Investigating the regioselective attachment of the lower ligand in Vitamin B_{12} biosynthesis

A thesis submitted by

Prathamesh Madhav Datar

In partial fulfilment of the degree of

BS - MS Dual Degree

IISER PUNE

Department of Chemistry,

Indian Institute of Science Education and Research, Pune

May, 2018

IISER PUNE

Certificate

*This is to certify that this dissertation entitled "**Investigating the regioselective attachment of the lower ligand 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 work carried out by "**Prathamesh Madhav Datar**" at "**IISER Pune**" under the supervision of "**Dr. Amrita Hazra, Assistant Professor, Department of Chemistry**" during the academic year **2017-2018.***

Prathamesh Datar

Signature of Student

Date 2nd May, 2018

Amrita Hazra

Signature of Supervisor

Date 2nd May '2018

IISER PUNE

Declaration

*I hereby declare that the matter embodied in the report entitled "**Investigating the regioselective attachment of the lower ligand in Vitamin B_{12} biosynthesis**" are the results of the work carried out by me at the **Department of Chemistry, IISER Pune** under the supervision of "**Dr. Amrita Hazra**" and the same has not been submitted elsewhere for any other degree.*

Prathyusha

Signature of Student

Date 2nd May, 2018

AmritaHazra

Signature of Supervisor

Date 2nd May '2018

Acknowledgements

The completion of this project would not have been possible without help from a lot of people. I would like to take this opportunity to thank everyone whom I am indebted to.

Firstly, I thank my supervisor, Dr. Amrita Hazra for her intellectual and moral support. Her keen interest in such a subtle topic and her important inputs have shaped this project into what it is today. I am also grateful to Dr. Amrita's funding agencies who have provided financial assistance for this project.

I am grateful to all the members of the Hazra lab-Yashwant, Yamini and Rupali for their valuable suggestions throughout the course of this project. Yashwant helped me a lot with the HPLC and FPLC. He was always available anytime I called him. I am really indebted to him. I am also thankful to Yamini for aiding me with molecular cloning, providing the *El*BzaC clone and her immense help in making the phylogeny tree.

CobT-BzaC interaction studies were performed in collaboration with Prof. Neil Marsh at the University of Michigan, Ann Arbor as a part of my summer internship in 2017. I am grateful to Prof. Marsh for hosting me and for his valuable inputs. I would like to thank Marsh lab members, especially Kyle for helping me with protein purification. I am thankful to Arti and Soumi for their help in performing native PAGE analysis as well as Marie and Somayesadat for teaching me how to perform size exclusion chromatography. Additionally, I thank the IISER - University of Michigan exchange program for the opportunity, the Department of Chemistry at the University of Michigan for financial support during my stay and INSPIRE student fellowship that covered a part of my travel expenses.

I take this opportunity to express my gratitude to my thesis advisor, Dr. Nishad Matange, who provided a more biological and evolutionary perspective to my predominantly chemical idea. His inputs made my project versatile and not just restricted to one discipline or line of thought.

Special thanks to members from Michiko Taga's Laboratory at the University of California, Berkeley – Kenny Mok for providing some important CobT plasmids; Amanda Shelton, Olga Sokolovskaya and Kristopher Kennedy for the providing a list of cobamide producers that was a major source for curating the list of CobT homologs required in making the phylogeny tree. Another special thanks to Dr. Ajeet Singh from NCL venture centre for performing the LC-MS analysis.

Finally, I would like to thank my parents and friends – Soumil and Mishika who kept me motivated throughout the journey and provided emotional support. I am also indebted to IISER, Pune for providing me such a wonderful platform where I could explore my capabilities and learn so much more than mere textbook knowledge.

Prathamesh Madhav Datar

March, 2018

List of abbreviations, gene and protein names

B_{12}	Vitamin B_{12} or Cobalamin
AIR	5'-phosphoribosyl-5-aminoimidazole
DMB	5, 6-dimethylbenzimidazole
Ado-cbi	5’-deoxyadenosyl cobinamide
5-OHBza	5-hydroxybenzimidazole
5-OMeBza	5-methoxybenzimidazole
5-OMe, 6-MeBza	5-methoxy, 6-methylbenzimidazole
DMB-RP	5, 6-dimethylbenzimidazole riboside phosphate
NaMN	Nicotinate mononucleotide
NMN	Nicotinamide mononucleotide
*Se*CobT	CobT homolog from *Salmonella enterica*
5-OHBza-RP	5-hydroxybenzimidazole riboside phosphate
6-OHBza-RP	6-hydroxybenzimidazole riboside phosphate
5-OMeBza-RP	5-methoxybenzimidazole riboside phosphate
6-OMeBza-RP	6-methoxybenzimidazole riboside phosphate
HPLC	High performance liquid chromatography
UV-Vis	Ultraviolet-visible
LC-MS	Liquid chromatography – mass spectrometry
*El*CobT	Nicotinate-nucleotide-5, 6-dimethylbenzimidazole phosphoribosetransferase (CobT) homolog from *Eubacterium limosum*
*Ec*CobT	CobT homolog from *Escherichia coli*
*Sm*CobU	Nicotinate-nucleotide-5, 6-dimethylbenzimidazole phosphoribosetransferase (CobU) homolog from *Sinorhizobium meliloti* (in aerobes, CobT is annotated as CobU)
*El*BzaC	BzaC homolog from *Eubacterium limosum*

*El*CobU	Adenosylcobinamide-phosphate guanylyltransferase (CobU) homolog from *Eubacterium limosum*
*El*CobS	Cobalamin 5'-phosphate synthase (CobS) homolog from *Eubacterium limosum*
RF cloning	Restriction free cloning
PCR	Polymerase chain reaction
Native PAGE	Native polyacrylamide gel electrophoresis
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
BSA	Bovine serum albumin
ESI-MS	Electron spray ionization – mass spectrometry

Abstract

Vitamin B_{12} (cobalamin), a member of the cobamide family of cofactors, is important for many organisms including humans, and facilitates diverse metabolic processes such as the biosynthesis and catabolism of amino acids, fatty acid metabolism, modification of tRNA, reductive dehalogenation and acetogenesis. The structure of Vitamin B_{12} and other cobamides consists of a tetrapyrrolic corrin ring, with a central cobalt ion, and an upper and a lower ligand. The upper ligand in cobamides is involved in the activity of the cofactor – catalyzing methyl transfer reactions and radical-based rearrangement reactions. Different cobamide cofactors vary mainly in the composition of their lower ligand which is covalently attached to the corrin ring via a nucleotide loop and may coordinate to the cobalt ion. 5, 6-dimethylbenzimidazole (DMB) is the lower ligand of Vitamin B_{12}.

The biosynthesis of cobamides can occur via an aerobic or an anaerobic route and typically consists of three main steps – the synthesis of the corrin ring, synthesis of the lower ligand and finally, their attachment to form the complete cobamide. The last step is initiated by CobT, an enzyme that couples the lower ligand to a ribophosphate derivative. This is further coupled to the nucleotide loop of the corrin ring with the enzymes CobU, CobS and CobC.

The anaerobic biosynthesis of DMB was studied in the anaerobic B_{12} producer *Eubacterium limosum* and found to derive from the purine biosynthesis pathway intermediate 5'-phosphoribosyl-5-aminoimidazole (AIR). Despite being a symmetric molecule, labelling studies in the *E. limosum* show that DMB is attached selectively via one nitrogen. This suggests that the mechanism of attachment of the lower ligand is regioselective in nature. In my thesis project, I focus on investigating the mechanism by which regioselectivity is conferred in the attachment of the lower ligand of B_{12}.

In *E. limosum*, enzymes of the *bza* operon sequentially convert AIR to DMB via three asymmetric precursors, 5-hydroxybenzimidazole, 5-methoxybenzimidazole and 5-methoxy, 6-methylbenzimidazole. The enzyme CobT, also a part of the *bza* operon in

E. limosum, attaches the lower ligand to the corrin nucleotide loop. To explain the selective attachment of one nitrogen in DMB, we hypothesized that either an asymmetric precursor of DMB may be the substrate for CobT, or that enzyme complex of the *bza* operon interacts with CobT, causing anisotropy in DMB biosynthesis that leads to the regioselectivity.

In this study, we show that the *E. limosum* CobT is regioselective for certain asymmetric benzimidazole ligands, and compared it with the activity of *Escherichia coli* CobT. Further, using bioinformatic analysis and the crystal structure of the *Salmonella enterica* CobT , we have identified key residues that are involved in the regioselective attachment of lower ligands and demonstrated that these residues affects the selectivity displayed by CobT. We also show that factors such as pH or enzyme-enzyme interactions have little or no effect on the regioselectivity of CobT.

Our study facilitates the understanding of the molecular basis of attachment of the lower ligand in cobamide biosynthesis. This will contribute to research on mechanisms by which microbes create cobamide diversity by varying the lower ligand. Our analysis not only deepens the understanding of the mechanism of CobT, but also makes it amenable to protein engineering so as to improve the production of Vitamin B_{12} and other cobamide cofactors.

Contents

List of Figures and Tables

Chapter 1

Introduction

1.1 Vitamin B_{12}: Structure

Cobalamin or Vitamin B_{12} (B_{12}) is an essential vitamin for many living organisms, including humans and is synthesized only by a subset of prokaryotes. It belongs to the cobamide family of co-factors and comprises of a central cobalt ion bound to a tetrapyrrolic corrin ring. An upper and a lower ligand are also bonded to the cobalt ion. Additionally, the lower ligand is covalently attached to the tetrapyrrolic ring via a nucleotide loop (Figure 1A) [1].

The upper ligand varies depending on the type of reaction being catalyzed. It is either a methyl group for reactions involving methyl transfers or a 5'-deoxyadenosine group to catalyze radical rearrangements [2]. Similarly, structurally diverse molecules feature

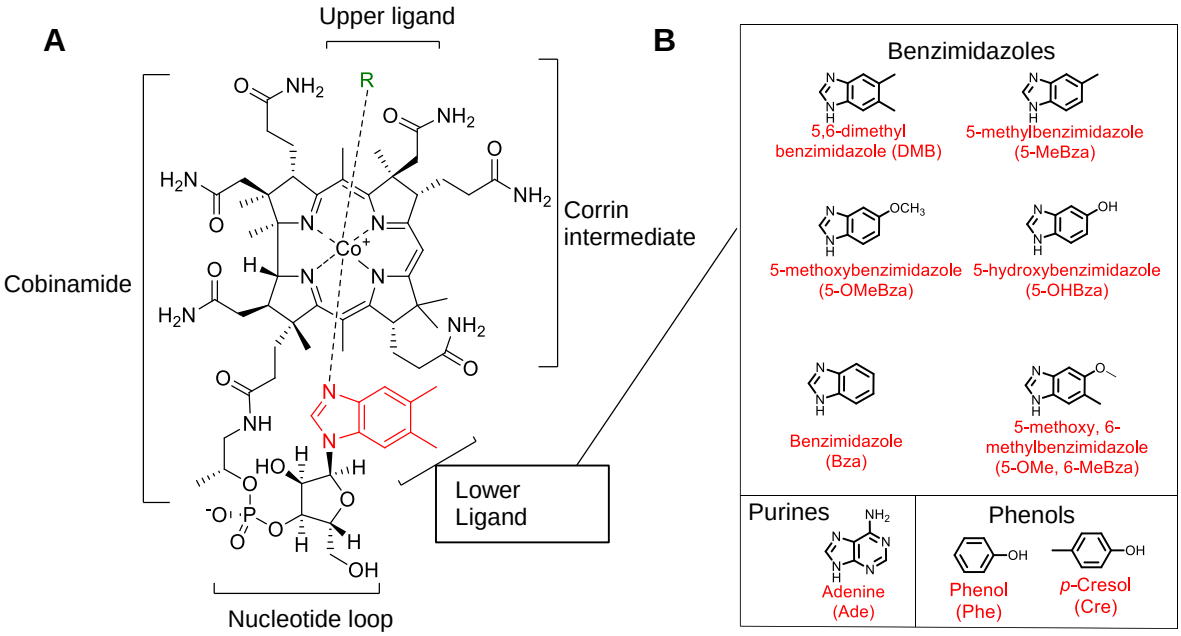

Figure 1: Cobamide structure and lower ligand diversity

(A) Cobalamin (B_{12}) structure consisting of a central cobalt ion and a tetrapyrrolic corrin ring. The upper ligand denoted by R can either be a methyl group or a 5'-deoxyadenosine group. The lower ligand, 5, 6-dimethylbenzimidazole (DMB, red) is covalently attached to the corrin ring via a nucleotide loop. (B) Lower ligand diversity seen in cobamides. Primarily three classes of small molecules feature as lower ligands for cobamides- benzimidazoles, purines and phenols.

as lower ligands in cobamides. These can be classified predominantly into three categories – benzimidazoles, purines and phenols (Figure 1B). For B_{12}, the lower ligand is 5, 6-dimethylbenzimidazole (DMB) [3].

1.2 Vitamin B_{12}: Biosynthesis

The biosynthesis of B_{12} can be divided into three parts – the synthesis of the corrin intermediate, 5'-deoxyadenosyl cobinamide (Ado-Cbi); synthesis of DMB and finally, the attachment of DMB to Ado-Cbi via a nucleotide loop [1]. Ado-Cbi and DMB can be synthesized both aerobically as well as anaerobically. Under aerobic conditions, DMB is synthesized by the breakdown of reduced flavin mononucleotide ($FMNH_2$) and is catalyzed by BluB [4]. The anaerobic biosynthesis pathway for DMB was recently discovered in *Eubcterium limosum* [5]. It comprises of the *bza* operon, whose gene products, BzaA, BzaB, BzaC, BzaD and BzaE, sequentially convert 5'-phosphoribosyl-5-aminoimidazole (AIR) to DMB through three intermediates: 5-hydroxybenzimidazole (5-OHBza), 5-methoxybenzimidazole (5-OMeBza) and 5-methoxy, 6-

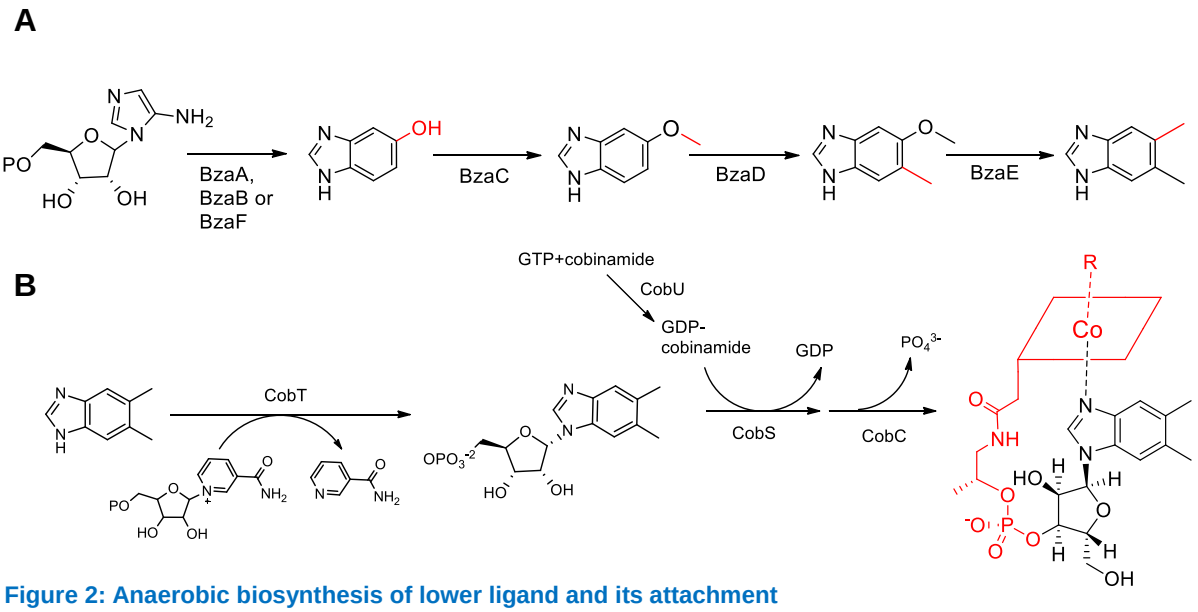

Figure 2: Anaerobic biosynthesis of lower ligand and its attachment

(A) Anaerobic biosynthesis of DMB by the gene products of the *bza* operon. BzaA and BzaB or BzaF convert aminoimidazole ribotide (AIR) to 5-OHBza. BzaC, BzaD and BzaE carry out successive methylations to obtain DMB. (B) Attachment of lower ligand and the cobinamide to form cobamide. CobT activates the lower ligand by forming a riboside phosphate (RP). CobU attaches a GDP group to cobinamide. CobS covalently links the lower ligand RP to cobinamide. CobC cleaves off the 5'- phosphate group of lower ligand RP.

methylbenzimidazole (5-OMe, 6-MeBza) (Figure 2A). In some organisms, the BzaA and BzaB are replaced by another homolog, BzaF [5].

