6MeBza lower ligand is a

challenging process, and hence could not be obtained yet. Next, to obtain the cobamide substrates, we used the guided biosynthesis and affinity purification using ED674 strain and immobilized BtuF protein, as described in methods section (Figure 5.7A). We fed ED674 cultures with 5-OHBza, 5-OMeBza, and DMB and purified the corresponding cobamides produced by the strain (Figure 5.7B). As expected, the ED674 strain makes pB_{12} as the native cobamide in absence of any added lower ligand (Figure 5.7B, 5.7C pink trace). When grown with exogenously supplied DMB, the strain produces cobalamin which shows UV-vis identical to cyanocobalamin (Figure 5.7B, 5.7C green trace). Next, when fed with 5-OHBza and 5-OMeBza we obtain [5-OHBza]Cba and [5-OMeBza]Cba showing characteristic UV-vis spectra (Figure 5.7B, 5.7C). We further validated the identity of the cobamides produced using LC-MS analyses (data not shown).

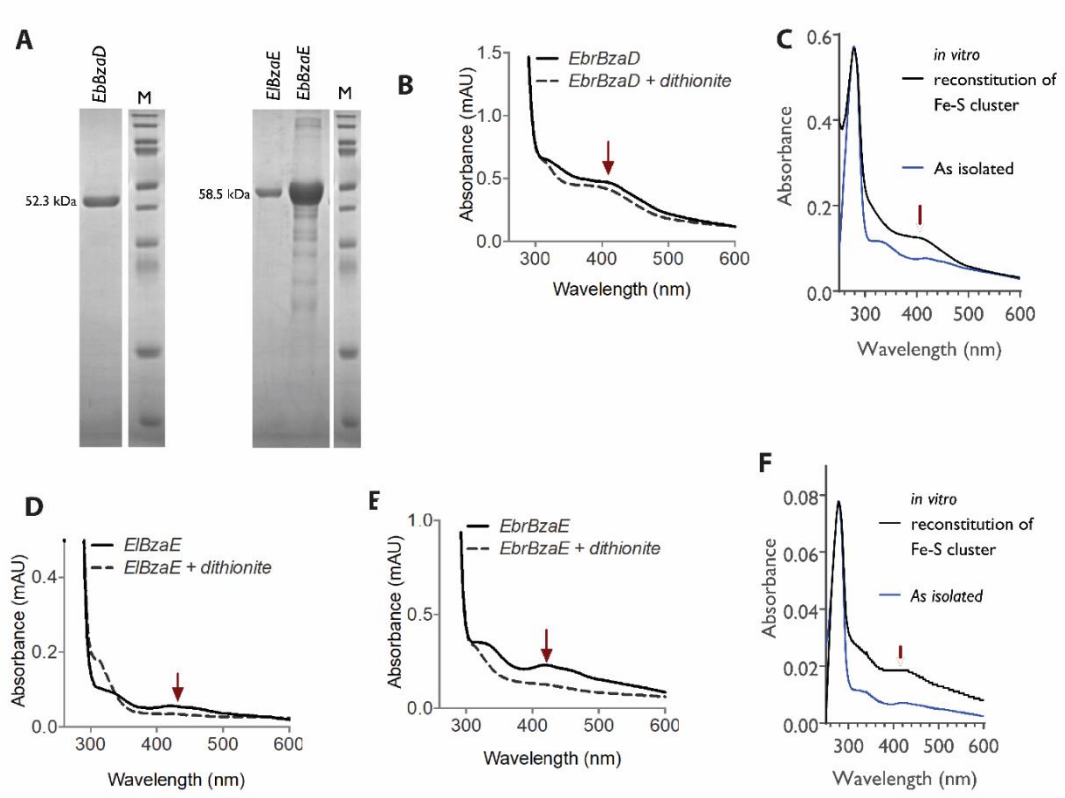


Figure 5.8 Primary characterization of BzaD and BzaE. A) SDS-PAGE gel showing purified *Eb*BzaD, *El*BzaE, and *Eb*BzaE. **B)** UV-Vis of purified *Eb*BzaD and dithionite treated *Eb*BzaD. **C)** UV-vis spectra of *Eb*BzaD purified as-isolated (blue trace) and the *Eb*BzaD subjected to in vitro reconstituted Fe-S cluster (black trace). **D)** and **E)** UV-Vis spectra of as-isolated *Eb*BzaE and *El*BzaE and subsequent dithionite treatment shows presence of a functional Fe-S cluster. **F)** The UV-vis spectra of as-isolated *Eb*BzaE (blue trace), and the *Eb*BzaE after the in vitro reconstitution of the Fe-S cluster (black trace).

Next, we moved onto obtaining purified enzymes BzaD and BzaE from *Eubacterium limosum* (*El*BzaD, *El*BzaE) and *Eubacterium barkeri* (*Eb*BzaD, *Eb*BzaE). We conducted protein overexpression under anaerobic conditions in *E.coli* BL21 (DE3) strains co-transformed with pDB1282 and the pTE28a vectors harboring *bzaD* and *bzaE* genes. Subsequently, the proteins were purified in an anoxic glove bag maintained at <70 ppm of oxygen. Unfortunately,

the *El*BzaD homolog fails to overexpress in *E.coli* BL21 (DE3) strain, and thus we use *Eb*BzaD as our only BzaD representative. The *El*BzaE and *Eb*BzaE overexpress and purify well (Figure 8A). The *Eb*BzaD, *El*BzaE, and *Eb*BzaE enzymes purify as light brown colored fraction, and under UV-Vis spectroscopy all three proteins show peaks at 420 nm which gets reduced upon addition of dithionite, thus validating with the predicted presence of Fe-S clusters. However, the ICP-MS analyses of the purified *Eb*BzaD, *El*BzaE, and *Eb*BzaE enzymes show less than 0.5 iron per protein, indicating poor *in vivo* reconstitution of the Fe-S cluster. We then conducted an *in vitro* reconstitution of Fe-S cluster using $FeCl_3$, Na_2S, and DTT (Protocol from Jennifer Bridwell-Rabb, personal communication). The *in vitro* reconstitution of the Fe-S cluster results in slightly better cluster occupancy than as-is purified enzymes (Figure 8B, 8C). However, further optimization and identification of the nature of the Fe-S cluster through spectroscopic methods remain to be done. We also conducted pilot experiments using intrinsic fluorescence of the proteins to probe whether BzaD and BzaE can bind to B_{12} and SAM, and we observe a decay in protein fluorescence upon addition of B_{12} and SAM. Further replicates of this assay are underway. Additionally, we have recently cloned and overexpressed *Dt*BzaE2, the novel and yet uncharacterized homolog of BzaE. Further, biochemical characterization will help ascertain the B_{12}, SAM, and Fe-S cluster binding properties of the homolog.

