Probing the mechanism and genomic context of CobT, the lower ligand activating enzyme in vitamin B_{12} biosynthesis

A Thesis

submitted to
Indian Institute of Science Education and Research Pune
in partial fulfilment of the requirements for the
BS-MS Dual Degree Programme
by
Sheryl Sreyas

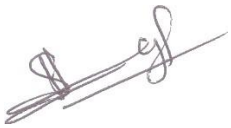
Indian Institute of Science Education and Research Pune
Dr. Homi Bhabha Road, Pashan, Pune 411008, INDIA.

March, 2020

Supervisor: Dr. Amrita Hazra

Certificate

This is to certify that this dissertation entitled "**Probing the mechanism and genomic context of CobT, the lower ligand activating enzyme in vitamin B_{12} biosynthesis**" towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Sheryl Sreyas at Indian Institute of Science Education and Research under the supervision of Dr. Amrita Hazra, Assistant Professor, Department of Chemistry, during the academic year 2019-2020.



Sheryl Sreyas
(IISER Pune)
20151015

Dr. Amrita Hazra
(IISER Pune)
Supervisor

Declaration

I hereby declare that the matter embodied in the report entitled "**Probing the mechanism and genomic context of CobT, the lower ligand activating enzyme in vitamin B_{12} biosynthesis**" are the results of the work carried out by me at the Department of Chemistry, Indian Institute of Science Education and Research, Pune, under the supervision of Dr. Amrita Hazra and the same has not been submitted elsewhere for any other degree.

Sheryl Sreyas
(IISER Pune)
20151025

Dr. Amrita Hazra
(IISER Pune)
Supervisor

Acknowledgement

The undertaking and completion of this thesis would have been impossible without the help of a number of people. First of all, I would like to thank my supervisor Dr. Amrita Hazra, who was with me at every single step giving me scientific guidance as well as moral support. Her keen insights and immense knowledge certainly have helped the project build up to become what it is at present. I would also like to thank her funding agencies for the financial support received to carry out the project.

I am extremely grateful for my lab members for their valuable inputs as well as for providing any help required. I am especially thankful to Yamini, who helped in each step of the project without who this project would not have reached where it is. She has provided me immense support and guidance and kept me going forward for which I would be eternally grateful. I also would like to thank Rupali, Yashwant, Ateek and Saswata for being such good colleagues and providing a supportive lab environment.

I also take this opportunity to turn to thank Dr. Nishad Matange for the helpful insights into the evolutionary and biological perspective and the engaging discussions which has helped me look at the project through a different perspective.

I would like to thank the LC-MS facility and NMR facility at IISER Pune for their technical support. I also thank IISER Pune Chemistry and Biology departments for providing the facilities required to undertake this project.

The completion of this project would not have been possible without the financial support from KVPY and IISER Pune.

Last but not least I would like to thank my parents and my brother for the immense moral support that kept me motivated throughout the duration of the project. I would also like to thank my friends and teachers for being with me throughout this journey.

Contents

List of figures and tables

Abstract

Vitamin B_{12} is a cobamide cofactor required by many organisms including humans for catalysis of various reactions. It is a structurally complex small molecule that consists of a tetrapyrrolic corrin ring, an upper ligand and a lower ligand, 5,6-dimethylbenzimidazole (DMB) attached via a nucleotide loop. NaMN: DMB phosphoribosyl transferase CobT is the enzyme that activates the lower ligand prior to its attachment. In this study we aim to understand the regiospecificity and substrate preference of CobT homologs from obligate anaerobes that carry out de novo biosynthesis of lower ligand of vitamin B_{12}. First based on gene neighbourhood of the encoding cobT gene, we classified the CobT homologs into three types, CobT1, CobT2 and CobT3. Then we reconstituted the activity of representative CobT1, CobT2 and CobT3 homologs and proved that only CobT1 homologs carry out a regiospecific activation of 5-OHBza. We studied the promiscuity in the substrate preference of the CobT homologs by characterizing their reactions with other naturally occurring lower ligands and proved that they are indeed catalytically promiscuous. Lastly, we studied the substrate preference of a representative CobT1 and CobT3 homolog proving that the most preferred substrate of CobT1 homolog is 5-OHBza whereas for the CobT3 homolog, it is DMB. In conclusion we establish that CobT1 homologs from obligate anaerobes differ vastly in their regiospecificity and substrate preference when compared to other CobT homologs. We also establish that an organism may possess multiple functional CobT enzymes which differ in activity. We also conclude that regardless of gene neighbourhood, CobT enzymes show promiscuity in their activity and can activate a variety of lower ligands.

1.Introduction

1.1 Vitamin B_{12} is a cobamide cofactor

Vitamin B_{12}, also known as cobalamin (Cbl) is an essential micronutrient that is required by many organisms including humans for metabolism. It belongs to the family of cobamide cofactors and is involved in the catalysis of a range of reactions like elimination, rearrangements, reductive dehalogenations and methyl group transfer reactions in primary metabolism[1,2].Cobamides are the largest non-polymeric biomolecules that consist of a tetrapyrrolic corrin ring that harbours a cobalt ion at the centre which coordinates with an upper axial ligand, and a lower axial ligand which is also attached to corrin ring via a nucleotide loop[3](Figure 1A).

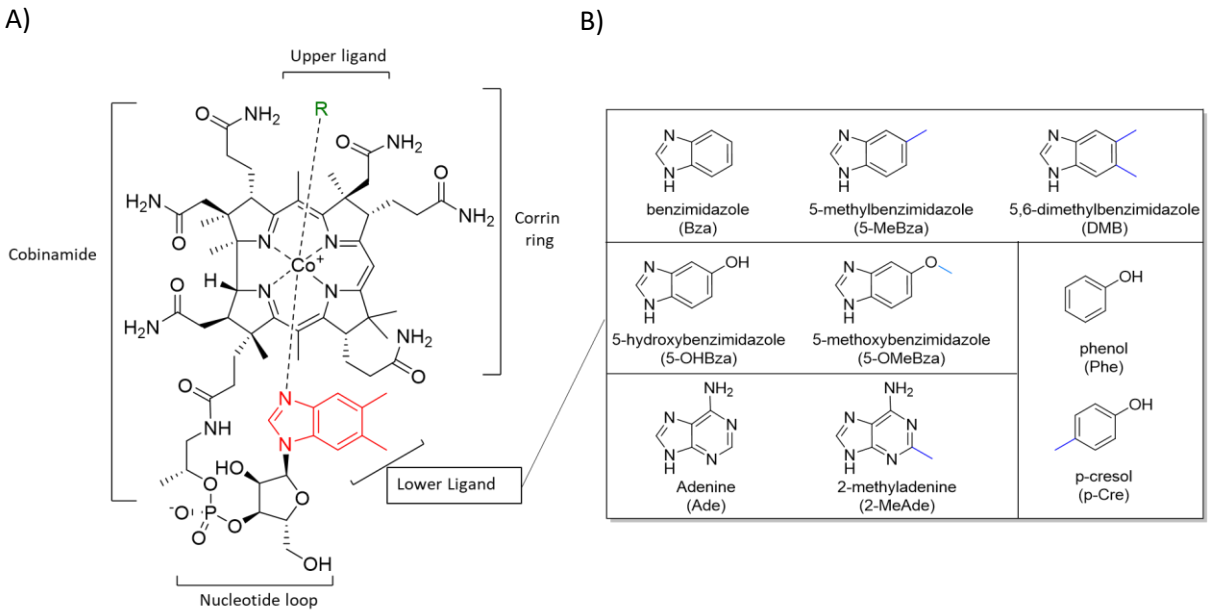


Figure 1: Cobamide structure and possible lower ligands

A) Structure of Vitamin B_{12} – consisting of the corrin ring harbouring the cobalt ion, the upper ligand responsible for the catalytic activity, and the lower ligand DMB attached to the corrin ring via the nucleotide loop. B) the various possible lower ligand of cobamides, which includes benzimidazoles, purines and phenols.

In the active form of the co-factor, the upper ligand that confers its reactivity can be either a 5'-deoxyadenosyl group (AdoCbl) or a methyl group (MeCbl). AdoCbl cofactor is mainly involved in radical rearrangement reactions (methyl-malonyl CoA mutase) and reduction reactions (ribonucleotide reductase). MeCbl is associated with transmethylation reactions (methionine synthase) which are part of processes like acetogenesis, methanogenesis and DNA methylations[1,2,4].

The lower ligand of the cobamides cobamide cofactors is one of the factors leading to cobamide diversity. Cobamides can incorporate structurally diverse small molecules like benzimidazoles, purines and phenols as the lower ligand[5,6](Figure 1B). Humans can solely utilize Vitamin B12 which is the cobamide characterised by 5,6-dimethylbenzimidazole (DMB) as the lower ligand[7] whereas other organisms including bacteria and algae have been found to use other cobamides as cofactors for reactions. For instance the eukaryotic algae and cyanobacteria have been found to use pseudocobalamin which is the cobamide with adenine as the lower ligand[8].

The lower ligand nucleotide can be coordinated to the cobalt ion (base-on) or dissociated (base-off) when the cobamide is bound to the enzyme. Usually in the base-off form of the cobamide a histidine from the protein is coordinated.

1.2 Biosynthesis of cobamides

The *de novo* biosynthesis of cobamide cofactors is restricted to a subset of prokaryotes and can occur via either an aerobic or an anaerobic pathway[9,10]. The biosynthesis of this structurally complex small molecule is equally elaborate and intricate.

The biosynthetic pathway of cobamides involve more than 30 enzymes for the synthesis and assembly of the structural components namely, the corrin ring and the lower ligand. It proceeds through independent synthesis of these components – the tetrapyrrolic corrin ring containing a cobalt ion is assembled to form the corrinoid, following which the lower axial ligand is synthesised and attached to the ring via a nucleotide loop[11] (Figure 1A). Aside from this some organisms also synthesise cobamides through the salvage pathway, where exogenous components are

transported into the cell via specialised transporter proteins and assembled together or remodelled[12].

There are two independent biosynthetic pathways described in literature, one that requires oxygen (aerobic) and the other which is sensitive to oxygen (anaerobic). These two pathways also differ in the timing of the cobalt ion insertion. The aerobic pathway requires molecular oxygen to facilitate ring contraction as well as for the lower ligand biosynthesis whereas the anaerobic pathway does not[11,13,14].

1.3 Lower ligand biosynthesis

While the overall pathway for synthesis of corrin ring remains similar in the two pathways, the lower ligand biosynthesis is very distinct in the oxygen dependent and oxygen sensitive pathways[11,13–15]. The aerobic biosynthesis of DMB is through the breakdown and rearrangement of reduced flavin mononucleotide and is catalysed by the enzyme, flavin destructase BluB which requires molecular oxygen for the reaction[16–18] (Figure 2A).

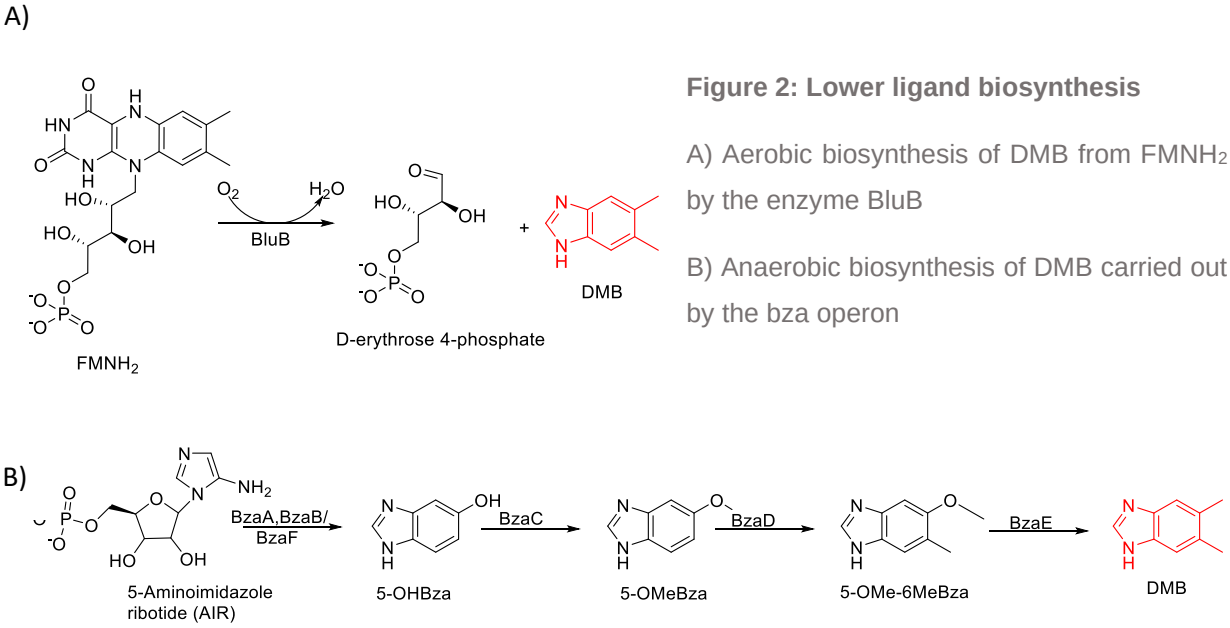


Figure 2: Lower ligand biosynthesis

A) Aerobic biosynthesis of DMB from $FMNH_2$ by the enzyme BluB

B) Anaerobic biosynthesis of DMB carried out by the bza operon

In the anaerobic biosynthesis of DMB, the gene products of the *bza* operon which contains the genes *bzaA, bzaB, bzaC, bzaD* and *bzaE* are responsible for the

conversion of the purine biosynthesis intermediate 5-aminoimidazole ribotide (AIR) into DMB[19]. In many organisms, *cobT* gene is also present within the operon. The significance of the CobT enzyme in the lower ligand biosynthesis is yet to be investigated. In some organisms *bzaA* and *bzaB* are replaced by a single gene homolog, *bzaF*. The pathway proceeds via the formation of three benzimidazole intermediates. BzaF, the gene product of *bzaF* has been shown to convert AIR into 5-OHBza, the first intermediate in the pathway[18,20]. Subsequent methylations of 5-OHBza is carried out to form DMB. Recent study in our lab has characterized BzaC, a SAM-dependent O-methyltransferase (the gene product of *bzaC*) which converts 5-OHBza into 5-OMeBza. The gene product of *bzaD* and *bzaE* are predicted to be methyl transferases that convert 5-OMeBza to 5-OMe-6-MeBza which is then converted to DMB[19] (Figure 2B).

It is interesting to note that these benzimidazole intermediates are found as naturally occurring lower ligands in cobamides. Recent studies have shown that the modularity of the bza operon may be one of the factors responsible for the occurrence of other benzimidazole derivatives as lower ligands. It has been shown that the lower ligands of cobamides produced by obligate anaerobes which possess bza operons can be corelated to the presence or absence of one or more of the three bza methyltransferases[10,19]. For instance, *Geobacter sulfurreducens* contain only *bzaF* and produces 5-hydroxybenzimidazoyl-cobamide[19], *Moorella thermoacetica* contains *bzaA*, *bzaB* and *bzaC* and produces 5-methoxybenzimidazoyl-cobamide[21,22] and *Eubacterium limosum* contains all the bza genes and produces cobalamin[5,19].