The enzymes CobT, CobU, CobS and CobC are involved in attachment of DMB to Ado-Cbi via the nucleotide loop (Figure 2B). CobT forms an α-DMB riboside phosphate (DMB-RP) by reacting DMB with nicotinate nucleotide (NaMN) or nicotinamide mononucleotide (NMN), CobU forms GDP-Ado-Cbi from Ado-Cbi and CobS attaches DMB-RP with GDP-Ado-Cbi to create cobalamin 5'-phosphate. CobC is the final enzyme in the pathway that cleaves off the phosphate group [6].

1.3 CobT and regioselectivity

Nicotinate-nucleotide-5, 6-dimethylbenzimidazole phosphoribosyl transferase, annotated as CobT, has remained a curious enzyme in the eyes of biochemists because of its ability to form an unusual α-glycosidic bond between ribose and DMB, unlike the β-glycosidic bond which is prevalent in biomolecules [7]. In organisms that produce cobamides aerobically, CobT is annotated as CobU [1].

The crystal structure for CobT from *Salmonella enterica* (*Se*CobT) allowed for elucidation of its reaction mechanism (Figure 3A) [8]. The proposed mechanism involves the E317 residue in the active site of *Se*CobT orients the lower ligand, (in this case, 5-OMeBza) such that the protonated nitrogen of 5-OMeBza, N3 is close to it. Later, the glutamate residue abstracts the proton over N3 and the anion tautomerizes to N1 which attacks C-1' of NMN, thus displacing the nicotinamide.

The imidazole N's of DMB can tautomerize [9]. While the tautomers of DMB are the same molecule, in asymmetric benzimidazoles like 5-OMeBza or 5-OHBza, these tautomers can exist as two separate constitutional isomers (Figure 3B). This suggests that if an asymmetric lower ligand, such as 5-OMeBza reacts with NMN/NaMN, one can obtain both α-5-OMeBza riboside phosphate (5-OMeBza-RP) and α-6-OMeBza riboside phosphate (6-OMeBza-RP) (Figure 3C). It is also possible that CobT displays regioselectivity in forming one product over the other.

Crofts and Hazra et al (2013) showed that two distinct isomeric products were formed in CobT reactions when 5-OHBza or 5-OMeBza were used as substrates [10]. The product isomers were characterized using high performance liquid chromatography

(HPLC), ultraviolet-visible absorption spectroscopy (UV-Vis), nuclear magnetic resonance spectroscopy (NMR) and liquid chromatography-mass spectrometry (LC-MS). It was observed that CobT homologs from different organisms had differences in regioselectivity. Interestingly, the UV-Vis absorption spectrum for each isomer is different. For 5-OMeBza-RP, the absorbance at 250 nm is greater than 280 nm, whereas for 6-OMeBza-RP, the opposite is true. Thus, they can be easily detected on an HPLC equipped with a UV-Vis spectrophotometer.

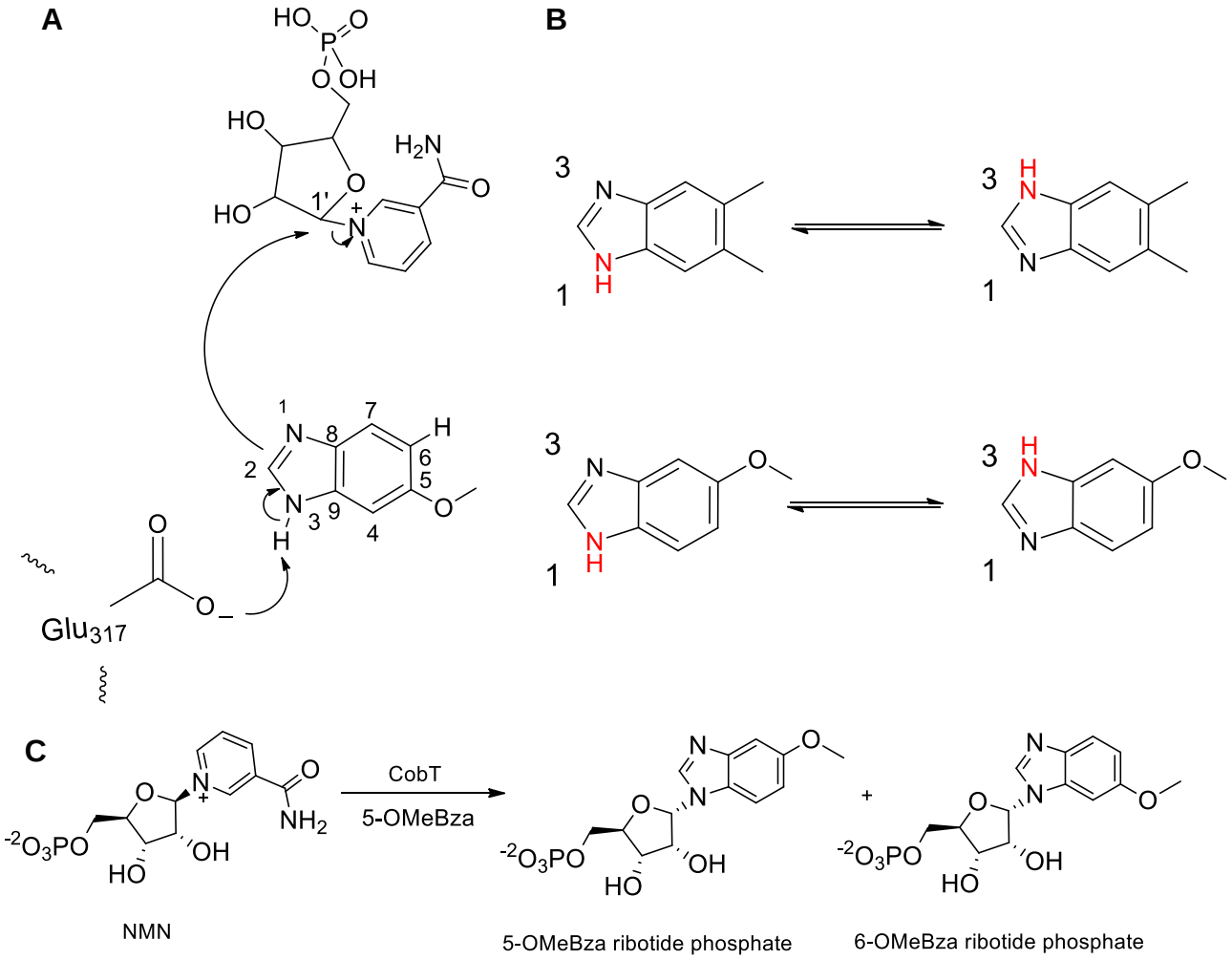

Figure 3: CobT reaction mechanism and possible products

(A) Reaction mechanism of *Se*CobT (as suggested by Cheong et al., 2002). E317 residue abstracts a proton from N3 of 5-OMeBza and N1 attacks C-1' of NMN to form 5-OMeBza-RP. (B) Symmetric tautomers of DMB but asymmetric tautomers of 5-OMeBza. (C) Two products possible if an asymmetric lower ligand, such as 5-OMeBza, reacts with NMN in the presence of CobT. In one case the proton would be on N1 and in the other case, it would be on N3.

1.4 Regiospecificity and regioselectivity

There exists confusion about the usage of the terms specificity and selectivity [11]. Historically, in a reaction where there exists a possibility of forming multiple products, if one product is formed in excess over the others, the reaction is called regioselective [11][12]. If one product is formed exclusively, and there is no scope of formation of the other product, because the mechanism is as such, then the reaction is called regiospecific [12][13].

Both tautomers of asymmetric lower ligands such as 5-OHBza or 5-OMeBza are constitutional isomers [13]. The reaction mechanism of CobT implies that each tautomer of the reactant would selectively give one product (such as 5-OMeBza with the proton on the N3 would exclusively make 5-OMeBza-RP and vice versa) (refer Figure 3A, B). Therefore, the reaction can be called regiospecific with respect to the tautomer that is used.

However, any asymmetric benzimidazole in solution would always have a mixture of both reactant tautomers and thus, there exists a possibility of both products being formed. Unless we are able to isolate each tautomer separately and then prove that a given tautomer exclusively forms a given product, the reaction cannot be called ‘regiospecific’. Thus, we have decided to use the more generic term ‘regioselectivity’ while studying the nature of CobT reactions.

Apart from CobT’s several enzymes have been reported to have a regioselective activity. An O-methyltransferase isolated from *Streptomyces* sp. KCTC 0041BP was found to display 100% regioselectivity in attaching the methyl group to the 4’-OH of flavonoids such as quercetin and luteolin [14]. Enhancing regioselectivity of enzymes can be a very useful bioengineering tool, with direct applications in biocatalysis and medicinal chemistry. With the help of single point mutations, the enzyme CYP260A1 from *Sorangium cellulosum* strain So ce56 was made regioselective for the hydroxylation of progesterone to obtain either 1α-hydroxy-progesterone or 17α-hydroxy-progesterone [15].

Such examples display how studying regioselectivity in enzymes is not only a worthwhile scientific exercise, but also has the potential to be used as an efficient biocatalysis tool.

Chapter 2

Project goals and scope

2.1 *Eubacterium limosum* and regioselectivity

In 1980-1990s, Renz and co-workers carried out a series of feeding studies with labelled C and N substrate on the anaerobic B_{12} producer, *Eubacterium limosum*, to identify the origin of each atom in DMB [9][17][18][26][27]. The discovery of the *bza* operon in *E. limosum* then allows us to correlate the findings from the feeding studies [5][22].

Identification of the source for each of the N's in anaerobic DMB biosynthesis demands a closer look at the biosynthesis of AIR. We observe that the source for one imidazole N of AIR is L-glutamine and for the other it is L-glycine [16]. AIR further gets converted to DMB and therefore we expect the source of the two N's for DMB to be L-glutamine and L-glycine. The labelling studies on DMB biosynthesis in *E. limosum* agree with this [17][18].

Additionally, Renz and co-workers observed that DMB attaches to ribose selectively with the N that originates from L – glycine. This was an intriguing result because DMB is a symmetric molecule. Thus the two forms owing to tautomerization between the nitrogen atoms are essentially the same and hence, should attach equally well. However, this did not seem to be the case. My thesis focuses on understanding the possible mechanisms of the regioselectivity in lower ligand attachment and attempts to understand its molecular basis.

2.2 Possible hypothesis and goals

2.2.1 Asymmetric lower ligand as CobT substrate

We hypothesized that it is not DMB, but one of its asymmetric precursors which acts as a substrate for CobT. We observe that there are in total three asymmetric precursors in DMB biosynthesis (Figure 2A). Bioinformatics on BzaC, BzaD and BzaE reveal that they are methyl transferases. Specifically, BzaC, BzaD and BzaE each has

an S-adenosyl methionine (SAM) binding site, and BzaD and BzaE additionally have a B_{12} binding site [5]. It is possible that for BzaD and BzaE, SAM is the methylating agent and a cobamide with bound lower ligand is the substrate, instead of a free base. Thus, 5-OHBza or 5-OMeBza are more plausible substrate choices for CobT than 5-OMe, 6-MeBza. They are attached to the corrin intermediate through CobU, CobT, CobS and CobC (as shown in Figure 2B). Eventually, BzaD and BzaE act upon the cobamide thus formed to convert the lower ligand into DMB.

If the CobT is responsible for the regioselectivity, the reaction of CobT with 5-OHBza or 5-OMeBza should result in the preferential formation of one of the two isomers, with 100% or near 100% formation of a single product, entirely because CobT has a very specific active site.

To verify this, we need to purify *Eubacterium limosum* CobT (*El*CobT) and reconstitute its activity. To validate the reaction of *El*CobT, we will test CobT homologs from other organisms reported in literature such as the CobT from *Sinorhizobium meliloti* (called *Sm*CobU), and ones that are not reported such as the CobT from *Escherichia coli* (*Ec*CobT)

2.2.2 Enzyme-enzyme interactions leading to selectivity

Another possibility is that enzymes involved in the attachment or biosynthesis of DMB interact with CobT and lead to selectivity, either by associating with CobT and affecting its active site or by transferring the synthesized lower ligand directly to CobT, thus maintaining any anisotropy in the biosynthesis.

To verify the second hypothesis, we need to purify BzaC, CobU and CobS from *Eubacterium limosum* (*El*BzaC, *El*CobU and *El*CobS) and study whether these proteins show any specific association. For this, we will perform native polyacrylamide gel electrophoresis (Native-PAGE) as well as size exclusion chromatography to analyze any interactions that may exist between these proteins.

2.2.3 Overall approach and scope

Our overall approach is to identify potential factors such as pH, enzyme interactions, and specific amino acid residues that lead to selectivity in different enzymes and studying the effect of the factors on an enzyme's regioselectivity. The scope of my project extends from probing the factors that confer molecular specificity in the biosynthesis of Vitamin B_{12} to understanding the factors that contribute towards the regioselective prowess of enzymes. In our case, altering regioselectivity and substrate specificity can be very useful in generating CobT enzymes that would eventually be able to attach different lower ligands, increasing the scope of native corrinoids that can be produced by an organism [3].