The progress so far with characterization of BzaD and BzaE, the last two methyltransferases of the *bza* operon includes optimization of purification of predicted substrates, and the enzymes with Fe-S cluster cofactors. The next steps involving reconstitution of the enzymatic activity of BzaD and BzaE are currently underway in our laboratory.

5.4 The Catch 2284: a discussion on DUF2284 and the last methylations in the *bza* operon pathway

Our comparative genomics analysis suggests a strong correlation between the role of DUF2284 in the function of BzaD and BzaE (or BzaE2) enzymes, which are predicted to catalyze the last steps in the *bza* operon pathway to make DMB, the lower ligand of B_{12}. The biochemical characterization of *El*BzaC, the BzaC homolog which contains DUF2284 show the presence of Fe-S cluster, and we observe a good cluster occupancy in the purified protein. However, the BzaC is an O-methyltransferase as seen through heterologous expression of the homologs and in vitro reconstitution of the *Mt*BzaC homolog. Additionally, using *Mt*BzaC and *Dt*BzaC we confirm that DUF2284 is not required for O-methylation reaction. Yet, the absence of activity with purified *El*BzaC is intriguing.

Further, the BzaD and BzaE proteins purify with Fe-S clusters as expected as well as appear to bind B_{12} and SAM as predicted based on domain architecture. Notably, the cluster

occupancy in the purified BzaD and BzaE homologs remain low (seen over multiple rounds of purification).

Given that heterologous expression of *bzaC-bzaD-bzaE* constructs in *E.coli* lead to formation of $B_{12,}$ we propose that the three methyltransferases are, in some way, conduct methylation in a concerted fashion. Thus, our current approach with studying the enzyme individually might not be the best way forward to study the mechanisms of DUF2284, BzaD, and BzaE. The characterization of DUF2284 requires insights from functional analyses with BzaD and BzaE, and *vice versa*- leading to a catch 22-esque dilemma for experiments with these set of enzymes.

In literature, small Fe-S cluster binding proteins such as DUF2284 are reported to be involved in a wide range of functions (Johnson et al. 2005). Examples include pyruvate formate-lyase activating enzymes that hosts the Fe-S cluster and SAM for generation of glycine radical in the pyruvate formate-lyase (Broderick et al. 1997), ferredoxins that store Fe-S clusters, Fe-S clusters that involved in substrate binding in radical-SAM enzymes, as well as in regulation of enzyme activity.

The predicted structures for three representative DUF2284 containing sequences namely, *Eubacterium limosum* BzaC, *Desulfofundulus thermosubterraneum* BzaX, and *Clostridioides difficile* DUF2284 were generated using AlphaFold hosted on UniProt (Jumper et. al., 2021) (Figure 5.9). The predicted structure of ElBzaC shows well-defined three domains corresponding to the dimerization, methyltransferase, and DUF2284 domains. In all three structures, the conserved cysteines appear to arrange in close proximity, and from an overview we can predict two sites for Fe-S cluster binding. Further systematic cysteine to alanine mutagenesis will precisely inform about the role of each of the cysteines in Fe-S cluster binding. Notably, all three structures indicate that the hydrophobic C-terminal end of the DUF2284 is likely to be buried away from surface, and nuances such as these can help in designing labelled constructs for studying protein-protein interactions between BzaD, BzaE, and DUF2284.

We have initiated a systematic experiment design to probe possible interaction between the DUF2284, BzaD, and BzaE utilizing vectors that allow co-expression of two or three proteins as a time. Alongside the BzaC, BzaD, and BzaE homologs from *Eubacterim barkeri*, we are also working with *Dt*BzaC, *Dt*BzaX, and *Dt*BzaE2 which allows precise analysis of DUF2284 as well as the new homolog BzaE2. Further analysis of the last methyltransferases of *bza* operon are bound to reveal new and exciting mechanisms for enzymes that bind B_{12}/radical-SAM cofactors, as well provide the last missing piece in the puzzle of B_{12} biosynthesis pathways.

5.5 Open avenues and interesting observations

5.5.1 Analysis of the phylogeny of DUF2284 beyond the *bza* operon

The domain of unknown function-DUF2284 is reported to be present in only bacteria and archaea, and this correlates well with the fact that B_{12} biosynthesis can only be conducted by a restricted set of prokaryotes (Goodacre, Gerloff, and Uetz 2014). To investigate the role of DUF2284 and test for its role in other biological systems, we acquired a large number of sequences for DUF2284 homologs from both archaea and bacteria and conducted a phylogeny analysis (Figure 5.9). For DUF2284s that were part of BzaC, only the corresponding sequence for the domain were used. In the resulting phylogeny tree, majority of the DUF2284 sequences which are a part of BzaC sequences appeared in a separate clade, and the products of *bzaX* in the *bza* operon clustered in a separate clade. Archaeal sequences formed a significantly independent clade, and two other less populated clades, except for one archaeal DUF2284 sequence that was situated in a small clade with bacterial sequences. The rest of all bacterial DUF2284 sequences formed separate clades and we see a pattern of taxonomical conservation within these sequences as well. We then generated multiple sequence alignments with sequences within each clade and we record the patterns in conservation of cysteines. We find that the first cysteine repeat from the N-terminal of the sequence (Figure 5.9, cysteine repeat #1), show a higher conservation with $CX_2CX_{7-8}C$ consensus. The variations in the distances between the cysteines in the second repeat, $CX_2CX_{2-6}CX_{3-12}C$, which is closer the C-terminal of the sequence majorly contributes to the differences between DUF2284 sequences across various clades (Figure 5.9, cysteine repeat #2). As an exception, the clades V which contains bacterial sequences lack the first cysteine repeat all together. In clade VI, we observe another conserved cysteine which results in a CX_2CX_7CC consensus. The sequences in clades V and VI contain a longer repeat #2 with an additional cysteine. Notably, the cysteine repeat #2 in the DUF2284 from BzaC sequences is shorter with a $CX_2CX_2CX_{4-8}C$ consensus, however the corresponding sequence from the BzaX sequences is a longer $CX_2CX_{2-6}CX_{12}C$. The phylogeny and sequence level differences in the two types of DUF2284 sequences from the *bza* operon hints towards a key role of BzaC-DUF2284 in the evolution of the operon. Overall, the differences in consensus sequences and the arrangement of cysteine repeats hints towards possible differences in protein structure and the way iron-sulfur clusters would be bound within these homologous DUF2284 sequences. Further analysis of archaeal and bacterial DUF2284 sequences that do not belong to *bza* operon will aid in understanding the general biochemical functions of this novel domain with yet unknown function.

