Characterisation of the Enzymes Involved in the Anaerobic Biosynthesis of Benzimidazolylcobamides

Samuel Merryn Jones

This thesis is submitted for the degree of Doctor of Philosophy

at the University of Kent, September 2024

Declaration

No part of this thesis has been submitted in support of an application for any degree or other qualification from the University of Kent, or any other University or Institution of learning.

Samuel Merryn Jones

September 2024

Abstract

Vitamin B_{12} is a complex metalloorganic cofactor involved in a variety of biochemical reactions that are essential for life. There are two routes by which its biosynthesis can be achieved: the oxygen independent (anaerobic) and oxygen dependent (aerobic), involving around 30 enzymatic steps. However, little biochemical characterisation has been conducted on the newly identified *bza* operon, which has been shown to encode enzymes capable of biosynthesising dimethylbenzamidizole (DMB), the α-ligand of vitamin B_{12}. Furthermore, other *bza*-like operons have been linked to the production of cobamides with alternatively functionalised benzimidazole α-ligands (benzimidazolylcobamides). This thesis describes the biochemical investigation of the enzymes involved in these Bza and Bza-like pathways, aiming to provide insight into how they operate.

Three different BzaC enzymes identified in Bza and Bza-like pathways have been recombinantly produced, purified and biochemically characterised. This involved reconstitution of BzaCs activity with a benzimidazolylcobamide substrate, investigation of their reaction mechanisms, substrate specificity, inhibition and 3D structure. Furthermore, an interesting BzaE enzyme, which potentially belongs to the newly identified B_{12} dependent radical SAM family, identified in a Bza-like pathway has been investigated.

This work identified that BzaC can catalyse regioselective SAM dependent methyl transfer to 5(6)-hydroxybenzimidazolylcobamide with differences seen in percentage conversion of substrate to product across the BzaCs investigated. Assays containing

5(6)-hydroxybenzimidazole cobamides with differing upper ligands (cyano, hydroxo, adenosyl) identified that the enzymes prefer an adenosylated substrate. Furthermore, the highest performing BzaC, *Dt*BzaC, was subject to kinetic analysis identifying a low micro molar affinity for its adenosylated substate (K_m = 11 μM), together supporting BzaC's role in cobamide biosynthesis.

The 3D crystal structure of the *Dt*BzaC dimer with SAH identified the SAM binding domain and in combination with substrate docking and structural comparison to similar enzymes allowed for a catalytic mechanism to be proposed, in which an active site histidine residue can deprotonate the substrate hydroxyl group, increasing its nucleophilicity, promoting SAM dependent methyltransfer.

In addition, biochemical analysis of *Dt*BzaE identified that it can catalyse the radical rearrangement of adenosyl 5(6)-methoxybenzimidazolylcobamide likely forming a hydroxymethyl product (Ado[5(6)-CH_2OH]Cba), deviating from the currently proposed reaction scheme for a BzaE enzyme.

Together the studies conducted in this thesis have contributed novel findings to the field, shifting the understanding of how these enzymes and pathway's function.

Table of contents

Table of figures

Tables

Abbreviations

5-aminoimidazole ribotide	AIR
5-aminolevulinic acid	5-ALA
5-deoxy-adenosine / adenosyl	Ado
5-hydroxybenzimidazole	5-OHBza
5-methoxy-6-methylbenzimidazole	5-OMe-6-MeBza
5-methoxybenzimidazole	5-OMeBza
5-methoxybenzimidazolylcobamide	[5-OMeBza]Cba
5-methylthioadenosine	MTA
5-OHBza-riboside	5-OHBza-R
5-OHBza-ribotide	5-OHBza-RP
Adenosine triphosphate	ATP
Adenosyl radical	Ado• / 5'-dA•
Adenosyl 5(6)-hydroxybenzimidazolylcobamide	Ado[5(6)-OHBza]Cba
Adenosyl 5(6)-methoxybenzimidazolylcobamide	Ado[5(6)-OMeBza]Cba
Adenosyl 5(6)-O-propargybenzimidazolylcobamide	Ado[5(6)-OPaBza]Cba
Adenosyl 5-hydroxy-6-methylbenzimidazolylcobamide	Ado[5-OH-6-MeBza]Cba
Adenosyl 5-hydroxymethylbenzimidazolylcobamide	Ado[5-CH_2-OHBza]Cba
Cobalt	Co
Cyano 5(6)- hydroxybenzimidazolylcobamide	CN[5(6)-OHBza]Cba
Cyano 5(6)- methoxybenzimidazolylcobamide	CN[5(6)-OMeBza]Cba
Desulfofundulus thermosubterraneus BzaC	*Dt*BzaC
Dimethylbenzamidizole	DMB
Dithiothreitol	DTT
Domain of unknown function	DUF

E. coli methylthioadenosine nucleosidase	*Ec*MTAN
Escherichia coli	*E. coli*
Eubacterium limosum Truncated BzaC	*El*BzaCT
High pressure liquid chromatography	HPLC
Hydroxo 5(6)- Hydroxybenzimidazolylcobamide	OH[5(6)-OHBza]Cba
Hydroxo 5(6)- Methoxybenzimidazolylcobamide	OH[5(6)-OMeBza]Cba
Hydroxo 5(6)-O-propargybenzimidazolylcobamide	OH[5(6)-OPaBza]Cba
Iron-sulphur cluster	Fe-S
Liquid chromatography linked mass spectrometry	LC-MS
Luria Bertani	LB
Methyl	Me
Methylcobalamin	MeCbl
Methylcobamide	MeCba
Microscale thermophoresis	MST
Molecular oxygen	O_2
Moorella thermoacetica BzaC	*Mt*BzaC
Nuclear magnetic resonance	NMR
Potassium cyanide	KCN
Propargyl-seleno-adenosyl-L-methionine	PaSe-AdoMet
Propargyl-seleno-methionine	PaSeMet
S-adenosyl-L-homocysteine	SAH
S-adenosyl-L-methionine	SAM
Sodium dodecyl-sulfate polyacrylamide gel electrophoresis	SDS-PAGE
Tetramethylrhodamine 5-Carboxamido-(6-Azidohexanyl)), 5-isomer	TAMRA azide
Trifluoroacetic acid	TFA

Acknowledgments

I want to take this opportunity to thank my supervisor Dr Andrew Lawrence for giving me the opportunity to join his lab. The last four years have been fantastic and I have learnt so much about vitamin B_{12}, enzymology and biochemistry from you. I feel like my confidence as a scientist has grown greatly under your supervision and for that I am very grateful. Looking forward to working with you more in the future.

I am also very grateful to Professor Martin Warren for allowing me to join his lab, attend their lab meetings and for all of the advice on my project. Thank you Dr Evelyne Deery for all of your guidance, friendship and support in the lab the last couple of years, I don't think I would have been able to achieve what I did without you. I would also like to thank my lab mates Amira, Rob and Kam for their friendship, support and great scientific conversations over the last few years, I wish you all the best.

In addition, I would like to show my appreciation to Dr Ivo Tews and his lab members at the University of Southampton for helping us to obtain and refine the crystal structure of *Dt*BzaC. Joern Werner at the University of Southampton for helping us to run the NMR of our cobamide sample. Kevin Howland at the University of Kent for all of his chromatography and mass spectrometry guidance from undergraduate and all the way through my PhD. I would also like to thank the SoCoBio DTP for allowing me to join their program and the BBSRC for funding the research.

I would also like to acknowledge my parents; who have both been there for me the whole way, thank you!

Chapter 1: Introduction

1.1 The pigments of life

The modified tetrapyrroles are generally macrocyclic molecules containing centrally coordinated metals, extensive side chains and diverse conjugations giving rise to a beautiful array of colours. They are involved in many essential biological processes such as respiration, photosynthesis and catalysis, all of which are fundamental for life (1), thus awarding them the term 'the pigments of life' (2). These modified tetrapyrroles include vitamin B_{12} (and accompanying cobamides), chlorophylls, hemes, sirohemes, coenzyme F_{430}, heme d_1, and bilins. All of which are derived from a common precursor, uroporphyrinogen III (**Figure 1.1**).

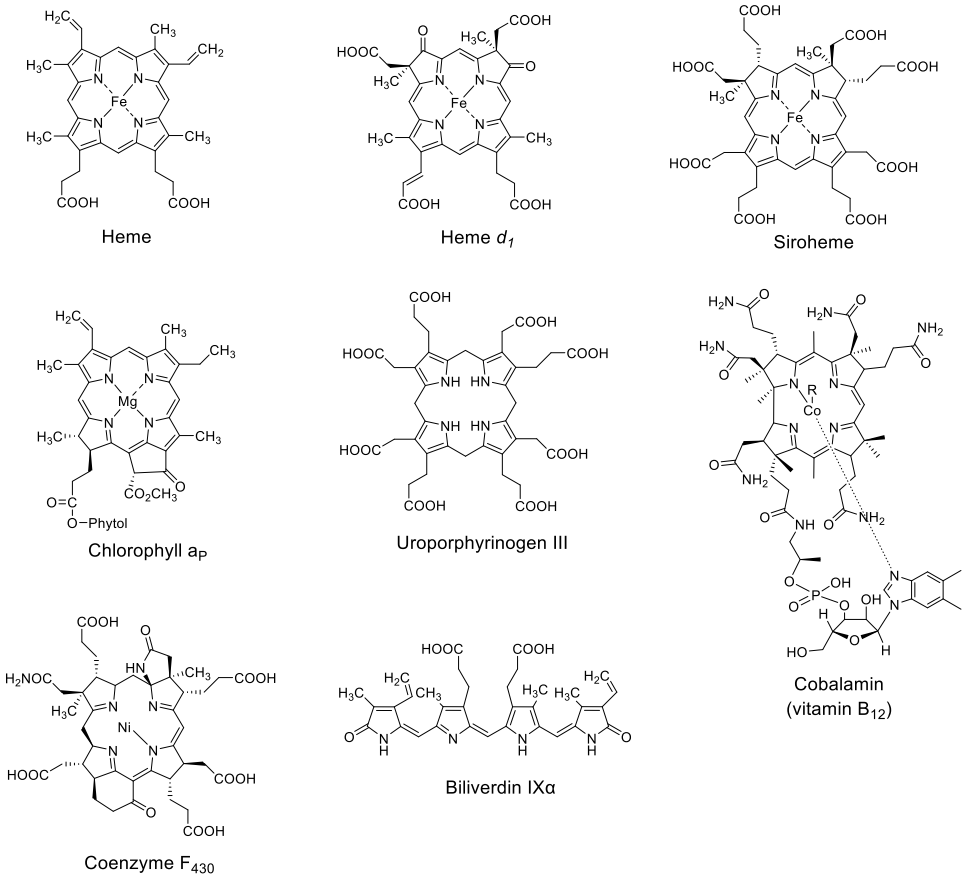


Figure 1.1 - The pigments of life.

Structures of the main modified tetrapyrroles, surrounding the common precursor uroporphyrinogen III. Adapted from (1).

1.1.1 Chlorophyll and bilins

The chlorophylls and the related bacteriochlorophylls have two main roles in photosynthesis, firstly to harvest solar energy (mediated by the electronic properties of the chlorin ring (conjugated bond system and magnesium)) and secondly to transfer this energy to reaction centres where the photochemistry for ATP generation and carbon dioxide fixation occurs (1) (**Figure 1.2**).

Bilins are also photosynthetic tetrapyrroles, however they are linear and formed from the enzymatic modification of biliverdin which is produced by the oxidative cleavage of the haem macrocycle (3). There are 4 main bilins called phycourobilin, phycoviolobilin, phycocyanobilin and phycoerythrobilin which vary in their conjugated bond systems. These form the chromophores of phycobiliproteins involved in light harvesting (4) (**Figure 1.2**) or the photosensors of sensor histidine kinases (1).

(A)

(B)

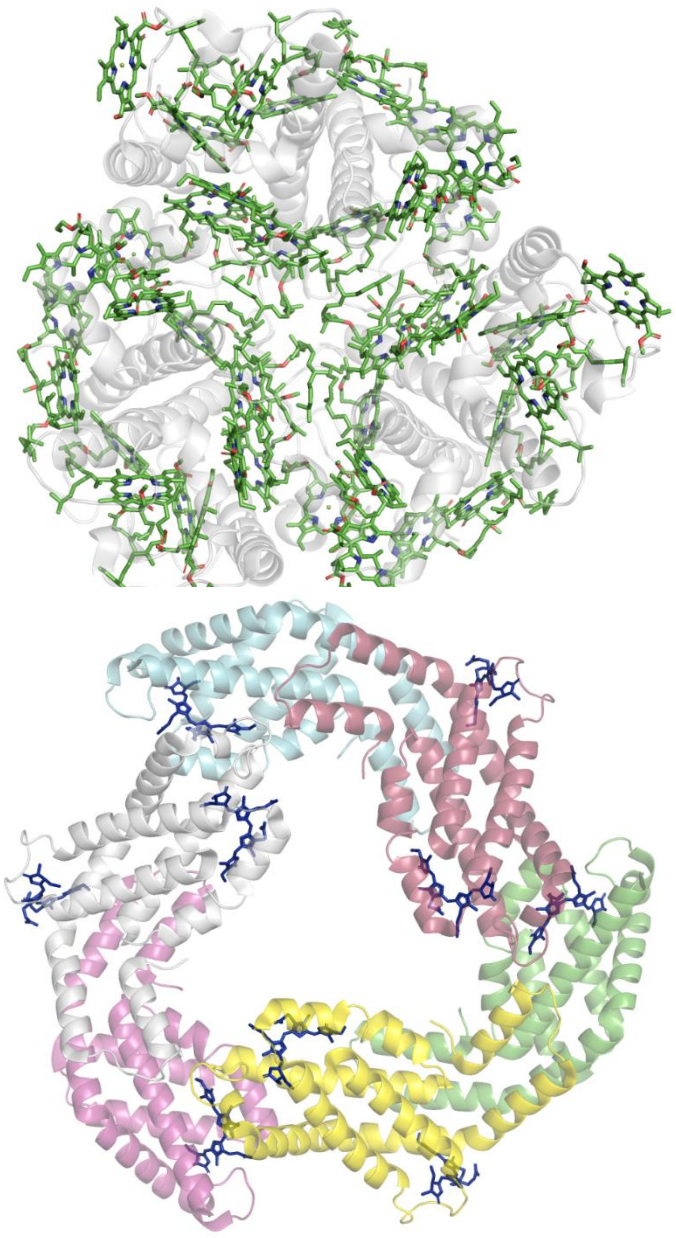


Figure 1.2 - Chlorophyll and bilin containing Light harvesting complexes.

(A) 2.5 Å crystal structure of the biological assembly of light harvesting complex II from *Pisum sativum* (PDB: 2BHW), Chlorophyll$_a$ and Chlorophyll$_b$ coloured green (5, 6) . **(B)** Crystal structure of the C-phycocyanin hexamer phycobiliprotein from the light harvesting antenna structure of *Microchaete diplosiphon* (PDB = 1CPC), each protein monomer is coloured differently and phycocyanobilin is coloured blue (7).

1.1.2 Haem

Haem is an iron containing tetrapyrrole, which has a wide range of functions in biological systems due to the multiple redox states that it can exist it. For example, it has a role in sensing gasses such as molecular oxygen (O_2), carbon monoxide (CO) and nitric oxide (NO). The most familiar example of this is the haem ferrous iron (Fe^{2+}) binding to oxygen in haemoglobin, which is stabilised by proximal and distal histidine residues (8–10), to allow for efficient oxygen transport (**Figure 1.3**). Furthermore, it can be utilised as an enzyme cofactor by catalases, cytochromes P_{450} and peroxidases and as an electron carrier in respiratory complexes (1).

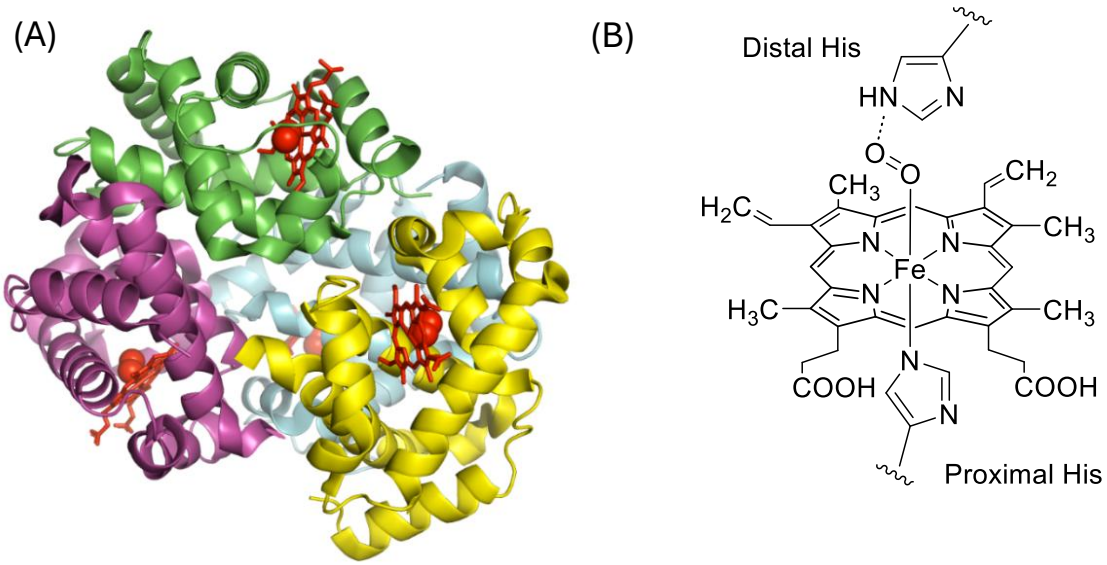


Figure 1.3 - O_2-haem binding in haemoglobin.