1.4 Nucleotide loop assembly pathway

The last step in the cobamide biosynthesis pathway is the assembly of the nucleotide loop attaching the lower ligand base to the corrinoid. There are four enzymes involved in this step namely, 1) Bifunctional adenosylcobalamin biosynthesis protein (CobU), 2) NaMN:DMB phosphoribosyltransferase (CobT), 3) Adenosylcobinamide-GDP ribazoletransferase (CobS) and 4) Adenosylcobalamin/ alpha-ribazole phosphatase (CobC)[11,12,23,24].

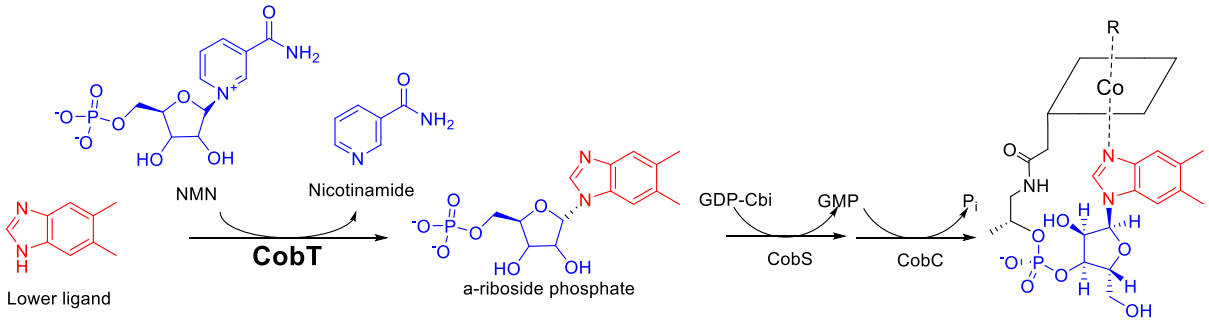


Figure 3: Lower ligand activation and attachment

CobT enzyme activates the lower ligand into an α-ribotide by the transfer of a phosphoribosyl moiety. CobS enzyme further condenses GDP-Cbi which is the product of CobU enzyme with the α-ribotide to form Cba-P and CobC enzyme removes the 5'-phosphate to give cobamide.

CobU converts the corrinoid intermediate adenosylcobinamide-phosphate (AdoCbi-P) to adenosylcobinamide-GDP (AdoCbi-GDP). The phosphoribosyltransferase CobT activates the free lower ligand by transferring a phosphoribosyl moiety from the cofactor to it to form an α-ribotide. CobS enzyme is responsible for the condensation of AdoCbi-GDP and the α-ribotide to form Adenosylcobalamin-5'-phosphate (AdoCba-P). The final step of the removal of the 5' phosphate of AdoCba-P to generate adenosylcobamide (AdoCba) is carried out by the CobC enzyme[23].

1.5 Promiscuity of the CobT enzyme

The phosphoribosyltransferase CobT activates the lower ligand prior to its attachment to form the cobamide. It is a fascinating enzyme as it creates a unique α–glycosidic bond between the nitrogen atom of the lower ligand and C1' of the ribose ring of the phosphoribosyl moiety[11,25,26]. Bacterial CobT enzymes have been shown to be promiscuous in terms of substrate specificity[27,28] as well as regiospecificity[29]. Though most of the work has been done on DMB as the lower ligand, numerous organisms possess purines, other benzimidazole derivatives and phenols as lower ligands[5,6]. CobT from *Salmonella enterica* has been shown to produce pseudocobalamin, with adenine as lower ligand[30] though it can make cobalamin when supplied with exogenous DMB in guided biosynthesis experiments[31]. It is hypothesized that the specialization of the CobT homolog may depend on the availability of the lower ligand[29] in its natural environment as well as the requirement by the organism[32]. In certain

cases, a cobamide with a particular lower ligand may be beneficial to an organism but production of a different cobamide may be detrimental to its survival. *Sulfurospirillum multivorans* requires nor-pseudocobalamin with adenine as the lower ligand for the reductive dehalogenase activity which is disrupted if nor-cobalamin with DMB as the lower ligand is produced[33].

Aside from the substrate promiscuity, promiscuity in terms of regioselectivity in the attachment of the lower ligand also leads to cobamide diversity[29]. Among the naturally occurring lower ligands some are symmetric and others asymmetric. The two nitrogen atoms of the imidazole ring in the asymmetric lower ligands are not equivalent. These lower ligands when activated can give two isomers as products depending on which nitrogen the phosphoribosyl moiety is transferred to. Some of the CobT homologs characterized in literature show catalytic promiscuity in terms of activating the asymmetric benzimidazoles as they can give both the isomeric products in varying ratios. Depending on the isomer formed two distinct constitutional isomers of the cobamide can be formed. The two constitutional isomers of a cobamide can affect the growth of the organism differently. Studies have shown that 6-methoxybenzimidazlyl-cobamide supports the growth of salmonella enterica better than 5-methoxybenzimidazoyl-cobamide[29].

Native regiospecificity and substrate preference of a CobT enzyme is an important part of its function. Promiscuity in these factors may positively affect the physiology or may be detrimental to the organism owing to the production of cobamides that cannot be used.

1.6 Project goals and scope

1.6.1. Regiospecificity of lower ligand attachment

The labelling studies carried out by Renz and others in the 1980s determined the origin of the each of the atoms in the lower ligand DMB produced by the obligate anaerobe *Eubacterium limosum*. It was reported that the two nitrogen atoms in the imidazole ring were derived from L-glutamine and L-glycine. It was surprising to see that the N atom derived from L-glycine was the one almost exclusively attached (more than 90%) to the ribose forming the a-glycosidic bond.[14,15,21,34] This is unexpected because the

two nitrogen atoms of the imidazole ring in DMB are equivalent due to resonance and hence has equal probability of forming the glycosidic bond. A possible hypothesis an alternate biosynthetic pathway where one of the asymmetric intermediates in the *bza* operon pathway namely, 5-OHBza, 5-OMeBza or 5-OMe-6-MeBza acts as the substrate for CobT and gets activated. Further methylations occur on the activated substrate[19]. Previous work in the lab characterized the CobT enzyme from the *bza* operon of *Eubacterium limosum* and proved that it carries out regiospecific activation of 5-OHBza but not of 5-OMeBza. The study undertaken in this thesis aims to understand if the CobT enzymes from the *bza* operon of other organisms also show a similar result in order to give further evidences for the possibility of the alternate biosynthetic pathway.

1.6.2 Classification of CobT homologs based on their gene neighbourhoods.

Previous preliminary data showed that *Eubacterium limosum* and *Moorella thermoacetica* possess two *cobT* genes in located in different gene neighbourhoods, one in the *bza* operon and one elsewhere. So, we study if there are other organisms possessing more than one *cobT* gene and to analyse their gene neighbourhoods to classify them.

1.6.3 The CobT homologs with different gene neighbourhoods may show difference in activity

The CobT homolog from the *bza* operon may be involved in the attachment of the lower ligand produced by the bza genes and the other CobT homolog may be involved in the attachment of lower ligands taken up from the environment. So, the regioselectivity and substrate preferences of the two CobT homologs may differ. So, in this thesis we also investigate the differences in the enzymatic activity of the different CobT homologs from the same organism.

1.6.4 CobT homologs from bza operon show difference in substrate preference

The CobT enzymes characterized in literature belong to either aerobes, facultative anaerobes or anaerobes that do not possess *bza* operon. Previous work in the lab that characterized CobT from the obligate anaerobe *Eubacterium limosum* which is reported to produce vitamin B_{12} shows that it carries out regiospecific activation of 5-OHBza but not 5-OMeBza. This observation shows that the CobT homolog is different from others already characterized in literature which do not show this regiospecificity. This led us to hypothesize that CobT enzymes from obligate anaerobes that contain bza operon may differ from CobT homologs of other organisms.

This difference may also be seen in their substrate preference. Previous studies have established that CobT enzymes regardless of the organism it is derived from show a preference for DMB[35]. These CobT homologs were from organisms among which there is none that possess *bza* operon or none that is reported to produce other benzimidazole lower ligands. We wish to study if this prevalent preference for DMB holds true in case of CobT homologs from obligate anaerobes possessing bza operon, namely *Moorella thermoacetica* and *Eubacterium barkeri*. *M*. *thermoacetica* is reported to produce 5-methoxybenzimidazoyl-cobamide and *E*. *barkeri* is predicted to produce cobalamin based on the bza operon architecture. Also based on our previous hypothesis that 5-OHBza may be the substrate for CobT in the anaerobic lower ligand biosynthesis leading to regiospecific lower ligand attachment, we wanted to see if the CobT1 homologs have an increased preference for 5-OHBza. We undertake a study to determine if these CobT homologs have a preference for the lower ligands they are predicted to produce.

1.6.5 Overall approach and scope

In this thesis, we characterise the in vitro activity of the various CobT homologs with different lower ligands to understand the regiospecificity of activation of these lower ligands with the enzymes. We carry out pairwise competition assays of different combinations of lower ligands to determine the substrate preference of the various CobT enzymes. We also undertake bioinformatic analysis of the protein sequences of the CobT homologs to better understand their difference or similarity in activity.

The study undertaken would help in understanding the anaerobic lower ligand biosynthetic pathway and the significant role of CobT in the same. Identifying the regiospecificity and substrate preferences of CobT enzymes can help in further study of the molecular basis of these. This can also aid in production of diverse cobamides with alternate lower ligands. Furthermore, understanding the activity of CobT homologs can help in deciphering the details of the cobamide diversity and cobamide exchange in microbial communities.

2. Materials and Methods

2.1 Materials

2.1.1 Chemicals: All media components and antibiotics for bacterial culture were obtained from HiMedia and TCI, respectively. All other chemicals were purchased from Sigma Aldrich unless otherwise specified. HPLC solvents were obtained from SD Fine Chemicals, and LC-MS grade methanol was procured from Sigma.

2.1.2 Reagents: Enzymes used for cloning were purchased from Takara and NEB. DNA and plasmid purification kits were obtained from Agilent and Qiagen.

2.1.3 Genomic DNA and plasmids: The plasmids harbouring *Se*CobT, was created by Dr. Amrita Hazra and Dr. Terence Crofts as described in previous literature[35]. Genomic DNA of *Eubacterium barkeri* was a gift from the Prof. Michi Taga at University of California, Berkeley. The genomic DNA for *Desulfobulbus mediterreans* was obtained from the RIKEN BRC through the National Bio-Resources Project of the MEXT/AMED, Japan.

2.2 Experimental methods

2.2.1 Molecular cloning for genes of interest

Genes for *Mt*CobT1, *El*CobT1 were amplified from the plasmids pKM076 and pKM077 containing the *bza* operons from *Moorella thermoacetica* and *Eubacterium limosum* respectively in the pTH1227 vector[19]. The other CobT homologs namely *El*CobT3, *Eb*CobT1, *Eb*CobT3 were amplified from the genomic DNA of the respective organisms. The genes were inserted into the desired expression vector pET28a by restriction free (RF) cloning[36]. Briefly, the gene of interest was amplified from source gDNA or plasmid by PCR and then a megaprimer was created by adding ~22-25 base pairs of flanking regions that were complementary to the expression vector were added to 3' and 5' ends of the gene. A PCR with the megaprimer and the empty

pET28a vector resulted in the insertion of the gene into the vector. The PCR product was digested by DpnI and transformed into chemically competent *E. coli* DH5a cells and resulting transformant colonies were selected on kanamycin resistance and screened for gene insertion using standard colony PCR protocols. Plasmids were extracted from the positive colonies and confirmed by DNA sequencing.

2.2.2 Protein overexpression and purification

The extracted plasmids were transformed into chemically competent *E. coli* BL(21)DE3 cells for protein overexpression. The transformants were inoculated in 10 ml LB broth and incubated at 37^{o}C for the primary culture. When the culture reached an OD_{600} of 0.6, 5 ml of it was added to a 1 L LB broth. At OD_{600} of 0.6, protein expression was induced using 0.5 mM IPTG and incubated at 25^{o}C for 18 hrs. The cells were harvested by centrifugation at 6000 rpm at 4^{o}C for 10 minutes and cell pellets were stored at −80 °C until purification.

For protein purification, the cell pellets were resuspended in lysis buffer (50 mM phosphate, 300 mM NaCl, 5 mM imidazole, pH 8, 100 µM PMSF). The resuspended cells were lysed by sonication (60% amplitude, 1s on 3s off cycle, 5 min). The lysate was centrifuged at 18000 rpm, 4^{o}C for 20 minutes and the clarified lysate was loaded onto a pre-equilibrated GE-HiTrap 5ml Ni-NTA column. A 50 ml wash with wash buffer (50 mM phosphate, 300 mM NaCl, 50 mM imidazole, pH 8) was given to remove unwanted proteins and the protein of interest was eluted in a buffer containing 50 mM phosphate, 300 mM NaCl, 250 mM imidazole, pH 8. The purified protein was desalted in 50mM phosphate, 150 mM NaCl and 0.05% beta-mercapto ethanol and stored with 12% glycerol at -80^{o}C.

2.2.3 Chemical synthesis of 5-OHBza

The 5-OHBza was synthesized as described previously with minor changes [37,38]. Briefly, 27 ml of 48% HBr was added to 600 mg 5-OMeBza. The reaction mixture was refluxed at 105^{o}C for 16 hrs and was monitored using TLC (chloroform: methanol: triethylamine: : 18 : 2 : 0.1). Upon indication of full conversion by TLC the reaction flask

was cooled on ice and was neutralized using NaOH. The quenched reaction was concentrated by rotatory evaporation. A silica slurry was made and the loaded onto a silica column. Washes of 2%, and 5% methanol in chloroform was given till TLC indicated that residual 5-OMeBza was fully removed. The compound was eluted with 10% methanol in chloroform, concentrated and dried by vacuum. The purity of the product was ascertained through ^{1}H NMR. The yield was calculated to be around 30%.

2.2.4 Enzyme assays for phosphoribosyltransferase activity of CobT homologs

For the characterization of the CobT homologs with the various lower ligands the reaction setup is as follows. 10 μM enzyme, 2 mM NMN, 500 μM lower ligand were added in 50 mM Tris buffer pH8.0 and 1 mM $MgCl_2$. Reactions were incubated at 25^{o}C for 48 hrs and quenched with 0.9% formic acid. For the quantitation of substrate turnover 5 μM enzyme and 1 mM lower ligand were used.

2.2.4 Analysis by HPLC and LC-MS

The enzymatic reactions were analysed by reverse phase HPLC on Agilent 1260 infinity II UHPLC system paired with a UV-Vis DAD detector and fluorescence detector (FLD). The column used was Agilent Zorbax Eclipse XDB 5□ 250mm x 4.60mm and the flowrate was 0.5 ml/min. The mobile phases used were 10 mM, ammonium acetate pH 6.5 (solvent A) and methanol (solvent B). Samples were analysed by the following method: 0-2 min 100% A; 2-25 min from 100% A to 70% B; 25-28 min 100% A.

LC-MS analysis were performed on ScieX X500R QTOF Mass spectrometer system attached to an Exion-LC series UHPLC. The analytes were separated on an Agilent Zorbax Eclipse Plus- C18 5μ 250mm x 4.60mm column. A flow rate of 0.5 ml/min was used throughout the method. The LC Method was used for separating analytes as described above. All MS analysis were conducted in +ESI mode with +4500 V ion-spray voltage, medium de-clustering potential (80V) and low collision energy (5V) at 400 °C and TOF-MS followed by targeted MRM-HR analysis was conducted using precursor ion mass and fragment ion mass for the analyte of interest. Data was

acquired and processed on ScieX-OS to obtain extracted ion chromatogram from the total ion chromatogram.