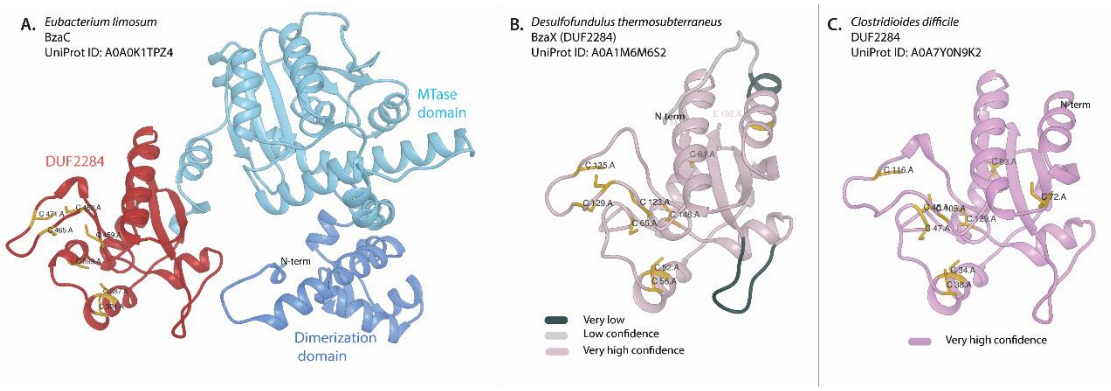
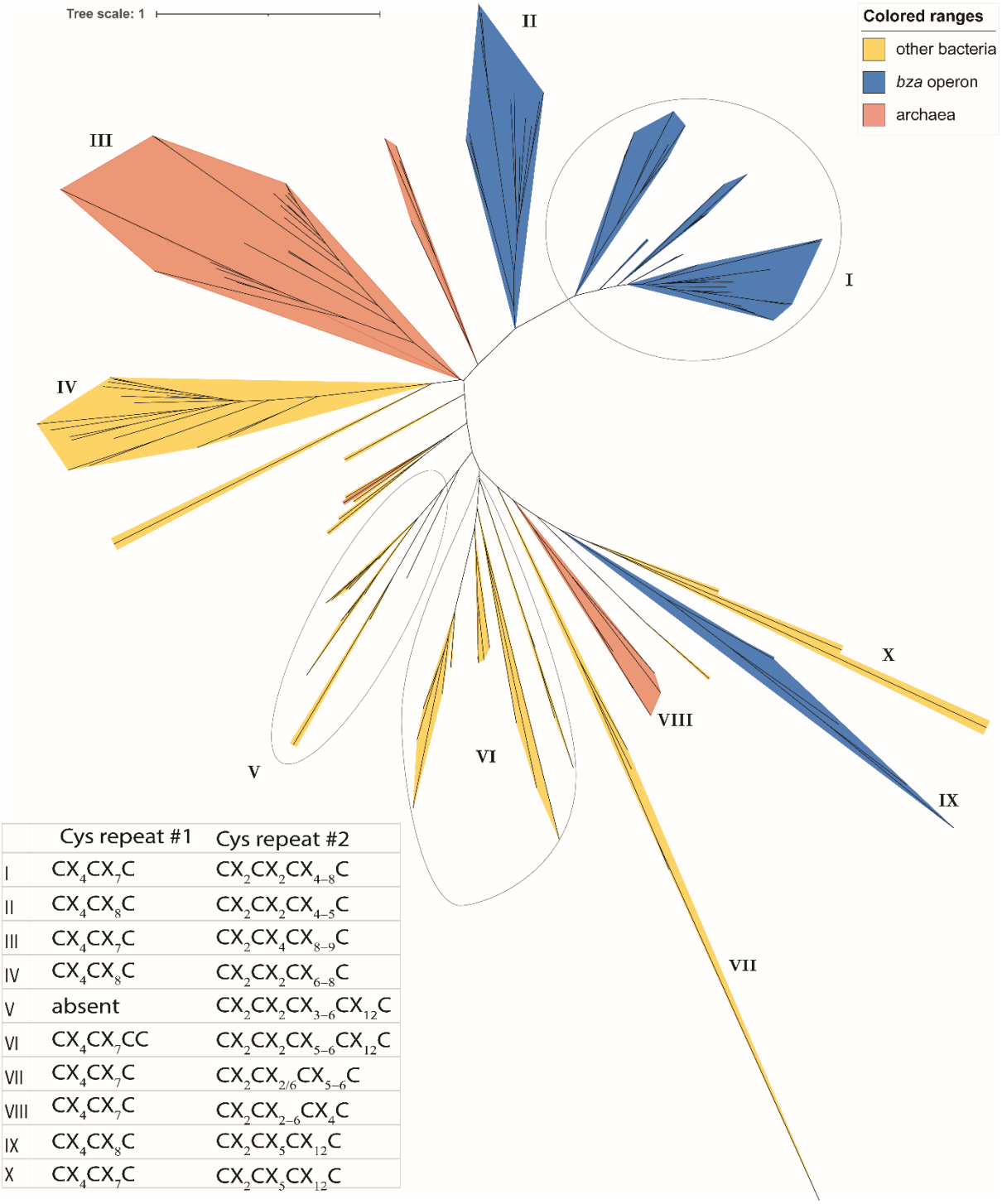


Figure 5.9 Predicted structures for three DUF2284 containing proteins. A) Predicted AlphaFold structure for *Eubacterium limosum* BzaC show the three conserved domains namely, dimerization domain (sky blue), methyltransferase (cornflower blue), and the DUF2284 (firebrick red). The cysteine residues are shown in golden yellow and marked with their residues number. **B)** Modeled structure of *bzaX* gene product encoding single domain DUF2284. The low confidence prediction regions are marked in light grey and grey, and high confidence regions in pink. The cysteine residues are colored in golden yellow. **C)** The DUF2284 sequence form *Clostridioides difficile* generated a very high confidence model and might be an interesting candidate for *in vitro* characterization of DUF2284 that exists outside the *bza* operon.