(A) Crystal structure of human oxyhaemoglobin (PDB ID = 2DN1) (9) with oxyhaem bound at each monomer. Haem and O_2 are displayed in red. **(B)** A schematic representation of O_2-haem binding and stabilisation in a haem binding domain of haemoglobin by the covalent bond between the proximal histidine and the central iron and the distal histidine with O_2.

1.1.3 Coenzyme F_{430}

Coenzyme F_{430} is a nickel containing tetrapyrrole which is a coenzyme involved in anaerobic methanogenesis and methane oxygenation mediated by the redox properties of the central nickel allowing for the reduction or oxidation of a substrate methyl group to form / consume methane. This is catalysed by an enzyme called Methyl-coenzyme M reductase (MCR) in anaerobic archaeal methanogens and methanotrophs (11, 12).

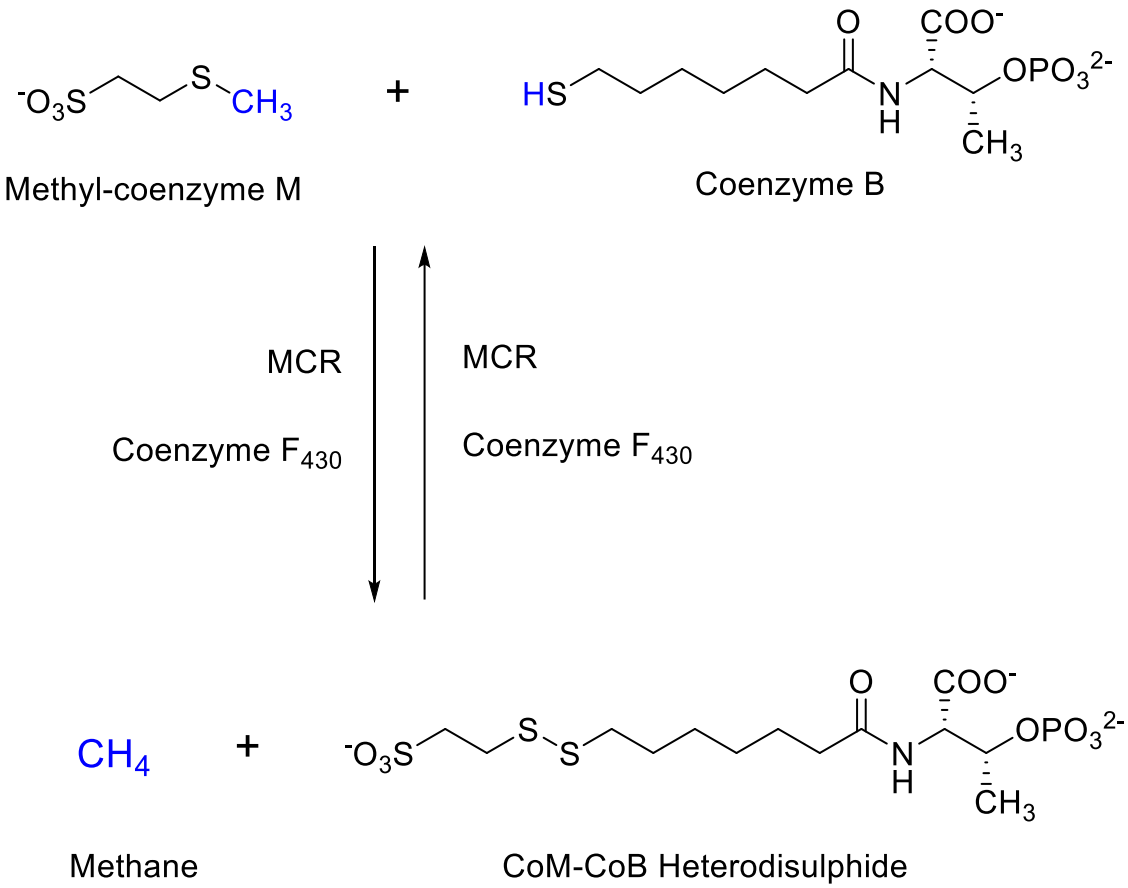


Figure 1.4 - The role of coenzyme F_{430} in Methyl-coenzyme M reductase catalysed methanogenesis / methane oxidation.

Adapted from (11).

1.1.4 Siroheme and heme d_1

Siroheme and heme d_1 are both iron containing cofactors, with siroheme being used by sulfite and nitrite reductases and heme d_1 by cytochrome cd_1 which is also a nitrite reductase.

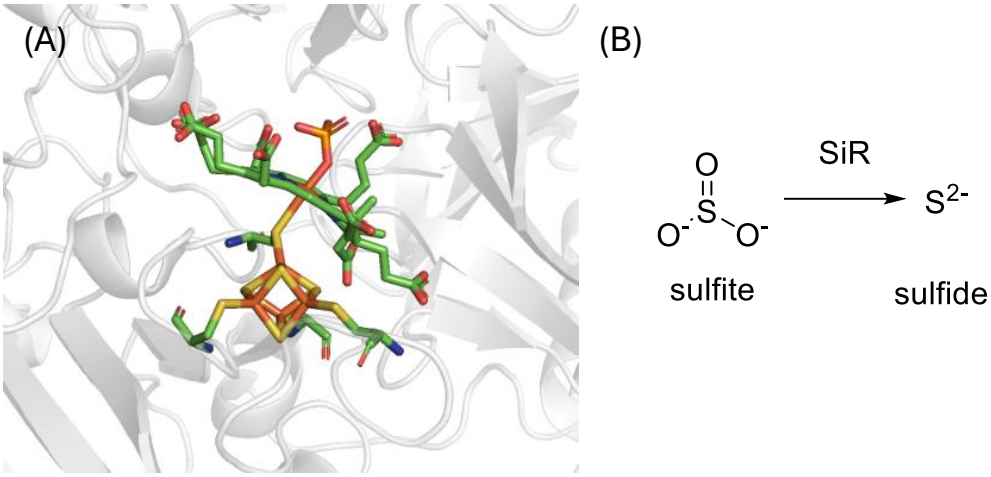


Figure 1.5 - Assimilatory sulfite reductase (SiR).

(A) Crystal structure of SiR (PDB ID: 1AOP) with the Siroheme bound to an iron sulphur cluster (orange and yellow) and an unexpected phosphate ion (red and orange) (13). **(B)** Schematic of the reaction catalysed by SiR.

1.2 Vitamin B_{12}

Vitamin B_{12} (cobalamin) is a complex molecule that is essential for human life. It is a modified tetrapyrrole which acts as a cofactor/coenzyme in a variety of biochemical reactions across the kingdoms of life such as isomerisation, methylation, and reductive halogenation. It was first identified via work carried out by Minot, Murphy and Whipple who demonstrated that the symptoms of pernicious anaemia could be

combated by adding liver to the diet (14, 15). This molecule was then isolated from liver samples and crystalised by Dorothy Hodgkin in 1956 (16) to reveal a cobalt containing, cyanolated, tetrapyrrole, which was named cyanocobalamin (17).

1.2.1 Structure

The structure of vitamin B_{12} can be split into 3 main components: a corrin ring containing a central cobalt ion, and alpha (α) and beta (β) axial ligands. The corrin ring is a modified tetrapyrrole which has undergone a ring-contraction process in which one of the carbon atoms (C20) which is used to connect the four pyrrole molecules has been removed, producing a lopsided macrocycle which coordinates a central cobalt ion. Furthermore, the cobalt ion is further coordinated by a lower ligand (α), attached to the corrin by an aminopropanol linked nucleotide loop, which in the biologically active form for eukaryotes is dimethylbenzamidizole (DMB). In other cobamides this lower ligand has been observed to be a variety of molecules such as benzimidazoles, phenol derivatives or purines (18). The upper ligand (β) in the originally isolated B_{12} molecule is a cyano group (due to the method of extraction), however in biological systems this group is either an adenosyl group (Ado), a methyl group (Me) or H_2O / OH (aqua / hydroxo group) (17, 19).

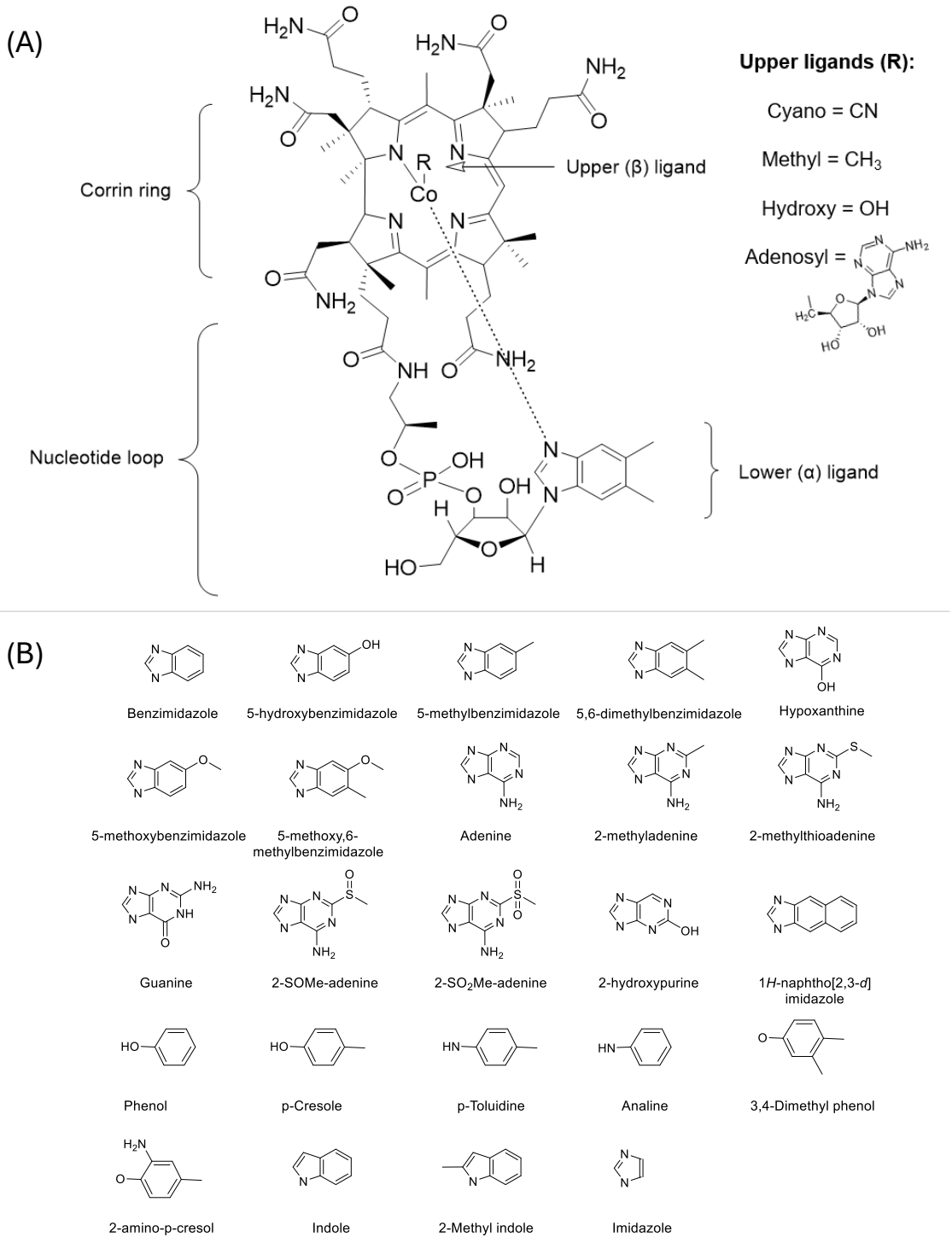


Figure 1.6 - Structure and ligand diversity of cobamides.

(A) Structure of vitamin B_{12}, with upper ligand variants indicated at the R position. **(B)** 24 molecules that have been identified in the lower ligand position of cobamides in different organisms.

1.2.2 Biological functions

Vitamin B_{12} can act as a cofactor or coenzyme in a variety of essential biochemical reactions. This biochemistry is mediated by the ability of the central cobalt ion to switch between different oxidation states (Co(III), Co(II) and Co(I)) and form cobalt – carbon bonds with the upper (β) axial ligands. Enzymes which utilise B_{12} can be split into four broad categories involving: isomerisation, methylation, s-adenosyl-L-methionine (SAM) processes and reductive halogenation. However, in eukaryotes only methionine synthase and methylmalonyl-CoA mutase use B_{12} (17).

1.2.3 α-ligand influence on biological activities

Biologically active cobamides can exist with either a 5-deoxy-adenosine (Ado) or methyl (Me) group upper ligand, which leaves Co in a 3^+ (III) oxidation state. This Co-C bond can undergo homolytic or heterolytic cleavage in B_{12}-dependent enzymatic reactions to produce Co(II) and an adenosyl radical (Ado·) or Co(I) and transfer of the methyl group, respectively, thus dictating the type of reaction catalysed by the enzyme (20).

1.2.4 Methylcobalamin dependent enzymes

Methylcobalamin (MeCbl) or methylcobamide (MeCba) dependent enzymes are important methyltransferases involved in energy generation in anaerobic organisms and methionine production from homocysteine. These methyl-transfer reactions

occur via S_N2 displacement mechanisms and require regeneration of the MeCba co-factor (facilitated by the super nucleophilic properties of cob(I)alamin)(21).

One of the best characterised examples of this is the methylcobalamin dependent methionine synthase (MetH) from *Escherichia coli* (*E. coli*). In this example, homocysteine (specifically, a thiol coordinated to a Zn(II)-ion) initiates nucleophilic attack on the methyl group of MeCbl, producing methionine and cob(I)alamin. The cob(I)alamin super nucleophile then attacks the methyl group of methyltetrahydrofolate producing methylcobalamin and tetrahydrofolate (Figure 1.7) (22).

This is an important reaction as methionine is required in a variety of central metabolic reactions such as for the generation of SAM which is required for an array of methylation reactions. Furthermore, disruptions result in the build-up of homocysteine and folate (trapped as methyltetrahydrofolate) which decreases the availability of tetrahydrofolates used in purine and thymidylate synthesis. This thus reduces protein synthesis and cell proliferation, producing megaloblastic anemia (17).

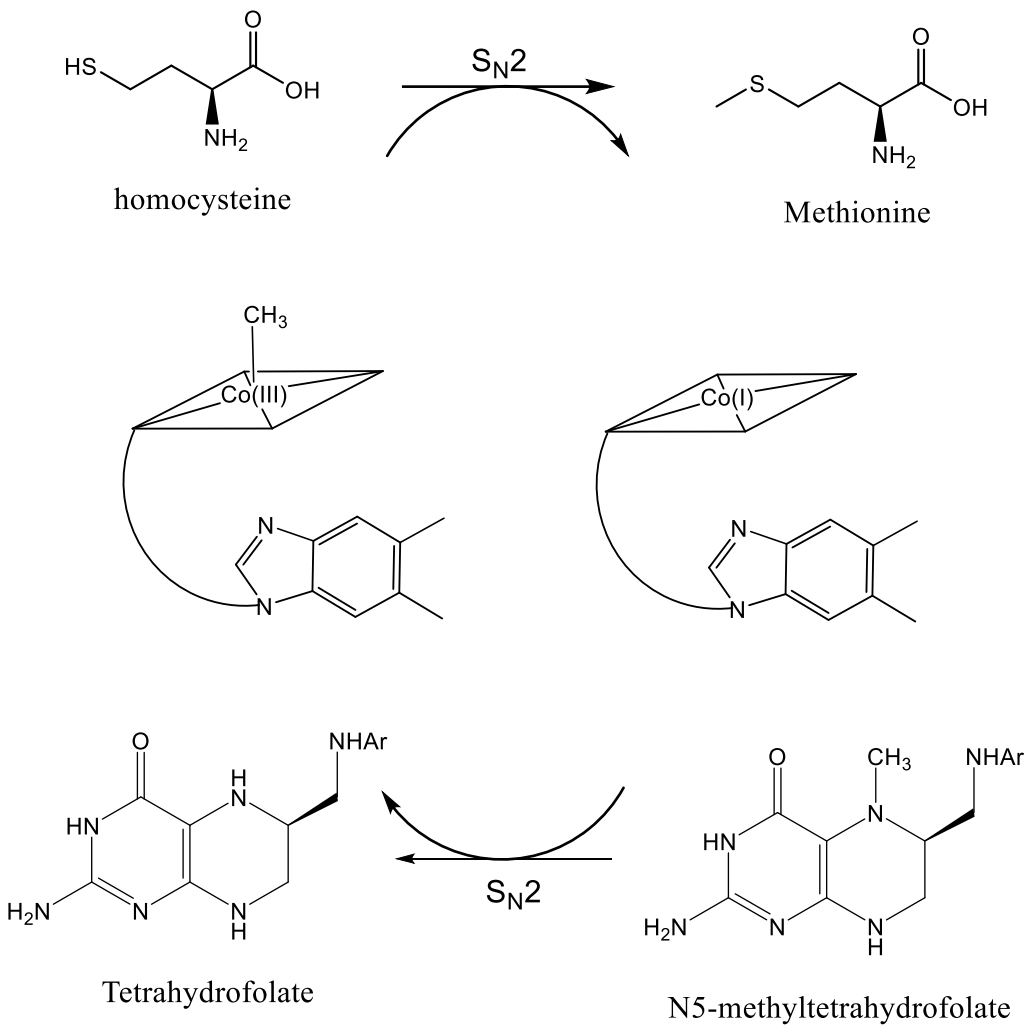


Figure 1.7 - MetH reaction scheme.