2.2.5 In vitro competition assays

For the pair wise competition assays for the lower ligands, enzyme reactions were carried out with the following reaction setup: 5 μM enzyme, 2 mM NMN, 1 mM of each lower ligand were added in 50 mM Tris buffer pH8.0 and 1 mM $MgCl_2$. The enzyme to substrate ratio was increased so as to obtain a large excess of the substrate so as to not bias the preference of the enzyme. Reactions were incubated at 25^{o}C and aliquots were taken at 0 min, 3 min, 10 min, and 6 h. Percentage conversion was calculated from the decrease in the area under the curve of the lower ligand by using a standard curve which was previously generated.

2.3 Theoretical methods

2.3.1 Phylogeny and gene neighbourhood analysis

The sequences of CobT homologs were obtained from a protein BLAST search at the NCBI website. The sequences used in the search were CobT from *Eubacterium limosum*, *Moorella thermoacetica*, and *Methanocaldococcus jannaschii*. The FASTA files of the CobT sequences were obtained from NCBI website by using and a unique eight letter code was assigned to each using Microsoft Excel[39,40]. The sequences were aligned using MUSCLE[41] and the alignment was viewed and edited in BioEdit[42]. The phylogeny tree was generated using CIPRES[43] and was viewed and edited in ITOL[44,45].

3. Results and Discussion

3.1 Phylogenetic analysis and gene neighbourhoods

A dataset comprising of 126 number of protein sequences of CobT homologs from organisms that produce, utilize or are predicted to produce cobamides was curated as described in the methods section. The organisms in the tree include archaea and bacteria which are further classified as aerobic bacteria, obligate anaerobic bacteria or, facultative anaerobic bacteria. Most obligate anaerobes harbouring a *bza* operon also shows a second cobT homolog and in such cases, all of the CobT sequences from an organism were added to the dataset. The sequences were aligned and a maximum likelihood tree was generated.

Next, the gene neighbourhoods of the each of the CobT homologs in the dataset was recorded and the CobT homologs were categorized based on their gene neighbourhoods (Figure 4). Briefly,

1. CobT1: When the encoding gene, *cobT* is a part of the *bza* operon found in obligate anaerobes

2. CobT2: If *cobT* gene is located within the *cob* operon (containing other cobamide biosynthesis genes. CobT2 are present in aerobes, anaerobes and facultative anaerobes.

3. CobT3: The homologs present in a neighbourhood which is not related to cobamide biosynthesis and are found in both aerobes and anaerobes.

Of the 59 obligate anaerobes possessing bza operon,14 have only cobT1, 6 have CobT1 and CobT2, and 18 have cobT1 and cobT3 while 19 have only cobT2/3. 2 organisms possess all three types of CobT. There are no organisms that possess a CobT2 and a CobT3 but not a CobT1.

The analysis shows that the archaeal CobT form a significantly distinct clade in the cladogram and shows limited sequence similarity in the alignment. Interestingly, our findings corroborate with findings by Jeter et al., that though the archaeal and cyanobacterial CobT proteins show less than 20% sequence similarity with bacterial

CobT proteins, that the catalytic Glutamate residue is conserved these (Figure 4, clade coloured red).

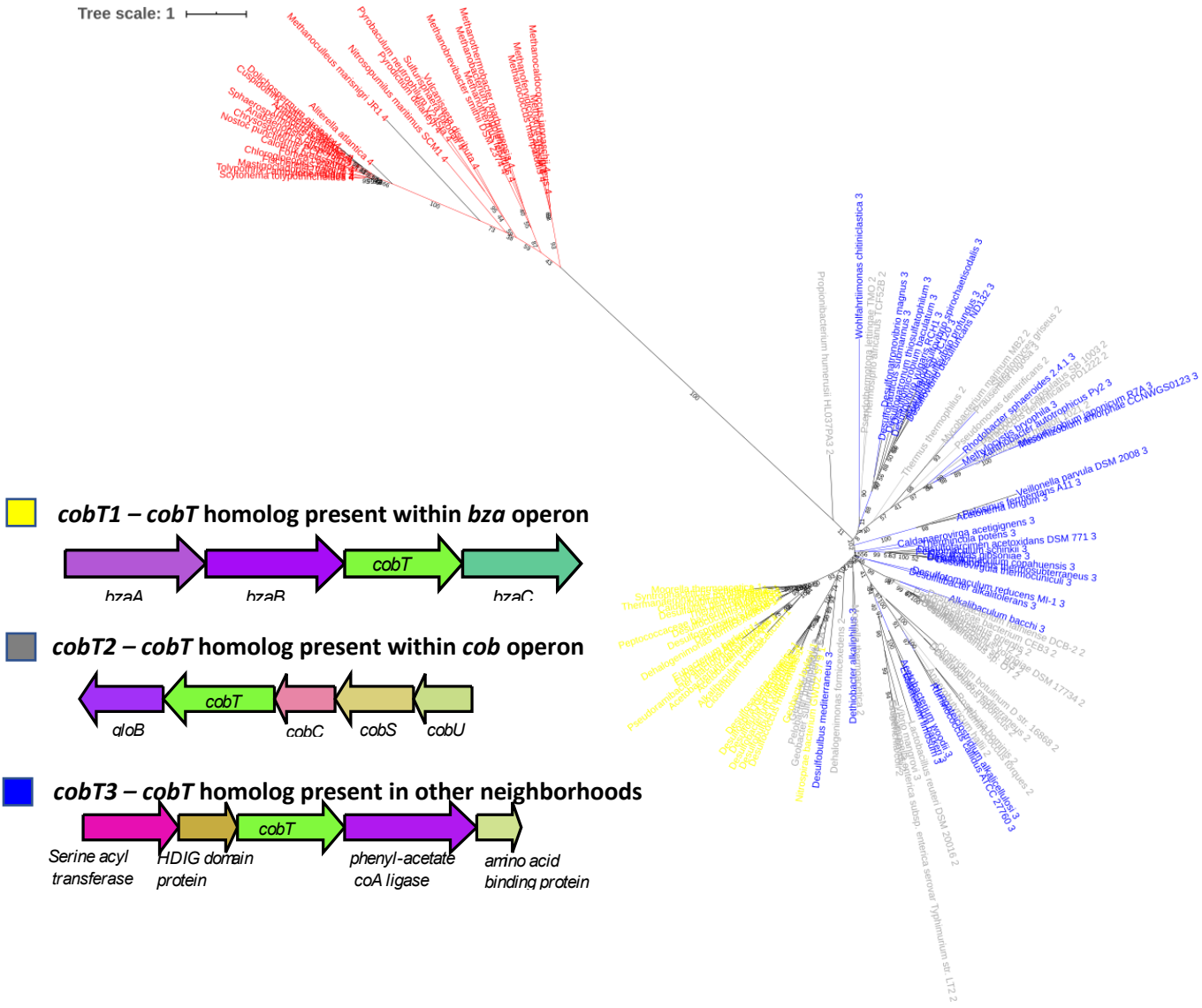


Figure 4: A phylogeny tree of CobT homologs from archaea and eubacteria

A radial form of the maximum likelihood tree consisting of 126 CobT homologs from archaea and eubacteria. Entries highlighted in A) red are archaeal homologs B) yellow are CobT1 homologs from bacteria C) grey are CobT2 homologs from bacteria and D) blue are CobT3 homologs from bacteria. We observe that Archaeal CobT homologs are significantly different from bacterial homologs.

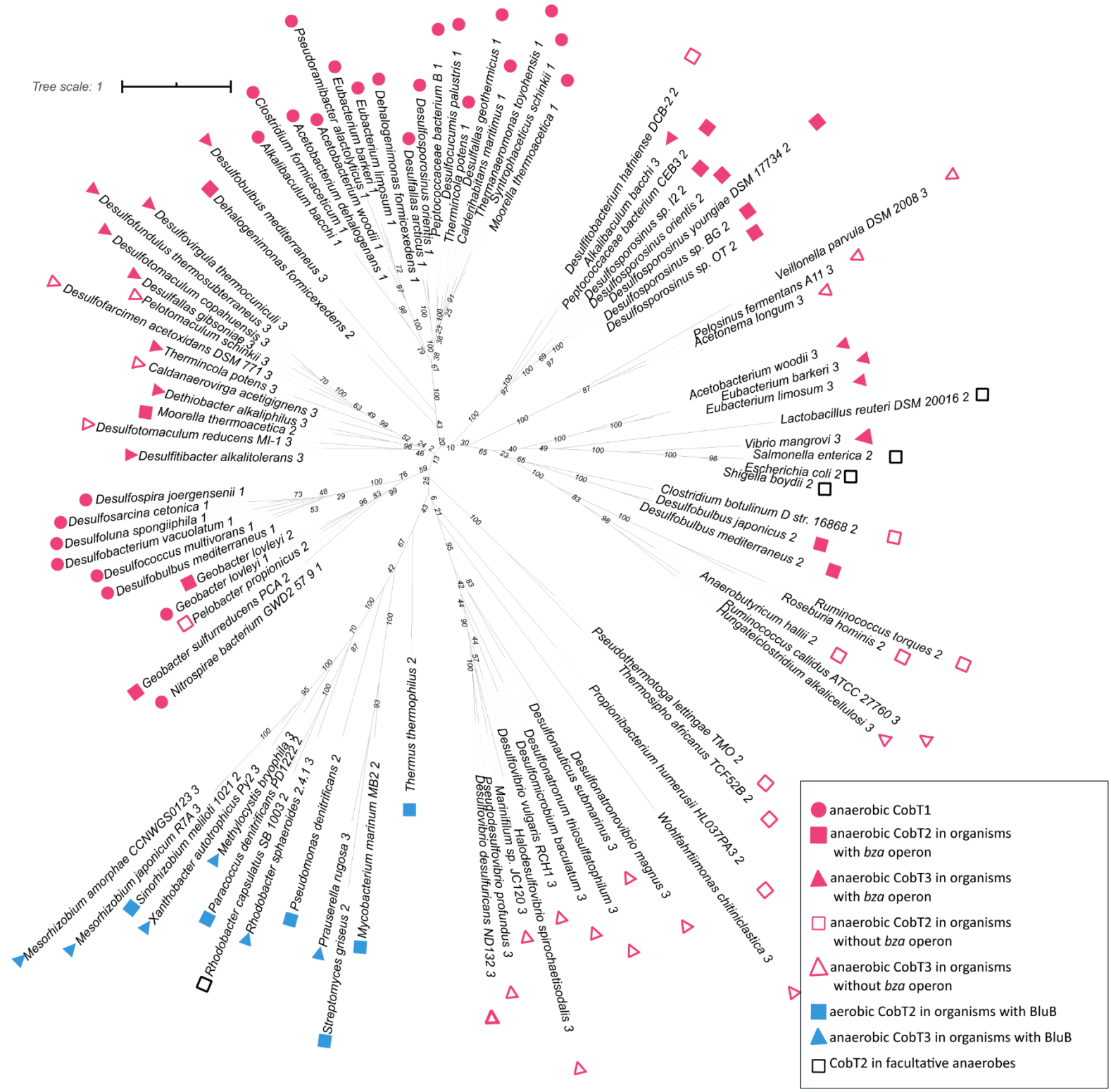


Figure 5: A phylogeny tree of CobT homologs from eubacteria

A radial form of maximum likelihood tree of 96 CobT homologs from Eubacteria, including aerobes, obligate anaerobes and facultative anaerobes. The entries in the tree has been classified on the basis of 1) aerobes, anaerobes and facultative anaerobes, 2) the type of CobT namely CobT1, CobT2 or CobT3 and 3) if they are present in an organism with bza operon in case of obligate anaerobes and if they are present in an organism with BluB. From the tree we see that CobT homologs from aerobes and facultative anaerobes form a clade significantly away from CobT homologs from obligate anaerobes. We also see clades forming on the basis of the type of CobT.

From the cladogram of Bacterial CobT homologs, it is apparent that among the obligate anaerobes which harbour multiple CobT we observe that instead of taxonomical conservation, the CobT homologs rather show conservation based on the gene neighbourhoods. The CobT1 homologs form a significantly separate clade (Figure 4, highlighted yellow) than CobT2 (Figure 4, highlighted grey) and CobT3 (Figure 4, highlighted blue) homologs. CobT2 and CobT3 homologs are more or less interspersed.

From the phylogeny tree of bacterial CobT homologs (Figure 4) we see that CobT1 (Figure 4, pink circle filled) clearly form a different clade from other CobT. We also see that though CobT2 and CobT3 are interspersed it is observed that CobT2 (pink triangle, filled) and CobT3 (pink square, filled) from organisms with *bza* operon group together as is the case with CobT2 (pink triangle, empty) and CobT3 (pink square, empty) homologs from obligate anaerobes without the *bza* operon. Among the homologs from each category we taxonomical conservation of protein sequences. We also see that CobT homologs from aerobes (blue) and facultative anaerobes (black) form a different clade from that of obligate anaerobes (pink).

Based on these observations, we hypothesize that CobT homologs of a category in different organisms may show similar activity and the different types of CobT present in the same organism may exhibit different catalytic activity.

For the comparative study of the enzymatic activity of the CobT homologs encoded by various gene neighbourhoods, a set of various CobT1, CobT2 and CobT3 were chosen. The chosen CobT homologs were CobT1 and CobT2 from *Moorella thermoacetica*, CobT1 and CobT3 from *Eubacterium barkeri* and *Eubacterium limosum*, CobT1, CobT2 and CobT3 from *Desulfobulbus mediterraneus,* and CobT1 from *Pseudoramibacter alactolyticus.* The CobT homologs characterized in the study

include CobT1, CobT2 and CobT3 homologs from obligate anaerobes (Figure 6). The study on the other mentioned CobT homologs is in progress.