	Cys repeat #1	Cys repeat #2
I	CX_4CX_7C	$CX_2CX_2CX_{4-8}C$
II	CX_4CX_8C	$CX_2CX_2CX_{4-5}C$
III	CX_4CX_7C	$CX_2CX_4CX_{8-9}C$
IV	CX_4CX_8C	$CX_2CX_2CX_{6-8}C$
V	absent	$CX_2CX_2CX_{3-6}CX_{12}C$
VI	CX_4CX_7CC	$CX_2CX_2CX_{5-6}CX_{12}C$
VII	CX_4CX_7C	$CX_2CX_{2/6}CX_{5-6}C$
VIII	CX_4CX_7C	$CX_2CX_{2-6}CX_4C$
IX	CX_4CX_8C	$CX_2CX_5CX_{12}C$
X	CX_4CX_7C	$CX_2CX_5CX_{12}C$

Figure 5.10 Phylogeny analysis of DUF2284 sequences. The blue shaded sequences belong to DUF2284 sequences from the *bza* operon. Other independent DUF2284 sequences from bacterial are archaeal origin are colored in yellow and orange, respectively. The DUF2284 sequences found as a part of BzaC sequences sort in clades I and II whereas the DUF2284 encoded as *bzaX* are present in clade IX. Most archaeal sequences form clades separate than the sequences of bacterial origin. The clades show some differences in the consensus sequences wherein the cysteine repeat # 1 show presence of three cysteines such as the distances between the residues are also highly conserved across the clades. The two exceptions to this observation are clade V in which the sequences lack the cysteines corresponding to repeat 1 and clade VI which has an additional fourth conserved cysteine. The cysteine repeat # 2, on the contrary, show larger variations across the clades arising from the differences in the distances between the cysteines.

5.5.2 Comparative genomics reveal new variants of *bza* operon

Our comparative genomics analysis to study the abundance of DUF2284 among the *bza* operon. We probed the gene neighborhood of *bzaC* in all organisms in NCBI that show presence of *bzaC* gene (as of 2019). We find <100 organisms that contain a BzaC homolog which corroborates with a contemporary study indicating that *bza* operon is a very rare gene cluster in nature (Shelton et al. 2018). Considering all the variants contributed by each of the six genes (*bzaAB, bzaF, bzaC, bzaD, bzaE,* and *cobT*), we find a total of 26 *bza* operon variants. However, several of these variants are unique and one-off examples. We have few observations worth mentioning (Table 5.1). First, in our dataset, five organisms are exceptions to the observed rule i.e., the DUF2284 coding gene is neither accompanied by *bzaD* nor *bzaE*. Four of these are *Acetobacterium* species that each harbor two potential BzaCs. In these organisms, one of the candidate BzaC is similar to *Mt*BzaC and is encoded by the *bza* operon and another candidate is similar to *El*BzaC which is present at an alternate location on the genomic DNA, often adjacent to a cobalamin riboswitch (Table 5.1). Second, we also observe some interesting variants that lack *bzaAB* / *bzaF* and *cobT* or only *cobT* within the operon and these genes are rather found at an alternate location in the genome sequence, adjacent to a cobalamin riboswitch (Table 1, row 87-92, marked with *). Third, few organisms such as *Dehalogenimonas formicexedens* NSZ 14 and *Calderihabitans maritimus* KKC1, also show randomized arrangement of the *bza* operon genes (Table 1, row 2 and 3). Four, an exceptional organism *Desulfosarcina cetonica* (Table 5.1, row 62) show fusion of two consecutive *cobT-bzaC* genes we name the gene as '*bzaT'*) which is followed by a *bzaX*, and *bzaD*, and *bzaE* genes. Five, *Desulfobacterium vacuolatum* (Table 5.1, row 41) contain the *bzaA-bzaB-cobT* under cobalamin riboswitch at a different gene locus than *bzaC-bzaD-bzaE* genes coded under another cobalamin riboswitch. We have initiated a project to test and validate the new *bza* operon using the same strategy that was used for discovery of the *bza* operon (Hazra et al. 2015). The *bza* operon from following organisms were amplified.

1. Desulfobacterium vacuolatum (bzaC-bzaD-bzaE)

2.Desulfotomaculum thermosubterraneum (bzaA-bzaB-cobT-bzaC-DUF2284-bzaE)

3.Desulfosarcina cetonica (bzaAB-cobTbzaC-DUF2284-bzaD-E)

4.Desulfobulbus mediterraneus (bzaA-bzaB-cobT-bzaC)

5. Pseudoramibacter alactolyticus (bzaA-bzaB-cobT-bzaC-bzaD-bzaE)

We plan to insert the genes in the pTH1227 vector that will allow heterologous expression of the operon in B_{12}-dependent methionine auxotroph *E.coli MG1655 ΔmetE* as well as the *E.coli MG1655 ΔmetEΔcobT* strains. The new bza operon variants might or might not be involved in biosynthesis of the 5-methylbenzimidazole and benzimidazole as lower ligands, which are known to be produced by anaerobic bacteria and the underlying the biosynthesis pathway is yet unknown. Further experimental explorations are required to verify the functional forms of these

bza operon variants. The identification of several variants of a gene cluster within the species mainly belonging to the two taxonomic groups, the firmicutes and deltaproteobacteria, make a strong case for further explorations into the evolution and inheritance of the *bza* operon.

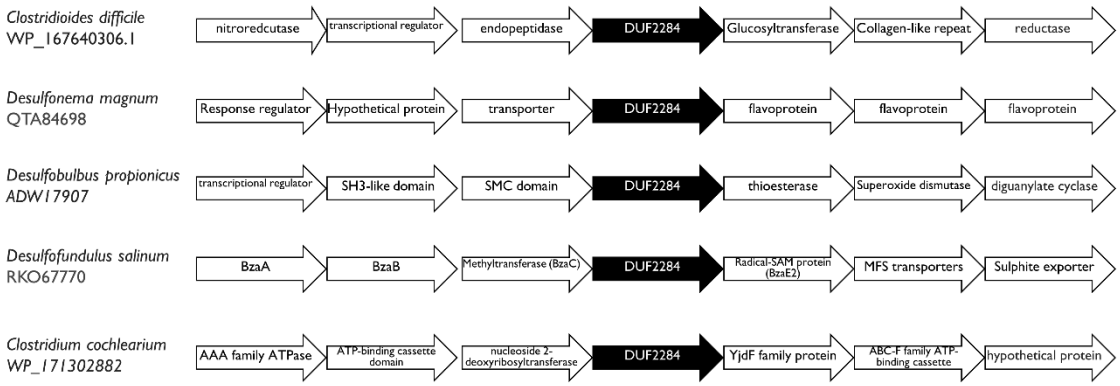


Figure 5.11 Genomic context of DUF2284 encoding genes. Manual analyses of over 300 DUF2284 sequences show that beyond the *bza* operon, the DUF2284 domain rarely localize with methyltransferases or radical-SAM enzymes. Here, gene neighborhood of DUF2284 (*bzaX*) was drawn for five randomly selected anaerobic bacterium as representative examples. The enzyme classification of each gene product to indicate possible functional context of the gene locus. The abundance of DUF2284 in bacteria and archaea demands high-throughput genomic context analyses tools to better understand the function of DUF2284 beyond the *bza* operon.