A schematic representation of the formation of methionine from the S_N2 nucleophilic substitution of homocysteine with CH_3 from methylcob(III)alamin. Cob(I)alamin is then recharged to methylcob(III)alamin using methyltetrahydrofolate. Adapted from (22).

1.2.5 Adenosylcobalamin dependent enzymes

The other major corrin upper ligand is 5'-deoxyadenosine. Enzymes utilising this co-factor engage in radical based reactions due to its ability to generate a 5'-deoxyadenosine radical (Ado·), which can then abstract a hydrogen from a substrate leading to complex rearrangements or removal of groups. The Ado· is then regenerated by abstraction of a proton from AdoH by the substrate radical, allowing the Ado· to recombine with cob(II)alamin to make adenosylcobalamin (AdoCbl) (**Figure 1.8**). The exact mechanism of the reaction varies on the substrate and the transformation required (23).

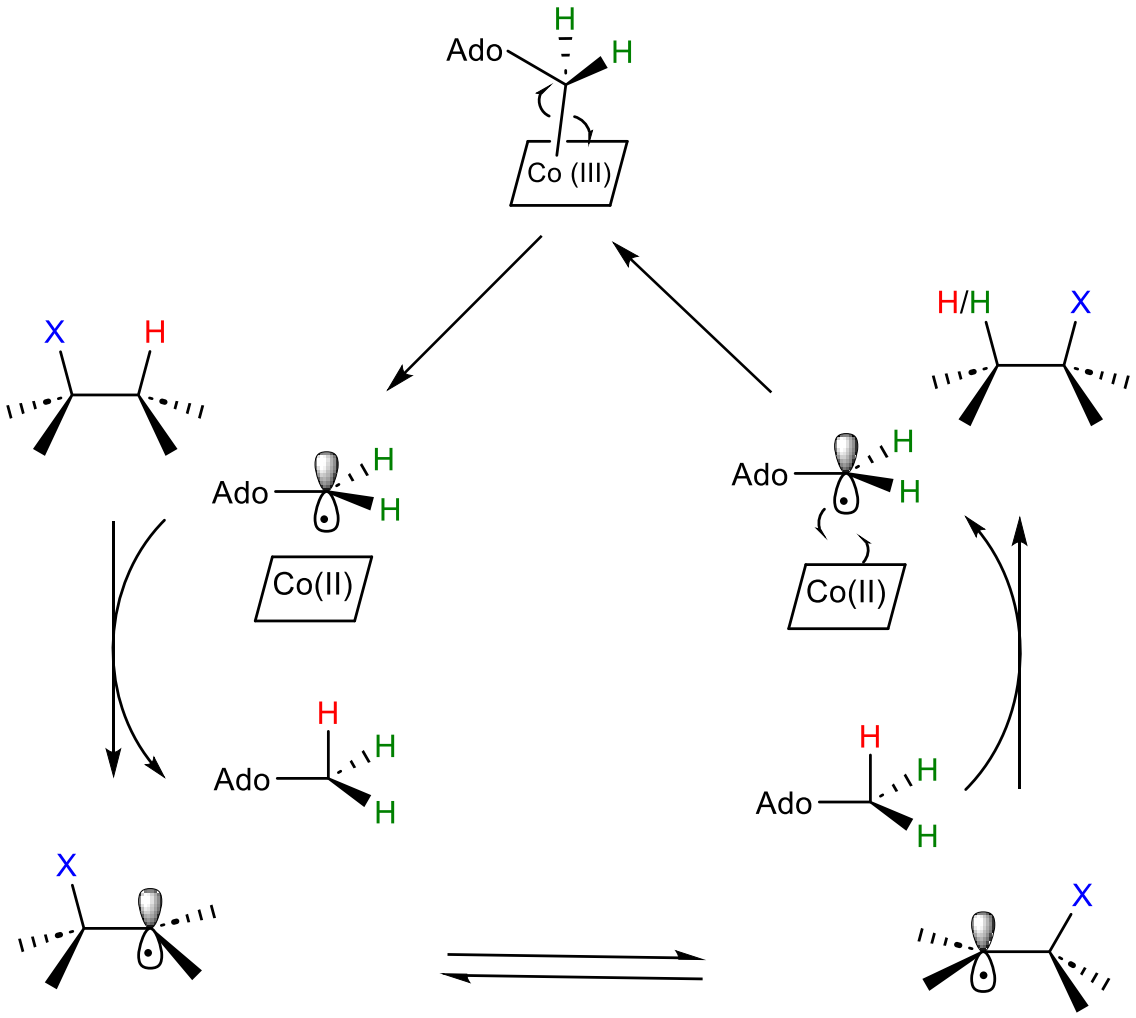


Figure 1.8 - Generalised mechanism of a AdoCbl dependent mutase catalysed rearrangement reaction.

Although this mechanism demonstrates how mutases work, it can also be applied to the eliminases and aminomutases (**Figure 1.9**). Furthermore, the mechanism for the class II ribonucleotide reductases is similar, but requires the production of an active site cysteinyl radical from the Ado· for its catalysis. Blue X = variable group to be moved. Adapted from (23).

These enzymes can be split into three major groups – (i) carbon-skeleton mutases (ii) eliminases and (iii) aminomutases. The carbon-skeleton mutases (i) work by migrating a hydrogen from one carbon to another within the substrate, eliciting a skeletal rearrangement. Examples of these are glutamate mutase (24) and methylmalonyl-CoA mutase (only AdoCbl dependent enzyme identified in animals characterised to date) (25). Similarly, the eliminases (ii) work by catalysing the rearrangements of ether 1,2-diols (e.g. Diol dehydratase and glycerol dehydratase), or 1,2-amino- alcohols (e.g. Ethanolamine ammonia-lyase) leading to the elimination of water or ammonia, producing aldehydes (23, 26, 27). Finally, the aminomutases (iii) catalyse the 1,2-migration of amino groups, utilising pyridoxal phosphate as a co-factor. For example, lysine-5,6-aminomutase (28). In addition to these subgroups one other enzyme group has been identified, the class II ribonucleotide reductases, which catalyses the reduction of ribonucleotides to deoxyribonucleotides (23).

Interestingly, the photolabile Co–C bond of AdoCbl has been harnessed by a transcription factor CarH to activate transcription of certain genes, which are required to mitigate light-induced damage, in the presence of light. The heterolytic cleavage of the Co-C bond upon light exposure leads to the disassembly of the AdoCbl - CarH tetramer assembly, which causes its disassociation from DNA, thus allowing transcription to occur (29, 30).

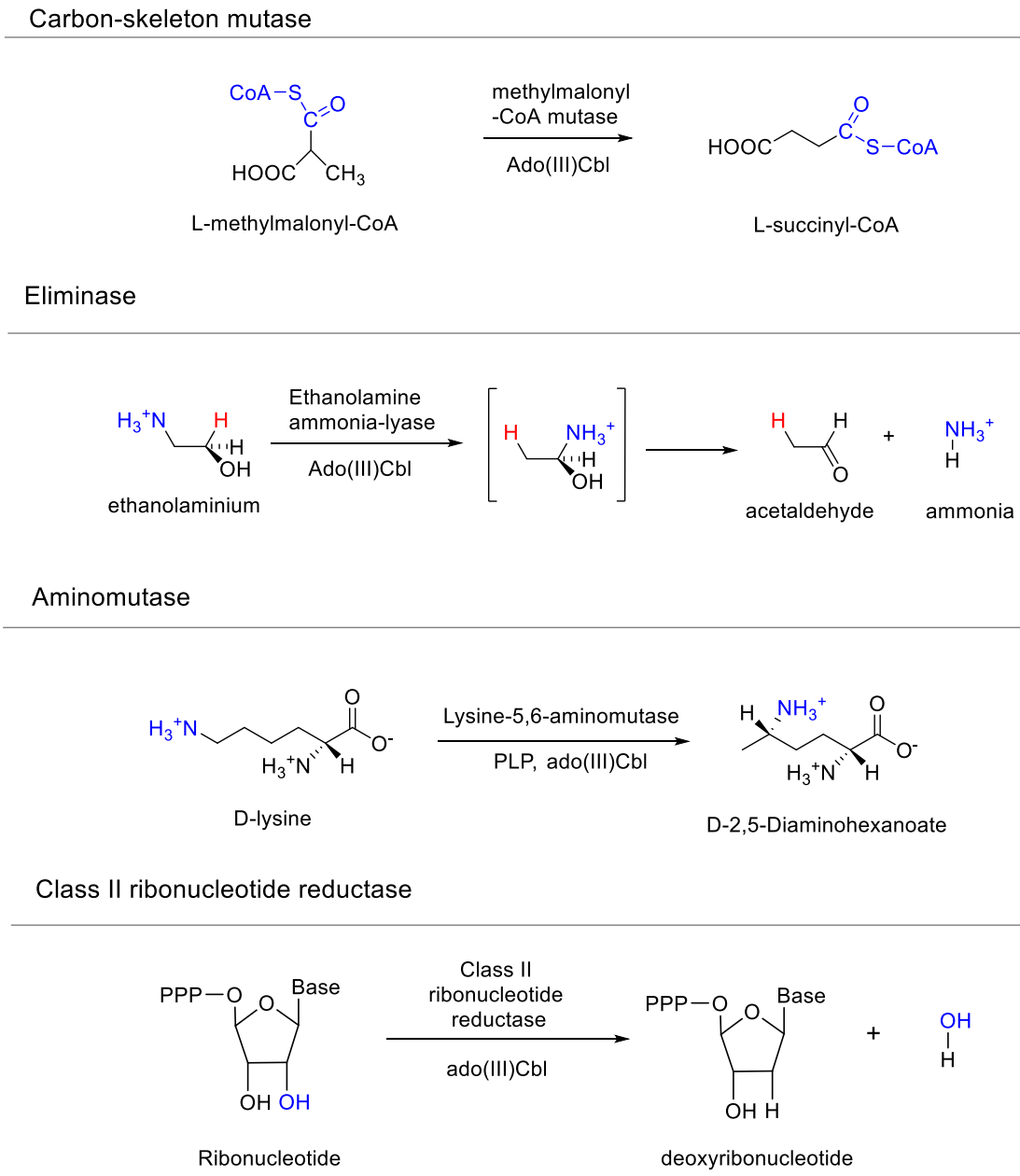


Figure 1.9 - The four classes of reactions catalysed by ado(III)Cbl dependent enzymes.

Carbon skeleton mutases exemplified by methylmalonyl-CoA mutase, eliminases exemplified by Ethanolamine ammonia-lyase, aminomutases by lysine-5,6-aminomutase and the class II ribonucleotide reductase.

1.3 Biosynthesis of the modified tetrapyrroles

1.3.1 Biosynthesis of uroporphyrinogen III

The biosynthesis of the tetrapyrroles occurs along a branched pathway, where they all originate from a common precursor; uroporphyrinogen III. There are two main routes for producing this molecule, which coincide with the production of 5-aminolevulinic acid (5-ALA). Some organisms (animals, fungi and some purple bacteria) (31) produce this molecule via a single enzymatic transformation, catalysed by ALA synthase, consisting of a decarboxylating condensation reaction between succinyl-CoA and glycine, named the C4 pathway (32). Other organisms (plants, archaea and most bacteria) (31) require a multi-step pathway (C5) starting with the conversion of Glutamate to Glutamyl-tRNA (33), subsequent transformation to glutamate-1-semialdehyde (GSA) by glutamyl-tRNA reductase (34). Glutamate 1-semialdehyde is then rearranged by Glutamate semialdehyde aminomutase to produce 5-ALA (35) (**Figure 1.10**).

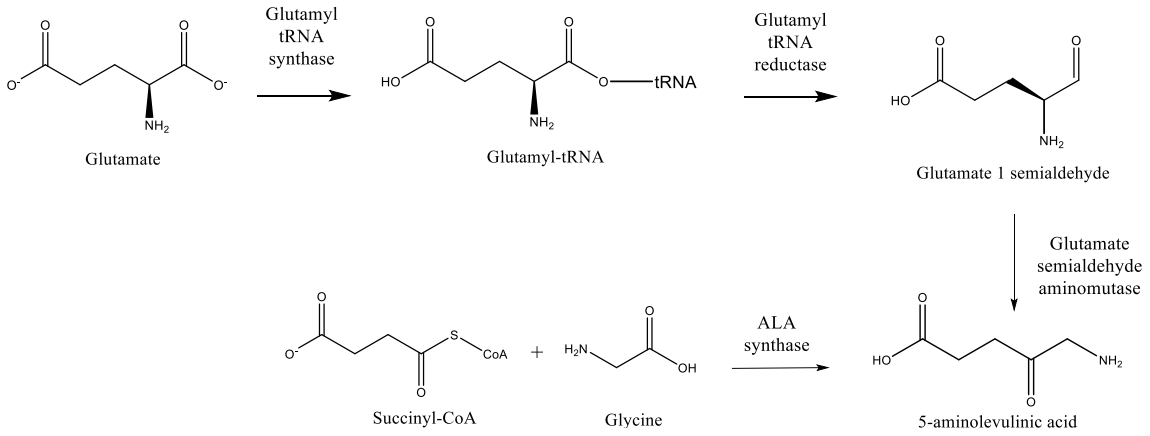


Figure 1.10 - C4 and C5 pathways for the biosynthesis of 5-ALA.

Schematic representation of the enzymatic pathways to produce 5-ALA, with C4 being on the bottom and C5 being displayed on the top.

From 5-ALA all organisms follow the same three step enzymatic pathway to uroporphyrinogen III. Firstly, an enzyme called porphobilinogen synthase (PbgS) catalyses a Knorr-type condensation reaction of two molecules of 5-ALA to produce the first pyrrole; porphobilinogen (PBG) (36). Four molecules of PBG are then converted to hydroxymethylbilane (HMB) via polymerisation catalysed by HMB synthase (HmbS) (37). Finally, uroporphyrinogen III synthase (UroS) inverts ring D of the bilane, which is important for substrate orientation in subsequent enzymatic steps as it produces asymmetry in the molecule, and catalyses the cyclisation of the bilane to produce uroporphyrinogen III (**Figure 1.11**) (1).

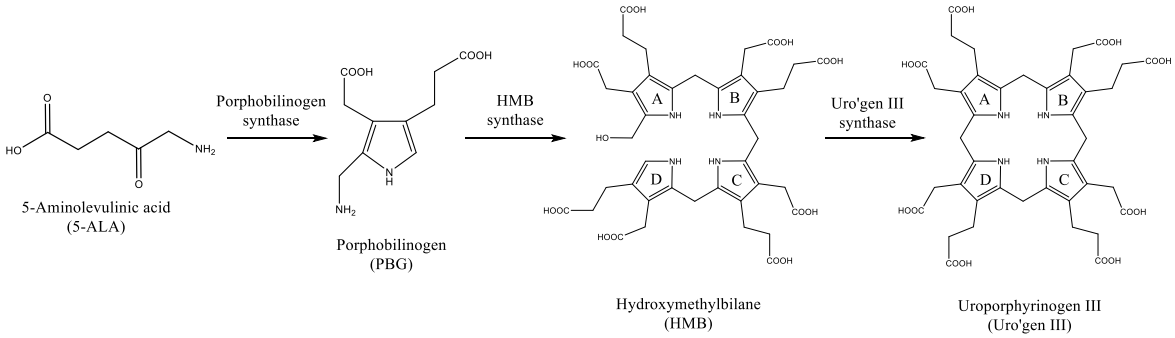


Figure 1.11 - Pathway for the conversion of 5-ALA to uro'gen III.

Schematic of the enzymatic conversion of PBG to HMB via HMB synthase and subsequent conversion to uro'gen III via uro'gen III synthase.

1.3.2 Tetrapyrrole biosynthetic branch point

The modifications of uro'gen III dictate the type of tetrapyrrole that will be subsequently produced. If the macrocycle is methylated at positions 2 and 7, forming an unstable intermediate named precorrin-2, the pathway is directed towards the production of cobamides, coenzyme F_{430} or siroheme. However, if the Uro'gen III carboxymethyl groups are decarboxylated, it directs biosynthesis towards chlorophylls and heme.