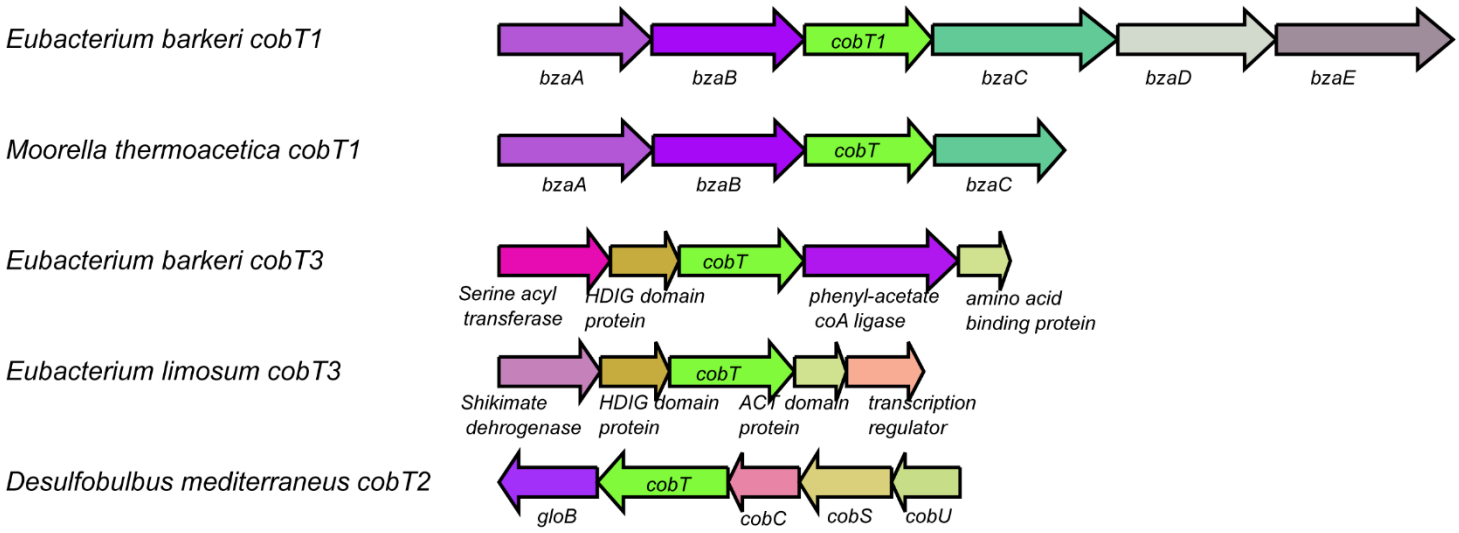


Figure 6: Gene neighbourhoods of the CobT homologs chosen for the study

The gene neighbourhoods of the *cobT1* from *Eubacterium barkeri* and *Moorella thermoacetica*, *cobT2* from *Desulfobulbus mediterraneus* and cobT3 from *Eubacterium* limosum and *Eubacterium* barkeri are annotated as obtained from the NCBI website.

3.2 Reconstitution of activity shows that all the chosen CobT homologs are functional phosphoribosyltransferases

Since there are multiple CobT homologs present within the genome of a single organism not all of them may be functional. The CobT homologs chosen in this study are CobT1 from *Moorella thermoacetica* (*Mt*CobT1), CobT1 from *Eubacterium bark*eri (*Eb*CobT1), CobT2 from *Desulfobulbus mediterraneus* (*Dm*CobT2), CobT3 from *Eubacterium limosum* (*El*CobT3), and CobT3 from *Eubacterium barkeri* (*Eb*CobT3). A sequence alignment of the chosen CobT enzymes and a well characterised CobT from *Salmonella enterica* (*Se*CobT) shows the presence of the active site residues and the catalytic glutamate residue indicating that all these different homologs may as well be functional phosphoribosyltransferases (Figure 7). We also see that the consensus sequence of Arg-Leu-Gly-Glu-Gly is conserved.

Figure 7: sequence alignment of CobT homologs: The sequence alignment of the chosen CobT homologs namely, *Mt*CobT1, *Eb*CobT1, *Dm*CobT2, *Eb*CobT3, *El*CobT3 with the well characterized *Se*CobT shows that there is conservation of active site residues, especially the catalytic glutamate residue marked with the red box.

The five CobT homologs namely *Mt*CobT1, *Eb*CobT1, *Eb*CobT3, *El*CobT3, *Dm*CobT2 were cloned, heterologously expressed in *E. coli* BL21(DE3) and purified by Nickel – NTA affinity chromatography as described in methods. To ascertain that each of these homologs are functional phosphoribosyltransferases their activity was reconstituted with DMB which is reportedly the preferred substrate of CobT enzymes and NMN. The reactions were analysed by HPLC and LC-MS.

It is reported in literature that *M*. *thermoacetica* makes cobamide with 5-OMeBza and *E. limosum* makes DMB as lower ligand. It is predicted from the *bza* operon architecture that *E. barkeri* makes DMB and *D. mediterraneus* makes 5-OMeBza.So, we were also curious if CobT from organisms which does not make DMB can also activate DMB *in vitro.* The reconstitution of activity of *Mt*CobT1 and *Eb*CobT1 shows that they are indeed active enzymes as they activate DMB to give a single product (Figure 8D, 8H) which was ascertained to be α-ribazole-5'-phosphate by LC-MS. The products show a fragment of m/z 147 corresponding to DMB (Figure 8E, 8I).

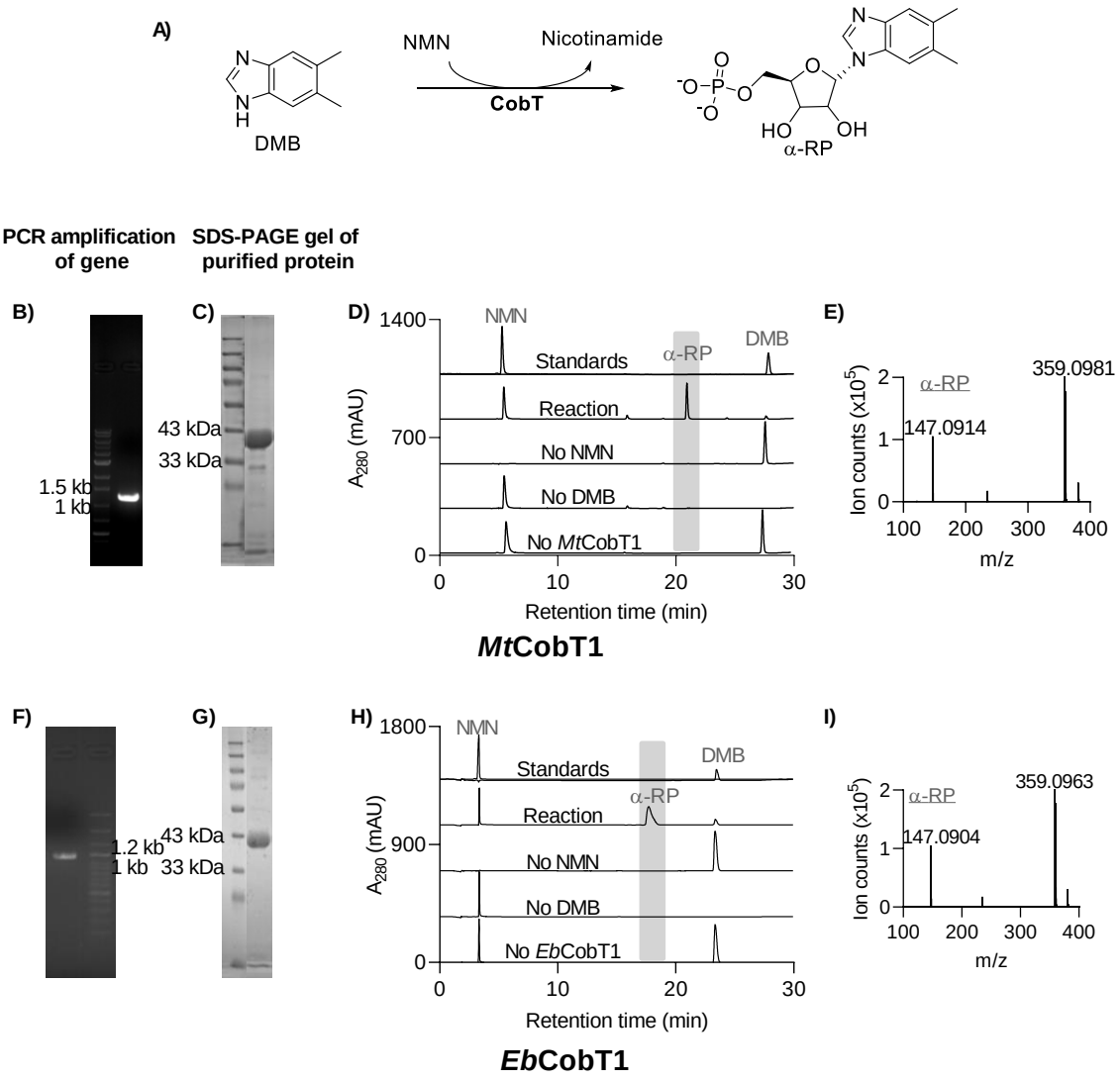


Figure 8: Reconstitution of activity of CobT1 homologs with DMB

In the figure A) shows the expected reaction of CobT enzyme with DMB to give the product α-ribazole-5'-phosphate (α-RP). B) shows the amplified gene for RF cloning, C) the purified protein D) the HPLC chromatogram of reconstitution of activity with DMB with the appropriate controls, and E) the mass spectrum of the product obtained for the homolog *Mt*CobT1. F), G), H) and I) shows the respective panels for *Eb*CobT1. We see that a single product is obtained in the case of both homologs which was ascertained to be α-RP from LC-MS analysis. The mass spectrum of the products shows a fragment with m/z of 147 corresponding to DMB.

The CobT3 homologs from *E. limosum* and *E. barkeri* were cloned (Figure 9 A, E), expressed and purified (Figure 9 B, F). both these organisms have a CobT1 homolog which is functional. So, we were curious to see if the CobT3 homolog was also a functional enzyme. The reconstitution of activity of *El*CobT3 and *Eb*CobT3 with DMB

shows that they are functional phosphoribosyltransferases yielding a product (Figure 9 C, G) with a mass spectrum corresponding to α-RP (Figure 9 D, H). This shows that the two organisms both have two different functional CobT enzymes. Only further in vivo studies can tell if they are truly functional in the organism and if their mode of actions are different.

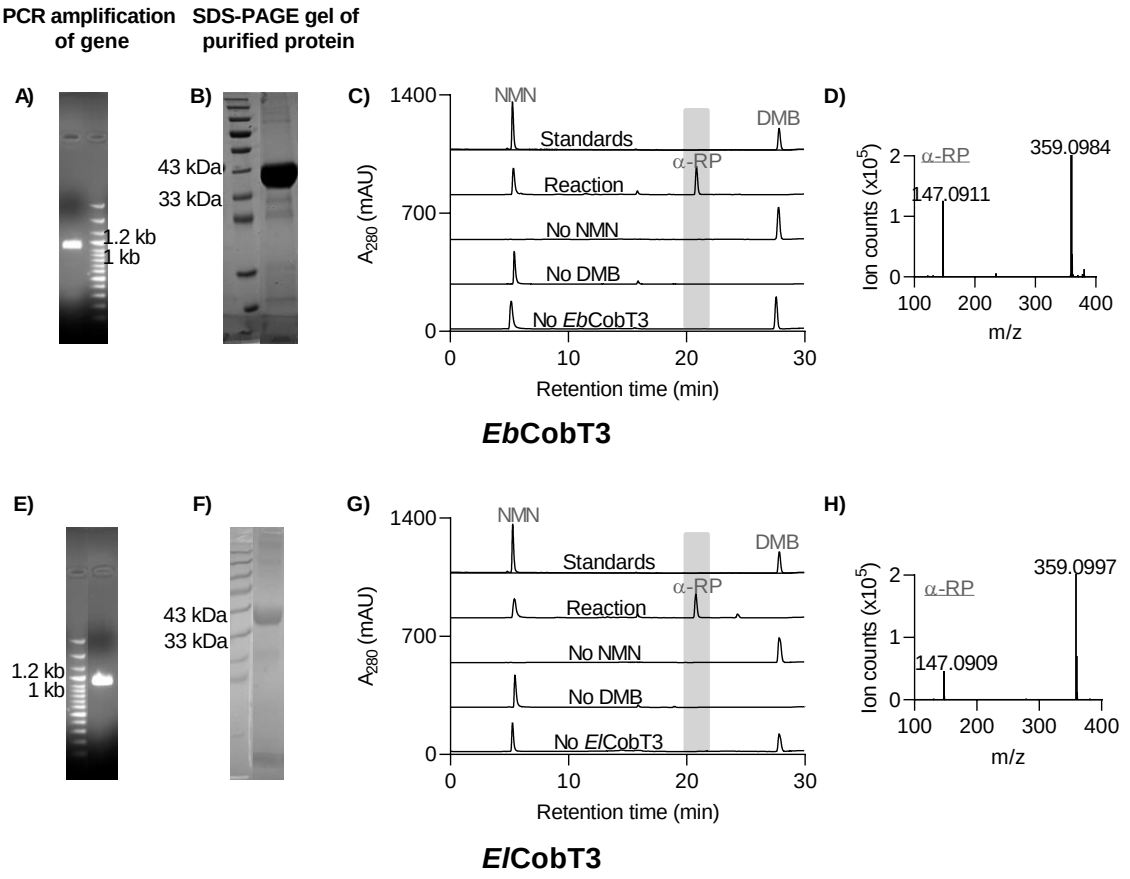


Figure 9: Reconstitution of activity of CobT1 homologs with DMB

In the figure A) and E) shows the PCR amplified genes for cloning, B) and F) shows the purified protein, C) and G) shows the HPLC chromatogram of the reaction of the enzymes with DMB and D) and H) shows the mass spectrum of the product obtained for *Eb*CobT3 and *El*CobT3 respectively. The results ascertain that both the homologs are functional phosphoribosyltransferases.

D. mediterraneus interestingly has four predicted CobT homologs, including a CobT1, a CobT2 and two CobT3 homologs according to data obtained from NCBI protein sequences database. One of the CobT3 has a Zn-dependent alcohol dehydrogenase domain in addition to conserved phosphoribosyltransferases domain. This leads us to think of a possibility that the additional CobT homologs aside from the CobT1 in *bza*

operon may be involved in other biosynthetic pathways. We have only cloned the CobT2 homolog from the organism which is seen to be a functional enzyme (Figure 10).

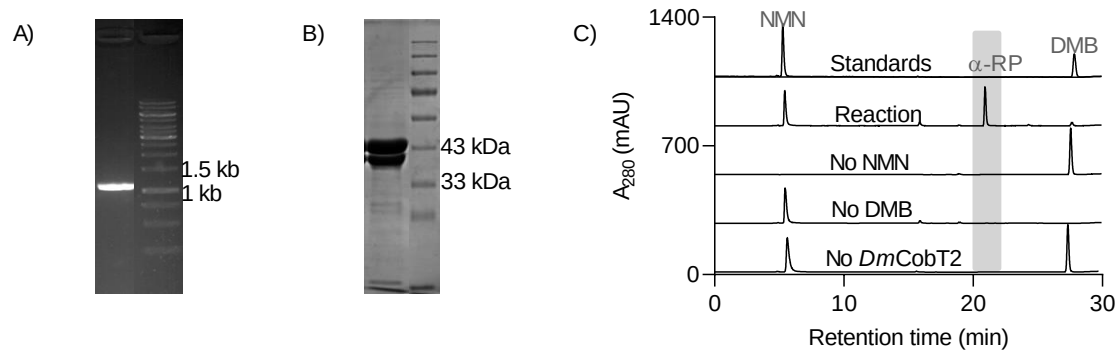


Figure 10: Reconstitution of activity of *Dm*CobT2 with DMB

A) Amplification of *Dm*cobT2 gene from the genomic DNA B) SDS-PAGE gel image showing purified *Dm*CobT2 protein C) *Dm*CobT2 is established as a functional phosphoribosyltransferase as it converts DMB into the product α-ribazole phosphate (α-RP)

From these experiments we see that irrespective of the gene neighbourhoods all the CobT homologs activated DMB and are functional. We see that an organism may have multiple functional CobT homologs. These homologs may be acting in different physiological contexts. For instance, the CobT1 homolog may be involved in the activation of the lower ligand produced by the bza operon the encoding gene is present in. The CobT2 and CobT3 homologs may also be acting on lower ligands procured from the surrounding environment and may be involved in the salvage pathway. Interestingly, we observed that the net turnover of DMB by different homologs differ slightly. Under similar reaction conditions for the different enzymes, we see that the *Eb*CobT3 and *El*CobT3 enzymes show full conversion of DMB in 48 hours whereas in case of *Mt*CobT1, *Eb*CobT1 and *Dm*CobT2 homologs we see unutilized DMB. This can be due to two reasons, the CobT1 and CobT2 have less preference for DMB compared to CobT3 and that their actual physiological substrate is not DMB or that CobT1 and CobT2 are less efficient enzymes compared to CobT3.