Table 5.1 (See on next page): Comparative genomics reveal diversity of *bza* operon. The genes neighboring the bzaC coding for the BzaC candidates used for phylogenetic analysis. The color-blocks represent presence of the genes- *bzaA*, *bzaB*, *bzaF*, *cobT*, *bzaD* and *bzaE*, in the neighborhood of ±5 genes surrounding the *bzaC* genes. In certain cases, *bzaAB*/*F* and *cobT* were found to be present at an alternate genomic location and those are indicated by an asterisk (*). Few genomic sequences that are in early stages of sequencing or assembly contain full BzaC sequence but the contigs or assemblies end in the middle of other *bza* operon genes are marked as # and are not included in the subsequent analysis. The O-methyltransferases sequence (marked &) was found downstream of a cobalamin riboswitch but no other *bza* genes were found it near it. The lower ligands are predicted on the basis of combination of the three methyltransferases present in each operon. The organisms for which the cobamides are characterized previously are shown in bold font.

#	Accession # BzaC	Organism	AdoCbl riboswitch	bzaA	bzaB	cobT	bzaC	DUF2284	bzaD	bzaE	Predicted lower ligand
1	WP_074617798	***Eubacterium limosum798 32_A2***									DMB **5**
2	WP_088552784	*Calderihabitans maritimus KKC1*									
3	WP_076003579	*Dehalogenimonas formicexedens NSZ 14*									
4	RJR20842	*Desulfobacteraceae bacterium SURF_67*									
5	RRF93047	*Eubacteriaceae bacterium EUB1*									
6	SFA12335	*Eubacterium aggregans SR12*									
7	SDX66345	***Eubacterium barkeri VPI 5359***									
8	SFP13941	*Eubacterium callanderi NLAE-zG225*									
9	WP_096918575	*Eubacterium maltosivorans YI*									
10	WP_118517165	*Eubacterium sp. AM05-23*									
11	OGW39895	*Nitrospirae bacterium GWD2_57_9*									
12	WP_006599075	*Pseudoramibacter alactolyticus ATCC 23263*									
13	CEO88753	*Syntrophaceticus schinkii*									
14	WP_084664742	*Thermanaeromonas toyohensis ToBE*									
15	WP_052219016	*Thermincola ferriacetica Z0001*									
16	WP_013119447	*Thermincola potens JR*									
17	KJR96872	*Peptococcaceae bacterium BRH_c4a*									
18	WP_014355847	***Acetobacterium woodii DSM 1030***									
19	SNR87440	*Anaerovirgula multivorans SCA*									
20	WP_015049717	*Thermacetogenium phaeum DSM 12270*									
21	RNC28551	*Can. Dichloromethanomonas elyunquensis RM*									
22	WP_089612495	*Dehalobacterium formicoaceticum DMC*									
23	WP_045572665	*Desulfosporosinus sp. I2*				!					
24	WP_075366261	*Desulfosporosinus sp. OL*				!					
25	WP_009618624	*Desulfosporosinus sp. OT*				!					
26	WP_007779263	*Desulfosporosinus youngiae DSM 17734*				*					
27	WP_068962968	*Desulfosporosinus sp. BG*				*					
28	SFQ95055	*Desulfallas geothermicum DSM 3669*				*					
29	WP_082058061	*Clostridium aceticum DSM 1496*									5-OMe,6-MeBza **4** & DMB **5**
30	AOY77160	***Clostridium formicaceticum ATCC 27076***									
31	WP_081562040	***Clostridium formicaceticum DSM 92***									
32	RJQ67621	*Desulfobacteraceae bacterium SURF_4*									
33	WP_100394655	*Desulfofaba hansenii P1*									
34	SFG63482	*Desulfotomaculum arcticum DSM 17038*									
35	WP_106823119	*Syntrophobacter sp. SbD1*									
36	WP_050741462	*Acetobacterium bakii DSM 8239*									
37	WP_028308597	*Desulfitibacter alkalitolerans DSM 16504*				*					
38	KUO53028	*Desulfitibacter sp. BRH_c19*				*					
39	WP_104373223	*Desulfocucumis palustris NAW5*		*	*						DMB **5**
40	KJS67751	*Peptococcaceae bacterium BICA1-7*		*	*						
41	WP_014184858	*Desulfosporosinus orientis DSM 765*		*	*						
42	WP_084068191	*Desulfobacterium vacuolatum DSM 3385*		*		*					
43	OGR23279	*Desulfobacterales bacterium RIFOXYA12*		*		*					
44	WP_015903462	*Desulfobacterium autotrophicum HRM2*		*		*					
45	WP_028582591	*Desulfobulbus japonicus DSM 18378*		*		*					
46	GBC62389	*Desulfonema ishimotonii Tokyo 01*		*		*					
47	WP_027354447	*Desulfosarcina sp. BuS5*		*		*					
48	WP_022669285	*Desulfospira joergensenii DSM 10085*		*		*					
49	WP_006521446	*Desulfotomaculum gibsoniae DSM 7213*		*	*	*					
50	WP_047828483	*Peptococcaceae bacterium CEB3*		*	*						