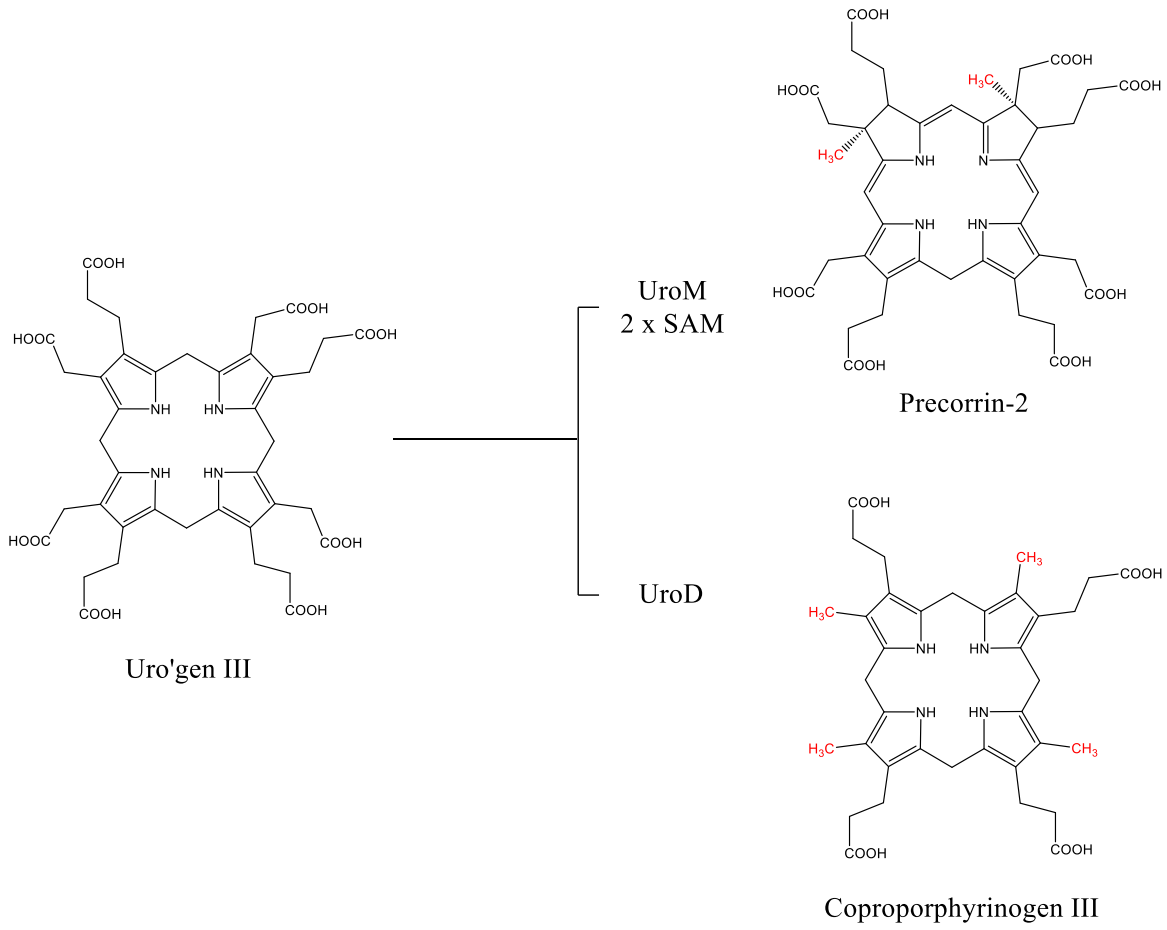


Figure 1.12 - Uro'gen III branch point.

Schematic of the two transformations that direct biosynthesis either to cobamides, coenzyme F_{430} or siroheme (UroM 2 x methylation) or towards heme and chlorophylls (UroD 4 x decarboxylation).

1.3.3 Biosynthesis of cobamides

The biosynthesis of vitamin B_{12} is complex and requires approximately 30 enzymatic transformations. This is a very energetically taxing process due to the requirement of a large number of enzymes and co-factors (e.g. ATP, NADPH, SAM). Furthermore, it is exclusively carried out by certain bacteria and archaea, meaning eukaryotes need to

obtain it in their diet for survival. There are two main pathways towards the production of B_{12}, the aerobic and anaerobic route, which differ in the requirement for O_2, the timing of cobalt insertion, and the method of ring contraction, with the anaerobic route being the most commonly utilised (1).

1.3.3.1 Construction of the corrin ring

For both pathways the transformation of uro'gen III into cobyrinic acid a,c-diamide (the corrin ring) requires the transfer of eight methyl groups derived from SAM molecules, an unusual ring contraction step where the methylated C20 carbon is extruded from the structure, a decarboxylation, six amidations, and insertion of the central cobalt. The main difference between the two routes, apart from the requirement / absence of O_2, is that in the aerobic pathway the cobalt is added after all the ring modifications have occurred, giving it the term 'late cobalt insertion'. Whereas, in the anaerobic pathway the cobalt insertion occurs early, just after precorrin II is oxidised to sirohydrochlorin. Overall, the same final structure is produced, however the order of transformations, the mechanisms required and the intermediates produced are different (1).

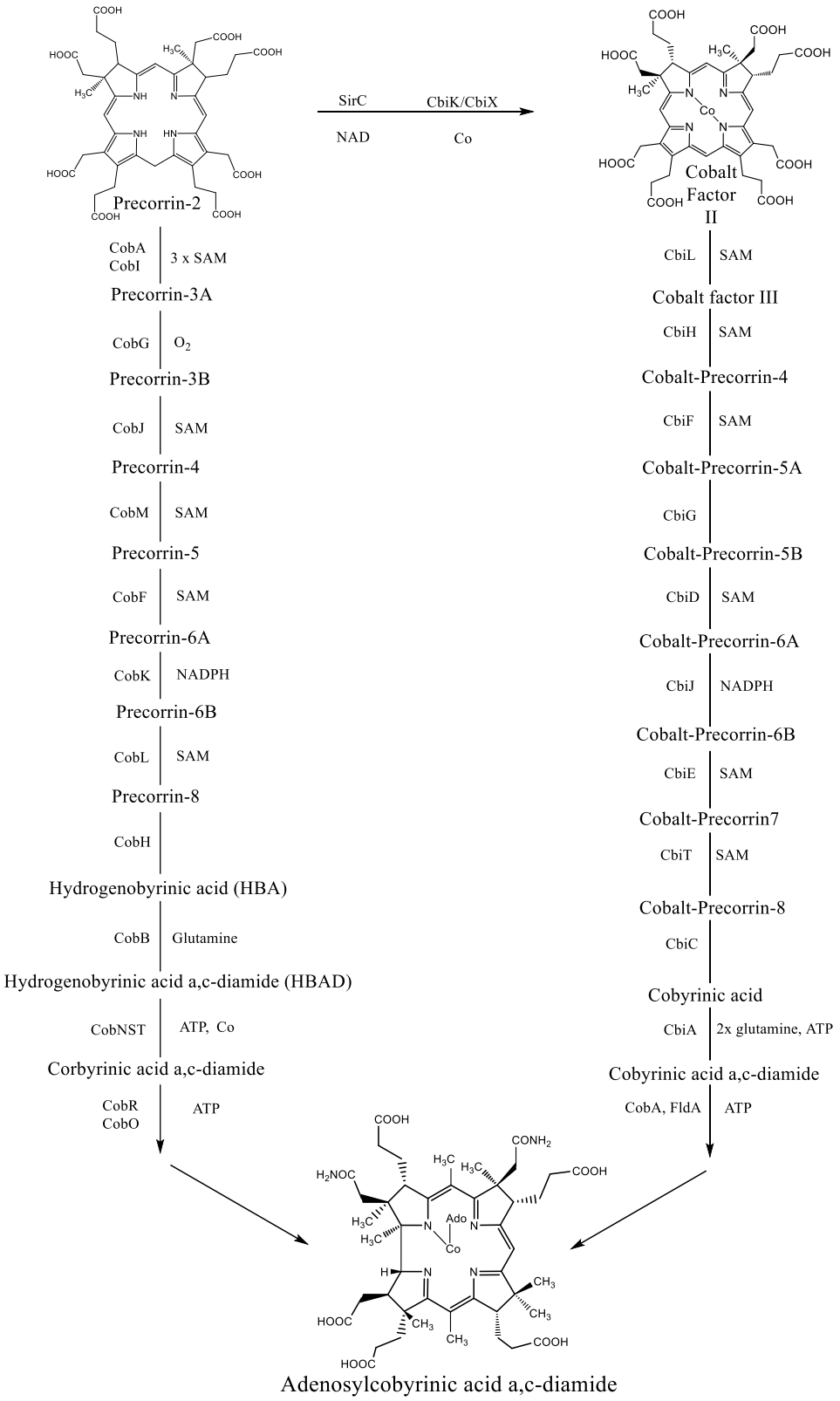


Figure 1.13 - Aerobic and anaerobic routes for the biosynthesis of Adenosylcobyrinic acid a,c-diamide.

The aerobic route is donated by the Cob abbreviation and the anaerobic by Cbi.

1.3.3.2 Addition of the α-axial ligand and nucleotide loop

Now that both pathways have produced a corrin ring (cobyrinic acid a,c-diamide), the central cobalt is adenosylated in the β-axial position giving rise to the upper ligand. The next step is to produce the nucleotide loop by the addition of an aminopropanol linker and the attachment of the α-axial (lower) ligand. Firstly, PduX (38) or BluE (39) catalyse the phosphorylation of threonine utilising ATP. This is then decarboxylated to produce aminopropanol phosphate and attached to the free carboxyl of adenosylcobyrinic acid by the action of CobD/CbiB and CobC to produce cobinamide (40–42). This is then primed for the attachment of the lower ligand via the attachment of GDP by CobP (aerobic) (43) or CobU (anaerobic) (44) or CobY (45) to the aminopropanol linker.

The DMB, or lower base equivalent in cobamides, is then activated for attachment to the cobinamide via the addition of a ribotide phosphate group from nicotinamide mononucleotide – catalysed by CobU/T, (46) producing an α-ribazole phosphate. CobV/S then catalyses the displacement of GDP with the α-ribazole phosphate to produce adenosylcobalamin phosphate (47). This phosphate is then removed via the action of CobC to produce the completed adenosylcobalamin or cobamide equivalent (1, 48) (**Figure 1.14**). 20 different bases have been observed in the lower ligand position in nature, which produces the structural diversity of the cobamide family (1).

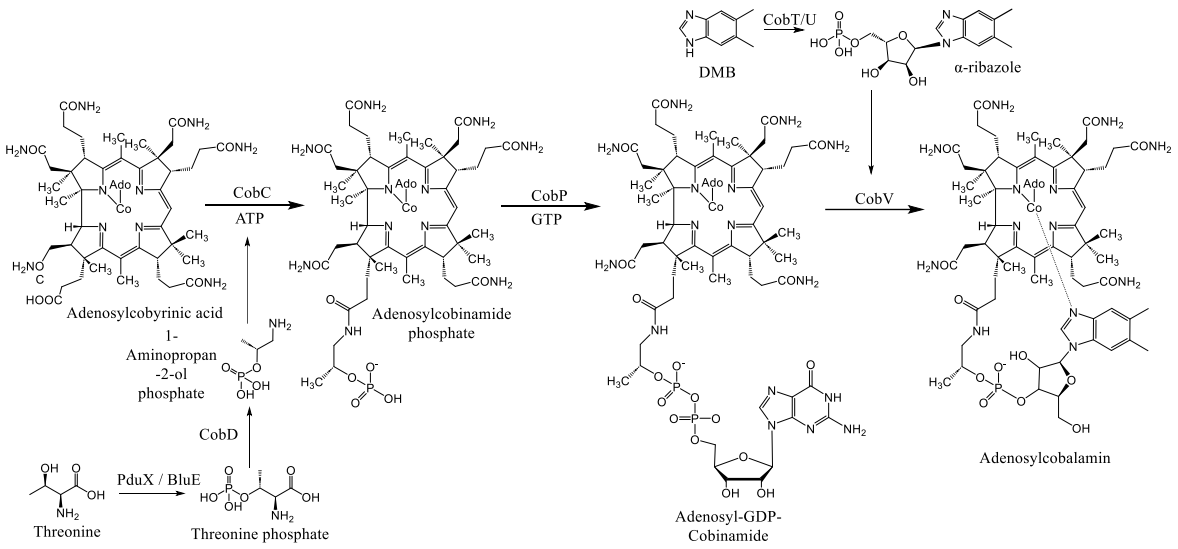


Figure 1.14 - Production of the nucleotide loop of vitamin B_{12}.

1.4 Biosynthesis of DMB

As mentioned above, the final steps of the biosynthesis of cobalamin and cobamides is the attachment of the α-axial (lower) ligand. In the case of the biosynthesis of cobalamin, DMB can be produced by two alternative pathways; the aerobic or anaerobic.

1.4.1 Aerobic pathway

In the aerobic pathway this is catalysed by an enzyme called BluB which carries out the fragmentation of reduced flavin mononucleotide ($FMNH_2$) using O_2 to produce DMB (49).

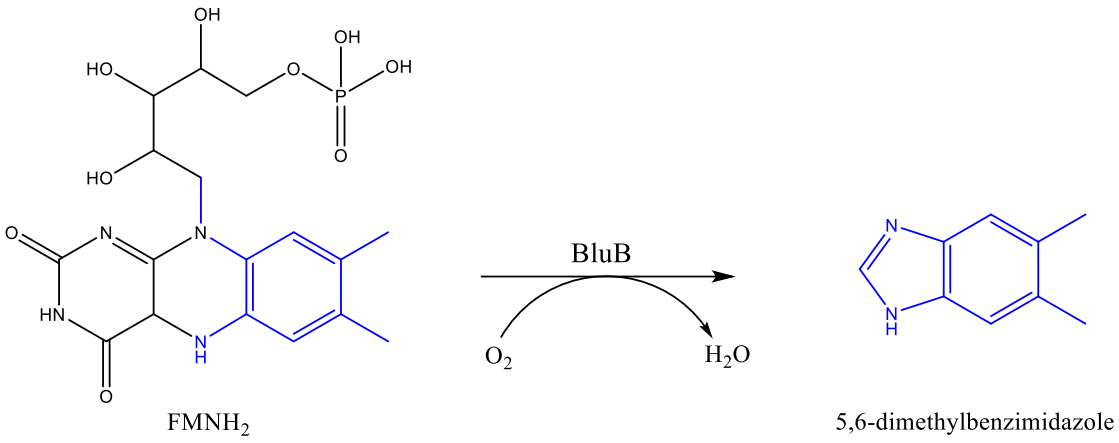


Figure 1.15 - BluB catalysed $FMNH_2$ fragmentation forming DMB.

1.4.2 Anaerobic pathway

The anaerobic pathway is more complicated due the inability to use O_2, and remained a mystery until 2015 when Hazara et al (50) first identified the genes required for the anaerobic biosynthesis of dimethylbenzamidizole (DMB) in *Eubacterium Limosum*. This was achieved via bioinformatic analysis of regions downstream from cobalamin riboswitches in the genome of *E. limosum* strain KIST612, identifying an operon containing 6 genes; *bzaA, B, C, D, E* and *cobT*, named the *bza* operon. *bzaA* and *bzaB* are annotated as *thiC*, which encodes a radical SAM enzyme involved in thiamine pyrophosphate biosynthesis. *cobT* encodes a phosphoribosyltransferase required for the activation of cobamide lower ligands (51, 52), *bzaC* encodes an enzyme annotated as a SAM dependent methyltransferase family protein, and *bzaD* and *E* are annotated as two B_{12} binding, Fe-S / SAM superfamily protein encoding genes. Due to the *bzaD* and *E* enzyme encoding genes having annotated B_{12} binding, iron-sulphur cluster (Fe-S) and SAM binding domains, it is hypothesised that they may belong to the newly identified B_{12} dependent radical SAM enzyme family (53, 54).

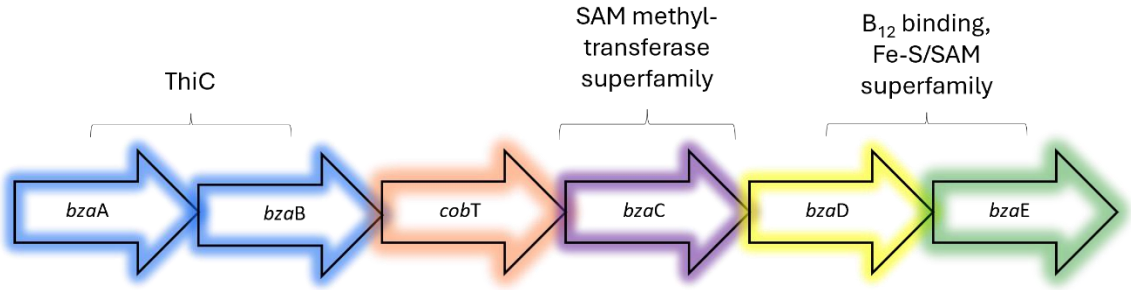


Figure 1.16 - *E.limosum bza* operon.

Schematic representation of the *bza* operon identified by Hazara et al 2015. It contains two genes with homology to *thiC* (A and B), followed by *cobT*, and then *bzaC* (SAM dependent methyltransferase family) and *bzaD* and *E* (two B_{12} binding Fe-S/SAM superfamily encoding genes).

The biochemical function of the operon was characterised by heterologous expression in *E. coli*. This model organism has an incomplete biosynthetic pathway for cobalamin, but does contain *cobUTSC* genes, thus allowing it to take lower ligand molecules such as purines or benzimidazoles and attach them to cobinamide to form complete cobamides or cobalamin itself. These cobamides can then be extracted, converted to their cyano form and analysed by liquid chromatography linked mass spectrometry (LC-MS) and nuclear magnetic resonance (NMR). Therefore, Hazara et al. 2015 expressed the *bza* operon in *E. coli*, grown anaerobically, feeding cobinamide and extracted and analysed the resulting cobamides, identifying cyanocobalamin. This therefore demonstrated the enzymes encoded by these genes were able to produce DMB.

To investigate the function of each enzyme, constructs of each gene separately and in sequential additions were produced and expressed. This firstly showed that BzaA and B were required to produce 5-hydroxybenzimidazole (5-OHBza) from 5-aminoimidazole ribotide (AIR). BzaC was then required to produce 5-methoxybenzimidazole (5-OMeBza), 5-methoxy-6-methylbenzimidazole (5-OMe-6-MeBza) production requires the addition of BzaD, and BzaE is required to produce DMB (**Figure 1.17**). Furthermore, this data correlates with the fact that *Moorella thermoacetica* has been shown to produce 5-methoxybenzimidazolycobamide [5-OMeBza]Cba and contains *bzaA*, *B*, *cobT* and *bzaC*.