3.3 Regiospecificity in the activation of asymmetric substrates differ in CobT homologs

As we previously described asymmetric lower ligands when activated by CobT can yield two different regio-isomers as products. This arises from the fact that the either of the two nitrogen atoms in the imidazole ring can be attached to the C1' of the ribose ring in the phosphoribosyl moiety. In case of symmetric substrates these two nitrogen atoms are equivalent due to resonance. This promiscuity in the activation has been regarded as a reason for cobamide diversity in organisms. To further see if the various types of CobT enzymes differ in their activity we studied their reactions with the asymmetric substrates 5-OHBza and 5-OMeBza which are also part of the *bza* operon pathway.

3.3.1 Only CobT1 homologs are regiospecific for 5-OHBza

Previous work from our lab[38] shows that CobT1 from *Eubacterium limosum* (*El*CobT1) when reconstituted with 5-OHBza yields a single product, α–5-OHBza riboside phosphate (5-OHBza-RP). Intrigued by this unique observation and to test if other CobT1 homologs encoded within *bza* operon exhibit regiospecificity for activation of 5-OHBza, we tested enzyme activity of *Eb*CobT1 and *Mt*CobT1 with 5-OHBza.

We see that for the reaction of *Mt*CobT1 and *Eb*CobT1 with 5-OHBza there is a single product formed which has a retention time of 10 min and 6 min on HPLC (Figure 11A, 11D) respectively. The product has a UV-Vis absorption spectrum which is similar to the absorption spectrum reported in literature for 5-OHBza-RP with $A_{250} > A_{280}$ (Figure 11B, 11F). To verify that this is indeed the product we analysed the reaction by LC-MS and the peak corresponding to 5-OHBza-RP shows mass fragment of m/z 135 corresponding to 5-OHBza (Figure 11C, 11F).

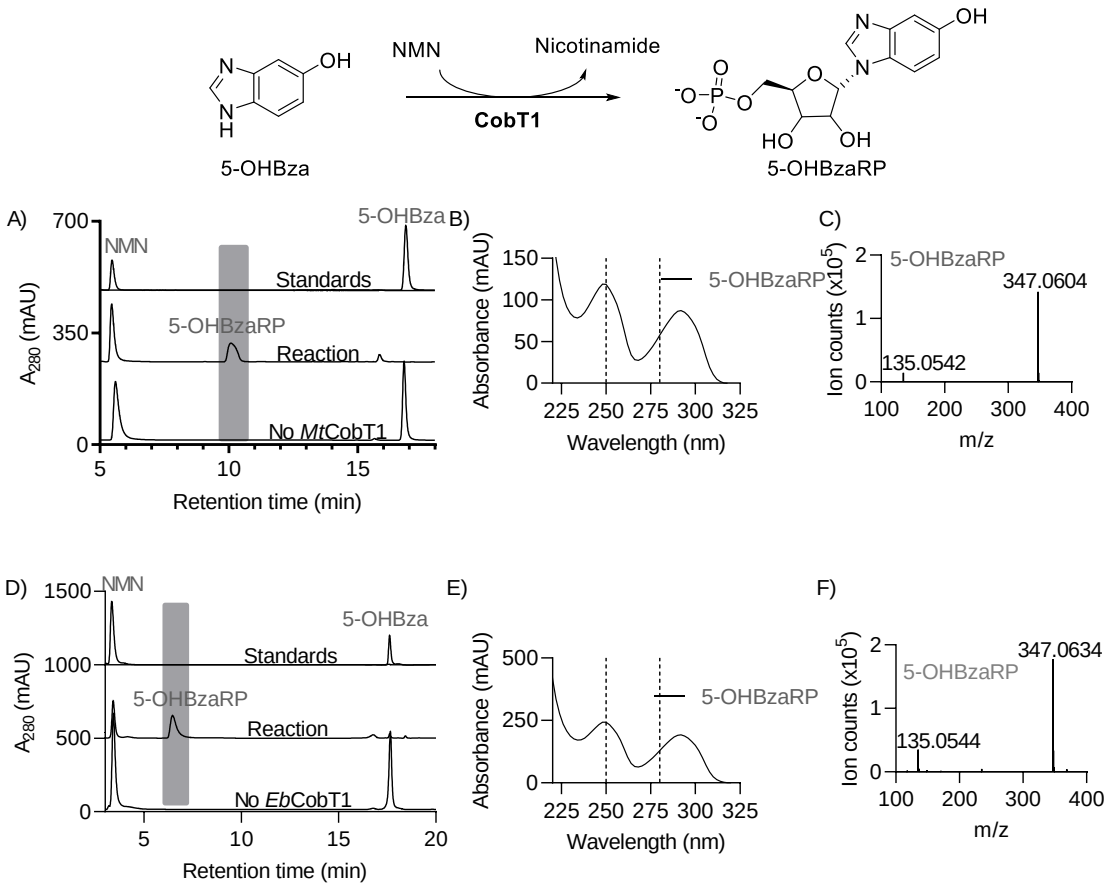


Figure 11: 5-OHBza activation by CobT1 homologs

HPLC chromatogram for 5-OHBza reaction with A) *Mt*CobT1 and D) *Eb*CobT1 which yields a single product. B) and C) are the UV absorption spectrum and Mass spectrum of the product of *Mt*CobT1 and E) and F) are the similar spectrum for the product of *Eb*CobT1. The reaction product in both cases was ascertained to be 5-OHBza based on the UV-Vis absorption spectrum with A250>A280. Both the mass spectrums also show a m/z of 135 corresponding to 5-OHBza.

Intrigued by the regiospecificity of the CobT1 homologs we were curious to see if the CobT3 homologs also show a similar activity. Based on the cladogram we see that CobT1 homologs tend to group together and are far from CobT3 and CobT2 homologs. We wanted to see if this reflects in the enzymatic activity as well.

In the case of *Eb*CobT3 and *El*CobT3 enzymes they yield two isomeric products, two new peaks emerge in the reaction at retention times 10 min and 11 min (Figure 11A, 11D, α-5-OHBza riboside phosphate (5–OHBza-RP) and, α–6-OHBza riboside phosphate (6–OHBza-RP). The two isomers differ in their UV absorption spectrum but show similar mass spectrum with a fragment of m/z 135 corresponding to 5-OHBza

(Figure 12C, 12F). In case of 5-OHBzaRP A250 > A280 whereas for 6-OHBzaRP A250 < A280 (Figure 12B, 12E). From the area under the curve we see that there is more of 6-OHBza-RP than 5-OHBza-RP and hence the enzyme has a regioselectivity for 6-OHBza-RP.

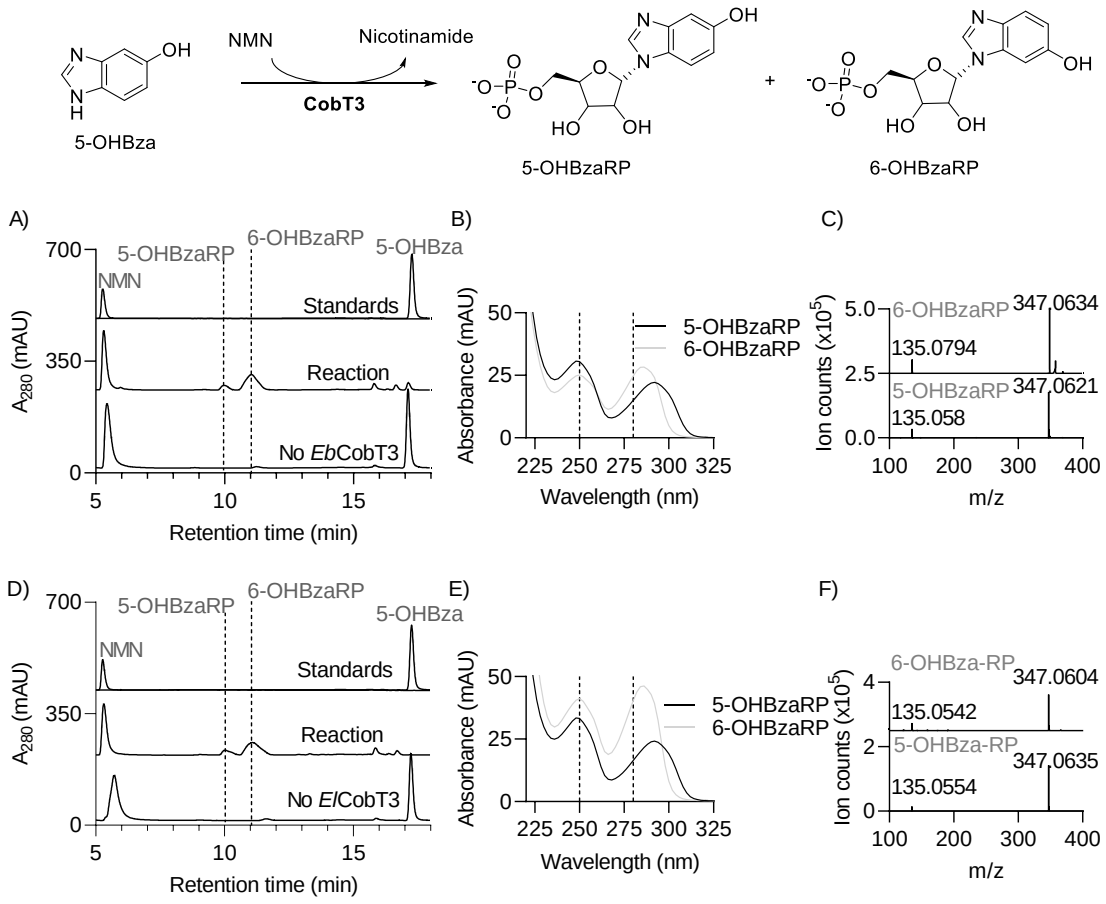


Figure 12: 5-OHBza activation by CobT3 homologs

HPLC chromatogram for 5-OHBza reaction with A) *Eb*CobT3 and D) *El*CobT3 which yields both the Regio-isomeric products, 5-OHBzaRP and 6-OHBzaRP. B) and C) are the UV absorption spectrum and Mass spectrum of the products of *Eb*CobT3 and E) and F) are the similar spectrum for the products of *El*CobT3

Since CobT2 and CobT3 homologs from obligate anaerobes are interspersed in the cladogram we expected *Dm*CobT2 to show similar activity like the CobT3 homologs. We indeed see that there are two isomers formed in the reaction but we also see that there is more of 5-OHBza-RP compared to 6-OHBza-RP (Figure 13). This leads us to propose that CobT2 are different from CobT3.

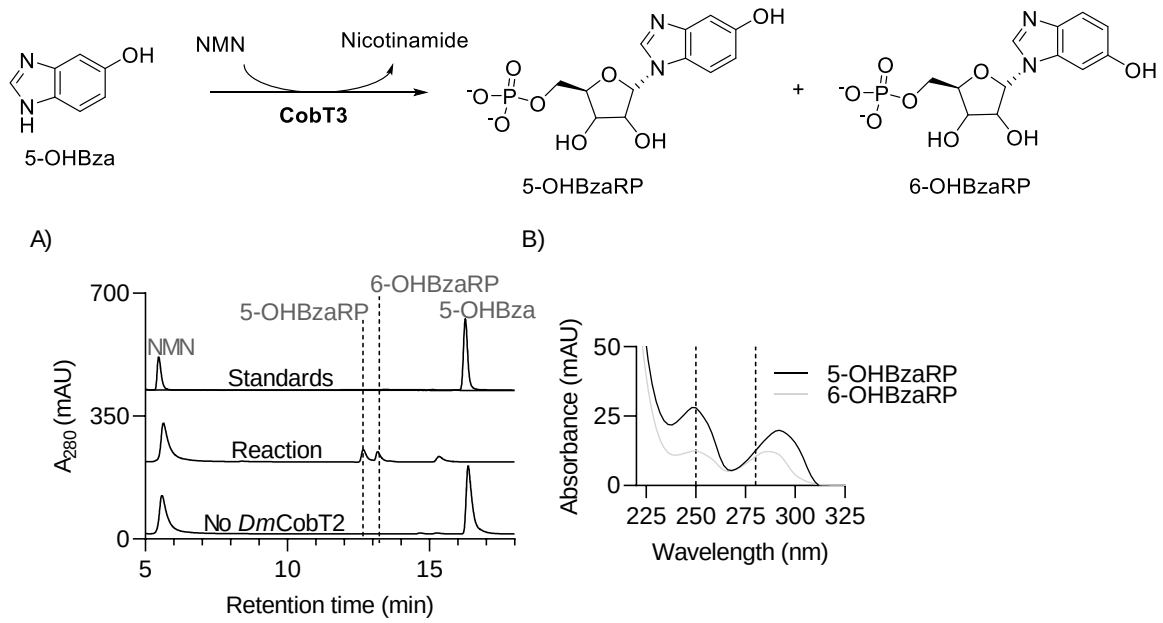


Figure 13: 5-OHBza activation by CobT2 homologs

A) HPLC chromatogram for 5-OHBza reaction with *Dm*CobT2 which yields both the Regio-isomeric products, 5-OHBzaRP and 6-OHBzaRP. B) UV absorption spectrum of the two products

From the reactions with 5-OHBza we can say that CobT1, CobT2 and CobT3 enzymes are different from each other in their catalytic activity. The regiospecificity in the activity of CobT1 homologs further encourages us to hypothesize that the actual substrate of the CobT1 enzyme in the pathway is 5-OHBza and that the CobT1 acts right after the conversion of AIR to 5-OHBza by BzaF. The subsequent Bza enzymes may act on the activated form of 5-OHBza leading to the regiospecific attachment of lower ligand in vitamin B_{12}. As further proof of this hypothesis current work in the lab has established that the dephosphorylated form of the product of CobT with 5-OHBza, i.e., 5-OHBza-R is a better substrate for the succeeding enzyme BzaC in *M. thermoacetica* whose predicted substrate is 5-OHBza.

3.3.2 All CobT homologs yield two products with 5-OMeBza

Since the CobT1 enzymes are regiospecific in activating 5-OHBza, we wanted to see if it shows a similar outcome in the case of 5-OMeBza, another asymmetric substrate also part of *bza* operon pathway. Also, it is reported that *M. thermoacetica* makes 5-

OMeBza-Cba[21,22] and not 6-OMeBza-Cba. So, we were curious if MtCobT1 gives a single product 5-OMeBza in the reaction which would explain this observation.