#	Accession # BzaC	Organism	AdoCbl riboswitch	bzaA	bzaB	cobT	bzaC	DUF2284	bzaD	bzaE	Predicted lower ligand
51	WP_084413176	*Desulfovirgula thermocuniculi DSM 16036*				*					Not predicted
52	WP_054935549	*Moorella glycerini NMP*				*					
53	WP_099323819	*Can. Kuenenia stuttgartiensis mbr1*				*					
54	OGU16002	*Geobacteraceae bacterium GWC2_53_11*				*					
55	RLC28695	*Deltaproteobacteria bacterium B10_G6 B10*			#					#	
56	OGR25629	*Desulfuromonadales bacterium GWD2_54_10*									
57	OGU04877	*Geobacteraceae bacterium GWC2_55_20*		*		*					
58	RMD79622	*Lentisphaerae bacterium J147*		*		*					
59	WP_020875905	*Desulfococcus multivorans DSM 2059*		*	*	*					
60	WP_071522194	*Moorella thermoacetica 194 DSM 12797*		*	*	*					
61	WP_113919373	*Alkalibaculum bacchi DSM 22112*		*	*	*			#		
62	ATW23833	*Peptococcaceae bacterium DCMF*				*			&		
63	WP_083456563	*Desulfosarcina cetonica*									DMB **5**
64	AEG13911	*Desulfotomaculum kuznetsovii DSM 6115*				*					
65	WP_073163425	*Desulfotomaculum australicum*				*					Not predicted
66	WP_066669469	*Desulfotomaculum copahuensis LMa1*				*					
67	WP_072871322	*Desulfotomaculum thermosubterraneum*				*					
68	WP_121452168	*Desulfofundulus salinum 435*				*					
69	WP_096892661	*Can. Scalindua japonica husup a2*				*					
70	OGW14547	*Nitrospinae bacterium RIFCSPLOWO2*				*					5-OMeBza **3**
71	SPF38744	*Can. Desulfosporosinus infrequens ShF1*				*					Not predicted
72	WP_007221343	*Can. Jettenia caeni*				*					
73	OHC01829	*Planctomycetes bacterium RIFCSPLOWO2*				*					
74	WP_011393222	***Moorella thermoacetica DSM 2955***									5-OMeBza **3**
75	WP_027355811	*Desulfofundulus thermocisternus DSM10259*				*					
76	KJS31251	*Desulfatitalea sp. BRH_c12*									
77	WP_028585410	*Desulfobulbus mediterraneus DSM 13871*									
78	WP_012471693	*Geobacter lovleyi SZ*									
79	SCX77818	*Desulfoluna spongiiphila AA1*									
80	RJP74872	*Desulfobacteraceae bacterium SURF_7*				*					
81	OQX63009	*Desulfococcus sp. 4484_241*				*					
82	WP_008519158	*Dethiobacter alkaliphilus AHT 1*				*					
83	KUO75528	*Desulfosporosinus sp. BRH_c37*				*					
84	OHB74475	*Planctomycetes bacterium RBG_16_41_13*				*					
85	WP_084504898	*Acetobacterium dehalogenans DSM 11527*									DMB **5**
86	WP_026394115	*Acetobacterium dehalogenans DSM 11527*									
87	WP_111889517	*Acetobacterium sp. KB-1*									
88	WP_111888193	*Acetobacterium sp. KB-1*									
89	OXS26005	*Acetobacterium sp. MES1*									
90	OXS27022	*Acetobacterium sp. MES1*									
91	WP_084633842	*Acetobacterium wieringae DSM 1911*									
92	WP_070369651	*Acetobacterium wieringae DSM 1911*									

Key:

*	gene present at a distant location
#	contig end
&	*bza* operon neighborhood not found

References:

Aitken, A, and M Learmonth. 1996. “Protein Determination by UV Absorption.” In Protein Protocols Handbook, edited by JM Walker.

Allen, Kylie D., and Susan C. Wang. 2014. “Initial Characterization of Fom3 from Streptomyces Wedmorensis: The Methyltransferase in Fosfomycin Biosynthesis.” Archives of Biochemistry and Biophysics 543: 67–73. https://doi.org/10.1016/j.abb.2013.12.004.

Bradford, MM. 1976. “A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding.” Anal. Biochem. 72: 248–254.

Bridwell-Rabb, Jennifer, Bin Li, and Catherine L. Drennan. 2022. “ Cobalamin-Dependent Radical S -Adenosylmethionine Enzymes: Capitalizing on Old Motifs for New Functions .” ACS Bio & Med Chem Au 2 (3): 173–86. https://doi.org/10.1021/acsbiomedchemau.1c00051.

Bridwell-Rabb, Jennifer, Aoshu Zhong, He G. Sun, Catherine L. Drennan, and Hung Wen Liu. 2017. “A B 12 -Dependent Radical SAM Enzyme Involved in Oxetanocin A Biosynthesis.” Nature 544 (7650): 322–26. https://doi.org/10.1038/nature21689.

Broderick, Joan B., Randall E. Duderstadt, Daniel C. Fernandez, Kristi Wojtuszewski, Timothy F. Henshaw, and Michael K. Johnson. 1997. “Pyruvate Formate-Lyase Activating Enzyme Is an Iron-Sulfur Protein.” Journal of the American Chemical Society 119 (31): 7396–97. https://doi.org/10.1021/ja9711425.

Chatterjee, Abhishek, Yue Li, Yang Zhang, Tyler L. Grove, Michael Lee, Carsten Krebs, J Squire, et al. 2008. “Reconstitution of ThiC in Thiamine Pyrimidine Biosynthesis Expands the Radical SAM Superfamily.” Nature Chemical Biology 4 (12): 1–19. https://doi.org/10.1038/nchembio.121.

Crofts, Terence S., Amrita B. Hazra, Jennifer LA Tran, Olga M. Sokolovskaya, Vadim Osadchiy, Omer Ad, Jeffrey Pelton, Stefan Bauer, and Michiko E. Taga. 2014. “Regiospecific Formation of Cobamide Isomers Is Directed by CobT.” Biochemistry 53 (49): 7805–15. https://doi.org/10.1021/bi501147d.

Deery, Evelyne, Andrew D. Lawrence, and Martin J. Warren. 2022. “Biosynthesis of Cobamides: Methods for the Detection, Analysis and Production of Cobamides and Biosynthetic Intermediates.” In Methods in Enzymology. Academic Press. https://doi.org/10.1016/BS.MIE.2022.01.013.

Edgar, Robert C. 2004. “MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput.” Nucleic Acids Research 32 (5): 1792–97. https://doi.org/10.1093/nar/gkh340.

Fyfe, Cameron D., Noelia Bernardo-García, Laura Fradale, Stéphane Grimaldi, Alain Guillot, Clémence Brewee, Leonard M.G. Chavas, Pierre Legrand, Alhosna Benjdia, and Olivier Berteau. 2022. “Crystallographic Snapshots of a B12-Dependent Radical SAM Methyltransferase.” Nature 602 (7896): 336–42. https://doi.org/10.1038/s41586-021-04355-9.