Chapter 3

Materials and methods

3.1 Materials

All chemicals and media components were obtained from Himedia, Rankem, Sigma-Aldrich, TCI or Fisher Scientific unless otherwise mentioned. Cloning vectors were a gift from Prof. Michiko Taga's laboratory at UC Berkeley.

3.2 Experimental methods

3.2.1 Molecular cloning

Genes were cloned from the genomic DNA of the respective organism for *El*CobU, *El*CobS and *Ec*CobT. *El*CobT was sub-cloned from pTH1227 vector into pET-28a(+) vector. All vectors were cloned using restriction free (RF) cloning.

Polymerase chain reactions (PCR) for DNA amplification were performed using PrimeSTAR GXL DNA polymerase kit from Takara. Conditions for PCR's were determined using Takara's manual. Primers were designed using SnapGene viewer and obtained from Sigma-Aldrich.

Mutants were prepared using forward primer for T7 promoter and reverse primer for the respective mutation site. A comprehensive table of all the primers used in this study can be found in Table 1.

Primer Name	Sequence
*Ec*cobT_F	5'-ATGCAAATACTTGCCGATTTACTGAATACG-3'
*Ec*cobT_R	5'-TAACTGTTCAAATCAGAAGTCGTAT-3'
*Ec*cobT_28a_F	5'-GGTGCCGCGCGGCAGCCATATGCAAATACTTGCCGATTTACTGAATACG-3'

*Ec*cobT_28a_R	5'-CGGAGCTCGAATTCGGATCCGCGTTAACTGTTCAAATCAGAAGTCGTAT-3'
*El*cobT_28a_F	5'-GGTGCCGCGCGGCAGCCATATGAGTGCATTAACTTTAAAAGAAGTTATTG-3'
*El*cobT_28a_R	5'-TGTCGACGGAGCTCGAATTCGGATCCGCGTTAAGCGAGGCCATTGGTGAC-3'
*El*_cobU_F	5'-ATGGGAAGCTTAGTATTTGTAACCGG-3'
*El*_cobU_R	5'-TTATTTCAATTTCATCGGAATCCCT-3'
*El*_cobU_28a_F	5'-GGTGCCGCGCGGCAGCCATATGGGAAGCTTAGTATTTGTAACCGG-3'
*El*_cobU_28a_R	5'-CGGAGCTCGAATTCGGATCCGCGTTATTTCAATTTCATCGGAATCCCT-3'
*El*_cobS_F	5'-ATGCGCAAAATACTGATCGCCATT-3'
*El*_cobS_R	5'-TCACGCTGCCAGAAGCATCC-3'
*El*_cobS_28a_F	5'-GGTGCCGCGCGGCAGCCATATGCGCAAAATACTGATCGCCATT-3'
*El*_cobS_28a_R	5'-CGGAGCTCGAATTCGGATCCGCGTCACGCTGCCAGAAGCATCC-3'
*El*CobT_F81S_R	5'-CTCCTGAGGAGAGGCCGAGAC-3'
*El*CobT_M89Q_R	5'-ATTGGCGACCTGCTGCATGGT-3'
*Sm*CobU_M93Q_R	5'-GTTCTCGACCTGCTGCGCCGTC-3'
*Se*CobT_SQ_wt_R	5'-ATTCGCCGCCTGAATCGCCGTCACGATTTTGGGCGAAACCGCTAC-3'

Table 1: List of primers

Following RF cloning, the PCR products were digested with DpnI (New England Biolabs) and transformed into *E. coli* DH5α chemically competent cells. Purified plasmids were obtained with the help of Qiagen mini-prep kit.

*El*BzaC was obtained from Yamini Mathur, IISER Pune. *Sm*CobU, *Sm*CobU F85S, *Sm*CobU M93Q and *Se*CobT S80F, Q88M were obtained through the courtesy of Michiko Taga at University of California, Berkeley.

3.2.2 Protein expression and purification

Purified plasmids were transformed into chemically competent *E coli* BL21 (DE3) cells. Transformed cells were spread on LB Agar petri plates containing 50µg/ml Kanamycin and incubated at 37°C overnight. Single colonies were inoculated in LB broth with 50µg/ml Kanamycin. Cultures were incubated at 37°C till an OD600 of about 0.6,

induced with 250µM isopropyl β-D-1-thiogalactopyranoside (IPTG) and incubated at 25^{o}C. After 16 hrs, cells were harvested by centrifugation at 6000 rpm. Pellets were resuspended in lysis buffer (50mM K_2HPO_4 + KH_2PO_4, 300mM NaCl, pH 8) and sonicated (1 sec on, 3 sec off at 60% amplitude, 7-10 mins). The sonicated lysate was centrifuged at 18,000 rpm for 45 min at 4^{o}C. Supernatant was loaded on a GE Histrap 5ml pre-packed Nickel affinity column and eluted on AKTA Pure chromatography system using 250mM imidazole. Pure fractions were identified using SDS-PAGE and pooled together. The protein was desalted in 50mM Tris-HCl buffer (pH 8) using BioRad Econo-Pac 10DG columns and stored at -80^{o}C with 0.05% beta-mercapto ethanol and 10% Glycerol.

3.2.3 Synthesis of 5-OHBza

27 ml of 40% HBr was added to 600 mg 5-OMeBza. The mixture was refluxed at 105^{o}C overnight and monitored using thin layer chromatography (MeOH:$CHCl_3$:$N(Et)_3$ = 4:16:0.1). The reaction mixture was neutralized with NaOH to achieve pH 8, rota-evaporated, extracted in Methanol and purified on a silica column using 5% MeOH in $CHCl_3$. 400 MHz ^{1}H NMR (ppm, splitting): 6.8, d (1H); 7, s (1H); 7.45, d (1H), 8, s (1H). 400 MHz ^{13}C NMR (ppm): 151.9, 142.1, 116.4, 112.4, 119.7.

3.2.4 Enzymatic reactions

All CobT only reactions consisted of 20µM CobT, 2mM NMN and 500µM lower ligand in 50mM Tris-HCl buffer at pH 8.5 with 10mM $MgCl_2$. For enzyme-enzyme interactions based reactions, 6µM *El*CobT, 2mM NMN and 500µM lower ligand were mixed in 50mM Tris-HCl, pH 7.5 with 10mM $MgCl_2$. Wherever applicable, bovine serum albumin (BSA) was added to a final concentration of 6, 60 and 600 µM. Similarly, *El*CobU and *El*BzaC were added to a final concentration of 6, 30 and 60 µM. For pH based experiments, NaH_2PO_4, Na_2HPO_4 and Na_3PO_4 were used to prepare 50mM solutions of pH 6.5, 7.5, 8.5 and 9.5. About 20 µM of *El*CobT was added, keeping the concentrations of NMN and lower ligands the same. All reactions were incubated at 25^{o}C for 4 hrs. Subsequently, they were quenched with 4% formic acid and centrifuged at 14,000 rpm for 45 min. The supernatants were analyzed using HPLC.

3.2.5 Purification of α-riboside phosphate products

The reactions were purified and analyzed on an Agilent 1260 Infinity-II series high performance liquid chromatography system equipped with UV-Vis diode array detector. For initial *El*CobT only reactions, *El*CobT pH based reactions and enzyme-enzyme interaction based reactions, a Phenomenex Luna (dimensions 250x4.6mm, 5µm C-18(2)) column was used for separation. For all other reactions, a Phenomenex Gemini (dimensions 250x4.6mm, 5µm NX-C18) column was used. Both columns were used at a flow rate of 1ml/min. Mobile phase consisted of: 10mM ammonium acetate (CH_3COONH_4) pH 6.5 (A) + MeOH (B). Samples were eluted using the following method: 0% B over 3 min, 0 – 15% B over 2 min, 15-50% B over 9 min, 50-70% B over 3 min

3.2.6 Liquid chromatography – mass spectrometric analysis

Wherever indicated, purified products were collected, lyophilized to about one-fifth the initial volume and submitted for LC-MS analysis at CAMS, NCL Venture Centre. Similarly, entire reaction mixtures too were submitted for LC-MS analysis.

3.2.7 Native PAGE

Native PAGE was performed on BioRad Mini-PROTEAN precast 4%-20% gradient polyacrylamide gels in Tris-Glycine buffer at a potential difference of 25-35 V. After 12-16 hrs, the gels were stained with coomassie brilliant blue dye and analyzed.

3.2.8 Size exclusion chromatography

Size exclusion chromatography was performed on GE Superdex 200 pre-packed column using AKTA Prime chromatography system. 50mM Tris-HCl buffer, pH 7.5 was used for elution at a flow rate of 0.3 ml/min.

3.2.9 SDS-PAGE

SDS PAGE was performed on samples using standard protocol [19].

3.3 Theoretical methods

3.3.1 Phylogeney tree

A list of FASTA formats for different CobT homologs was curated using NCBI database. An eight character unique code was obtained for each entry in the list using Microsoft Excel. The sequences were then aligned using MUSCLE and viewed/edited using BioEdit. The BioEdit file was uploaded as a .phylip4 file on CIPRES gateway. Phylogenetic tree was generated using RaXML HPC2 on XSEDE. The tree was viewed/edited using MEGA7 and iTOL [20].

3.3.2 Modelling protein structures

Protein structures for *El*CobT, *Ec*CobT and *Sm*CobU were modelled using PHYRE^2 [21]. The structures were modelled on the templates of crystal structures 6B5F (*Yersinia enterocolitica* 8081), 1JHP (*Salmonella enterica)* and 1J33 (*Thermus thermophilus*) respectively.

Chapter 4

Results and discussions

4.1 Is *El*CobT regioselective?

Based on previous data, if NMN reacts with an asymmetric lower ligand in the presence of any CobT, there is a possibility of obtaining two isomers, both of which can be separated on an HPLC [10]. We cloned, heterologously expressed *El*CobT in *E. coli* and subsequently purified it using metal affinity chromatography (Figure 4A and B). Later, we reconstituted its activity with DMB and analyzed the reaction on HPLC. As expected, only one product peak was observed corresponding to DMB-RP (Figure 4C).

To test the regioselectivity of *El*CobT we examined its activity with 5-OMeBza. It was observed that the reaction yielded two products in presence of *El*CobT, as observed on the HPLC chromatogram (Figure 5A). The ultraviolet-visible (UV-Vis) absorption spectra for the products obtained with 5-OMeBza were identical to 5-OMeBza-RP and 6-OMeBza-RP reported previously (Figure 5B) [10]. Additionally, liquid

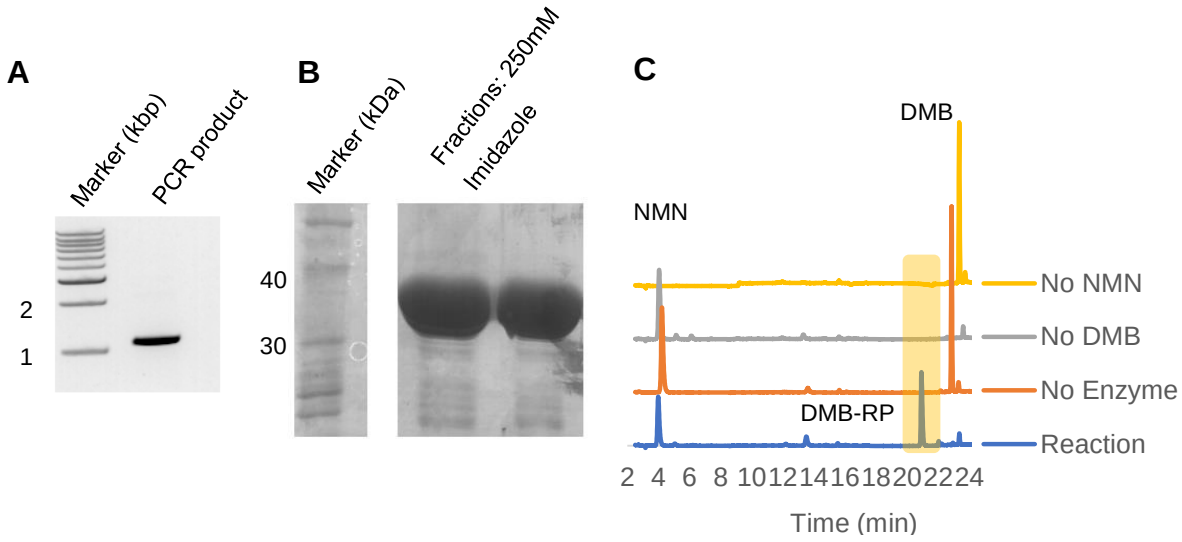

Figure 4: Cloning, purification and activity of *El*CobT

(A) Agarose gel of *El*CobT gene amplified using PCR. (B) SDS-PAGE of purified *El*CobT using Ni metal affinity chromatography. (C) HPLC chromatogram for *El*CobT activity with DMB and appropriate controls. Peaks corresponding to NMN, DMB and the product DMB-RP have been labelled.

chromatography-mass spectrometry (LC-MS) analysis detected accurate masses corresponding to the reactants and products (Figure 6A, B and C). The analysis strongly suggested that *El*CobT does not have exclusive regioselectivity for 5-OMeBza. (The above study was performed prior to the commencement of my MS thesis project.)

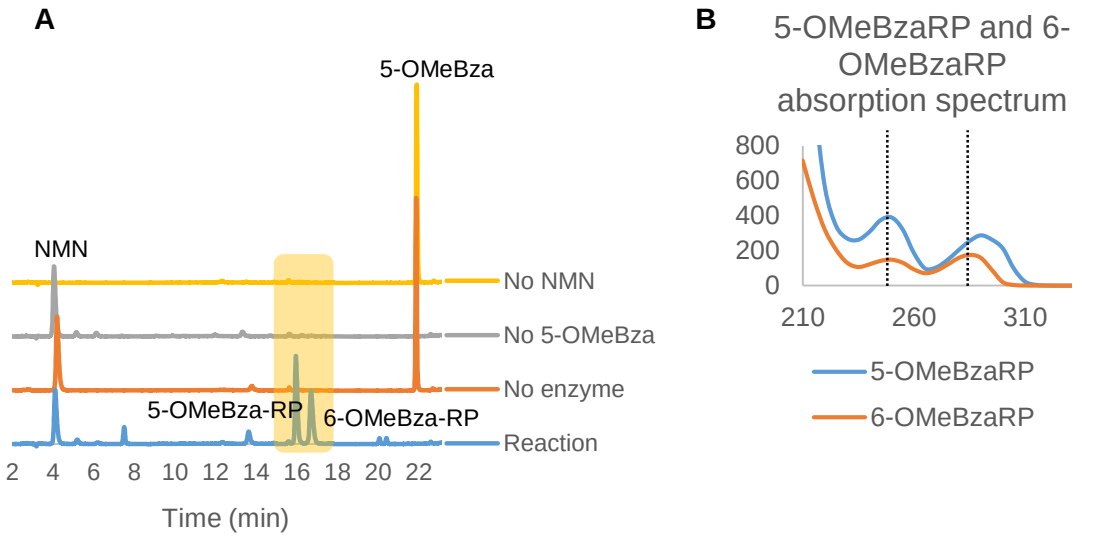

Figure 5: *El*CobT activity with 5-OMeBza

(A) HPLC chromatogram for *El*CobT activity with 5-OMeBza and appropriate controls. Peaks corresponding to NMN, 5-OMeBza and the product isomers have been labelled. (B) Absorption spectrum of both product peaks. We observe that the products have distinct absorption with A250>A280 for 5-OMeBza-RP and A250<A280 for 6-OMeBza-RP (Crofts and Hazra et al, 2014).