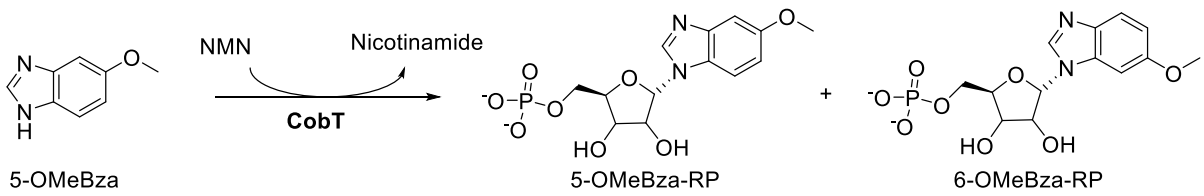


Figure 14: CobT reaction with 5-OMeBza yielding two regio-isomeric products

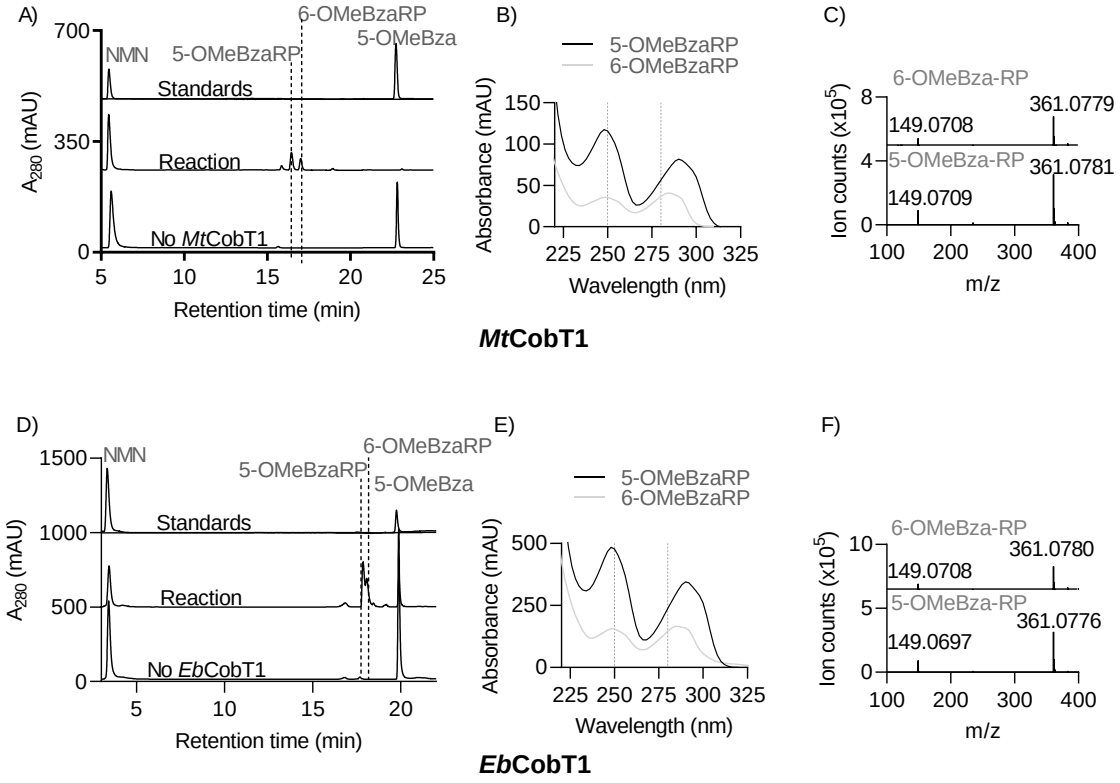


Figure 15: 5-OMeBza activation by CobT1 homologs

A) Shows the HPLC Chromatogram B) shows the UV-Vis absorption spectrum and C) shows the mass spectrum of the products of the reaction of 5-OMeBza with *Mt*CobT1. D), E) and F) shows the respective panels for *Eb*CobT1. Both the CobT1 gives two products in the reaction that differ in the absorption spectrum. The mass spectrum of the products shows a fragment of m/z of 149 corresponding to 5-OMeBza.

The CobT1 enzymes, *Mt*CobT1 and *Eb*CobT1 reacts with 5-OMeBza yielding both products, α–5-OMeBza riboside phosphate (5–OMeBzaRP) and, α–6-OMeBza riboside phosphate (6–OMeBzaRP) (Figure 14, 15A, 15D).The Absorption spectrum of the two isomers are different as 5-OMeBzaRP has $A_{250} > A_{280}$ and 6-OMeBzaRP has $A_{250} < A_{280}$ (Figure 15B, 15E). The products show identical mass spectrum and also a fragment of m/z 149 corresponding to 5-OMeBza confirming that they are constitutional isomeric products (Figure 15C, 15F).

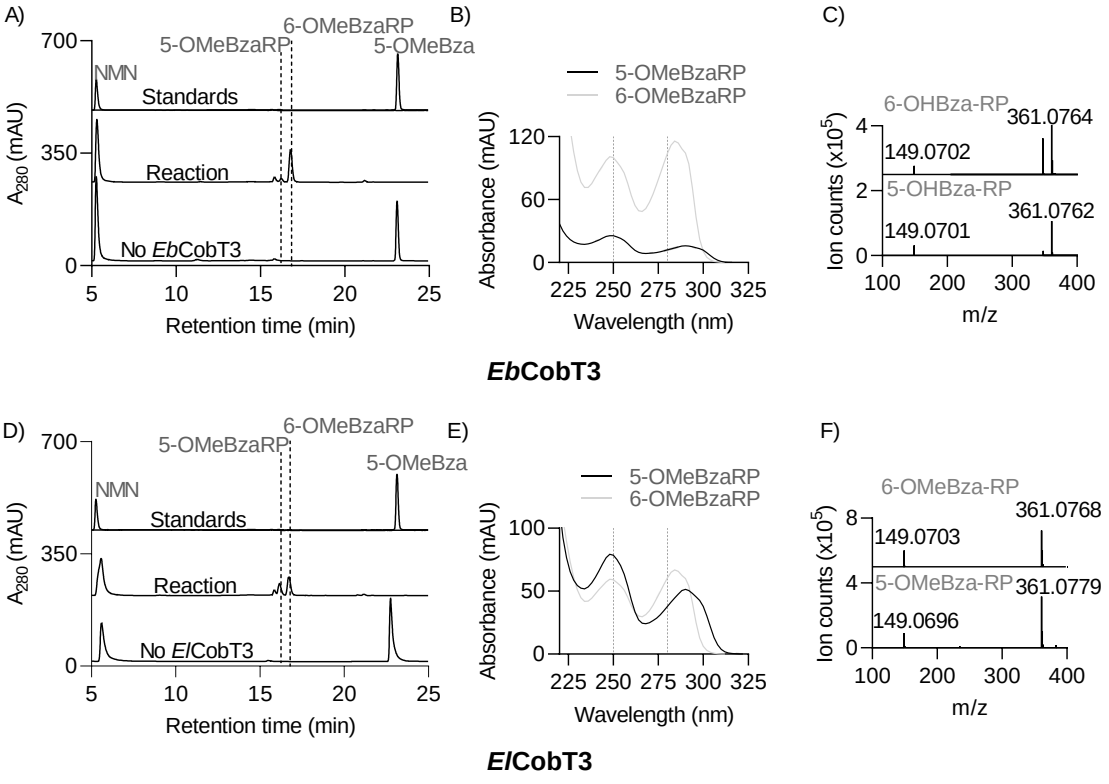


Figure 16: 5-OMeBza activation by CobT3 homologs

A) Shows the HPLC Chromatogram B) shows the UV-Vis absorption spectrum and C) shows the mass spectrum of the products of the reaction of 5-OMeBza with *Eb*CobT3. D), E) and F) shows the respective panels for *El*CobT3. Both the CobT1 gives two products in the reaction that differ in the absorption spectrum. The mass spectrum of the products shows a fragment of m/z of 149 corresponding to 5-OMeBza.

In the reaction of EbCobT3 and ElCobT3 with 5-OMeBza, we see that they give both the isomers, 5-OMeBza-RP and 6-OMeBza-RP as products (Figure 16A, 16D). This was also ascertained by the UV-Vis absorption spectrum and mass spectrum (Figure 16 B,C,D,E).

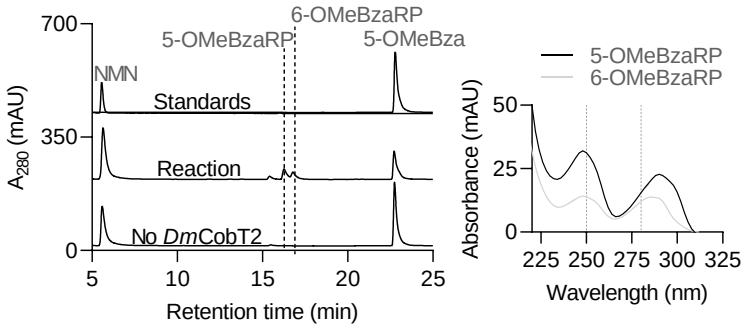


Figure 17: 5-OMeBza activation by *Dm*CobT2

HPLC Chromatogram and UV-Vis absorption spectrum of reaction of DmCobT2 with 5-OMeBza

DmCobT2 in the reaction with 5-OMeBza also shows the two isomeric products differing in their absorption spectra (Figure 17).

Overall, from the reactions, we see that all CobT1, CobT2 and CobT3 homologs yield two products and are not regiospecific in the activation of 5-OMeBza. In the case of MtCobT1 this is surprising because of it is reported to produce a single isomer of the cobamide with 5-OMeBza as the lower ligand. This substantiates our claim that 5-OHBza is the physiological substrate of CobT and leads to the regiospecific lower ligand attachment. But there is also the possibility of the formation of a single isomer inside the cell or the substrate preference of the succeeding enzymes in the nucleotide loop assembly pathway that takes up only one particular isomer. These possibilities have to be evaluated by *in vivo* studies and the characterization of the other enzymes in the pathway.

We interestingly observe that the ratios of the two isomers formed in case of the various types of CobT enzymes are slightly different. In case of CobT1 and CobT2 homologs the area under the curve of the products imply that there is more of 5-OMeBza-RP compared to 6-OMeBza-RP where as in the case of CobT3 homologs 6-OMeBza isomer is higher.

3.4 Substrate promiscuity of CobT homologs

Apart from DMB, 5-OHBza and 5-OMeBza several other benzimidazoles, purines and phenolic molecules also act as the lower ligands in naturally occurring cobamides. Furthermore, benzimidazoles found environmental samples can be incorporated into cobamides by microorganism. The ability of create a plethora of cobamides can be ascribed to the substrate promiscuity exhibited by CobT. We tested if CobT from the three categories can exhibit substrate promiscuity and differences in the regiospecificity of other lower ligands.

We tested the activity of the CobT homologs with other naturally occurring lower ligands, 5-MeBza, Bza and Adenine. Adenine was chosen as it is one of the default lower ligand because of its abundance in physiological conditions in bacteria that do not carry out de novo biosynthesis of lower ligand like *Salmonella enterica* and also because pseudocobalamin (cobamide with adenine as lower ligand) is a cofactor reported in literature.

In the reaction of *Mt*CobT1, *Dm*CobT2 and *El*CobT3 with 5-MeBza we see a product peak at retention time 18 min (Figure 18A). The mass spectrum of the peak also shows a fragment of m/z 133 corresponding to 5-MeBza. The reaction of the three CobT homologs with Bza yields a product peak with retention time 15 min (Figure 18B) and shows a mass fragment of m/z 119 corresponding to Bza.

With adenine as the substrate only *Mt*CobT1 and *El*CobT3 shows a conversion. It was surprising to observe that with *Mt*CobT1 adenine yielded a single product with retention time of 11 min but with *El*CobT3 there were two products formed with retention times 9 min and 11 min (Figure 18C). All the product peaks yielded similar mass spectrum which shows m/z 136 corresponding to adenine (Figure 18E, 18F). It seems like peak 2 in case of *El*CobT3 and the peak 3 in case of *Mt*CobT1 are similar products whereas peak 1 in case of *El*CobT3 is a different regio-isomer. This may imply that MtCobT1 is regiospecific for adenine as well. But we can only make this claim after deciphering the structure of these different products and validating the observation. Overall, this experiment proves that all CobT homologs regardless of their gene neighbourhoods are promiscuous in terms of their catalytic activity be it in substrate preference or regiospecificity in the activation of the different substrates.

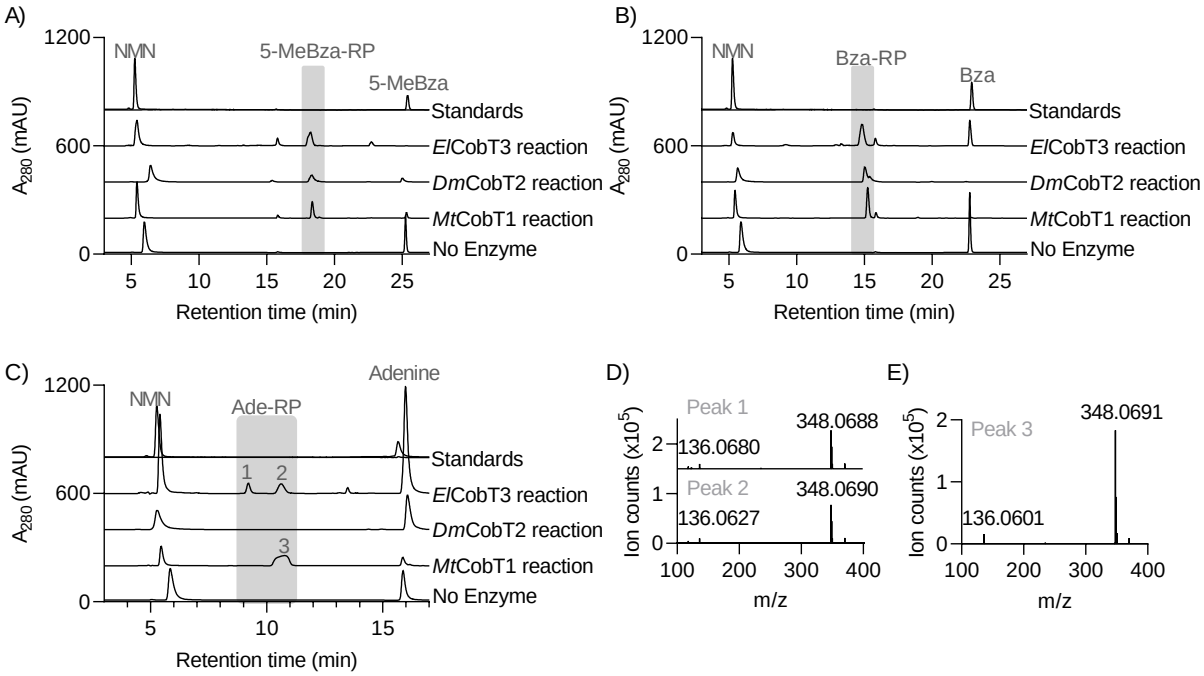


Figure 18: Activation of other lower ligands by CobT homologs

HPLC chromatogram showing the reaction of ElCobT3, DmCobT2 and MtCobT1 with A) 5-MeBza, B) Bza and C) Adenine. We see that all the three CobT enzymes activate 5-MeBza and Bza to give the corresponding products, 5-MeBza-RP and Bza-RP. In case of adenine only ElCobT3 and MtCobT1 gives products. D) Mass spectrum of the peaks corresponding to the products of the adenine reaction matches with the expected product spectrum with a fragment of m/z 136 corresponding to adenine.