Goodacre, Norman F, Dietlind L Gerloff, and Peter Uetz. 2014. “Protein Domains of Unknown Function Are Essential in Bacteria.” MBio 5 (1): e00744-13. https://doi.org/10.1128/mBio.00744-13.Editor.

Hall, T.A. 1999. “BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT.” Nucleic Acids Symposium 41: 95–98. https://doi.org/citeulike-article-id:691774.

Hazra, Amrita B., Andrew W. Han, Angad P. Mehta, Kenny C. Mok, Vadim Osadchiy, Tadhg P. Begley, and Michiko E. Taga. 2015. “Anaerobic Biosynthesis of the Lower Ligand of

Vitamin B12." Proceedings of the National Academy of Sciences of the United States of America 112 (34): 10792–97. https://doi.org/10.1073/pnas.1509132112.

Johnson, Deborah C., Dennis R. Dean, Archer D. Smith, and Michael K. Johnson. 2005. "Structure, Function, and Formation of Biological Iron-Sulfur Clusters." Annual Review of Biochemistry 74 (1): 247–81. https://doi.org/10.1146/annurev.biochem.74.082803.133518.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kim, Hak Joong, Reid M. McCarty, Yasushi Ogasawara, Yung Nan Liu, Steven O. Mansoorabadi, Jake Levieux, and Hung Wen Liu. 2013. "GenK-Catalyzed C-6' Methylation in the Biosynthesis of Gentamicin: Isolation and Characterization of a Cobalamin-Dependent Radical SAM Enzyme." Journal of the American Chemical Society 135 (22): 8093–96. https://doi.org/10.1021/ja312641f.

Kim, Jin Yong, Joo Won Suh, Suk Ho Kang, Thi Huyen Phan, Si Hyung Park, and Hyung Jin Kwon. 2008. "Gene Inactivation Study of GntE Reveals Its Role in the First Step of Pseudotrisaccharide Modifications in Gentamicin Biosynthesis." Biochemical and Biophysical Research Communications 372 (4): 730–34. https://doi.org/10.1016/j.bbrc.2008.05.133.

Knox, Hayley L., Percival Yang Ting Chen, Anthony J. Blaszczyk, Arnab Mukherjee, Tyler L. Grove, Erica L. Schwalm, Bo Wang, Catherine L. Drennan, and Squire J. Booker. 2021. "Structural Basis for Non-Radical Catalysis by TsrM, a Radical SAM Methylase." Nature Chemical Biology 17 (4): 485–91. https://doi.org/10.1038/s41589-020-00717-y.

Knox, Hayley L., Erica K. Sinner, Craig A. Townsend, Amie K. Boal, and Squire J. Booker. 2022. "Structure of a B12-Dependent Radical SAM Enzyme in Carbapenem Biosynthesis." Nature 602 (7896): 343–48. https://doi.org/10.1038/s41586-021-04392-4.

Letunic, Ivica, and Peer Bork. 2016. "Interactive Tree of Life (ITOL) v3: An Online Tool for the Display and Annotation of Phylogenetic and Other Trees." Nucleic Acids Research 44 (W1): W242–45. https://doi.org/10.1093/nar/gkw290.

Miller, Mark A., Wayne Pfeiffer, and Terri Schwartz. 2010. "Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees." 2010 Gateway Computing Environments Workshop, GCE 2010.

Pierre, Stéphane, Alain Guillot, Alhosna Benjdia, Corine Sandström, Philippe Langella, and Olivier Berteau. 2012. "Thiostrepton Tryptophan Methyltransferase Expands the Chemistry of Radical SAM Enzymes." Nature Chemical Biology 8 (12): 957–59. https://doi.org/10.1038/nchembio.1091.

Seravalli, Javier, Stephen W Ragsdale, Dennis Dean, and Patricia Dossantos. n.d. "Pulse Chase Studies of the Synthesis of Acetyl-Coenzyme a By Carbon," 5–6.

Shelton, Amanda N., Erica C. Seth, Kenny C. Mok, Andrew W. Han, Samantha N. Jackson, David R. Haft, and Michiko E. Taga. 2018. "Uneven Distribution of Cobamide Biosynthesis and Dependence in Bacteria Predicted by Comparative Genomics." ISME Journal 13 (3): 789–804. https://doi.org/10.1038/s41396-018-0304-9.

Sinner, Erica K., Rongfeng Li, Daniel R. Marous, and Craig A. Townsend. 2022. "ThnL, a B12-Dependent Radical S-Adenosylmethionine Enzyme, Catalyzes Thioether Bond Formation in Carbapenem Biosynthesis." Proceedings of the National Academy of Sciences 119 (34): e2206494119.

Stamatakis, Alexandros. 2014. "RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies." Bioinformatics 30 (9): 1312–13. https://doi.org/10.1093/bioinformatics/btu033.

Towns, John, Tim Cockerill, Maytal Dahan, Ian Foster, Kelly Gaither, Andrew Grimshaw, Victor Hazlewood, et al. 2014. “XSEDE: Accelerating Scientific Discovery.” Computing in Science and Engineering 16 (October): 62–74. https://doi.org/10.1109/MCSE.2014.80.

Wang, Bo, Anthony J. Blaszczyk, Hayley L. Knox, Shengbin Zhou, Elizabeth J. Blaesi, Carsten Krebs, Roy X. Wang, and Squire J. Booker. 2018. “Stereochemical and Mechanistic Investigation of the Reaction Catalyzed by Fom3 from Streptomyces Fradiae, a Cobalamin-Dependent Radical S-Adenosylmethionine Methylase.” Biochemistry 57 (33): 4972–84. https://doi.org/10.1021/acs.biochem.8b00693.

Werner, Williard J., Kylie D. Allen, Kaifeng Hu, Gregory L. Helms, Brian S. Chen, and Susan C. Wang. 2011. “In Vitro Phosphinate Methylation by PhpK from Kitasatospora Phosalacinea.” Biochemistry 50 (42): 8986–88. https://doi.org/10.1021/bi201220r.

Zhong, Aoshu, Yu Hsuan Lee, Yung Nan Liu, and Hung Wen Liu. 2021. “Biosynthesis of Oxetanocin-A Includes a B12-Dependent Radical SAM Enzyme That Can Catalyze Both Oxidative Ring Contraction and the Demethylation of SAM.” Biochemistry 60 (7): 537–46. https://doi.org/10.1021/acs.biochem.0c00915.