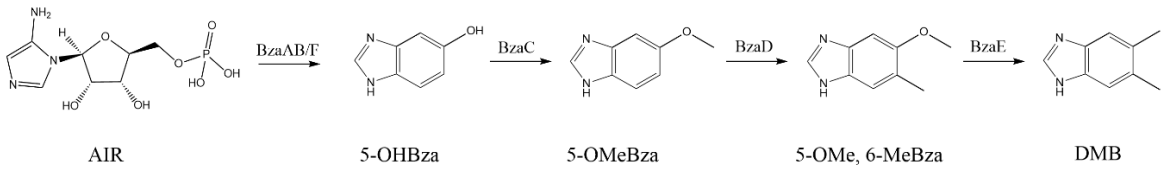


Figure 1.17 - Proposed pathway for the anaerobic biosynthesis of DMB.
Schematic representation of the proposed Bza enzyme catalysed biotransformations required for the conversion of AIR into DMB by the actions of BzaA, B, C, D and E.

1.4.2.1 Alternative *bza* operons

Since the identification of the *bza* operon of *E. limosum*, additional *bza*-like operons have been identified in different organisms with different gene compositions such as *Desulfuromonas acetoxidans* with just contains *cobT* and *bzaF* and *Morella*

thermoacetica which contains *bzaA*, *B*, *T* and *C* (55). Furthermore, an interesting operon from *Desulfofundulus thermosubterraneus* which contains *bzaA*, *B*, *C*, *X* (DUF domain containing) and *E2* (an unusual *bzaE*). These therefore present routes by which anaerobic organisms may produce benzimidazole lower ligand variants via different biochemical mechanisms.

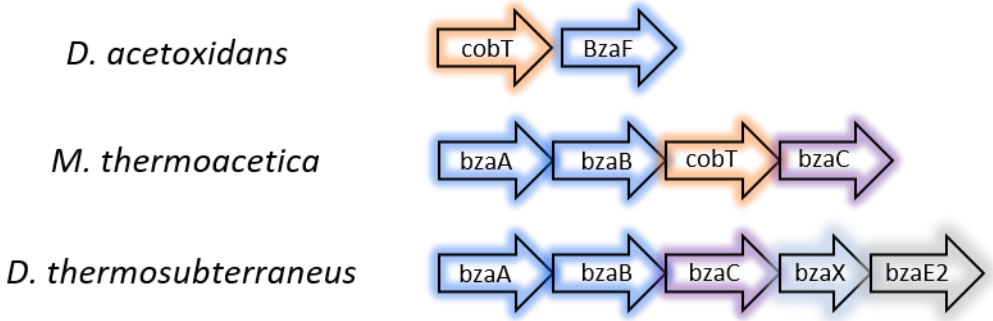


Figure 1.18 - Alternative *bza* operons.

1.5 Aims of the thesis

The aim of this thesis is to provide additional biochemical insight into how the anaerobic pathways for the biosynthesis of benzimidazolylcobamides work. Thus far the enzymes that catalyse the conversion of AIR into DMB have been identified, and via heterologous expression the pathway intermediates have been identified, however there is little biochemical detail on how these enzymes catalyse their reactions. Furthermore, there has been little biochemical investigation of how these alternative pathways operate and what products they might produce.

The work discussed in this thesis primarily focuses on the characterisation of the SAM dependent methyltransferase BzaC from different *bza* operons and begins characterisation of the BzaE2 enzyme. To allow for this characterisation a diverse set

of experimental techniques were employed such as protein purification, enzyme activity reconstitution, reaction kinetics, ligand binding affinity analysis, structural biology and analytical chemistry. This has resulted in an increased understanding of how these enzymes function on a mechanistic level and have provided a different perspective on how these pathways may operate within the producing strains.

Chapter 2: Materials and methods

2.1 Chemical and reagents

Most chemicals and reagents were obtained from Sigma-Aldrich (Merck), Doug Discovery (for benzimidazole's) or Fisher scientific. Restriction enzymes were bought from Promega. PD10 columns and chelating Sepharose were obtained from Cytiva. The S-adenosyl-L-Methionine used in this thesis, unless stated otherwise, was S-(5'-Adenosyl)-L-methionine p-toluenesulfonate salt from Sigma-Aldrich (A2408).

2.2 Bacterial strains and plasmids

2.2.1 Bacterial strains

Bacterial strains were provided by Dr Evelyne Deery (Warren lab, University of Kent) and Dr Andrew Lawrence.

Table 2.1 - List of bacterial strains used in this thesis.

Name	Description	Source
Escherichia coli		
DH10β	F– *mcr*A Δ(*mrr-hsdRMS-mcrBC*) ϕ80*lacZ*ΔM15 Δ*lacX74* *recA1* *endA1* *araD139* Δ (*ara-leu*)*7697*	Thermo Scientific

	galU galK λ– *rpsL*(Str^R) *nup*G	
BL21 Star (DE3)pLysS-*btu*B	F$^-$*omp*T *hsd*SB (r_{B^-}, m_{B^-}) *gal dcm rne*131 (DE3) pLysS-*btu*B (CamR)	Dr Evelyne Deery / Invitrogen
ED674	MG1655 with *cobA-I-G-J-F-M-K-L-H-B-W-N-S-T-Q-J-D-bluE-C-bluF-P-U-cbiW-V-E-btuR-R*.	Dr Evelyne Deery
Rosetta 2 pLysS pET3a-*hmat2a*-I117A	Rosetta 2 pLysS with *human methionine adenosyltransferase II-α* with the I117A mutation and a His_6-tag.	Dr Andrew Lawrence

2.2.2 Plasmids

Plasmids were either a gift from Dr Susanne Schroeder, Dr Andrew Lawrence, Dr Evelyne Deery or Yamini Mathur (Indian Institute of Science Education and Research (IISER) Pune) or produced via restriction enzyme cloning in this work. All pET14b vectors used in this work had been modified with an SpeI site inserted between the NdeI and BamHI sites.

Table 2.2 - List of plasmids used in this thesis.

Name	Description	Source
pET14b-*ElbzaCT*	*E. limosum bzaC* with a C-terminal truncation removing the DUF 2284 domain under control by a T7 promotor and a N-terminal His_6-tag.	Dr Susanne Schroeder and Dr Andrew Lawrence
pET14b-*MtbzaC*	Codon optimised for *E. coli Moorella thermoacetica bzaC* under control by a T7 promotor and a N-terminal His_6-tag.	This work
pET28a-*DtbzaC*	*Desulfofundulus thermosubterraneus bzaC* under control by a T7 promotor and a N-terminal His_6-tag.	Dr Yamini Mathur

pET14b-*RccobU*	*Rhodobacter capsulatus cobU* under control by a T7 promotor and a N-terminal His_6-tag.	Dr Evelyne Deery
pET28a-*Ecmtan*	*E. coli methylthioadenosine nucleosidase* under control by a T7 promotor and a N-terminal His_6-tag.	Dr Yamini Mathur
pET14b-*ElbzaE*	*E. limosum bzaE* under control by a T7 promotor and a N-terminal His_6-tag.	Dr Susanne Schroeder
pET28a-*DtbzaE2*	*Desulfofundulus thermosubterraneus bzaE* under control by a T7 promotor and a N-terminal His_6-tag.	Dr Yamini Mathur

2.3 Microbiology techniques

2.3.1 Sterilisation

All media and solutions were either sterilised via autoclaving at 121 °C, 1 bar positive pressure for 20 minutes or filter sterilised using a 0.2 µm or 0.45 µm filter.

2.3.2 Media for *E. coli* growth

Luria Bertani (LB) medium was used either for liquid growth or with 1.5 % (w/v) agar for solid growth. Additionally, 2TY was occasionally utilised for radical SAM enzyme production. For ED674 growth, a modified phosphate buffered M9-minimal media with various supplements was utilised (M9-YE).

Table 2.3 - List of medias used for bacterial cultures in this thesis.

Name	Component	w/v (%)
LB	Tryptone	1
	Yeast extract	0.5
	NaCl	0.5
	dH_2O	
2YT	Tryptone	1.6

	Yeast extract NaCl dH_2O	1 0.5

Name	Component	Amount per litre
M9-YE	Yeast extract 10 x M9 salts 50 % glycerol 0.1 M $CaCl_2$ 1 M $MgSO_4$	10 g 100 mL 10 mL 1 mL 2 mL

The 10 x M9 salts (6 % (w/v) Na_2HPO_4, 3 % KH_2PO_4, 1 % NH_4Cl, and 0.5 % NaCl in H_2O) 50 % glycerol, 0.1 M $CaCl_2$, and 1 M $MgSO_4$ were all made up and autoclaved separately and added to the 900 mL 10 g yeast extract solution.

2.3.3 Bacterial storage

700 µL starter culture was supplemented with 300 µL 50 % glycerol (15 % final concentration) and stored at -80 °C.

2.3.4 Starter cultures

2.3.4.1 DH10β and BL21

5 or 10 mL LB solutions in tubes with 5 x aeration were supplemented with respective antibiotics, inoculated with cells, and incubated at 28 – 37 °C, 150 – 180 rpm for 16-24 hr.

2.3.4.2 ED674 growth

5 or 10 mL LB solutions in tubes with 5 x aeration were supplemented with 0.2 % glucose, inoculated with a few colonies, and incubated at 28 - 30 °C, 150 – 180 rpm for 16-24 hr.

2.3.5 1 L cell cultures

2.3.5.1 BL21

1 L LB in 2.5 L baffled flasks were supplemented with respective antibiotics and inoculated with 5-10 mL starter culture and incubated at 28 – 37 °C, 150 – 180 rpm until OD_{600} ~ 0.6 – 1 was reached. Cultures were then induced with 400 μM IPTG and incubated at 19°C, 150 – 180 rpm for 16-24 hr.

2.3.5.2 ED674

1 L M9-YE in 2.5 L non-baffled flasks were inoculated with 5-10 mL starter culture and incubated at 28 – 30°C, 150 rpm until OD_{600} ~ 1 was reached. Cultures were then supplemented with 50 mg $CoCl_2$, relevant concentration of benzimidazole, induced with 400 µM IPTG and incubated at 28 - 30°C, 150 rpm for 16-24 hr.

2.3.6 Antibiotics

Table 2.4 - List of antibiotics used in this thesis.

Antibiotic	solvent	Working concentration (µg / mL)
Ampicillin	H_2O	100
Kanamycin	H_2O	30
Chloramphenicol	Ethanol (100 %)	34

2.3.7 1 L Cell harvest

Cultures were centrifuged at 4000 rpm, 20 min, 4-8°C, supernatant discarded, and pellets resuspended in buffer (30 mL per 1 L culture). ED674 pellets were resuspended in 20 mM Tris-HCL pH 8, 100 mM NaCl and BL21 pellets were resuspended in binding buffer.

2.4 Molecular biology techniques

2.4.1 *E. coli* competent cell production

50 mL LB in a 250 mL baffled flask, supplemented with respective antibiotics, was inoculated 1:100 (starter culture : media) with DH10β or BL21 Star (DE3)pLysS-btuB starter culture and incubated at 37 °C, 180 rpm until OD_{600} 0.5 was reached. Cells were then incubated on ice for 10 minutes and subsequently centrifuged at 3000 rpm, 10 minutes, 4 °C, discarding supernatant. Pellets were then resuspended in 17 mL of ice cold, filter sterilised 0.1 M $CaCl_2$, 15 % glycerol and incubated on ice for 20 minutes. Cells were then centrifuged at 3000 rpm, 10 minutes, 4 °C, the supernatant was discarded, and the pellets were resuspended in 2 mL ice cold, filter sterilised 0.1 M $CaCl_2$, 15 % glycerol. 50 – 100 µL aliquots were then produced in sterile Eppendorf tubes and stored at -80 °C.

2.4.2 *E. coli* transformation

20 – 50 µL DH10β or BL21 Star (DE3)pLysS-btuB cells were mixed 10:1 or 25:1 with purified plasmid (cells : plasmid) and incubated on ice for 20 minutes. Cells were then heat shocked by incubated at 42 °C for 40 – 60 seconds, incubated on ice for 2 minutes, resuspended in 4-5 x LB and incubated at 37 °C, static for 1 hour. Cells were then plated on LB agar supplemented with respective antibiotics and incubated at 37°C overnight.

2.4.3 Plasmid amplification and purification

Plasmid DNA was amplified via production of 5 mL starter cultures of *E. coli* transformed with the plasmid of interest. Cells were then harvested via centrifugation at 4000 rpm, 10 minutes at 4 °C and supernatant discarded. Plasmid DNA was extracted and purified from the cell pellets using the Qiagen mini prep kit, following the manufacturers protocol.

2.4.4 DNA restriction digest

Plasmid and DNA fragments were digested at 37 °C for 1-2 hr using 1 unit of NdeI and SpeI GQ per µL of DNA in 1 x buffer D.

2.4.5 Agarose gel electrophoresis

DNA fragments supplemented with 1 x loading dye were separated on a 1 % agarose gel in TAE buffer (40 mM Tris-HCl pH 8.0, 10 mM acetic acid, 1 mM EDTA) containing ethidium bromide. Gels were run at 75 V for 45 minutes. Gels were then visualised and imaged using UV or blue light. DNA size was estimated using the Promega 1 Kb DNA ladder.

2.4.6 DNA purification

2.4.6.1 Gel extraction

DNA was purified from agarose gels using the QIAquick Gel Extraction Kit following the manufacturers protocol.

2.4.6.2 Clean up

To purify DNA fragments from restriction digests the gel extraction protocol above was followed with the following modifications. Firstly, DNA was diluted 1:5 (DNA: PB) with buffer PB, applied to a QIAquick spin column and centrifuged for 1 minute. The supernatant was discarded and the gel extraction protocol was followed starting from step 10.

2.4.7 DNA ligations

Plasmid and insert DNA were ligated at room temperature for 2 hours or overnight at 4 °C using 0.3 units of T4 DNA ligase per µL of DNA in 1 X ligase buffer.

2.4.8 DNA sequencing

Sequencing of plasmid was carried out using GENEWIZ sequencing services following their sample preparation guidelines using their T7 (5'-3' TAATACGACTCACTATAGGG)

and T7 term (5'-3' GCTAGTTATTGCTCAGCGG) primers. Resulting sequencing files were then analysed in SnapGene Viewer (56).

2.4.9 Plasmid map generation

Plasmid maps were produced in Benchling using their in-silico restriction enzyme cloning process (57).

2.5 Protein purification

2.5.1 Cell lysis

Resuspended cell pellets of BL21 DE3 pLysS-btuB containing the protein of interest were sonicated at 60 % amplitude, 30 seconds on 30 seconds off, for 4-5 minutes per 30 mL cells. Samples were then centrifuged at 18, 000 rpm, 4-8 °C for 20 minutes.

2.5.2 Immobilized Metal Affinity Chromatography (IMAC)

A variety of different sized columns were utilised to purify His_6-tagged protein from cell lysates. Generally, Sepharose resin was applied to an empty column, washed with water, and charged with 50 mM $NiSO_4$. The columns were then washed with binding buffer (20 mM Tris-HCL pH 8, 500 mM NaCl, 5 mM imidazole) and supernatant was applied to the column. The columns were then sequentially washed with binding buffer

and unique buffer (60 mM Imidazole) and finally eluted using elution buffer (500 mM Imidazole), collecting 1-2 mL fractions.

2.5.3 PD10 chromatography

PD10 columns were washed with water and equilibrated with 35 mL protein buffer (20 mM Tris-HCL pH 8, 100 mM NaCl). 2.5 mL protein sample was then applied to the column and then eluted with 3.5 mL protein buffer.

2.5.4 Size exclusion chromatography

Chromatography was carried out on a AKTA Pure FPLC using a Superdex™ 200 increase 10/300 GL column and an isocratic gradient of 20 mM Tris-HCL pH 8, 100 mM NaCl at 0.5 mL / min for 1.5 column volumes. 500 µL fractions were collected in a 48 well plate.

A standard curve to allow for protein molecular weight estimation was produced by making and running three standard solutions. Firstly, a solution of blue dextran was used to analyse the void volume (V_o). Next two protein mixtures 1: Carbonic Anhydrase from bovine erythrocytes, Alcohol Dehydrogenase from yeast and Apoferritin from horse spleen (29, 150 and 443 kDa) and 2: Albumin, Bovine Serum, β-Amylase from sweet potato and Thyroglobulin, bovine (66, 200, 669 kDa) were produced and run separately. The traces produced were used to analyse the elution volume (V_e) and the V_e/V_o was plotted against protein molecular weight to produce a standard curve. The

following equation for an exponential curve was then used to calculate the molecular weight of the protein of interest:

$$y = ae^{bx}$$

2.5.5 Protein storage

Purified protein was supplemented with a final concentration of 10 % glycerol and stored at either -20 or - 80°C.

2.5.6 Sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Protein samples were mixed with laemmli sample buffer and boiled for 5-10 minutes at 95-100°C. Samples and ladder (PageRuler™ Unstained Protein Ladder) were then loaded onto a 12 % SDS-PAGE gel and ran at 200 V through the stacking gel and 150 V through the resolving gel. The gels were then stained with Coomassie blue, de-stained with water or de-staining solution and then imaged.

2.5.7 Anaerobic protein purification

Proteins containing iron-sulphur clusters were purified and assayed in a Belle technologies Glovebox with O_2 levels below 20 ppm using buffers that had been purged of oxygen using argon before placing inside the box.