We also tried out the activity of *Mt*Cobt1 with an unnatural substarte 5-ClBza. MtCobT1 activates 5-CLBza to give a product with a mass spectrum having a fragment ion of m/z 153 corresponding to 5-ClBza (Figure 19).

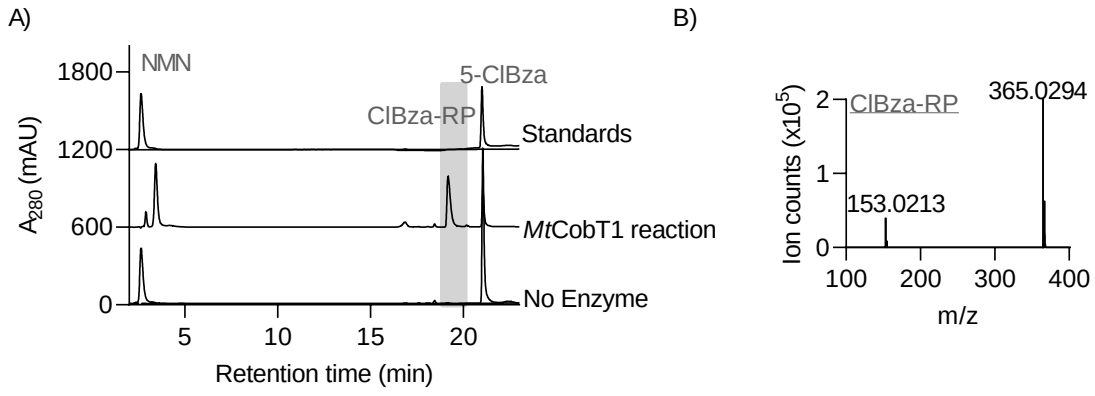


Figure 19: Activation of 5-ClBza by MtCobT1 yielding a product with expected mass spectrum

3.5 CobT1 homologs differ in their substrate preference compared to other CobT

Previous literature that studied the substrate preference of CobT enzymes from six bacteria including aerobic, anaerobic and facultative anaerobic bacteria that differ in the native cobamides produced had concluded that DMB is the most preferred substrate of all these CobT homologs. This disproved the hypothesis that the affinity of CobT enzyme may depend on the lower ligand of the native cobamide produced by the organism[28].

The characterization of CobT homologs in obligate anaerobes possessing bza operon undertaken in this thesis revealed that *Mt*CobT1 and *Eb*CobT1, the CobT1 homologs are significantly different from CobT2 homologs, namely *Dm*CobT2 and CobT3 homologs, *Eb*CobT3 and *El*CobT3 in their catalytic activity especially in the regiospecificity of activation of asymmetric lower ligands. This led us to hypothesize that CobT1 homologs show a difference in affinity towards different lower ligands when compared to CobT2 and CobT3 homologs. Also based on our previous hypothesis that 5-OHBza may be the substrate for CobT in the anaerobic lower ligand biosynthesis leading to regiospecific lower ligand attachment, we wanted to see if the CobT1 homologs have an increased preference for 5-OHBza. Another possibility is that the affinity of CobT may depend on the lower ligand produced by the bza operon, for instance MtCobT1 may have higher affinity for 5-OMeBza.

Based on previous literature, we decided to carry out pair-wise competition reactions of the substrates we are interested in to determine the relative preference of substrate of each CobT homolog. The lower ligands chosen as substrates were 5-OHBza, 5-OMeBza, DMB and Adenine. The possibility of encountering these substrates for a CobT enzyme is higher because the three benzimidazoles are part of the anaerobic lower ligand biosynthesis pathway and adenine is one of the default lower ligands in organisms that do not make benzimidazoles because of its abundance in physiological system. The competition reactions were conducted with equimolar concentrations of the substrates in all possible combinations under steady state conditions and the average of two replicates was taken. The reactions were quenched at different time points, 3min, 10 min and 6h and the percentage conversion was determined by decrease in the amount of substrate present. To calculate the amount of substrate,

standard curves were generated for each individually by calculating the HPLC peak areas.

The CobT homologs chosen for the study were *Se*CobT, *Mt*CobT1 and *Eb*CobT3. *Se*CobT was chosen as it is the most thoroughly characterized CobT in literature and its preference of the various substrates is reported in literature. Based on gene neighbourhood *Se*CobT can be classified as a CobT2.

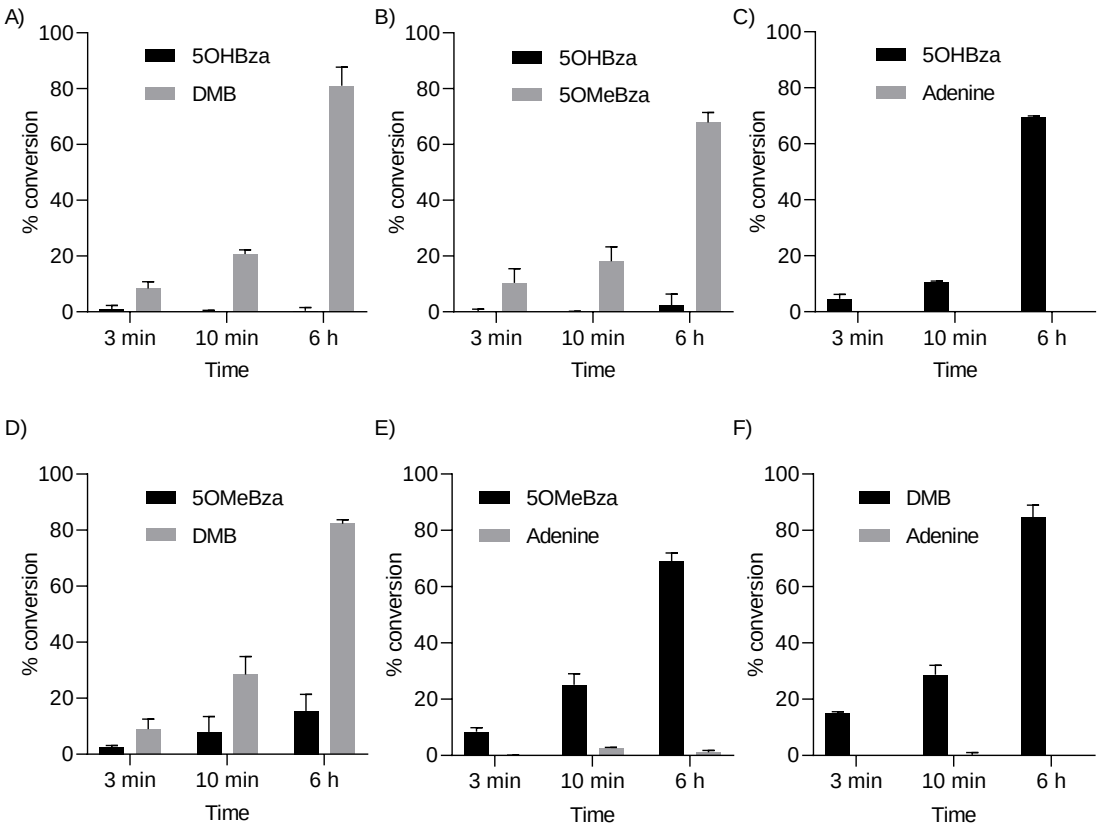


Figure 20: Pair wise competition reactions of *Se*CobT with the various lower ligands

Percentage conversion of each lower ligand in the pair wise competition reactions of all possible combinations of 5-OHBza, 5-OMeBza, DMB and Adenine at the time points, 3 min, 10 min and 6 h. A) 5-OHBza and 5-OMeBza, B) 5-OHBza and DMB, C) 5-OHBza and Adenine, D) 5-OMeBza and DMB, E) 5-OMeBza and Adenine and F) DMB and Adenine. The results showed that the preference of *Se*CobT is as follows, DMB>5-OMeBza>5-OHBza>Adenine. The data shown in figure is an average of two replicates.

When *Se*CobT reacted with a mix of 5-OHBza and 5-OMeBza we see that there is higher conversion of 5-OMeBza at all the three timepoints, 3 min, 10 min and 6 h (Figure 20A). This indicates that *Se*CobT has greater preference for 5-OMeBza

compared to 5-OHBza. Similarly, from the other combinations of lower ligands we see that DMB is preferred compared to 5-OHBza (Figure 20B), 5-OHBza is preferred compared to Adenine (Figure 20C), DMB is preferred compared to 5-OMeBza (Figure 20D), 5-OMeBza is preferred compared to adenine (Figure 20E) and DMB is preferred compared to Adenine (Figure 20F). The overall order of preference of the lower ligands by *Se*CobT is as follows, DMB > 5-OMeBza > 5-OHBza >Adenine. This agrees with the results obtained from previous studies on the preference of substrates by *Se*CobT[28]. So, we further carried out the competition assays with *Mt*CobT1 and *Eb*CobT3.

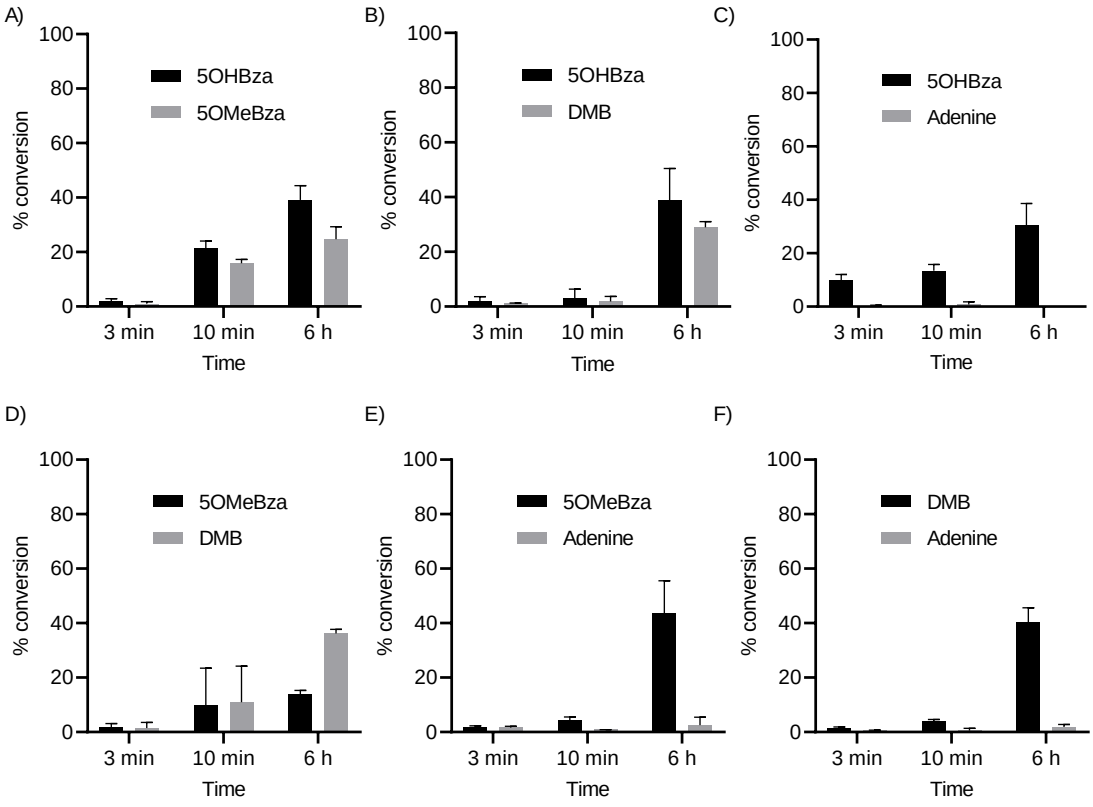


Figure 21: Pair wise competition reactions of *Mt*CobT1 with the various lower ligands

Percentage conversion of each lower ligand in the pair wise competition reactions of all possible combinations of 5-OHBza, 5-OMeBza, DMB and Adenine at the time points, 3 min, 10 min and 6 h. A) 5-OHBza and 5-OMeBza, B) 5-OHBza and DMB, C) 5-OHBza and Adenine, D) 5-OMeBza and DMB, E) 5-OMeBza and Adenine and F) DMB and Adenine. The results showed that the preference of *Mt*CobT1 is as follows, 5-OHBza>DMB>5-OMeBza >Adenine. The data shown in figure is an average of two replicates.

With *Mt*CobT1 we see that the overall turnover of the enzyme is lower compared *Se*CobT. The low percentage conversion in the initial time points make the values more error prone and hence it is difficult to deduce any inference from the initial time points.

Based on the 3 min and 10 min time point there is no clear preference of any substrate over the other in case of *Mt*CobT1. So, we used the 6 h time point to infer the affinity of the enzyme to the different substrates. We see that 5-OHBza is preferred compared to 5-OMeBza (Figure 21A), 5-OHBza is preferred compared to DMB (Figure 21B), 5-OHBza is preferred compared to adenine (Figure 21C), DMB is preferred compared to 5-OMeBza (Figure 21D), 5-OMeBza is preferred compared to adenine (Figure 21E) and DMB is preferred compared to adenine (Figure 21F). The overall substrate preference of *Mt*CobT1 is as follows, 5-OHBza > DMB > 5-OMeBza > adenine.

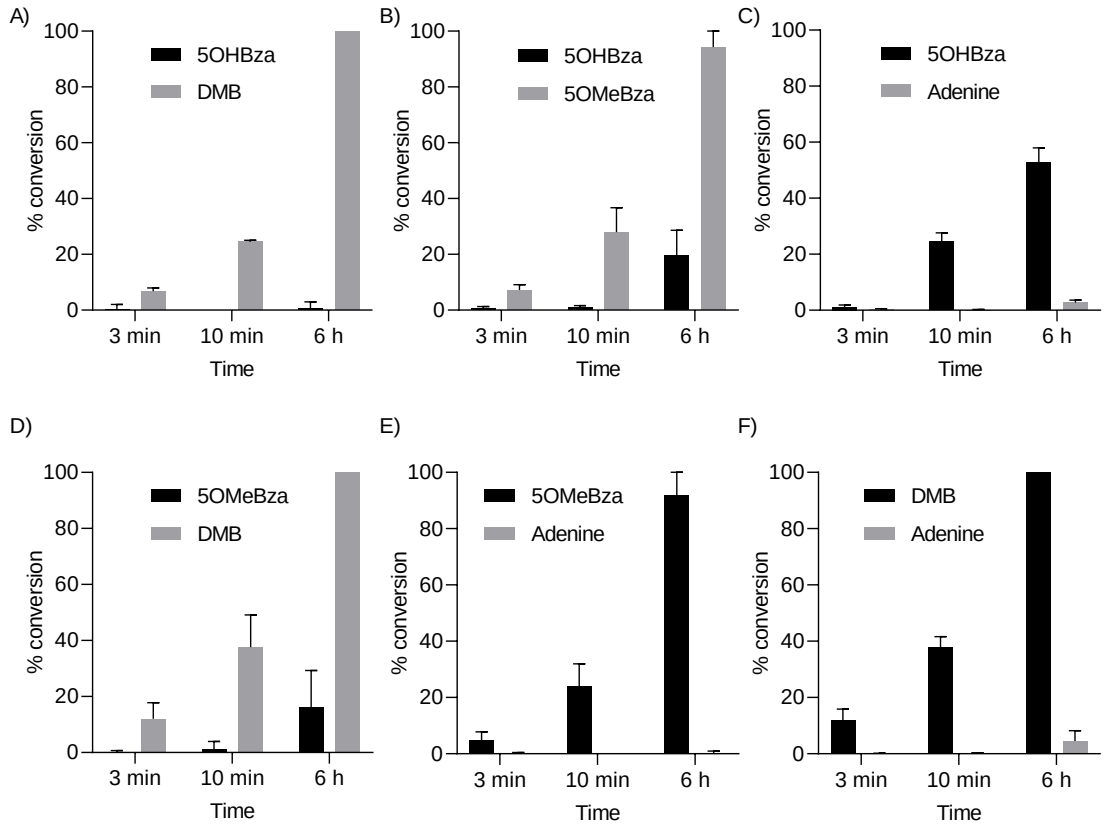


Figure 22: Pair wise competition reactions of *Eb*CobT3 with the various lower ligands

Percentage conversion of each lower ligand in the pair wise competition reactions of all possible combinations of 5-OHBza, 5-OMeBza, DMB and Adenine at the time points, 3 min, 10 min and 6 h. A)

5-OHBza and 5-OMeBza, B) 5-OHBza and DMB, C) 5-OHBza and Adenine, D) 5-OMeBza and DMB, E) 5-OMeBza and Adenine and F) DMB and Adenine. The results showed that the preference of *Eb*CobT3 is as follows, DMB>5-OMeBza>5-OHBza>Adenine. The data shown in figure is an average of two replicates.