Later, we chemically synthesized 5-OHBza from 5-OMeBza and tested its activity with *El*CobT. The HPLC chromatogram displays formation of a single peak. The UV-Vis absorption spectrum corresponds to that of α-5-OHBza riboside phosphate (5-OHBza-RP) reported earlier (Figure 7A and B) [10]. However, nothing can be commented on the regioselectivity for 5-OHBza, as there are two possibilities – either *El*CobT is indeed regioselective for 5-OHBza and thus only one product isomer is obtained in its reaction with NMN or both the isomers are formed but not resolved in the HPLC chromatogram. Though the UV-Vis absorption spectrum for the peak corresponded to 5-OHBza-RP one can argue that the spectrum for α-6-OHBza riboside phosphate (6-OHBza-RP) is merged with it and cannot be distinguished because it is produced in a lesser amount.

To prove that the 5-OHBza product peak corresponds to a single product, we would need to analyze the reaction of 5-OHBza and NMN in the presence of a CobT homolog

which produces both product isomers for 5-OHBza. The HPLC analysis would need to be performed under the same conditions as that of *El*CobT.

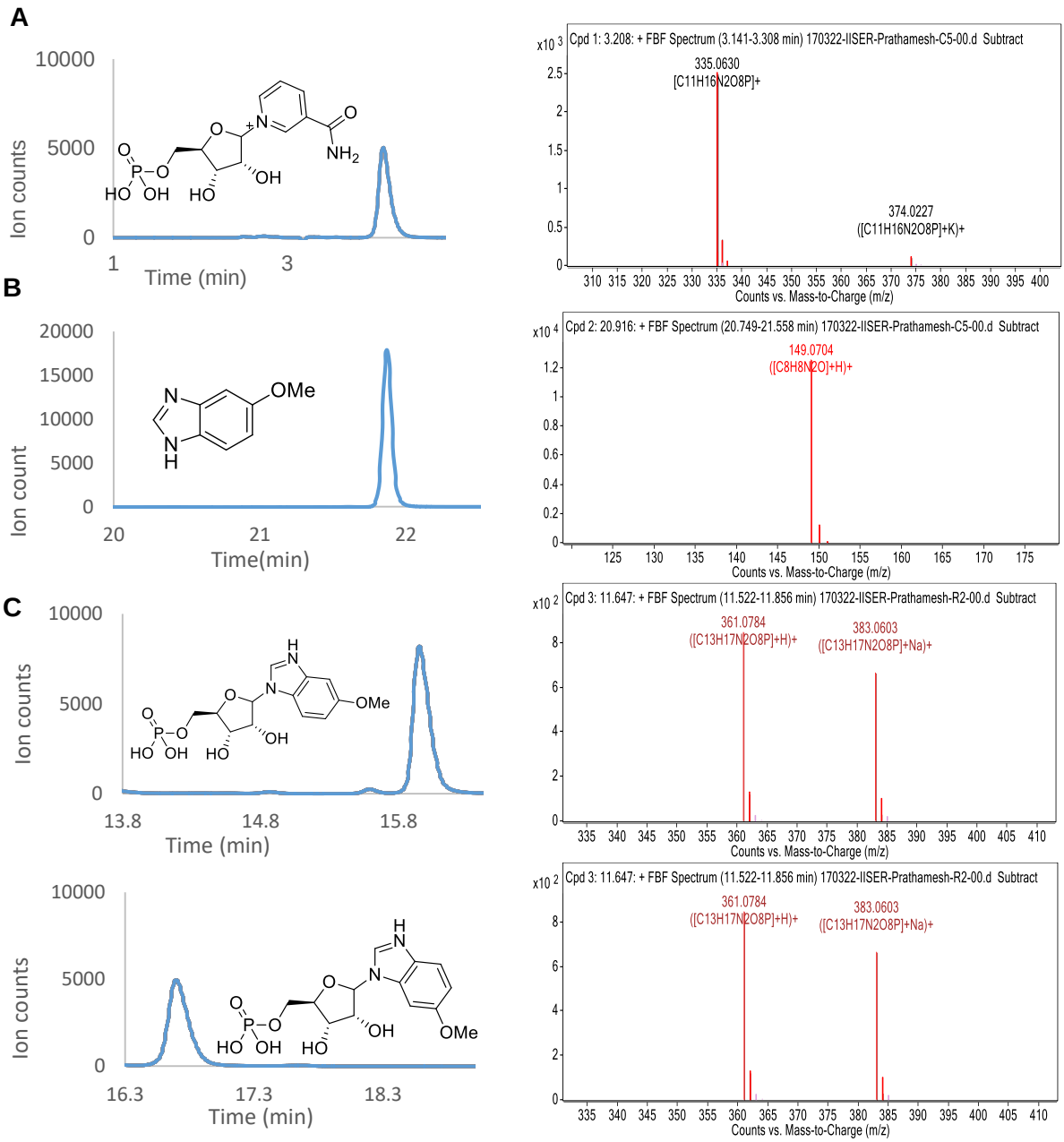

Figure 6: LC-MS analysis of *El*CobT activity with 5-OMeBza

Extracted ion chromatograms (EIC) and electron spray ionization-mass spectrometry (ESI-MS) analysis of NMN (A), 5-OMeBza (B) and products from the reaction of *El*CobT with 5-OMeBza(C).

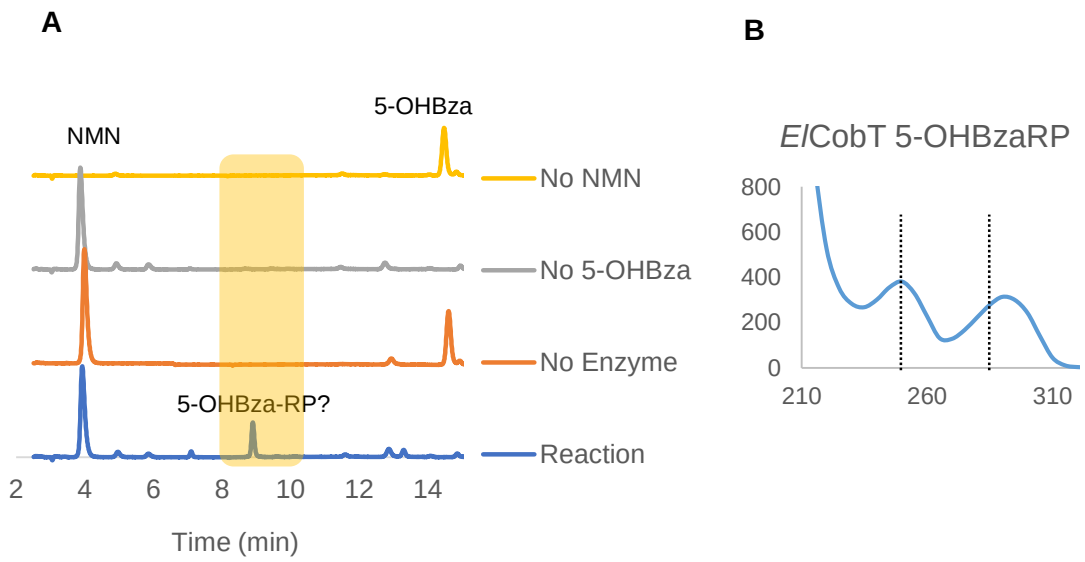

Figure 7: *Ef*CobT activity with 5-OHBza

(A) HPLC chromatogram for *Ef*CobT activity with 5-OHBza and appropriate controls. Peaks corresponding to NMN, DMB and the product have been labelled. (B) Absorption spectrum of the product peak. We observe that the product has distinct absorption with A250>A280, which most likely corresponds to 5-OHBza-RP. However, it is possible that the other peak is merged and not visible.

4.2 *Ec*CobT proves the regioselectivity of *Ef*CobT

In order to verify that *Ef*CobT indeed provides a single product with 5-OHBza, we cloned *Ec*CobT, expressed it and purified it using metal affinity chromatography (Figure 8A and B). On reconstitution if its activity with 5-OHBza and subsequent HPLC analysis, we observed two distinct peaks, with their UV-Vis absorption spectra corresponding to 5-OHBza-RP and 6-OHBza-RP respectively (Figure 8C and D). Comparing the HPLC chromatogram for *Ef*CobT's activity with 5-OHBza, it is clear that only one product isomer is produced with *Ef*CobT, the one that elutes at the same retention time as 5-OHBza-RP. Presence of a 6-OHBza-RP peak cannot be detected in HPLC, even when the reaction is repeated multiple times.

In order to verify the peaks from the *Ec*CobT reaction as product isomers, they were manually collected, lyophilized to approximately one fifth of their initial volumes and analyzed using LC-MS. Both product isomers were successfully detected with their correct masses, along with those of the reactants (Figure 9A, B and C).

This analysis suggests that *El*CobT is indeed regioselective for 5-OHBza, with 100% regioselectivity, as detected by HPLC. In order to completely prove that the regioselective attachment seen in DMB is indeed because of CobT, we would need to prove that the N1 of 5-OHBza, that gets attached to ribose to form 5-OHBza-RP (refer Figure 3A) originates from L-glycine. This would require reconstituting the activity of BzaA and BzaB with AIR carrying the correct N^{15} label.

Though BzaA and BzaB have not yet been characterized, their homologous enzyme, BzaF from *Desulfuromonas acetoxidans* has been reconstituted [22]. If we follow both the L-glutamine and the L-glycine N's through the proposed reaction mechanism, we can identify the source of each N on 5-OHBza. Even if it is not absolute evidence, the above analysis suggests that N1 of 5-OHBza indeed comes from L-glycine. Should 5-OHBza-RP act as a substrate for BzaC the label would be carried forward. Reconstitution of BzaC activity is currently underway and would provide evidence for this.

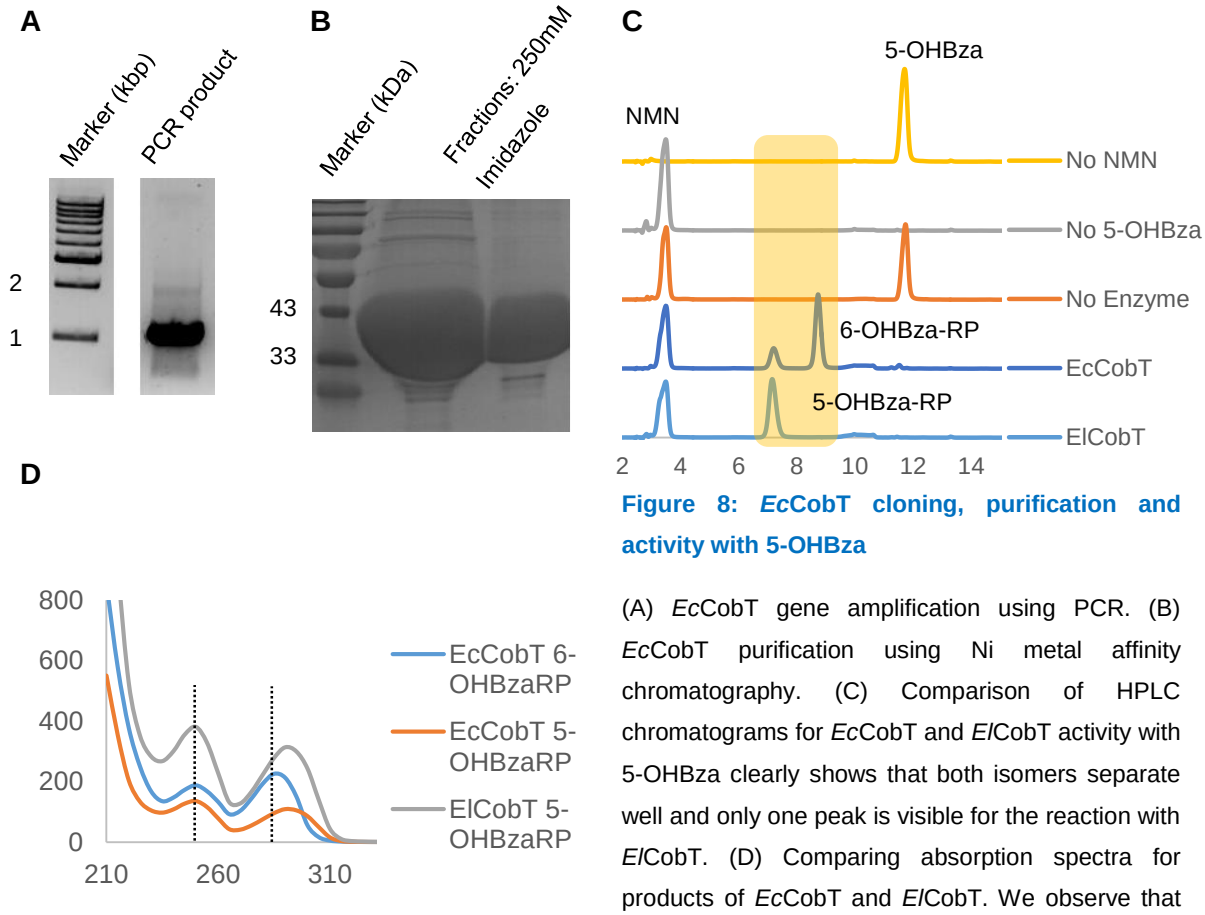

Figure 8: *Ec*CobT cloning, purification and activity with 5-OHBza

(A) *Ec*CobT gene amplification using PCR. (B) *Ec*CobT purification using Ni metal affinity chromatography. (C) Comparison of HPLC chromatograms for *Ec*CobT and *El*CobT activity with 5-OHBza clearly shows that both isomers separate well and only one peak is visible for the reaction with *El*CobT. (D) Comparing absorption spectra for products of *Ec*CobT and *El*CobT. We observe that the isomer formed with *El*CobT is 5-OHBza-RP.