2.5.8 Protein quantification

2.5.8.1 A_{280} estimation

Protein samples were diluted in buffer to produce an absorbance signal of less than 1 at 280 nm in a UV-VIS spectrometer. Absorbance values were then dilution adjusted at concentration was estimated using the following equation:

$$A_{280} = \text{Concentration (M)} \times \text{Extinction coefficient } (M^{-1} cm^{-1}) \times \text{Path Length (cm)}$$

Values were converted to mg / mL when required by using the following equation:

$$\text{mg/ml protein concentration} = \text{Molecular mass (Da)} \times \text{Concentration (M)}$$

2.5.8.2 Bradford assay

For proteins with a low abundance of aromatic residues a Bradford assay was utilised for more accurate concentration estimations. Firstly, a serial dilution of Bovine serum albumin (BSA) from 1 – 0.125 mg / mL was produced, mixed with 1 x Bradford dye and incubated at room temperature for 5 minutes. Absorbance values for each standard was then recorded at 595 nm in a UV-VIS spectrometer to produce a standard curve. The protein of interest was then mixed with 1 x Bradford dye and analysed as above. Protein concentration was then estimated using the standard curve using the equation:

$$y = mx + c.$$

2.6 Protein biochemistry

2.6.1 End-point BzaC assays

Typically, BzaC end-point assays contained 10 µM BzaC (*Dt*, *Mt* and *El*BzaCT), 1 mM SAM, 20 µM [5(6)-OHBza]Cba (Ado, Hydroxo or Cyano) in 20 mM Tris-HCL pH 8, 100 mM NaCl. Furthermore, controls lacking specific components of interest were set up simultaneously. Reactions were incubated at 30 - 37 °C for 16-24 hours before being quenched with 1 % acetic acid glacial, centrifuged at 13,000 rpm for 5 minutes to remove precipitated protein and analysed via HPLC. These reaction setups were modified, as described in the respective texts, for certain experiments.

2.6.2 Fluorimetry

Firstly, *Dt*BzaC's intrinsic tryptophan fluorescence was tested by exciting 10 µM *Dt*BzaC in 20 mM Tris-HCl pH 8, 100 mM with 280 nm light and recording emission signal from 300-400 nm in a fluorimeter.

Next, a 10-point SAM (in DMSO) titration (0.1-2.46 mM) and a 10-point ATP (in mQ) (0.2-4.74 mM) were carried on 10 µM *Dt*BzaC using the excitation and emission signals above, mixing and waiting 1 minute to allow for equilibrium to be reached before each reading. Furthermore, titrations of the same volume additions were carried out using DMSO and mQ as controls. Emission values at 347 nm for the SAM and ATP titrations were dilution and control adjusted. ΔF values were then calculated by subtracting each titration point from the apo protein signal and plotted against ligand concentration.

Nonlinear curve fitting using the total binding model ($y = (B_{max} * x) / (K_d + x) + NS * x$) was then carried out in Jupyter notebook using the python package SciPy. To calculate the SAM dependent signal change, the nonlinear regression curve of ATP was subtracted from SAM and plotted. The data was then fitted to the total binding model to obtain K_d and B_{max} values.

2.6.3 Differential scanning fluorimetry (DSF)

2.6.3.1 5'-deoxy-5'-thionucleosides, BzaC binding assay

50 µL reactions in triplicate containing 4 µM BzaC, 10 X SYPRO™ orange and a range of ligand concentrations in 50 mM Tris-HCl pH 7.5, 100 mM NaCl were set up on ice and then incubated at room temperature for 10 minutes. Reactions were then incubated in a QuantStudio 3 Real-time PCR system using a cycle method of; (1) Ramp (2.5 °C/s) to 40 °C and hold for 2 minutes (2) Ramp (0.05 °C/s) to 95 °C, recording signal at 558 ± 11 using 520 ± 10 excitation energy, then holding for 2 minutes.

2.6.3.2 Cobamide, BzaC binding assay:

50 µL reactions in triplicate containing 4 µM *Dt*BzaC, 80 µM Nile red and a range of ligand concentrations in 50 mM Tris-HCL pH 8, 100 mM NaCl were set up on ice and then incubated at room temperature for 10 minutes. Reactions were then incubated in a QuantStudio 3 Real-time PCR system using a cycle method of; (1) Ramp (2.5 °C/s) to

40 °C and hold for 2 minutes (2) Ramp (0.05 °C/s) to 95 °C, recording signal at 623 ± 14 using 580 ± 10 excitation energy, then holding for 2 minutes.

2.6.3.3 $K_{d\ apparent}$ quantification:

Firstly, ΔT_ms for BzaC with various ligand concentrations were determined via Boltzmann curve fitting using the Thermal Shift Assay–Curve Rapid and Automatic Fitting Tool (TSA-CRAFT) webserver (58). Then $K_{d\ apparent}$ values were calculated by plotting ΔT_m against [ligand] and using nonlinear regression curve fitting and the equation (One site -- Total binding, $Y=B_{max}*X/(K_d+X) + NS*X + Background$) on GraphPad prism 5.

2.6.3.4 K_d determination via isothermal analysis:

BzaC 5'-deoxy-5'-thionucleosides affinities (K_d) were determined using the FoldAffinity webserver (59, 60). Firstly, fluorescence curves were fit using the local model and ΔCP (kcal/K/mol) set to zero. Unfolded fraction fitting was then carried out specifying 4 μM BzaC and using the temperature range just above apo protein T_m. This provided the K_d values and asymmetric 95 % confidence interval (CI) values for the curve fit at each temperature. The most accurate K_d value was then decided by selecting the value with the lowest CI variance.

2.6.4 Microscale thermophoresis (MST)

A 4 µM stock of *Dt*BzaC was produced by diluting *Dt*BzaC in 1 x PBS-T and a 50 nM solution of RED-tris-NTA dye in 1 x PBS-T was made. Next, a 16-point serial dilution of *Dt*BzaC in 1 x PBS-T was performed and each sample was subsequently supplemented with a final concentration of 25 nM dye. The reactions were then incubated at RT for 30 minutes before being loaded into the capillaries for analysis at 40 % LED/excitation power and medium MST power in a NanoTemper Monolith machine. An EDTA test was carried out repeating the sample set up above with the addition of 16.6 mM EDTA pH 7.4 to each condition.

2.6.5 *Dt*BzaC kinetics

2.6.5.1 Continuous assay

50 µL reactions (in duplicate) containing 5 µM *Dt*BzaC, 25 µM Ado[5(6)-OHBza]Cba, 50 µM *Ec*MTAN, 0.092 units XOD, 0.022 units HRP, 0.1 mM Amplifu Red, 5 mM $MgCl_2$, and a range of SAM or SAH concentrations in 20 mM Tris-HCL pH 8, 100 mM NaCl were produced. Samples were incubated at 28 °C, for 45 minutes before initiation of reactions via injection of *Dt*BzaC. Signal was then recorded at 544 nm excitation, 620-10 nm emission every 1.5 minutes for 60 minutes.

2.6.5.2 Stopped assay

2.6.5.2.1 Ado[5(6)-OHBza]Cba analysis

100 µL reactions in duplicate containing 1 µM *Dt*BzaC, 1 mM SAM, 50 µM *Ec*MTAN and a 5-point serial dilution of Ado[5(6)-OHBza]Cba from 400 – 12.5 µM (plus a 0 µM control) in 20 mM Tris-HCL pH 8, 100 mM NaCL were produced. Samples were pre-incubated at 37 °C for 10 – 15 minutes before being initiated by addition of *Dt*BzaC. Reactions were then incubated at 37 °C for 0, 7.5 and 15 minutes before being quenched with 2 % acetic acid glacial. Samples were then exposed to light for 30 minutes, centrifuged at 13, 000 rpm, 5 minutes and supernatants were analysed via reversed-phase HPLC (method 4).

2.6.5.2.2 SAM analysis

100 µL reactions in duplicate containing 1 µM *Dt*BzaC, 50 µM *Ec*MTAN, 75 µM Ado[5(6)-OHBza]Cba and a 7-point serial dilution of SAM from 2400 – 37.5 µM (plus a 0 µM control) in 20 mM Tris-HCL pH 8, 100 mM NaCL were produced. Samples were pre-incubated at 37 °C for 10 – 15 minutes before being initiated by addition of *Dt*BzaC. Reactions were then incubated at 37 °C for 0, 7.5 and 15 minutes before being quenched with 2 % acetic acid glacial. Samples were then exposed to light for 30 minutes centrifuged at 13, 000 rpm, 5 minutes and supernatants were analysed via reversed-phase HPLC (method 5).

2.6.6 Protein x-ray crystallography

2.6.6.1 Crystalisation, diffraction and model building

*Dt*BzaC dimer, SAH containing crystals were obtained via mixing 13.6 mg / mL *Dt*BzaC dimer with 5 x SAH and setting up crystallisation trails using both the JCSG and MORPHEUS screens. Crystal screens were put down using the Oryx crystallisation robot producing 400 nL sitting drops with 3 different ratios of protein to screen buffer (1:1 protein : screen, 2:1 protein screen, 1:2 protein : screen). Crystals were then picked and mixed with cryoprotectant and stored. Samples were then analysed at the European Synchrotron Radiation Facility (ESRF). For 1 crystal that diffracted to 2.17 Å an initial model was produced via molecular replacement with an AlphaFold 2 model.

2.6.6.2 Structure refinement

The *Dt*BzaC structure was refined using COOT (61) and the Collaborative Computational Project Number 4 (CCP4) (62) software to perform multiple rounds of manual adjustment and Refmac (63) cycles.

2.6.6.3 Protein 3D structure visualisation and analysis

Protein structures were visualised, analysed and figures were produced using PyMOL (64). Additional analysis was also carried out using PDBsum (65).

2.6.7 Molecular docking

Firstly, the 3D structure of Ado[5-OHBza]Cba was produced by modifying the DMB of adenosylcobalamin taken from the crystal structure of *Thermotoga maritima* Ribonucleotide Reductase, NrdJ, in complex with dTTP and adenosylcobalamin (3O0N) (66) to 5-OHBza using PyMOL. This structure was then prepared for docking via production of a pdbqt file using AutoDock tools. Furthermore, the refined *Dt*BzaC dimer 3D structure was prepared for docking in AutoDockTools (67) by deleting the waters and adding hydrogens and charges and saving in pdbqt format. In addition, the grid box (area defined for docking) was produced around chain A's open active site. Docking was then carried out using the AutoDock Vina (68) command line tool with exhaustiveness set to 8.

2.7 Analytical chemistry

2.7.1 Cobamide alkaline hydrolysis and purification of alpha-ribazoles

Cobamide hydrolysis was carried out following L. Malalasekara et al. 2022 (69). ~20-180 µg Cobamide was mixed 1:1 (volume: volume) with 5 M NaOH and incubated at 90 °C for 75 minutes, shaking. The hydrolysates were then incubated at 4 °C for 30 minutes and neutralised via addition of 1:2 5 M HCL: hydrolysate. The alpha-ribazole products were then de-phosphorylated by diluting hydrolysate 1:1.5 with 0.3 M Ammonium Acetate pH 8.8, adjusting pH to 8.5- 9, adding 5 units Alkaline phosphatase intestinal and 1 mM $MgCl_2$ and incubating at 35-37 °C for 16-24 hours. Alpha-ribazoles were then purified from the hydrolysates via Boronic acid chromatography. 500 µL

boronic acid resin was washed with 5 column volumes (CV) of mQ and equilibrated with 3 CV 0.3 M Ammonium Acetate pH 8.8. The hydrolysates were mixed 1:2 with 0.3 M Ammonium acetate pH 8.8 and flown through the column. The column was then washed with 3 CV 0.3 M Ammonium acetate pH 8.8 and eluted with 4 CV 0.1 M Formic Acid. Elution's were then concentrated, if required, under oxygen flow or by vacuum centrifugation. Samples were then analysed by reversed-phase HPLC and/or LC-MS.

2.7.2 High pressure liquid chromatography (HPLC)

A variety of different C18 columns, solvents and gradients were utilised for the separation of relevant molecules.

2.7.2.1 Method 1

Separation of cobamides produced in BzaC end point activity assays were carried out on an ACE 5AQ 150 x 2.1 mm C18 column using 0.1 % Trifluoroacetic acid (TFA) (A) and acetonitrile (B) mobile phase at 0.2 mL / min, 30 °C. The following gradient was utilised: 0-1 min 10 % B, 1-21 min 50 % B, 21-23 min 100 % B, 23-25 min 100 % B, 25-27 min 10 % B with a 10 min post time. Signals were recorded at 360 nm.

2.7.2.2 Method 2

Separation of 5'-deoxy-5'-thionucleosides produced in BzaC end point activity assays were carried out on an Agilent Zorbax SB-C18 5 µm 4.6 x 150 mm column using 0.1 %

TFA (A) and acetonitrile (B) mobile phase at 1 mL / min, 30 °C. The following gradient was utilised: 0-1 min 0% B, 1-15 min 10 % B, 15-20 min 30% B, 20-25 min 50 % B, 25-30 min 100% B, 30-35 min 100 % B, 35-40 min 0% B. Signals were recorded at 260 nm.

2.7.2.3 Method 3

Chromatography was carried out using an ACE 5AQ 125 x 4.6 mm C18 column, 0.1 % TFA and 100 % Acetonitrile at 30 °C, 1 mL / min. The following gradient was utilised; 0-1 min 5 % B, 1-11 min 90 % B, 11-12 min 5 % and a post time of 5 minutes for re-equilibration.

2.7.2.4 Method 4

Chromatography was carried out using an ACE 5AQ 150 x 2.1 mm C18 column, 0.1 % TFA and 100 % Acetonitrile at 30 °C, 0.2 mL / min. The following gradient was utilised; 0-5 min 0% B, 5-45 min 80% B, 45-46 min 100% B, 46-51 min 100% B, 51-53 min 0% B and a post time of 10 minutes. Signals were collected at 360 nm.

2.7.2.5 Method 5

Chromatography was carried out using Agilent Zorbax SB-C18 5 µm 4.6 x 150 mm column using 0.1 % TFA (A) and methanol (B) at 1 mL / min, 30 °C. The following gradient was utilised: 0-1 min 20 % B, 1-10 min 60 % B, 11-12 min 20 %, with a 5-minute post time. Signals were recorded at 360 nm.

2.7.2.6 Method 6

This method was adapted from Mathur et al. 2020 (55) and was carried out using an XTerra® RP_{18} 3.5 µm 4.6 x 150 mm column and a solvent system consisting of 10 mM Ammonium Acetate pH 6.5 (A) and 100 % methanol (B) at 0.5 mL / min at 25°C. The following gradient was utilised: 0-1 min 5% B, 1-25 min 70% B, 25-28 min 100 % B, 28-30 min 100 % B, 30-32 min 0 % B, with a post time of 5 minutes. Signals were recorded at 260 nm.

2.7.2.7 Method 7

Chromatography was carried out using an Agilent Zorbax SB-C18 5 µm 4.6 x 150 mm column using 0.1 % TFA (A) and methanol (B), 1 mL / min, 30 °C. The following gradient was utilised: 0-1 min 5 % B, 1-11 min 90% B, 11-12 min 5 % B with a 5-minute post time. Signals were recorded at 360 nm.

2.7.2.8 Method 8

This method was adapted from Mathur et al. 2020 (55) and was carried out using an Agilent Zorbax SB-C18 5 µm 4.6 x 150 mm column, 10 mM Ammonium Acetate pH 6.5 (A), 100 % Methanol (B) at 0.5 mL / min, 30 °C. The following gradient was utilised: 0-1 min 5 % B, 1-25 min 70 % B, 25-28 min 100 % B, 28-30 min 100 % B, 30-32 0 % B with a post time of 5 minutes. Signals were recorded at 260 nm.

2.7.2.9 Method 9

Chromatography was carried out on an ACE 5AQ 150 x 2.1 mm C18 column, 0.1 % TFA (A), 100 % acetonitrile (B) at 0.2 mL / min 30°C. The following gradient was utilised: 0-5 min 0 % B, 5-45 min 80 % B, 45-50 min 100 % B, 50-55 min 100 % B, 55-60 min 0 % B, 60-65 min 0 % B with a 10-minute post-time. Signals were recorded at 360 nm.

2.7.2.10 Method 10

Chromatography was carried out on an ACE 5AQ 150 x 2.1 mm C18 column, 0.1 % TFA (A), 100 % acetonitrile (B) at 0.2 mL / min 30°C. The following gradient was utilised: 0-3 min 5 % B, 3-40 min 100 % B, 40-50 min 100 % B, 50-55 min 5 % B with a post time of 10 minutes. Signals were recorded at 360.

2.7.2.11 Method 11

Chromatography was carried out on an ACE 5AQ 150 x 2.1 mm C18 column, 0.1 % TFA (A), 100 % acetonitrile (B) at 0.2 mL / min 30°C. The following gradient was utilised: 0-3 min 5 % B, 3-20 min 90 % B, 20-22 90% B, 22-25 min 5 % B with a post time of 5 minutes. Signals were recorded at 360 nm.

2.7.3 Mass spectrometry

A Bruker micrOTOF-Q II mass spectrometer linked to a HPLC was used for all LC-MS experiments, using electrospray ionisation (positive) and auto MS/MS for fragmentation.

2.7.4 Reverse phase chromatography

Manual RP18 columns were used to purify a variety of cobamides including [5-OHBza]Cba and [5-OMeBza]Cba (with a variety of upper ligands).

Generally, RP18 resin is applied to an empty column and washed with 100 % Methanol, 50 % Methanol, and then 0.1 % TFA. Sample then applied to column, washed with 20 mM Tris – HCL pH 8, 8M Urea and water. Sample then eluted using 50 % methanol (collecting desired coloured fractions). Pressure applied to top of column during chromatography.

2.7.5 Nuclear magnetic resonance (NMR)

Ado[5(6)-OHBza]Cba was re-suspended in 10 mM Kpi in D_2O pD 7.4 to a concentration of ~ 1.3 mM. 1D ^{1}H (with W5 water suppression), ^{1}H-^{13}C HSQC, ^{1}H-^{13}C HMBC, 300ms ROESY and 80ms TOCSY analysis were carried out on the samples using a Bruker 700 MHz prodigy NMR spectrometer. Data analysis was carried out using The Collaborative Computational Project for NMR (CCPN) (70) TopSpin and ACD/Labs software packages.