From the competition assays with *Eb*CobT3 (Figure 22) we get to know that the overall substrate preference of the enzyme is as follows, DMB > 5-OMeBza > 5-OHBza >Adenine.

From the pair wise competition assay we see that the most preferred substrate of *Mt*CobT1 is 5-OHBza whereas for *Se*CobT and *Eb*CobT3 it is DMB. This is significant as it shows that MtCobT1 is an enzyme vastly different from any other characterized CobT. This is also interesting because the native lower ligand produced by the bza operon in *Moorella thermoacetica* is 5-OMeBza, but MtCobT1 prefers 5-OHBza. This further boosts our hypothesis that the actual physiological substrate of CobT1 enzymes may be 5-OHBza. Low turnover of CobT1 also corroborates with the low turnover of BzaC the subsequent enzyme in *bza* operon and may imply that the *bza* operon enzymes in general may have less efficient enzymatic activity.

4. Conclusions

In this thesis we undertook studies in order to characterize the not well explored anaerobic CobT enzymes from organisms possessing the *bza* operon responsible for lower ligand biosynthesis in obligate anaerobes. Based on previous literature and work carried out in our lab we discovered that many obligate anaerobes possessing bza operon may have a *cobT* present within the operon and another *cobT* gene present elsewhere. By further looking into the gene neighbourhoods of the *cobT* genes we discovered that they may be of three types and classified the CobT homologs based on the genes neighbouring the corresponding *cobT* gene. The three types CobT are namely, CobT1 where the encoding gene is present within the *bza* operon, CobT2 where the encoding gene is present within the *cob* operon and CobT3 where the encoding gene is present in a neighbourhood which seems to have no relation to cobamide biosynthesis. Organisms may possess one, two, or all three types of CobT.

A phylogenetic analysis of CobT homologs from archaea and Eubacteria revealed that archaeal CobT enzymes are drastically different from bacterial CobT enzymes though all of them possess the catalytic glutamate residue. We also saw that CobT1 homologs form a significant clade far from other CobT enzymes indicating that they may be vastly different in catalytic activity.

Based on previous work in lab we hypothesized that CobT1 homologs are regiospecific for the activation of 5-OHBza an intermediate in the *bza* operon pathway also the proposed substrate for BzaC enzyme. By reconstituting the activity of *Mt*CobT1 and *Eb*CobT1 we proved their regiospecificity in the activation of 5-OHBza. We also proved that the two CobT homologs from the same organism show difference in catalytic activity by studying *Eb*CobT1 and *Eb*CobT3. *Eb*CobT1 regiospecifically activates 5-OHBza, but *Eb*CobT3 does not.

We also studied the promiscuity in the substrate preference of the CobT1, CobT2, and CobT3 homologs and learned that they can indeed activate other lower ligands like 5-MeBza and Bza. We also saw that CobT1 and CobT3 can convert adenine into Ade-RP whereas CobT2 did not. It was surprising to see that MtCobT1 which is

regiospecific for 5-OHBza was also regiospecific for adenine whereas *El*CobT3 was not regiospecific for either of the substrates yielding two regio-isomers as products.

Lastly, we studied the substrate preference of *Mt*CobT1 and *El*CobT3 and compared it to the substrate preference of known CobT enzymes in literature. We saw that *Mt*CobT1 has highest preference 5-OHBza unlike any other CobT characterized in literature all of which prefer DMB. *El*CobT3 showed a substrate preference similar to *Se*CobT wherein the most preferred substrate was DMB. Adenine was the least preferred substrate for all the CobT enzymes.

Overall, we see that there are multiple CobT homologs present in obligate anaerobes and they differ in their gene neighbourhoods. An organism can possess more than one functional CobT enzyme and each of them may have differences in the regiospecific activation of lower ligands and their substrate preference. We also see that CobT enzymes are promiscuous and can use unnatural substrates. This can be exploited to produce novel unnatural cobamides which may differ from existing cobamides. Comparison of the sequences of the homologs and further study on factors governing regiospecificity can be carried out which can help tune the CobT enzymes and further produce desirable isomers of cobamides. Similarly tuning the substrate preference to produce desirable lower ligands in cobamides is also to be explored.

References

1. Roth, J. R., Lawrence, J. G. & Bobik, T. A. COBALAMIN (COENZYME B12): Synthesis and Biological Significance. *Annu. Rev. Microbiol.* **50**, 137–181 (1996).

2. Banerjee, R. & Ragsdale, S. W. The Many Faces of Vitamin B 12 : Catalysis by Cobalamin-Dependent Enzymes . *Annu. Rev. Biochem.* **72**, 209–247 (2003).

3. Banerjee, R. The Yin-Yang of cobalamin biochemistry. *Chemistry and Biology* vol. 4 175–186 (1997).

4. Giedyk, M., Goliszewska, K. & Gryko, D. Vitamin B12 catalysed reactions. *Chemical Society Reviews* vol. 44 3391–3404 (2015).

5. Stupperich, E., Eisinger, H.-J. & Schurr, S. Corrinoids in anaerobic bacteria. *FEMS Microbiol. Lett.* **87**, 355–360 (1990).

6. Allen, R. H. & Stabler, S. P. *Identification and quantitation of cobalamin and cobalamin analogues in human feces*. http://www.ajcn.org/.

7. STUPPERICH, E. & NEXØ, E. Effect of the cobalt-N coordination on the cobamide recognition by the human vitamin B12 binding proteins intrinsic factor, transcobalamin and haptocorrin. *Eur. J. Biochem.* **199**, 299–303 (1991).

8. Helliwell, K. E. *et al.* Cyanobacteria and Eukaryotic Algae Use Different Chemical Variants of Vitamin B12. *Curr. Biol.* **26**, 999–1008 (2016).

9. PERLMAN, D. Microbial synthesis of cobamides. *Adv. Appl. Microbiol.* **1**, 87–122 (1959).

10. Shelton, A. N. *et al.* Uneven distribution of cobamide biosynthesis and dependence in bacteria predicted by comparative genomics. *ISME J.* **13**, 789–804 (2019).

11. Mattes, T. A., Deery, E., Warren, M. J. & Escalante-Semerena, J. C. Cobalamin Biosynthesis and Insertion. in *Encyclopedia of Inorganic and*

Bioinorganic Chemistry 1–24 (John Wiley & Sons, Ltd, 2017). doi:10.1002/9781119951438.eibc2489.

12. Escalante-Semerena, J. C. Conversion of cobinamide into adenosylcobamide in bacteria and archaea. *Journal of Bacteriology* vol. 189 4555–4560 (2007).

13. Moore, S. J. & Warren, M. J. The anaerobic biosynthesis of vitamin B12. in *Biochemical Society Transactions* vol. 40 581–586 (2012).

14. Renz, P., Endres, B., Kurz, B. & Marquart, J. *Biosynthesis of vitamin B12 in anaerobic bacteria Transformation of 5-hydroxybenzimidazole and 5-hydroxy-6-methylbenzimidazole into 5,6-dimethylbenzimidazole in Eubacterium limosum*. vol. 217 (1993).

15. Renz, P. Riboflavin as precursor in the biosynthesis of the 5,6-Dimethylbenzimidazole-moiety of vitamin B $_{12}$. *FEBS Lett.* **6**, 187–189 (1970).

16. Campbell, G. R. O. *et al.* Sinorhizobium meliloti bluB is necessary for production of 5,6-dimethylbenzimidazole, the lower ligand of B12. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 4634–4639 (2006).

17. Taga, M. E., Larsen, N. A., Howard-Jones, A. R., Walsh, C. T. & Walker, G. C. BluB cannibalizes flavin to form the lower ligand of vitamin B12. *Nature* **446**, 449–453 (2007).

18. Hazra, A. B., Ballou, D. P. & Taga, M. E. Unique Biochemical and Sequence Features Enable BluB to Destroy Flavin and Distinguish BluB from the Flavin Monooxygenase Superfamily. *Biochemistry* **57**, 1748–1757 (2018).

19. Hazra, A. B. *et al.* Anaerobic biosynthesis of the lower ligand of vitamin B12. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 10792–10797 (2015).

20. P. Mehta, A. *et al.* Anaerobic 5-Hydroxybenzimidazole Formation from Aminoimidazole Ribotide: An Unanticipated Intersection of Thiamin and Vitamin B12 Biosynthesis. *J. Am. Chem. Soc.* **137**, 10444–10447 (2015).

21. WURM, R., RENZ, P. & HECKMANN, G. THE BIOSYNTHESIS OF THE VITAMIN B $_{12}$ -ANALOG 5-ETHOXYBENZIMIDAZOLYLCOBAMIDE IN CLOSTRIDIUM THERMOACETICUM. *FEMS Microbiol. Lett.* **7**, 11–13 (1980).

22. Irion, E. & Ljungdahl, L. Isolation of Factor IIIm Coenzyme and Cobyric Acid Coenzyme plus Other B12 Factors from Clostridium thermoaceticum. *Biochemistry* **4**, 2780–2790 (1965).

23. Maggio-Hall, L. A. & Escalante-Semerena, J. C. *In vitro synthesis of the nucleotide loop of cobalamin by Salmonella typhimurium enzymes*. *PNAS October* vol. 12 www.pnas.org (1999).

24. O'toole, G. A., Trzebiatowski, J. R. & Escalante-Semerenas, J. C. *THE JOURNAL OF BIOIDCICAL CHEMISTRY The cobC Gene of Salmonella typhimurium Codes for a Novel Phosphatase Involved in the Assembly of the Nucleotide Loop of Cobalamin**. vol. 269 (1994).

25. Cheong, C. G., Escalante-Semerena, J. C. & Rayment, I. Capture of a labile substrate by expulsion of water molecules from the active site of nicotinate mononucleotide:5,6-Dimethylbenzimidazole phosphoribosyltransferase (CobT) from Salmonella enterica. *J. Biol. Chem.* **277**, 41120–41127 (2002).

26. Trzebiatowski, J. R. & Escalante-Semerena, J. C. *Purification and Characterization of CobT, the Nicotinate-mononucleotide:5,6-Dimethylbenzimidazole Phosphoribosyltransferase Enzyme from Salmonella typhimurium LT2* The resulting ribotides were incor-porated into cobamides that were differentially utilize*. http://www.jbc.org/ (1997).

27. Cheong, C.-G., Escalante-Semerena, J. C., Rayment, I., Escalante-Semerena, J. & Hall, F. *Structural Investigation of the Biosynthesis of Alternative Lower Ligands for Cobamides by Nicotinate Mononucleotide:5,6-Dimethylbenzimidazole Phosphoribosyltransferase (CobT) from Salmonella enterica. †, ‡ Downloaded from*. http://www.jbc.org/ (2001).

28. Hazra, A. B., Tran, J. L. A., Crofts, T. S. & Taga, M. E. Analysis of substrate specificity in CobT homologs reveals widespread preference for DMB, the lower axial ligand of vitamin B12. *Chem. Biol.* **20**, 1275–1285 (2013).

29. Crofts, T. S. *et al.* Regiospecific formation of cobamide isomers is directed by CobT. *Biochemistry* **53**, 7805–7815 (2014).

30. Anderson, P. J. *et al.* One pathway can incorporate either adenine or

dimethylbenzimidazole as an α-axial ligand of B12 cofactors in Salmonella enterica. *J. Bacteriol.* **190**, 1160–1171 (2008).

31. Keck, B. & Renz, P. Salmonella typhimurium forms adenylcobamide and 2-methyladenylcobamide, but no detectable cobalamin during strictly anaerobic growth. *Arch. Microbiol.* **173**, 76–77 (2000).

32. Keller, S. *et al.* Selective utilization of benzimidazolylnorcobamides as cofactors by the tetrachloroethene reductive dehalogenase of Sulfurospirillum multivorans. *J. Bacteriol.* **200**, (2018).

33. Keller, S. *et al.* Exogenous 5,6-dimethylbenzimidazole caused production of a non-functional tetrachloroethene reductive dehalogenase in Sulfurospirillum multivorans. *Environ. Microbiol.* **16**, 3361–3369 (2014).

34. Munder', M., Vogt', J. R. A., Vogler', B. & Renz', P. *Biosynthesis of vitamin B12 in anaerobic bacteria Experiments with Eubacterium limosum on the incorporation of D-[I-' 3Clerythrose and [' 3C]formate into the 5,6-dimethylbenzimidazole moiety*. *Eur. J. Biochem* vol. 204 (1992).

35. Hazra, A. B., Tran, J. L. A., Crofts, T. S. & Taga, M. E. Analysis of Substrate Specificity in CobT Homologs Reveals Widespread Preference for DMB, the Lower Axial Ligand of Vitamin B12. *Chem. Biol.* **20**, 1275–1285 (2013).

36. Van Den Ent, F. & Löwe, J. RF cloning: A restriction-free method for inserting target genes into plasmids. *J. Biochem. Biophys. Methods* **67**, 67–74 (2006).

37. Crofts, T. S., Seth, E. C., Hazra, A. B. & Taga, M. E. Cobamide structure depends on both lower ligand availability and CobT substrate specificity. *Chem. Biol.* **20**, 1265–1274 (2013).

38. Datar, P. Investigating the regioselective attachment of the lower ligand in Vitamin B12 biosynthesis. (IISER Pune, 2018).

39. Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Wheeler, D. L. GenBank. *Nucleic Acids Res.* **33**, (2005).

40. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

41. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

42. Hall, T. A. BIOEDIT: A USER-FRIENDLY BIOLOGICAL SEQUENCE ALIGNMENT EDITOR AND ANALYSIS PROGRAM FOR WINDOWS 95/98/ NT. (1999) doi:10.14601/PHYTOPATHOL_MEDITERR-14998U1.29.

43. Miller, M. A., Pfeiffer, W. & Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. in *2010 Gateway Computing Environments Workshop, GCE 2010* (2010). doi:10.1109/GCE.2010.5676129.

44. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242-5 (2016).

45. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).