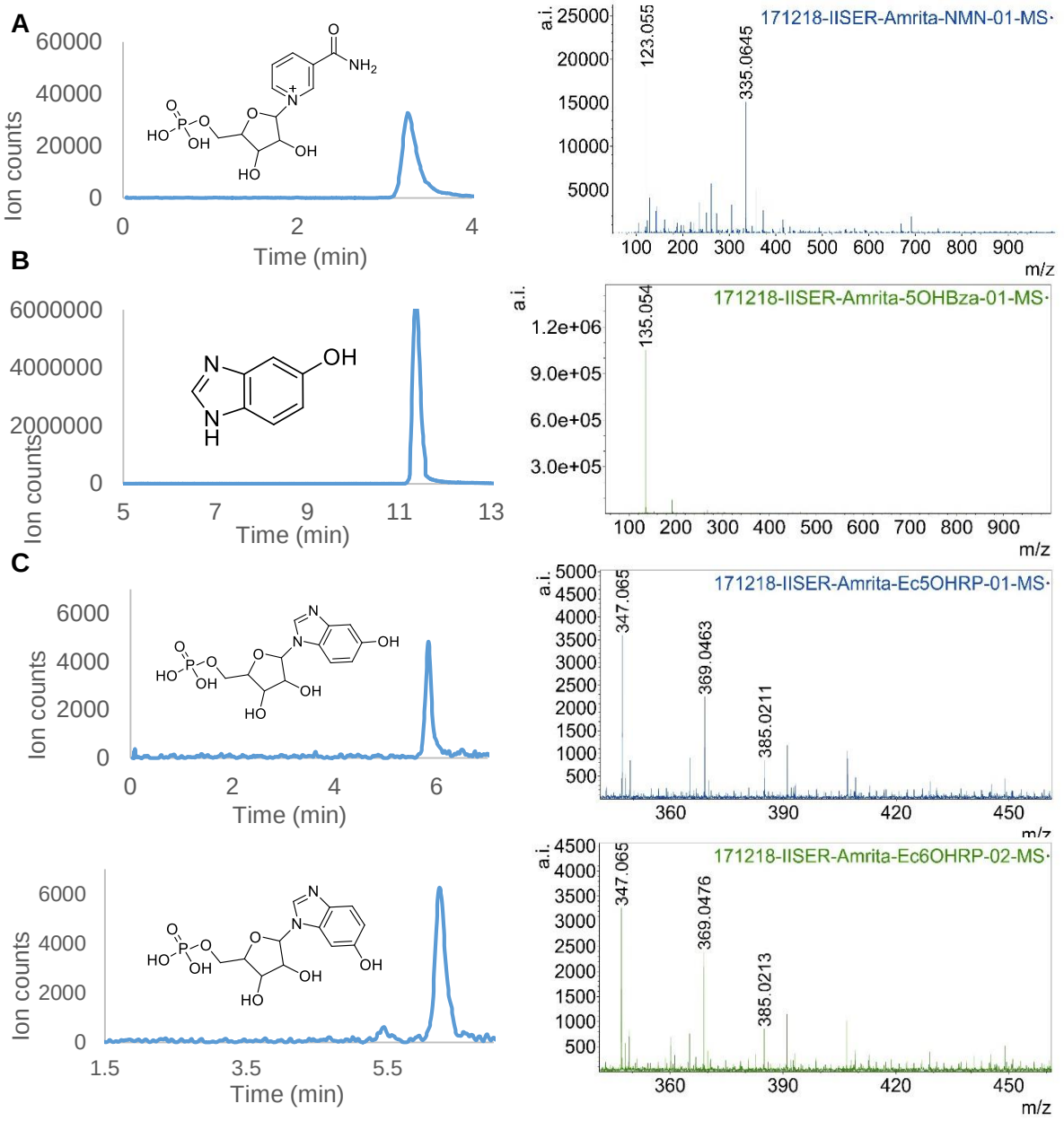

Figure 9: LC-MS analysis of *Ec*CobT activity with 5-OHBza

Extracted ion chromatograms (EIC) and electron spray ionization-mass spectrometry (ESI-MS) analysis of NMN (A), 5-OHBza (B) and products from the reaction of *Ec*CobT with 5-OHBza(C).

In addition to 5-OHBza, we also reconstituted the activity of *Ec*CobT with 5-OMeBza. To our surprise, we observed that only 5-OMeBza-RP was predominantly obtained, with a very tiny peak for 6-OMeBza-RP (Figure 10A and B). The area under the peak for 5-OMeBza-RP is about 47-50 times more than that of 6-OMeBza-RP on the HPLC chromatogram. Comparing that to *El*CobT, the area under the peak for 5-OMeBza-RP is only 1.3-1.4 times as high as 6-OMeBza-RP.

Intrigued by this phenomenon, we decided to study the molecular factors that govern this regioselectivity.

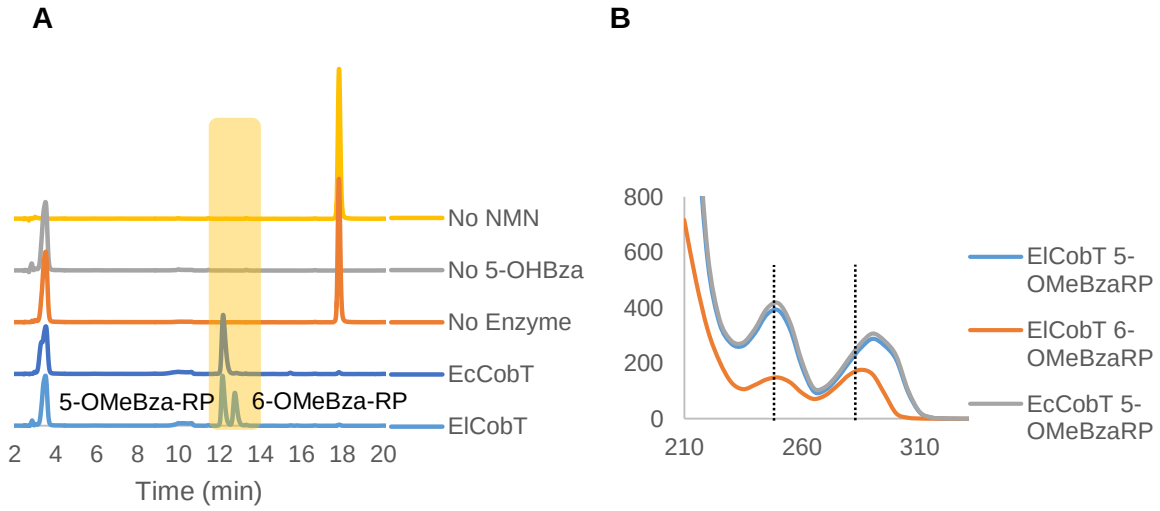

Figure 10: *Ec*CobT activity with 5-OMeBza

(A) Comparison of HPLC chromatograms for *Ec*CobT and *El*CobT activity with 5-OMeBza shows that *Ec*CobT is selective for 5-OMeBza. (B) Comparing absorption spectra for products of *Ec*CobT and *El*CobT. We observe that the isomer formed with *Ec*CobT is 5-OMeBza-RP.

4.3 Studying the active site of CobT

In order to identify the molecular factors, we studied the crystal structure of *Se*CobT, available on the PDB database [8]. We can see the 5-OMeBza moiety bound to the active site of *Se*CobT (Figure 11A). The catalytic residue, glutamate at 317th position, E317 is about 2.4 A^{o} away from the imidazole N. In addition to that, two other residues are very close to the –OMe group of 5-OMeBza. They are serine at 80th position, S80 and glutamine at 88, Q88, each at a distance of 3.7 and 3.8 A^{o} respectively. These residues can influence the orientation of 5-OMeBza in the active site and thus, influence the selectivity observed in product formation.

To examine what the corresponding residues are in other organisms, we generated a model for the nicotinate-nucleotide-5, 6-dimethylbenzimidazole phosphoribosyl transferase (CobU) from *Sinorhizobium meliloti* (*Sm*CobU) and aligned it with the crystal structure of *Se*CobT (Figure 11B). We observe that the residue corresponding to S80 from *Se*CobT is phenylalanine at 85th position, F85 and for Q88, it is methionine at 93rd position, M93.

Further, the primary sequences of *El*CobT, *Ec*CobT, *Se*CobT and *Sm*CobU were aligned. Upon alignment, we noticed that for *El*CobT and *Sm*CobU, the residues are F and M whereas for *Ec*CobT and *Se*CobT, they are S and Q (Figure 11C).

In order to determine whether the residues we identified have an effect on the regioselectivity of CobT, we purified and reconstituted the activity for *Sm*CobU, *Sm*CobU F85S and *Sm*CobU M93Q.

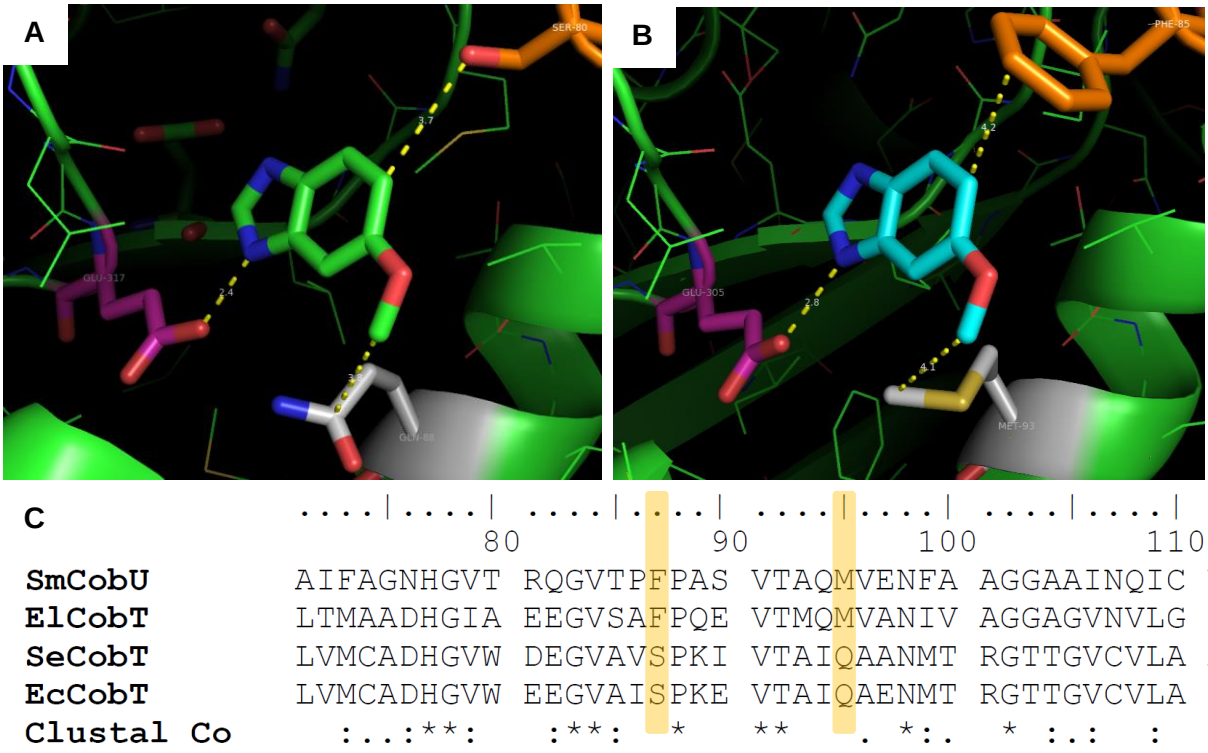

Figure 11: Crystallographic and sequence analysis of CobT homologs

(A) Crystal structure of *Se*CobT with 5-OMeBza at the active site. Crystal structure obtained from PDB (1JHP) and visualized using PyMOL [8]. Catalytic residue E317 is coloured in Magenta. Residues S80 (Orange) and Q88 (White) are close to the –OCH_3 group and may affect its binding. (B) 3D model for *Sm*CobU (generated on PHYRE^2 [21] using 1J33) aligned with 1JHP. The catalytic residue E305 (Magenta) and the residues near the –OCH3 group: F85 (Orange) and M93 (White) are highlighted. (C) Primary sequence alignment of *Se*CobT, *Ec*CobT, *Sm*CobU and *El*CobT. Highlighted residues correspond to S80 and Q88 from *Se*CobT. (Sequences aligned using MUSCLE and viewed using BioEdit.).

4.4 Would altering key residues change regioselectivity?

*Sm*CobU, *Sm*CobU F85S and *Sm*CobU M93Q were obtained from Taga lab, University of California, Berkeley and were over-expressed and purified using Ni affinity chromatography (Figure 12A, B and C). Their activities were reconstituted using 5-OHBza and 5-OMeBza (Figure 12D and E). A very interesting phenomenon was observed. *Sm*CobU F85S was much less regioselective in formation of 5-OHBza-RP and 5-OMeBza-RP as compared to the wild type (wt) enzyme and the M93Q mutant (Figure 12F, G). The ratio of area under the peak for 5-OHBza-RP and 6-OHBza-RP decreased from about 7 in wt and M93Q to 2.7 in the F85S mutant. For 5-OMeBza, the change was much greater, from about 15 in wt and M93Q to 1.7 in the F85S mutant.

This analysis clearly displays the importance of the F85 residue in maintaining regioselectivity for *Sm*CobU. The decrease in selectivity for formation of 5-OMeBza-RP and 5-OHBza-RP in the F85S mutant can be explained by the protein model we have generated. A mutation from F to S creates more space in the active site, allowing for reduction in the steric clashes between the substrate and the enzyme. Therefore, the substrate can bind more freely in either directions (Figure 11B). Since the steric clashes will be more pronounced in 5-OMeBza than 5-OHBza, the relative ratio of 5-OMeBza-RP:6-OMeBza-RP is higher than 5-OHBza-RP:6-OHBza-RP for wt and M93Q.

Further, we wished to see whether mutating both residues in one enzyme would allow it to mimic the other enzyme. For that purpose, we expressed and purified *Se*CobT S80F, Q88M double mutant (Figure 13A) (clone was courtesy of Taga lab, University of California, Berkeley) and reconstituted its activity with both lower ligands. Comparing the activity with the wild type enzyme [10], we observe that the selectivity has changed drastically for 5-OHBza (Figure 13B, C). The 6-OHBza-RP peak is barely visible, similar to what we see for *Sm*CobU's activity with 5-OHBza.