Table I: List of Plasmids created and used in the duration of the thesis

S.No.	Plasmid name	Plasmid backbone	Insert source/ name	Restriction sites	Notes
1	pYMH001	pET28a	*Eubacterium limosum, bzaC*	NdeI and BamHI	None
2	pYMH002	pET28a	*Moorella thermoacetica, bzaC*	NdeI and BamHI	None
3	pYMH003	pET28a	*Moorella thermoacetica, bzaC*	NdeI and BamHI	Insertion of DUF2284 domain
4	pYMH004	pET28a	*Eubacterium limosum, bzaC*	NdeI and BamHI	deletion of DUF2284 domain
5	pYMH005	pET28a	*Eubacterium limosum, bzaD*	NdeI and BamHI	None
6	pYMH006	pET28a	*Eubacterium limosum, bzaE*	NdeI and BamHI	None
7	pYMH007	pET28a	*Eubacterium limosum, bzaC*	NdeI and BamHI	Cloned only the DUF2284 domain
8	pYMH008	pET28a	*Eubacterium barkeri, bzaC*	NdeI and BamHI	None
9	pYMH009	pET28a	*Eubacterium barkeri, bzaD*	NdeI and BamHI	None
10	pYMH010	pET28a	*Eubacterium barkeri, bzaE*	NdeI and BamHI	None
11	pMBS011	pET28a	*Eubacterium limosum, cobT-bzaC*	NdeI and BamHI	None
12	pMBS012	pET28a	*Moorella thermoacetica, cobT-bzaC*	NdeI and BamHI	None
13	pMBS013	pET15b	*Escherichia coli K-12 MG1655, metK*	NdeI and XhoI	None
14	pMJ1208-1	pET19b	*Methanocaldococcus jannaschii DSM 2661, mat*	NdeI and BamHI	None
15	pYMH014	pET28a	*Moorella thermoacetica, bzaC*	NdeI and BamHI	deletion of dimerization domain
16	pYMH015	pET28a	*E.coli, mtnN*	NdeI and BamHI	methylthioadenosie nucleosidase from *E.coli*
17	pPMD016	pET28a	*E.coli, cobT*	NdeI and BamHI	
18	pPMD017	pET28a	*Eubacterium limosum, cobT*	NdeI and BamHI	
19	pSSH018	pET28a	*Moorella thermoacetica, cobT*	NdeI and BamHI	
20	pPMD1	pET28a	*Eubacterium limosum, cobT*	NdeI and BamHI	
21	pPMD2	pET28a	*Eubacterium limosum, cobS*	NdeI and BamHI	
22	pPMD3	pET28a	*Eubacterium limosum, cobU*	NdeI and BamHI	

23	pPMD4	pET28a	*Escherichia coli, cobT*	NdeI and BamHI	
24	pPMD5	pET28a	*Escherichia coli, cobU*	NdeI and BamHI	
25	pPMD6	pET28a	*Escherichia coli, cobC*	NdeI and BamHI	
26	pPMD11	pET28a	*Salmonella enterica, cobT*	NdeI and BamHI	
27	pSSH019	pET28a	*Desulfotomaculum thermosubterraneum , bzaC*	NdeI and BamHI	
28	pSSH020	pET28a	*Desulfotomaculum thermosubterraneum , cobT*	NdeI and BamHI	
29	pSSH021	pET28a	*Desulfotomaculum thermosubterraneum , bzaE2*	NdeI and BamHI	
30	pSSH022	pET28a	*Desulfotomaculum thermosubterraneum , bzaX*	NdeI and BamHI	
31	pSSH023	pET28a	*Desulfobulbous mediterraneous, cobT1*	NdeI and BamHI	
32	pSSH024	pET28a	*Desulfobulbous mediterraneous, cobT2*	NdeI and BamHI	
33	pSSH025	pET28a	*Desulfobulbous mediterraneous, cobT3*	NdeI and BamHI	
34	pYMH026	pET28a	*Eubacterium barkeri, cobT1*	NdeI and BamHI	
35	pYMH027	pET28a	*Eubacterium barkeri, cobT3*	NdeI and BamHI	
36	pYMH028	pET28a	*Aphanomyces astaci*	NdeI and BamHI	
37	pYMH029	pET28a	*Aphanomyces invadans*	NdeI and BamHI	
38	pYMH030	pET28a	*Chytriomyces confervae*	NdeI and BamHI	
39	pAVH029	pET28a	*Dunaliella salina*	NdeI and BamHI	
40	pYMH033	pET28a	*Desulfobulbous mediterraneous, bzaC*	NdeI and BamHI	
41	pYMH034	pET-Duet	*Eubacterium barkeri, bzaC*	BamHI and NotI	*bzaC* inserted in MCS site-1
42	pYMH035	pET-Duet	*Eubacterium barkeri, bzaD*	BamHI and NotI	*bzaD* inserted in MCS site-1
43	pYMH036	pET-Duet	*Eubacterium barkeri, bzaD*	NdeI and XhoI	*bzaD* inserted in MCS site-2
44	pYMH037	pET-Duet	*Eubacterium barkeri, bzaE*	NdeI and XhoI	*bzaE* inserted in MCS site-2

45	pVSGH038	pCA24N_2	*Eubacterium barkeri, cobT1*	NdeI and BamHI	For expression in *E.coli* MG1655 *ΔcobT ΔmetE*
46	pVSGH039	pCA24N_2	*Eubacterium barkeri, cobT3*	NdeI and BamHI	For expression in *E.coli* MG1655 *ΔcobT ΔmetE*
47	pVSGH040	pCA24N_2	*Desulfobulbous mediterraneous, cobT1*	NdeI and BamHI	For expression in *E.coli* MG1655 *ΔcobT ΔmetE*
48	pVSGH041	pCA24N_2	*Desulfobulbous mediterraneous, cobT2*	NdeI and BamHI	For expression in *E.coli* MG1655 *ΔcobT ΔmetE*
49	pVSGH042	pCA24N_2	*Desulfobulbous mediterraneous, cobT3*	NdeI and BamHI	For expression in *E.coli* MG1655 *ΔcobT ΔmetE*
50	pYMH043	pTH1227	*Desulfobulbous mediterraneous, bza operon*	KpnI and SphI	For expression in *E.coli* MG1655 *ΔcobT ΔmetE*
51	pYMH044	pTH1227	*Desulfotomaculum thermosubterraneus, bza operon*	KpnI and SphI	For expression in *E.coli* MG1655 *ΔcobT ΔmetE*
52	pYMH045	pET28a	*Eubacterium limosum, cobT-bzaC-bzaD-bzaE*	NdeI and BamHI	