2.8 Bioinformatics

2.8.1 Sequence alignments

Sequence alignments of genes or proteins were carried out using Clustal Omega (71) and visualised in Jalview (72).

2.8.2 3D structure alignments

PDBeFold (73) was utilised for the alignment of the determined *Dt*BzaC structure with other structures in the PDB to identify structures with homology.

2.8.3 Chemical structures

Chemical structures were drawn in ChemDraw.

Chapter 3: Biochemical characterisation of the SAM dependent methyltransferase BzaC

3.1 Introduction

3.1.1 SAM dependent methyltransferases

As highlighted in the biosynthesis of the different tetrapyrroles, methylations are important chemical transformations for diversifying and guiding natural product biosynthesis. They also play important roles in biological systems such as epigenetic and protein regulation, cell signalling, in substrate recognition, and are important modifications required for chemical scaffold modifications for the production of pharmaceuticals with diverse functions (74, 75). In nature these methylations usually occur via SAM dependent methyltransferases which catalyse S_N2 and S_N2-like methyltransfer of the electrophilic SAM methyl group to oxygen (O) (74, 75) (**Figure 3.1**), nitrogen (N), carbon (C) and sulphur (S) on substrates of interest (74).

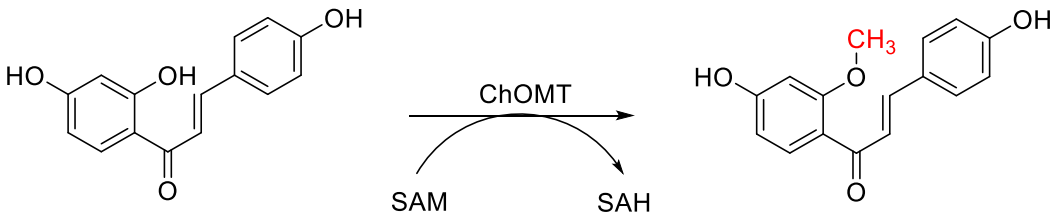


Figure 3.1 - Chalcone O-methyltransferase reaction scheme.

3.1.2 Previous characterisation of BzaC

Bioinformatic sequence investigation of BzaCs from different organisms by Mathur et al. 2020 (55) identified two main domain architectures, a dimerisation domain and a

methyltransferase domain. This methyltransferase domain makes BzaC a member of the class I methyltransferases, all of which use SAM as a methyl donor, and contains the classical GxGxG SAM binding motif. Furthermore, *El*BzaC contains an extra domain of unknown function (DUF) 2284 at its C-terminus, containing two cysteine rich motifs (CX_3CX_7C and $CX_2CX_2CX_5C$) consistent with iron-sulphur cluster formation. This DUF domain is present and fused to BzaC in some bacteria (e.g. *El*BzaC), present but separately translated to BzaC in others (e.g. *Desulfofundulus thermosubterraneus*) or absent (e.g. *Moorella thermoacetica*).

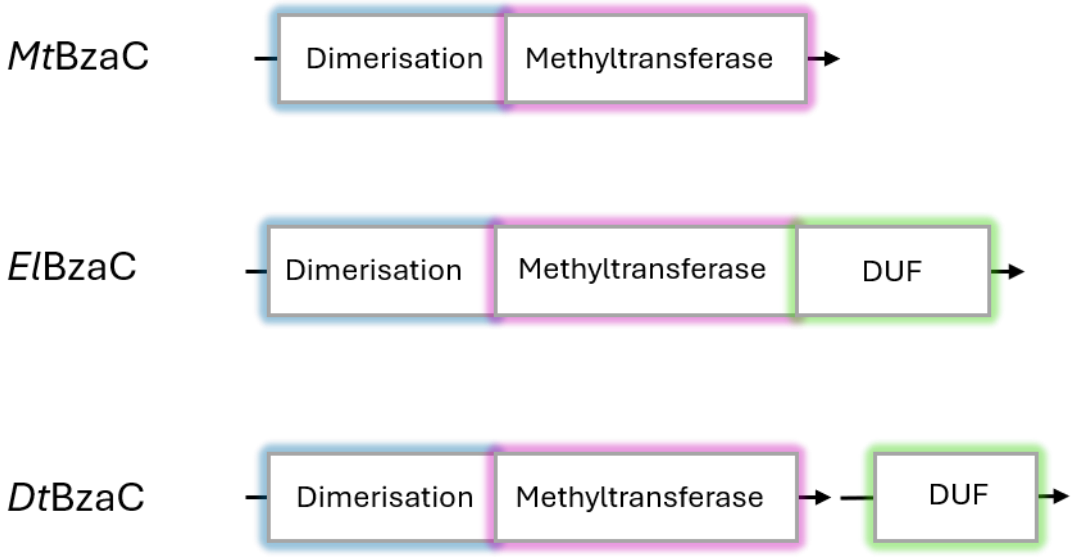


Figure 3.2 - BzaC domain architecture.

Adapted from Mathur at al. 2020.

Biochemical characterisation then established that *Moorella thermoacetica* BzaC (*Mt*BzaC) was able to catalyse SAM dependent methyl-transfer to 5-OHBza to form 5-OMeBza in vitro (**Figure 3.3**), supporting the findings of Hazara et al. 2015 (50), however only producing ~ 10 % product in an optimised system.

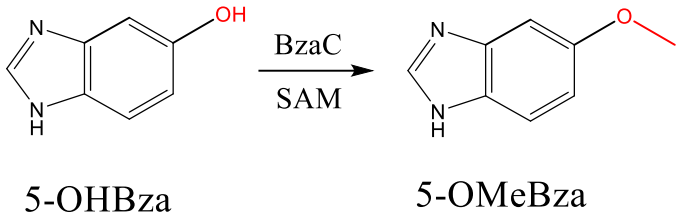


Figure 3.3 - BzaC catalysed SAM dependent methylation of 5-OHBza to produce 5-OMeBza.

Additionally, the study identified that *Moorella thermoacetica* CobT (*Mt*CobT) catalyses the regiospecific activation of 5-OHBza by transfer of the ribose phosphate group of nicotinamide mononucleotide (NMN) to N1 of 5-OHBza, forming 5-OHBza-ribotide (5-OHBza-RP), which is 'activated' and ready for attachment to cobinamide. Furthermore, this enzyme demonstrated some promiscuity by activating 5-OMeBza also, losing some regioselectivity producing both 5- and 6- isomers due to the asymmetric benzimidazole nitrogens being able to participate in the reaction (**Figure 3.4**).

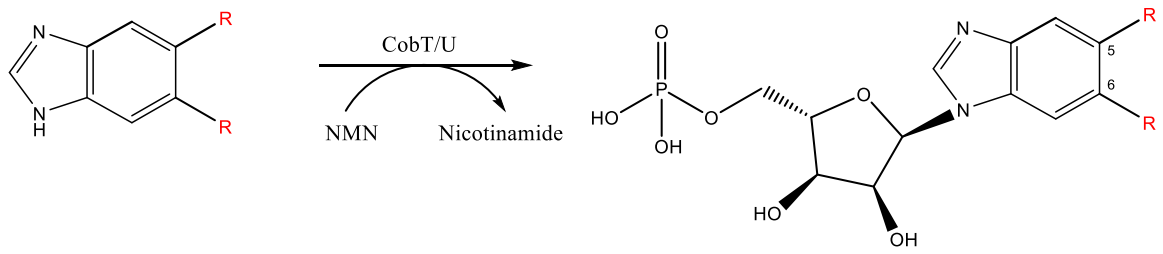


Figure 3.4 - CobT/U reactions.

Schematic representation of benzimidazole activation by CobT. Variable groups (R) are identified in red.

The small amounts of product formed and the placement of BzaC after CobT in the operon suggested the ribotide intermediate may be the preferred substrate of BzaC. Reconstitution of activity with 5- and 6-OHBza-RP surprisingly produced no 5 or 6-OMeBza-RP, however some 5- and 6-OMeBza-R (5- and 6-OHBza-riboside) was identified, thus suggesting chemical substrate dephosphorylation may have allowed for activity. The experiment was therefore repeated with 5- and 6-OHBza-R which showed the production of 5- and 6-OMeBza-R, with more 5- product produced than 6, demonstrating the ability of BzaC to catalyse regioselective methyltransfer to ribosides. Interestingly the specific activity with 5-OHBza-R was 5 times higher than that of 5-OHBza. Thus, suggesting that the riboside intermediate is more likely the physiological substrate of BzaC.

This data then led to the revision of the current hypothesised pathway (Hazara et al. 2015 (50)) where 5-OHBza is first activated by *Mt*CobT, de-phosphorylated by *Mt*CobC and then methylated by *Mt*BzaC to produce 5-OMeBza-R (**Figure 3.5**).

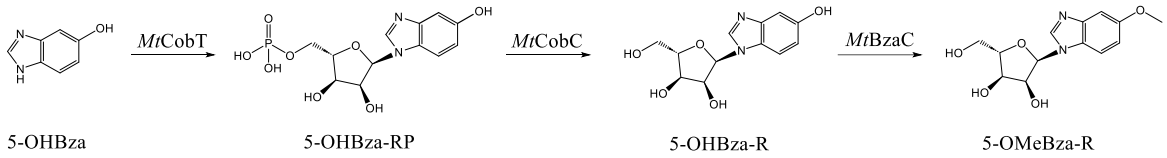


Figure 3.5 - Hypothesised biosynthesis of 5-OMeBza-R from 5-OHBza via *Mt*CobT, C and *Mt*BzaC.

Due to the absence of a B_{12} binding domain in BzaC, the hypothesis that the 5-OHBza could be attached at the lower ligand position of a cobamide (forming 5-hydroxybenzimiazolylcobamide) and then acted upon by BzaC was dismissed. However, due to the low activity of BzaC with 5-OHBza and its increased activity with the riboside intermediate (but still low), it suggests the possibility that the preferred substrate may be 5-hydroxybenzimiazolylcobamide, with BzaC binding the cobamide in a novel manner.

3.1.3 Chapter aims

The aim of the work carried out in this chapter was to explore the ability of BzaC enzymes to catalyse SAM dependent methyltransferase to a cobamide substrate, probe the biochemistry of the reaction and its ability to be manipulated to produce novel cobamide analogues.

3.2 Results

3.2.1 Cloning, production and purification of BzaC enzymes

Three BzaC enzymes from different organisms containing *bza* operons of varying composition and length were selected for analysis. Namely, *Moorella thermoacetica* BzaC (*Mt*BzaC), *Eubacterium limosum* Truncated BzaC (DUF domain removed), (*El*BzaCT) and *Desulfofundulus thermosubterraneus* BzaC (*Dt*BzaC) (Figure 3.6). These BzaCs were selected due to differences in the presence of a DUF domain and BzaD and E enzymes in their respective pathways, as we wanted to assess potential impacts on BzaC activity, to infer more about the biosynthetic pathways. The DUF domain of *El*BzaC was removed to allow for assays under aerobic conditions, due to the oxygen sensitivity of the potential iron-sulphur cluster.

Each pathway contains BzaA, B, C, however the later enzymes are different. The *El*BzaC pathway contains the fused DUF domain, D, and E and produces cobalamin ([DMB]Cba), the *Mt*BzaC pathway has no DUF protein and no later enzymes and is known to make 5-methoxybenzimidazolylcobamide (55). The *Dt*BzaC pathway contains a separately translated DUF2284 domain containing protein (BzaX) and an enzyme named BzaE2 (which has homology to both BzaD and E, but higher homology to E), and the product of the organism is unknown.

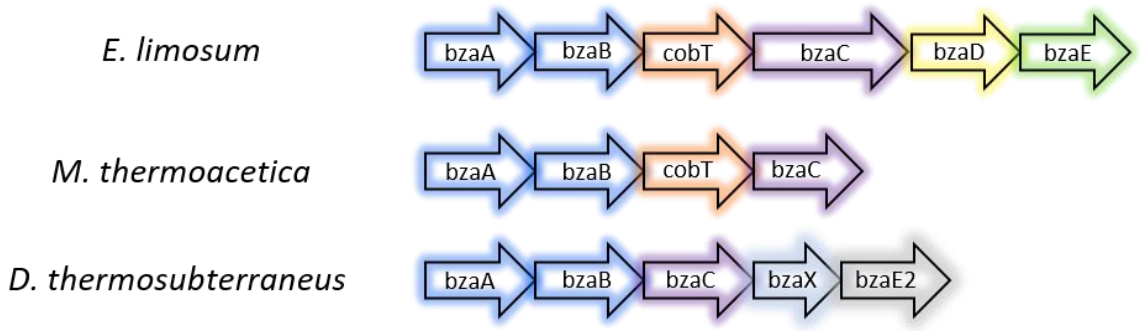


Figure 3.6 - *bza* operon structure of each BzaC chosen for study.
All gene sizes are generic apart from *bzaC* and X which have been size adjusted to demonstrate the increased size of the *El*bza*C* methyltransferase DUF fusion.

pET14b-*ElbzaCT* (**Supplementary figure 1**) was provided to me as a gift from Dr Andrew Lawrence and pET28a-*DtbzaC* (**Supplementary figure 2**) was provided to me as a gift from Dr Yamini Mathur. An *E. coli* codon optimised *MtbzaC* gene fragment was synthesised by Integrated DNA technologies (IDT) and subsequently cloned into a modified pET14b vector between SpeI and NdeI sites using restriction enzyme cloning.

To do this a plasmid called pET14b-*ElbzaE* (which is composed of the modified pET14b vector and the gene encoding *El*BzaE) (**Supplementary figure 3**) and the synthesised DNA were digested using NdeI and SpeI. The pET14b vector was then purified via separation on an 1 % agarose gel and subsequent isolation from the gel via cutting of the band and gel extraction. The *Mt*BzaC encoding gene fragment was purified from the restriction enzyme mixture following the PCR clean up protocol. Now both purified digested vector and insert had been isolated they were ligated together using T_4 ligase (producing pET14b-*MtbzaC*) (**Supplementary figure 4**) and transformed into DH10β cells. DH10β-pET14b-*MtbzaC* starter cultures were then produced to allow for the

amplification and subsequent purification of pET14b-*MtbzaC*. Isolated constructs were then subjected to restriction digestion using NdeI and SpeI and analysed on a 1 % agarose gel to assess successful insertion (**Figure 3.7**). Constructs were then sent for sequencing which identified the correct insertion of the gene of interest with no mutations present. One construct was then selected for future experiments.

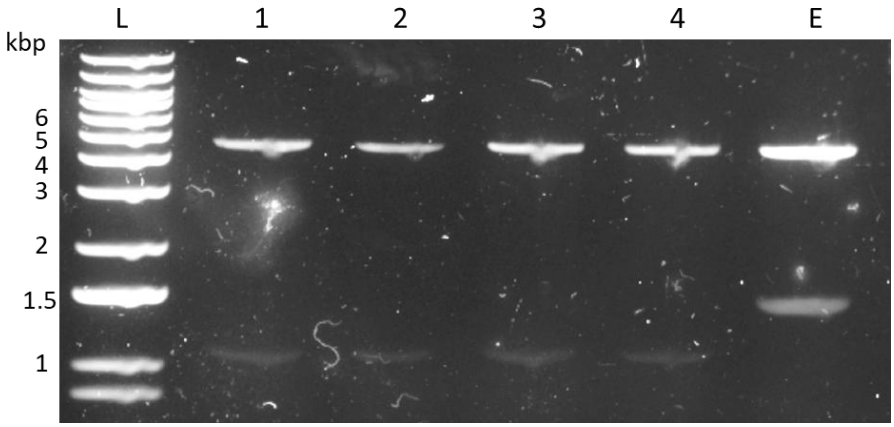


Figure 3.7 - pET14b-*MtbzaC* production.

1 % agarose gel of restriction enzyme digests of pET14b-*MtbzaC* constructs (1,2,3,4) and pET14b-*ElbzaE* (original plasmid used to obtain vector) (expected insert sizes *MtbzaC* = 1.1 kbp, pET14b, *ElbzaE* = 1.5 kbp and vector (pET14b) = 4.6 kbp).

Each enzyme was recombinantly produced in *BL21* (DE3) *pLysS-btuB* and purified via IMAC and size exclusion chromatography to isolate dimer species of each enzyme. As seen in **Figure 3.8** bands on a 12 % SDS-PAGE gel of expected sizes were observed for each enzyme, indicating correct production and purification. Furthermore, both *El*BzaCT and *Mt*BzaC showed high levels of purity with little additional bands present, whereas *Dt*BzaC purity was lower. Due to the bands primarily being lower molecular weight (MW) it suggests that this could be due to C-terminal proteolysis/degradation. In addition, gel filtration identified that each enzyme predominantly forms a homodimer in solution, which is expected due to the presence of a dimerisation domain in each protein. However, varying abundances of higher MW proteins were identified for each enzyme, suggesting differential tendencies to aggregate, with the most homogenous being *El*BzaCT.