This indicates that mutating both residues of one enzyme to that of the other enzyme can indeed allow the mutated enzyme to mimic the other enzyme. In order to quantify the exact change in regioselectivity we would need to analyze the reactions of *Se*CobT

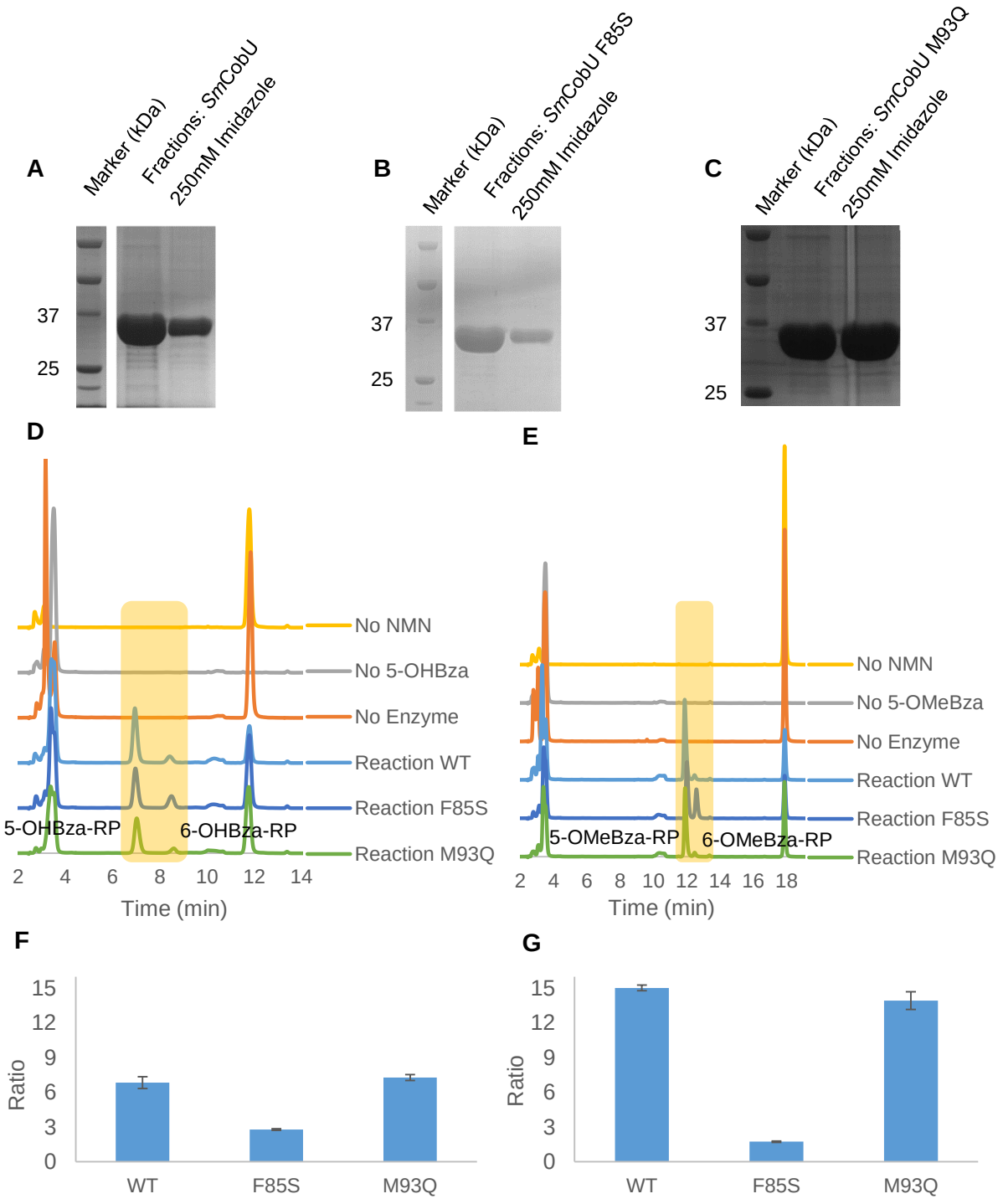

Figure 12: *Sm*CobU(T), *Sm*CobU F85S and *Sm*CobU M93Q purification and activity

SDS-PAGE gel for metal affinity based purification of *Sm*CobU (A), *Sm*CobU F85S (B) and *Sm*CobU M93Q (C). Comparison of activities of *Sm*CobU and mutants with 5-OHBza (D) and 5-OMeBza (E). We observe that the F85S mutant has lost selectivity. (F) Ratio of area under the peaks for 5-OHBza-RP and 6-OHBza-RP for *Sm*CobU wt, *Sm*CobU F85S and *Sm*CobU M93Q. (G) Ratio of area under the peaks for 5-OMeBza-RP and 6-OMeBza-RP for *Sm*CobU wt, *Sm*CobU F85S and *Sm*CobU M93Q. We observe that the regioselectivity has reduced significantly for the F85S mutant for both 5-OHBza and 5-OMeBza. Error bars indicate standard deviation.

and *Se*CobT S80F, Q88M using the same HPLC conditions. Therefore, we are

currently cloning the *Se*CobT wt enzyme from the double mutant. We are also cloning *El*CobT F81S and M89Q to study their change in regioselectivity with the wt enzyme.

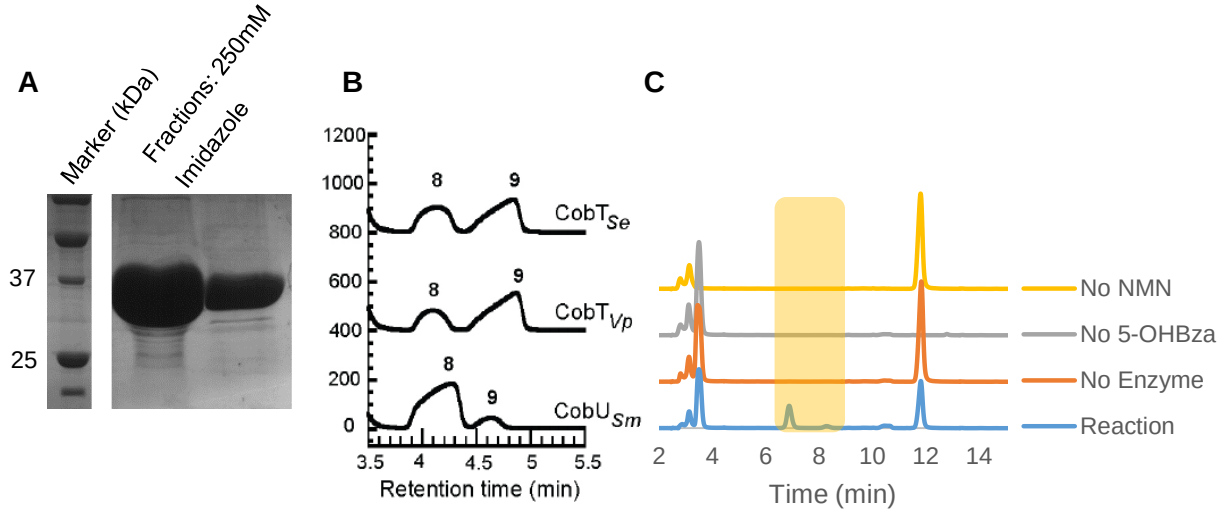

Figure 13: *Se*CobT S80F, Q88M purification and activity with 5-OHBza

(A) SDS-PAGE gel for metal affinity based purification of *Se*CobT S80F, Q88M. (B) SeCobT activity with 5-OHBza (top one labelled CobT$_{Se}$) (Adapted from Crofts et al, *Biochemistry*, **2014**, 53: 7805−7815) We observe two peaks corresponding to 5-OHBza-RP and 6-OHBza-RP. (C) Activity of *Se*CobT S80F, Q88M with 5-OHBza. Only one peak is predominantly observed, corresponding to 5-OHBza-RP with a very tiny peak corresponding to 6-OHBza-RP.

4.5 pH and regioselectivity

In order to examine whether pH has any effect on regioselectivity of *El*CobT, its activity with 5-OMeBza was monitored under five different pH conditions – 6.5, 7.5, 8.5, 9.5 and 10.5. The reactions were analyzed using HPLC. Ratios for areas under the peaks of 5-OMeBza-RP and 6-OMeBza-RP were calculated for each pH condition (Figure 14A and B).

It was observed that the reactivity of *El*CobT was higher at a higher pH, as reported earlier [23]. However, no significant change was observed in the ratios of 5-OMeBza-RP and 6-OMeBza-RP with changing pH. This shows that the regioselectivity of *El*CobT does not vary with pH.

The error bars are particularly high for the conditions tested at pH 10.5. Since phosphates were used as buffers for this study, the lower solubility of NaH_2PO_4 and Na_3PO_4 in water might be a reason for a drastic variation in pH. It is also possible that the buffering capacity of phosphates is low at 10.5. A better buffer can be used,

however, in order to maintain uniformity in the types of ions in water over the entire pH range, we chose phosphates.

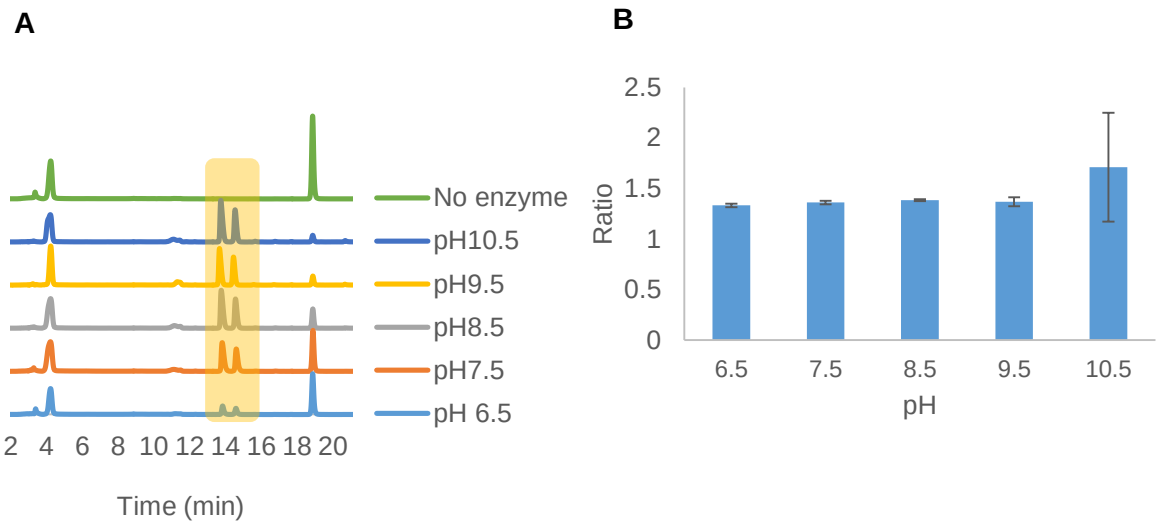

Figure 14: *El*CobT activity with 5-OMeBza at different pH values

(A) HPLC chromatogram for *El*CobT activity with 5-OMeBza at different pH. We observe that the reactivity increases with pH. (B) Mean ratio of area under the peaks corresponding to 5-OMeBza-RP and 6-OMeBza-RP at different pH. Error bars indicate standard deviation with three sets of values. We observe that the change in ratios with pH is not significant.

4.6 Studying enzyme-enzyme interactions

Enzyme-enzyme interactions were studied at the University of Michigan, Ann Arbor as a part of my summer internship conducted in Prof. Neil Marsh's lab from June, 2017 to July, 2017. The following enzymes were considered for enzyme-enzyme interactions – *El*CobT, *El*CobS, *El*CobU, and *ElBzaC*. However, *El*CobS, *El*CobU did not express at a variety of different conditions. Therefore, we designed enzyme-enzyme interaction studies for *El*CobT and *El*BzaC only. *Ec*CobT was used to compare the differences between *El*CobT + *El*BzaC and *Ec*CobT + *El*BzaC mixtures.

A preliminary native PAGE was performed on an equimolar mixture of *El*CobT and *El*BzaC incubated at 25^{o}C for 1 hr. A SDS-PAGE was performed for the same samples and compared with the native gel. (Figure 15A and B). It was observed that *El*BzaC aggregates to a large extent and forms higher order oligomers, making it difficult to detect any visible change in the mobility of *El*CobT + *El*BzaC when mixed together.

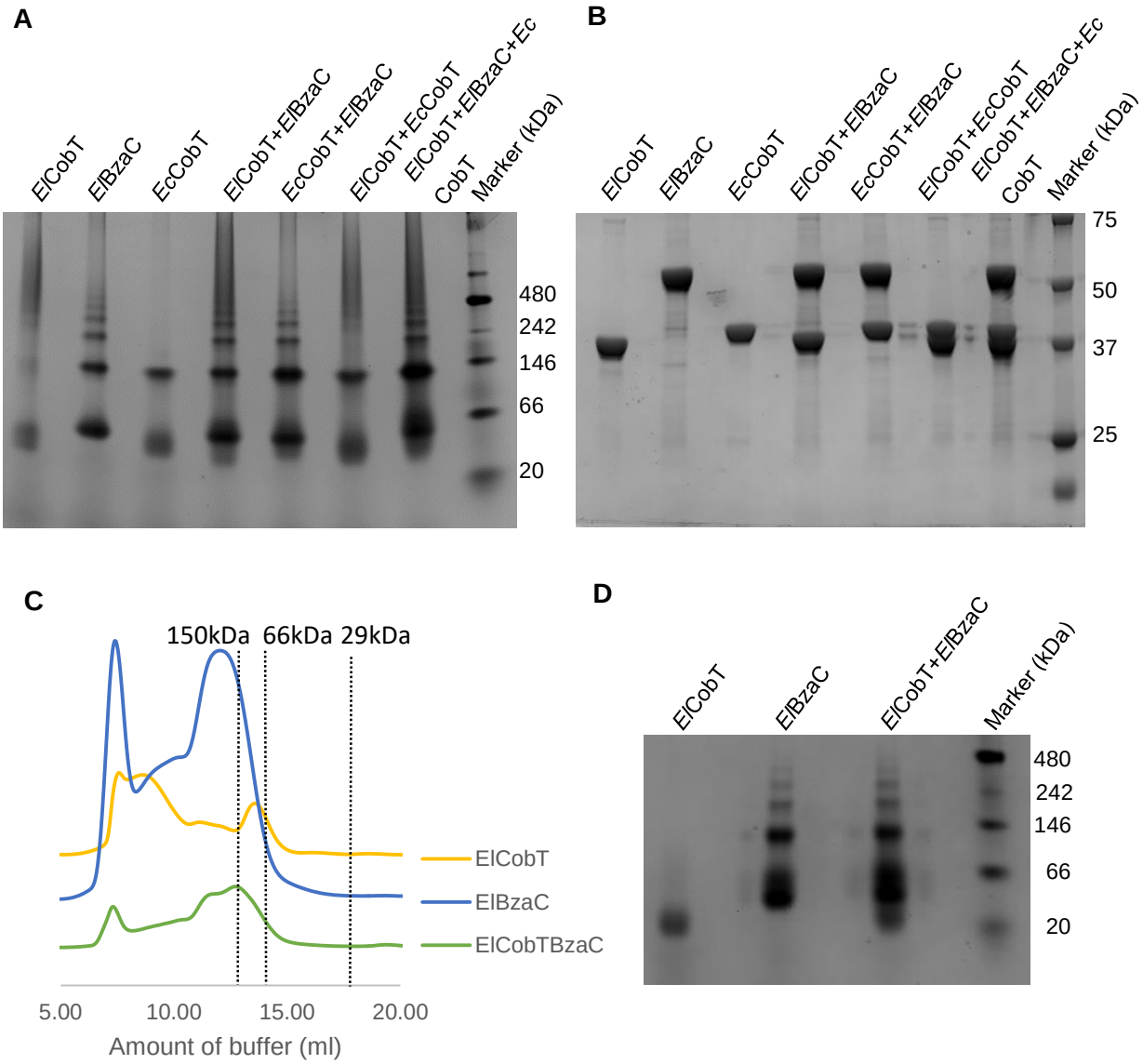

Figure 15: Studying enzyme-enzyme interactions between *El*CobT and *El*BzaC

(A) Native PAGE gel of purified *El*CobT, *El*BzaC, *Ec*CobT and their various combinations incubated at 25ºC for one hour. (B) SDS-PAGE gel of the same samples as A. (C) Size exclusion chromatogram of purified *El*CobT, *El*BzaC and *El*CobT+*El*BzaC incubated at 25ºC for one hour. (D) Native PAGE gel of monomeric fractions for *El*CobT and *El*BzaC purified using size exclusion chromatography and incubated at 25ºC for one hour.

Size exclusion chromatography of the same samples was equally inconclusive. Many peaks were observed in the chromatogram for *El*BzaC, most of which were present in *El*BzaC + *El*CobT as well (Figure 15C). This made it difficult to identify any peak shifts or occurrences of new peaks.

To overcome this problem, peaks for the monomeric units of *El*BzaC and *El*CobT were identified using native PAGE and collected. The monomeric fractions were then mixed together in an equimolar ratio and incubated at 25^{o}C for 1 hr. Native PAGE analysis of these samples indicated a strong tendency of *El*BzaC for oligomerization, despite the fact that a monomeric fraction was collected for analysis (Figure 15D). We thus concluded that size exclusion chromatography and native PAGE were incapable of detecting existence of a significant interaction (if any) between *El*BzaC and *El*CobT.

4.7 Enzyme-enzyme interactions and regioselectivity

Since we essentially wanted to study whether enzyme-enzyme interactions have an impact on regioselectivity, we decided to use HPLC as a read-out. The enzyme to be studied can be incubated with *El*CobT in stoichiometric amounts and the ratios of areas under the peaks for 5-OMeBza-RP and 6-OMeBza-RP can be compared with and without the addition of the enzyme being studied.

To ensure that the addition of an arbitrary enzyme does not change *El*CobT's regioselectivity, we incubated bovine serum albumin (BSA) in stoichiometric ratios of 1:1, 10:1 and 100:1 with *El*CobT and compared the ratios of 5-OMeBza-RP and 6-OMeBza-RP (Figure 16A and B). No significant change was observed. However, we noticed that at BSA:*El*CobT=100:1, the reactivity of *El*CobT decreases.

Later, similar reactions were studied with *El*CobT + *El*BzaC, *El*CobT + *El*CobU and *El*CobT + *El*CobU + *El*BzaC. Since very high concentrations of BSA had reduced *El*CobT's reactivity, *El*CobU and *El*BzaC were added in stoichiometric ratios of 1:1, 5:1 and 10:1 with *El*CobT. For the reactions involving all three enzymes, both *El*CobU and *El*BzaC were added in ratios of 1:1:1, 5:5:1 and 10:10:1 with *El*CobT.

The ratio of 5-OMeBza-RP:6-OMeBza-RP was calculated for each case and compared with *El*CobT only control (Figure 17C, D and E). No significant change in ratios was observed. These results indicated that enzyme-enzyme interactions, if any, did not have any effect on the selectivity of *El*CobT, at least for the conditions tested.

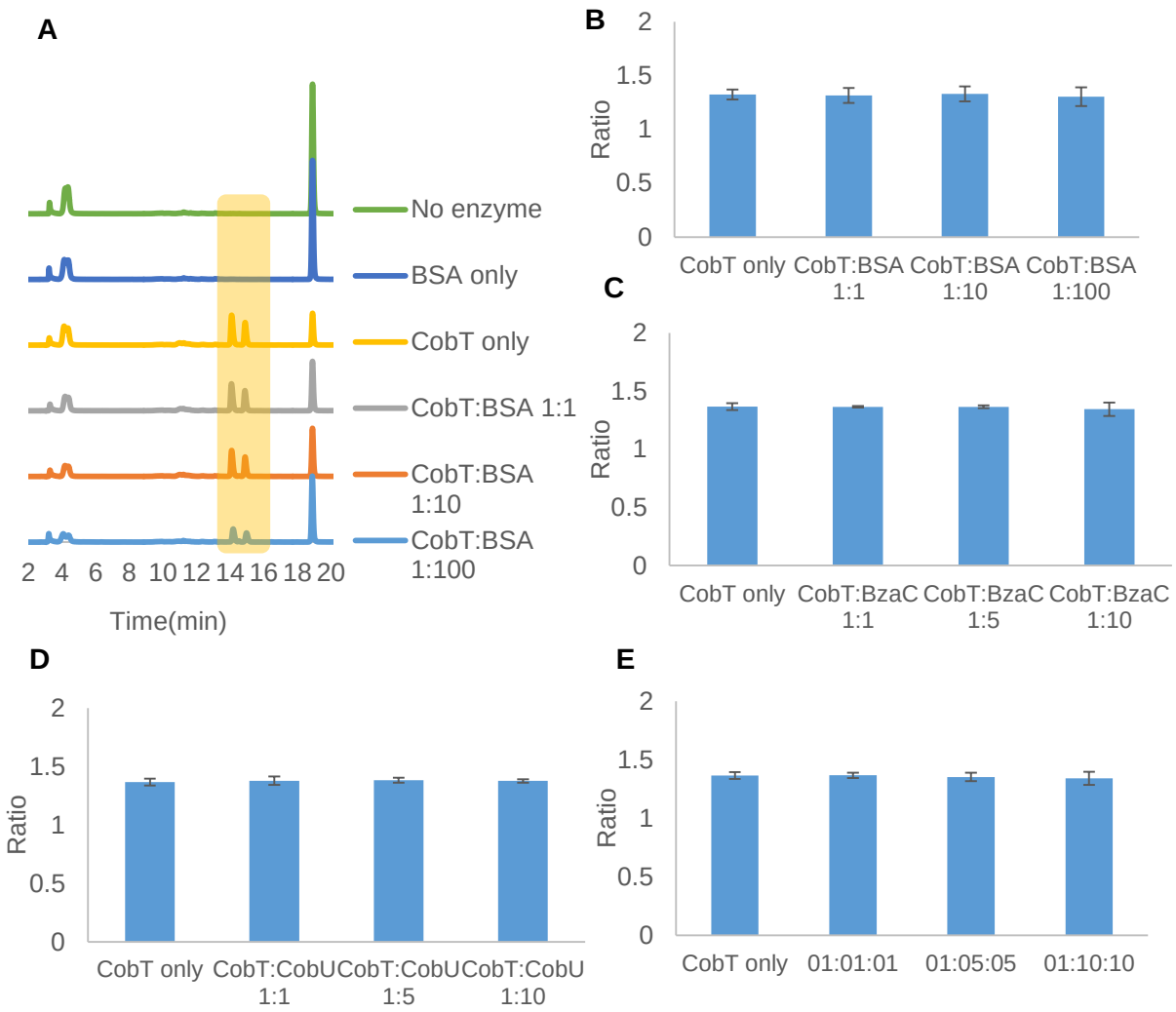

Figure 16: *El*CobT activity with 5-OMeBza in presence of other proteins

(A) HPLC chromatogram for *El*CobT activity with 5-OMeBza in presence of BSA. Not much change is observed in regioselectivity. (B) Mean ratio of area under the peaks corresponding to 5-OMeBza-RP and 6-OMeBza-RP at different stoichiometric ratios of *El*CobT and BSA (B), *El*BzaC (C), *El*CobU (D) and *El*CobU+*El*BzaC (E). Error bars indicate standard deviation with three sets of values. We observe no significant change in regioselectivity in presence of other proteins.

4.8 Phylogenetic analysis

A phylogenetic analysis can work as a very powerful tool. It can help us determine whether organisms that fall within the same clade have similar phenotypes and draw generalizations.

Since we had identified two key residues that have an effect on the regioselectivity, we wished to know how varied these residues are in other cobamide producers. Therefore, we curated a list of 92 CobT sequences from various organisms whose native cobamides have either been characterized or predicted. We used the list to generate a phylogeny tree (Figure 17).

A few interesting observations can be made from this tree. Firstly, all the Archaea clade together to form a separate branch. *Nostoc punctiforme,* a cyanobacterium, is the only non-Archaea in this branch [24]. All organism names whose CobT enzymes were a part of the current study have been coloured red. We observe that *E. coli* and *S. enterica typhimurium LT2* clade together. *S. meliloti* also clades with other organisms of the same family [25].

Eubacterium limosum, however, clades very closely with *Acetobacteroim woodii*, a B_{12} producer that also has the *bza* operon and *Moorella thermoacetica* which has an incomplete *bza* operon that produces 5-OMeBza as the lower ligand [5]. Other species of the *Eubacterium* genus (green background) clade away from *Eubacterium limosum* and with some species of the genus *Clostridium*. This suggests that the CobT homologs from *Eubacterium limosum*, *Acetobacterium woodii* and *Moorella thermoacetica*, which are a part of their *bza* operons might have spread via horizontal gene transfer among themselves. Therefore, these CobT homologs may have similar activity and show similar regioselectivity with asymmetric lower ligands. We compared the 'key residues' that have been identified through this study among these CobT homologs. Interestingly, the equivalent of S80 from *Se*CobT is F in *E. limosum* and *A.woodii* but in *M. thermoacetica* it is replaced by C. Structurally, Cys is similar to Ser except that the –OH group is replaced by an –SH. It would be interesting to see what changes this brings to the regioselectivity of *M. thermoacetica* CobT.

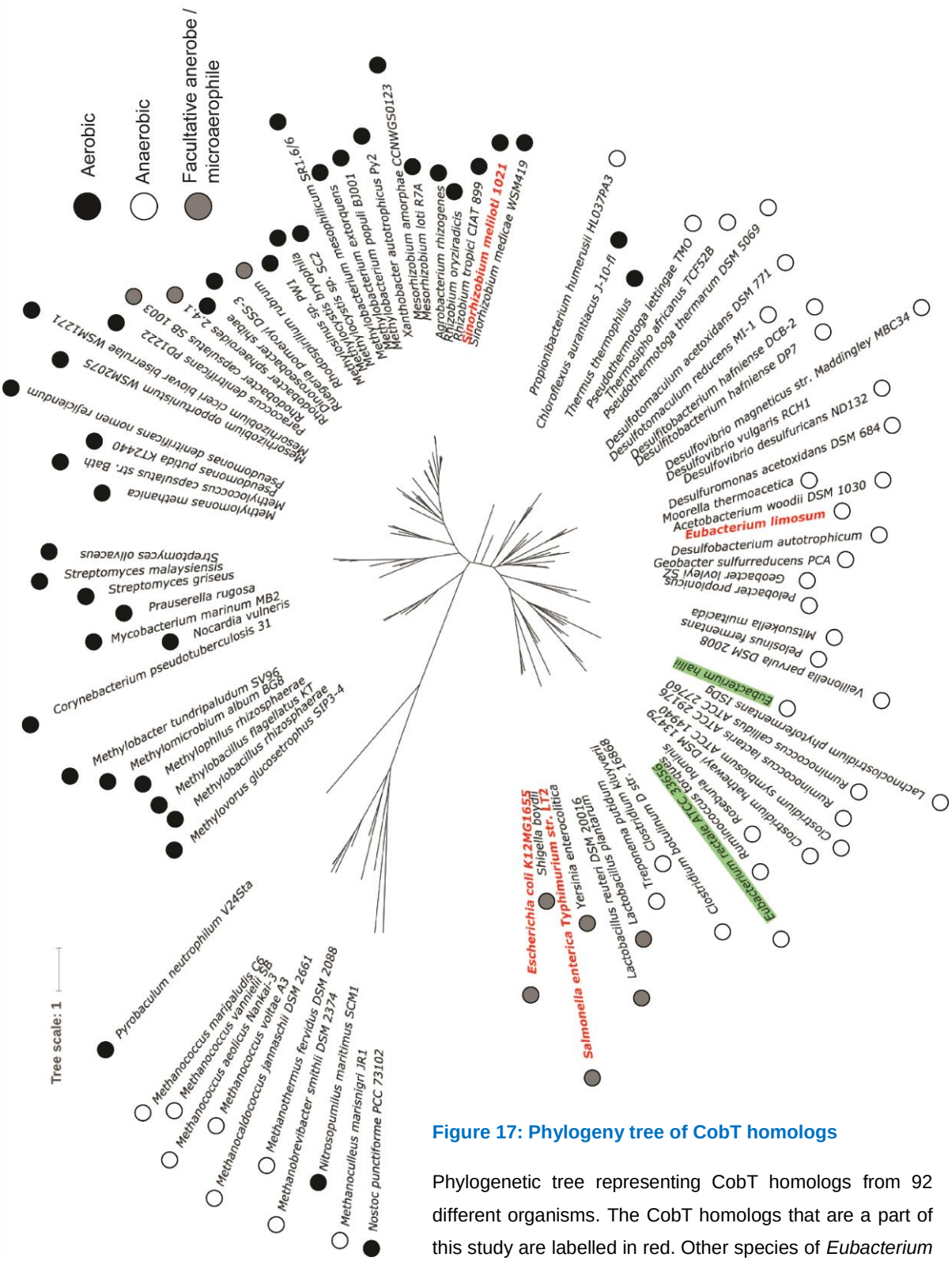

Figure 17: Phylogeny tree of CobT homologs

Phylogenetic tree representing CobT homologs from 92 different organisms. The CobT homologs that are a part of this study are labelled in red. Other species of *Eubacterium* genus are highlighted in green. Organisms are annotated as aerobic or anaerobic based on their culture conditions mentioned on ATCC website and relevant references.

Additional support for the argument that the CobT homologs which are a part of the *bza* operon in *E. limosum*, *A. woodii* and *M. thermoacetica* may have spread through horizontal gene transfer as comes from the fact that each of these organisms have a second CobT homolog reported in their genome. The other CobT homologs (hereafter referred to as CobT2) have never been studied before. In order to assess their role in regioselectivity and their relatedness to the CobT homologs that are a part of the bza operon, we created a new phylogeny tree with CobT2 homologs for *E. limosum*, *A. woodii* and *M. thermoacetica* (Figure 18). Interestingly, *E. limosum* and *A. woodii* CobT2 homologs clade together and closer to facultative anaerobes and microaerophiles. *M. thermoacetica* CobT2 clades with organisms of the family *Peptococcaceae*.

In order to understand more about the role of CobT2 homologs, we made a comprehensive list of the gene neighbourhoods for all CobT homologs that are a part of the tree. Later, we broadly classified the gene neighbourhoods into four categories – cobinamide biosynthesis genes, lower ligand attachment genes, lower ligand biosynthesis genes and miscellaneous genes (non-cobamide biosynthesis genes). We annotated each CobT homolog based on the four categories of genes that are present in their neighbourhood. (The annotations were made using the 'datasets' feature available in iTOL. Since such a feature is not available in a radial tree, a circular tree was used instead.)

We observe that for CobT2 homologs in *E. limosum*, *A. woodii* and *M. thermoacetica*, the gene neighbourhood consists of miscellaneous (non-cobamide producing) genes. This indicates that they probably do not play a role in cobamide biosynthesis and therefore their regioselectivity may not affect the regiochemistry of the lower ligand attachment. Another possibility is that the CobT homolog that is a part of the *bza* operon is involved in lower ligand attachment when the lower ligand is produced by the organism whereas the CobT2 homolog is responsible for attaching lower ligands that are internalized by the organism from its surrounding. Nevertheless, it would be worthwhile to study the regioselectivity of the CobT2 homologs.

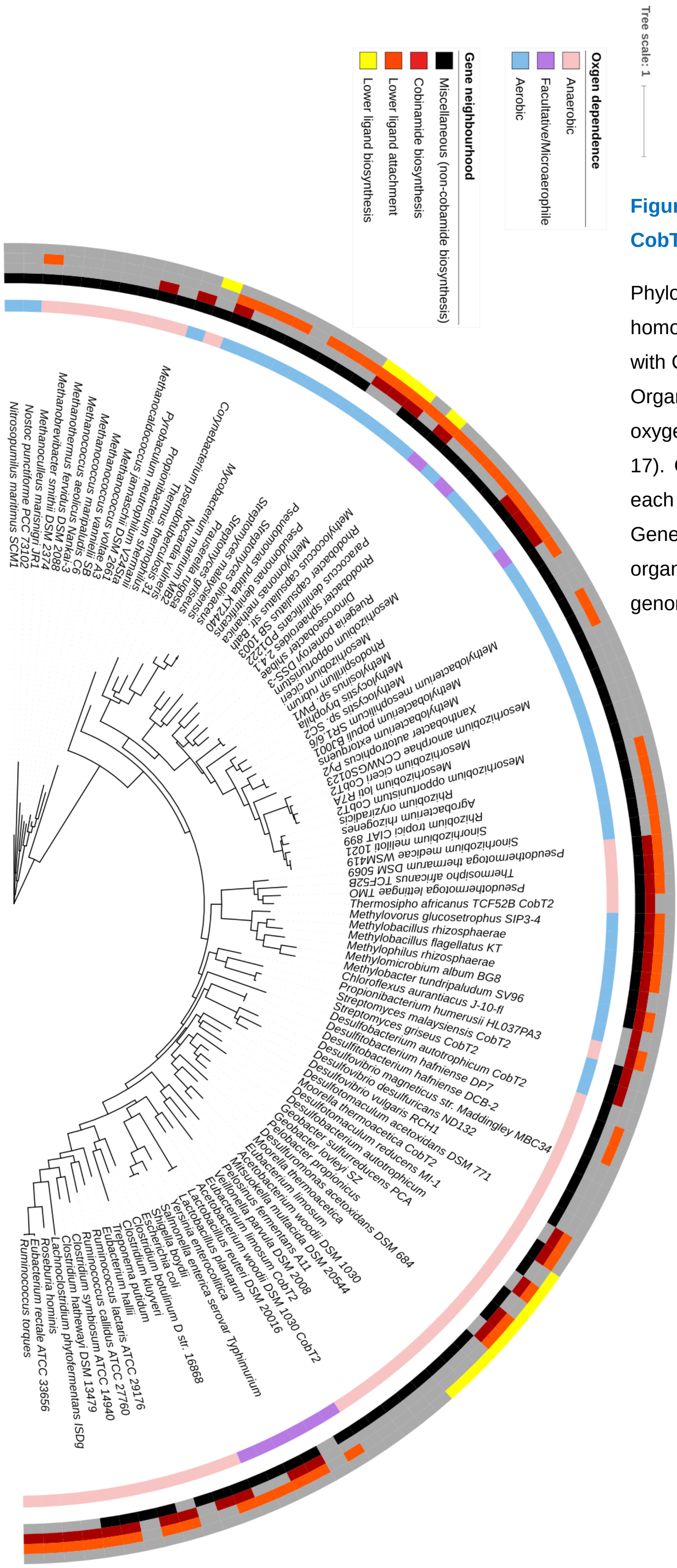

Figure 18: Phylogeny tree with CobT2 homologs

Phylogenetic tree representing CobT homologs from 92 different organisms with CobT2 homologs of 10 organisms. Organisms are classified based on their oxygen dependence as earlier (Figure 17). General gene neighbourhoods of each organisms are also mentioned. Gene neighbourhoods were studied for organisms using their respective genomes available on NCBI database.

Chapter 5

Conclusions and future directions

The goal of this project was to determine the cause for regioselective incorporation of the lower ligand, DMB in vitamin B_{12}. We wished to know why DMB, a symmetric molecule, attaches to ribose only through the N that originates from L-glycine. We speculated that the enzyme that attaches DMB to ribose, CobT is responsible for the regioselectivity either because of a very specific active site or because it interacts with other enzymes. We considered the possibility that CobT is regioselective in attaching an asymmetric DMB precursor, such as 5-OHBza or 5-OMeBza and the asymmetry is carried forward to DMB.

In this study, we have proved that *El*CobT is regioselective for the formation of 5-OHBza-RP from 5-OHBza. Reconstituting the activity of BzaA and BzaB from *E. limosum*, the enzymes that putatively synthesize 5-OHBza, would prove whether the benzimidazole nitrogen that *El*CobT attaches to ribose indeed originates from L-glycine. Similarly, reconstituting the activity of *El*BzaC would prove that the label is carried forward. Such an understanding of the enzyme activities would help us determine the correct sequence in which these enzymes appear in the biosynthetic pathway.

As a part of this study, we have successfully characterized the activity and regioselectivity of *Ec*CobT, another novel CobT that had never been studied before.

The regioselectivity of *El*CobT was studied with two of the three precursors of DMB. The argument provided for not studying the activity of 5-OMe, 6-MeBza was that the enzymes BzaD and BzaE have a B_{12} binding sight, indicating that a cobamide is already formed before they act on the lower ligand. However, it would still be useful to study the regioselectivity of CobT with 5-OMe, 6-MeBza. In this study we observed that the change in regioselectivity was more pronounced in 5-OMeBza as compared to 5-OHBza for *Sm*CobU and its mutants. This can be attributed to the steric bulk of the –OMe group as compared to the –OH group. Similarly, the high steric bulk at both 5 and 6 positions of 5-OMe, 6-MeBza would allow for greater interactions with adjacent

CobT residues leading to significant changes in regioselectivity even with smaller perturbations in the active site.

Further, we discovered that the regioselectivity of *El*CobT is not affected by pH. It also remains unaffected by the presence of *El*CobU and/or *El*BzaC, even in much higher stoichiometric ratios.

Our attempt to study enzyme-enzyme interactions between *El*CobT and *El*BzaC using native-PAGE and size exclusion chromatography was inconclusive. These methods were unable to detect any sort of interactions (if present) between the two enzymes. Enzyme-enzyme interaction studies were performed with *El*BzaC as it is immediately adjacent to *El*CobT. However, since no significant change was observed in regioselectivity, the study can also be performed with *El*BzaD and *El*BzaE. A variety of iterations can be performed, such as adding the enzymes in different combinations as well as adding the respective substrates for different enzymes.

In our effort to understand the molecular factors that govern selectivity of CobT, we identified two key residues in the active site of *Se*CobT, S80 and Q88. These residues are near the benzene ring of the benzimidazole and thus can influence the orientation in which it binds via interactions with the side chain attached to the benzene. The corresponding residues in *El*CobT and *Sm*CobU are F and M. We have proved that modifying the key residues in *Sm*CobU has an effect on the enzyme's regioselectivity. Specifically, it appears that modifying the phenylalanine (F85) reduces the selectivity of the enzyme. Similarly, mutating both residues of *Se*CobT to those of *Sm*CobU allowed the double mutant to mimic *Sm*CobU in terms of regioselectivity displayed for 5-OHBza.

If we observe the active site of *Se*CobT in detail, we can identify two more residues that are close to the benzene ring of the benzimidazole. They are valine at the 84th position (V84) and methionine at the 177th position (M177). Studying the regioselectivity of CobT as a function of these residues would also be interesting.

Using a list of 92 CobT sequences from different organisms, we generated a phylogeney tree to gain more insights into CobT activity. We discovered that *El*CobT clades with other organisms that have the *bza* operon (either complete or incomplete) and not with other species from the same genus (*Eubacterium*). Additionally, the CobT2 homologs from *E. limosum*, *A. woodii* and *M. thermoacetica* clade differently

from the CobT homologs that are a part of the *bza* operon indicating that they may have different regioselectivity.

Through the study, we have demonstrated the importance of the enzyme reaction conditions and the active site residues in determining the regioselectivity of CobT homologs. With the help of crystal structures and phylogenetic analysis, we can identify more such residues that would have an effect on regioselectivity.

By modifying key residues, we are essentially able to alter how CobT binds to a given tautomer of an asymmetric benzimidazole. With further studies, we can engineer a CobT to improve its binding and reactivity with a non-native lower ligand and incorporate it into an organism to make non-native cobamides. Such a system can be used for large scale production of uncommon cobamides that are required by the industrial sector.

References

1. Roth, JR; Lawrence, JG and Bobik, TA; Cobalamin (Coenzyme B_{12}): Synthesis and Biological Significance. *Annu. Rev. Microbiol.,* **1996**, *50*, 137–81.
2. Banerjee, R and Ragsdale, SW; The many faces of B_{12}: catalysis by Cobalamin dependent enzymes. *Annu. Rev. Biochem*. **2003**, *72*, 209–247.
3. Crofts, TS; Seth, EC; Hazra, AB and Taga, ME; Cobamide structure depends on both lower ligand availability and CobT substrate specificity. *Chemistry & Biology*, **2013**, *20*, 1265–1274.
4. Taga, ME; Larsen, NA; Howard-Jones, AR; Walsh, CT and Walker, GC; BluB cannibalizes flavin to form the lower ligand of vitamin B_{12}. *Nature*, **2007**, *446*, 449-453.
5. Hazra, AB; Han, AW; Mehta, AP; Mok, KC; Osadchiya, V; Begley, TP and Taga, ME; Anaerobic biosynthesis of the lower ligand of vitamin B_{12}. *Proc. Nat. Acad. Sci. USA*, **2015**, *112*, 34: 10792-10797.
6. Escalante-Semerena, JC; Conversion of Cobinamide into Adenosylcobamide in Bacteria and Archaea. *Jou. Bacteriology*, **2007**, *189*, 13: 4555–4560.
7. Friedmann, HC and Harris, DL; The Formation of α-Glycosidic S-Nucleotides by a Single Displacement Trans-IV-glycosidase *J. Biol. Chem.* **1965***, 240,* 406-412.
8. Cheong, CG; Escalante-Semerena, JC and 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.* **2002**, *277,* 41120-41127.
9. Renz, P; Investigations on the biosynthesis of the 5,6-dimethylbenzimidazole moiety of vitamin B_{12}; In *Vitamin B_{12}, and B_{12}-Proteins* by Bernhard Krautler, Duilio Arigoni and Bernard T. Golding, 119-130, Wiley Verlag GmbH, 1998.
10. Crofts, TS; Hazra, AB; Tran, JLA; Sokolovskaya, OM; Osadchiy, V; Ad, O; Pelton, J; Bauer, S and Taga, ME. *Biochemistry*, **2014,** *53*, 7805-7815.

11. Anslyn, EV and Dennis Dougherty, *Modern Physical Organic Chemistry*, Mill Valley, California: University science books, 2006, pg. 319-322.
12. Hassner, A; Regiospecificity. A useful terminology in addition and elimination reactions. *J. Org. Chem.*, **1968,** *33***,** 2684-2686.
13. Clayden, J; Nick Greeves; Stuart Warren and Peter Wothers; *Organic Chemistry*, Oxford: Oxford University Press, 2001, pg. 607, 523-525.
14. Darsandhari, S; Dhakal, D; Shrestha, B; Parajuli, P; Seo, JH; Kim, TS; Sohng, JK; Characterization of regioselective flavonoid *O*-methyltransferase from the *Streptomyces* sp. KCTC 0041BP. *Enzyme Microb. Technol.* **2018**, *113*, 29-36.
15. Khatri, Y; Jozwik, IK; Ringle, M; Ionescu, IA; Litzenburger, M; Hutter, MC; Thunnissen, AMWH and Bernhardt, R. Structure-based engineering of steroidogenic CYP260A1 for stereo and regioselective hydroxylation of Progesterone, *ACS Chem. Biol.*, **2018**, Article ASAP.
16. McMurry, JE and Tadhg Begley; *Organic Chemistry of Biological Pathways*, Englewood, Colorado: Roberts and Company, 2005, pg. 321-323.
17. Lamm, L; Hörig, JA; Renz, P and Heckmann, G; Biosynthesis of vitamin B_{12} in anaerobic bacteria: Mode of incorporation of glycine into the 5, 6-dimethylbenzimidazole moiety in *Eubactevium limosum, Eur J Biochem*, **1980**, *109*,115–118.
18. Vogt, JR and Renz, P; Biosynthesis of vitamin B_{12} in anaerobic bacteria: Experiments with *Eubactevium limosum* on the origin of the amide groups of the corrin ring and of *N-3* of the 5,6-dimethylbenzimidazole part; *Eur J Biochem*, **1988**, *171*, 655–659.
19. Sambrook, J and Russell, DW; SDS-Polyacrylamide gel electrophoresis of proteins, *CSH Protoc.;* **2006**.
20. Letunic, I and Bork, P; Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees, *Nucleic Acids Res*, **2016**, *44*, 242-5.
21. Kelley, LA; Mezulis, S; Yates, CM; Wass, MN and Sternberg, MJ. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, **2015**, *10*, 845-858.
22. Mehta, AP; Abdelwahed, SH; Fenwick, MK; Hazra, AB;Taga, ME; Zhang, Y; Ealick, SE and Begley, TP; Anaerobic 5-hydroxybenzimidazole formation from

aminoimidazole ribotide: An unanticipated intersection of thiamin and vitamin B_{12} biosynthesis. *J. Am. Chem. Soc.*, **2015**, *137*, 10444−10447.

23. Trzebiatowski, JR and Escalante-Semerena, JC; Purification and characterization of CobT, the nicotinate mononucleotide 5, 6-dimethyl benzimidazole phosphoribosyl transferase enzyme from *Salmonella typhimurium* LT2. *J. Biol. Chem.* **1997**, *272*, 17662-17667.
24. Meeks, JC; Elhai, J; Thiel, T; Potts, M; Larimer, F; Lamerdin, J; Predki, P and Atlas, R; An overview of the genome of *Nostoc punctiforme*, a multicellular, symbiotic cyanobacterium. *Photosynth Res.* **2001**; *70*, 85-106.
25. Young, JM; Kuykendall, LD; Martínez-Romero, E; Kerr, A and Sawada, H; A revision of Rhizobium Frank 1889, with an emended description of the genus, and the inclusion of all species of Agrobacterium Conn 1942 and Allorhizobium undicola de Lajudie et al. 1998 as new combinations: Rhizobium radiobacter, R. rhizogenes, R. rubi, R. undicola and R. vitis. *Int J Syst Evol Microbiol*, **2001**, *51*, 89-103.
26. Munder, M; Vogt, JRA; Vogler, B and Renz, P; Biosynthesis of Vitamin B_{12} in anaerobic bacteria: Experiments with *Eubacterium limosum* on the incorporation of D-[1-^{13}C]erythrose and [^{13}C]formate into the 5,6-dimethylbenzimidazole moiety, *Eur. J. Biochem.*, **1992**, *204*, 679-683.
27. Vogt, JRA; Lamm-Kolonko, L and Renz, P; Biosynthesis of Vitamin B_{12} in anaerobic bacteria: Experiments with *Eubacterium limosum* and D-erythrose ^{14}C-labeled in different positions, *Eur. J. Biochem*. **1988**, *174*, 637-640.