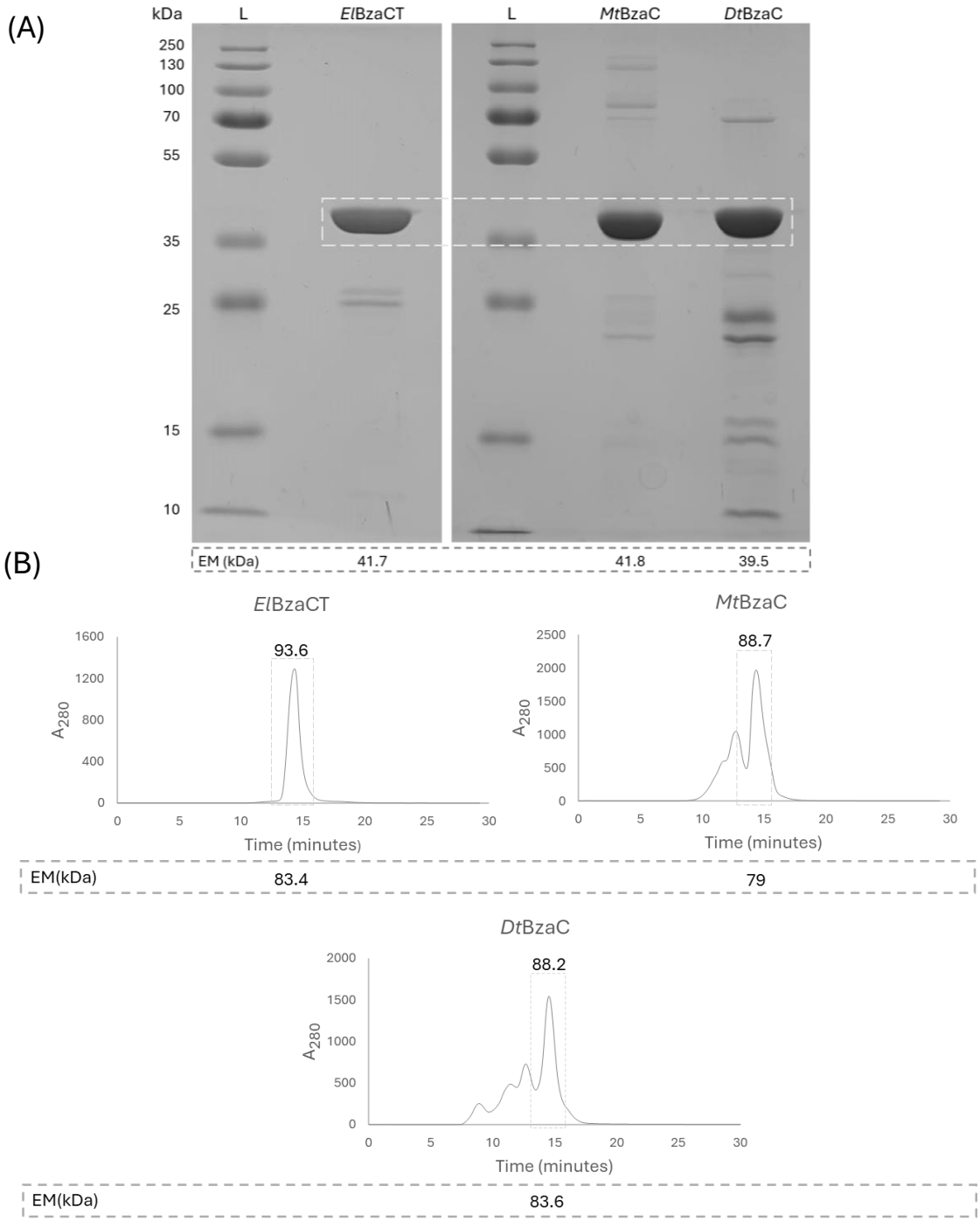


Figure 3.8 - BzaC purification.

(A) 12 % SDS-PAGE gels loaded with 5 µg of purified protein, BzaC bands indicated with grey dashed box. **(B)** Gel filtration profiles of each enzyme with caluclated masses of the dimer species indicated. Expected mass (EM) of each protein shown under respective figures.

3.2.2 Guided biosynthesis of adenosyl 5(6)-hydroxybenzimidazolylcobamide

To enable the investigation of the hypothesis that BzaC is able to catalyse methyl-transfer to a cobamide substrate, adenosyl 5(6)-hydroxybenzimidazolylcobamide (Ado[5(6)-OHBza]Cba) was produced via guided biosynthesis. This was achieved using an engineered strain of *E. coli* (ED674, generated by Dr Evelyne Deery) containing the genes required for the biosynthesis of AdoCbi-GDP and the attachment of supplemented bases, such as benzimidazoles, at the lower ligand position. The substrate is referred to as Ado[5(6)-OHBza]Cba due to the potential presence of both 5- and 6- isomers in the preparation, however it is represented as the 5-isomer in schematics for simplicity. The source of these potential isomers is explained and their presence is investigated in chapter 4.

Firstly, to optimise the production of Ado[5(6)-OHBza]Cba a small-scale trial of production was carried out supplementing 50 mL ED674 cultures with a series of 5-hydroxybenzimidazole (5-OHBza) concentrations (0 – 80 mg /L). The cobamides were then purified via immobilised BtuF (an *E. coli* periplasmic B_{12} binding protein bound to a nickel column) and relative abundance was assessed via HPLC (method 1). The optimal concentration of 5-OHBza identified for a non-induced system was 20 mg /L and 40 mg/ L for an induced system (IPTG).

The process was then scaled up to produce quantities required for biochemical assays. After isolation via a Ni-BtuF column the cobamide was further purified by RP-18 chromatography and dried via vacuum centrifugation. Purity of the product was then assessed via LC-MS (Figure 3.9) (HPLC method 9) and the amount of product quantified via UV-VIS spectroscopy. This confirmed the production of Ado[5(6)-

OHBza]Cba as an m/z of 784.34 ($[M+H]^2$) was observed (expected = 784.315) and a lower ligand fragment m/z of 347.012 ([M+H]) (expected = 347.06). Furthermore, it produced amounts viable for biochemical assays (~ 0.2 mg /L ED674) at high levels of purity (~90 % from the integration of the HPLC chromatogram recorded at 360 nm).

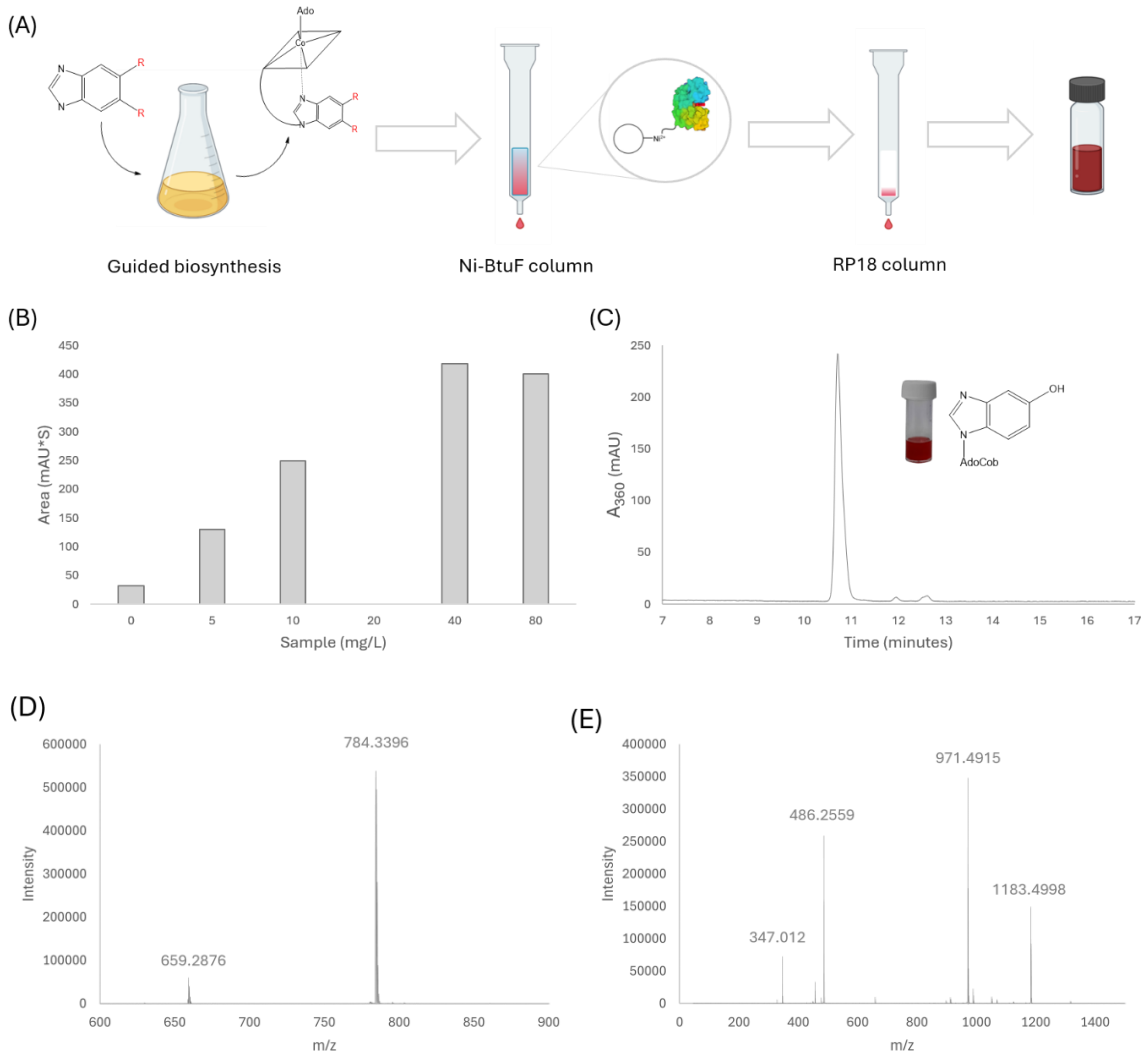


Figure 3.9 - Guided biosynthesis of cobamides.

(A) Schematic representing the overall process of production and purification of cobamides via ED674. **(B)** Bar chart of relative abundances of Ado[5(6)-OHBza]Cba produced in the optimisation trial utilising induction of gene expression. **(C, D, E)** HPLC, MS and MS/MS spectra of purified ado[5(6)-OHBza]Cba.

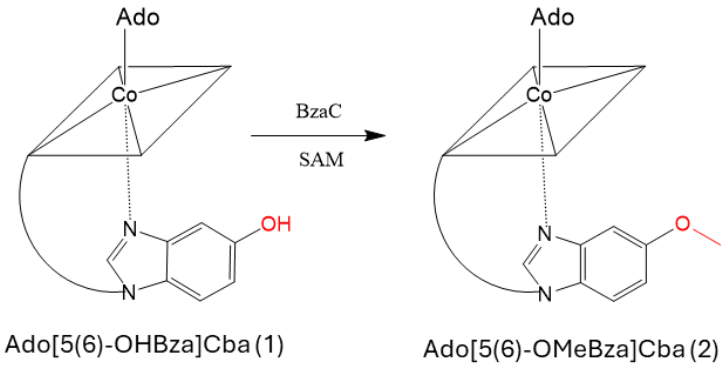
3.2.3 BzaC end-point activity assays

To assess the ability of BzaC to catalyse SAM dependent methyl-transfer to a cobamide substrate, end-point BzaC assays containing 10 µM BzaC (*Dt*, *Mt* or *El*BzaCT), 1 mM SAM, 20 µM Ado[5(6)-OHBza]Cba in 20 mM Tris-HCl pH 8, 100 mM NaCl were set up and incubated at 37 °C for 17 hr 30 minutes. Reactions were then quenched with 1 % glacial acetic acid and analysed via HPLC (method 1) and LC-MS (HPLC method 9). In each reaction (except from the controls lacking each component individually) a new peak with an increased retention time was observed. LC-MS characterisation identified this product to be adenosyl 5(6)-methoxybenzimidazolylcobamide (Ado[5(6)-OMeBza]Cba) with a m/z of 791.35 ($[M+H]^{2+}$)(expected m/z= 791.325). MS/MS fragmentation further confirmed the methylation occurred on the benzimidazole lower ligand (observed m/z = 361.09 ([M+H]), expected = 361.07).

Due to the reaction being non-radical it is most likely that methyl-transfer occurred to the nucleophilic oxygen at the 5-position of the benzimidazole, rather than any sp^2 or sp^3 hybridised carbon or nitrogen on the benzimidazole skeleton. This is further supported by the NMR data shown in Mathur et al. 2020 (55) demonstrating the BzaC catalysed production of 5-OMeBza from 5-OHBza.

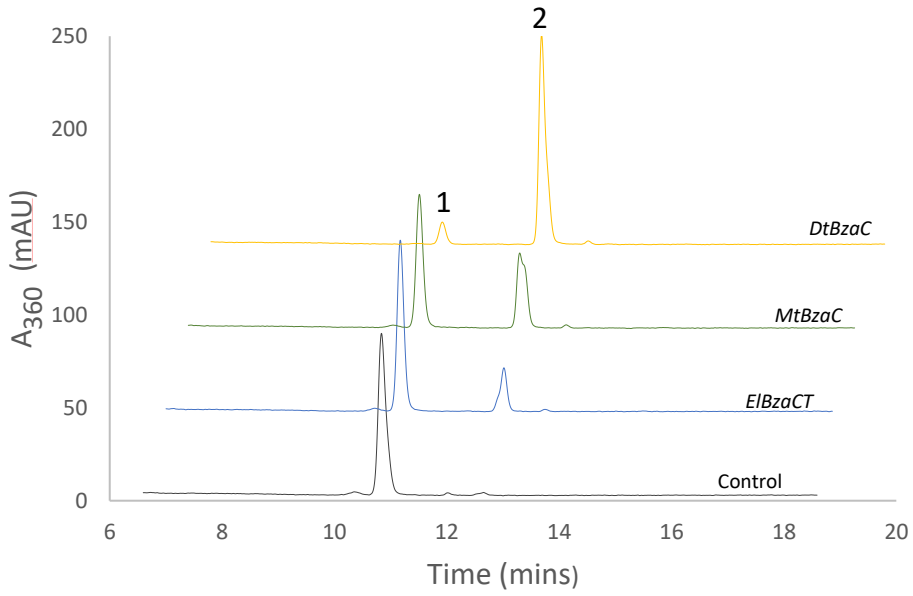
Interestingly, the percentage conversion of substrate to product varied between each enzyme, with the greatest conversion catalysed by *Dt*BzaC (91 %), demonstrating the ability of the enzyme to do more than one round of catalysis. This difference could be contributed to by various factors such as differences in substrate affinity, stability, or inhibition profiles.

(A)

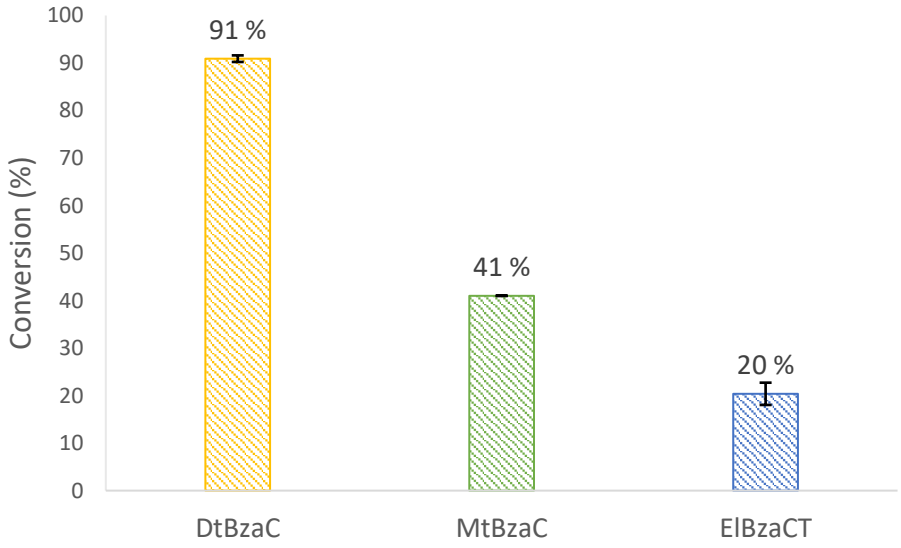


Ado[5(6)-OHBza]Cba (1) Ado[5(6)-OMeBza]Cba (2)

(B)



(C)



(C)

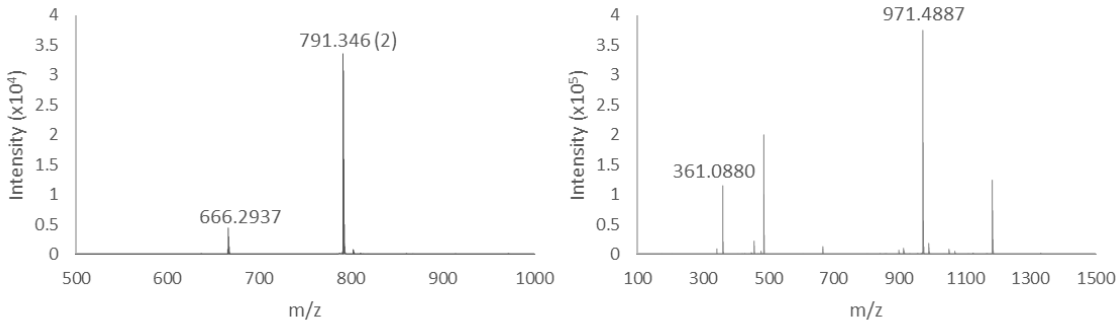


Figure 3.10 - BzaC end-point assays.

(A) Simplified schematic of the BzaC reaction. **(B)** HPLC chromatogram at 360 nm of each BzaC reaction (control = reaction minus *Dt*BzaC) showing the SAM dependent BzaC catalysed O-methylation of Ado[5(6)-OHBza]Cba (1) to produce Ado[5(6)-OMeBza]Cba (2). **(C)** Bar chart showing the average % conversion of substrate to product by each enzyme. Error bars represent the standard deviation, n = 2. **(D)** MS and MS/MS spectra of Ado[5(6)-OMeBza]Cba produced in a BzaC reaction.

3.2.4 Investigation of BzaC inhibition

To investigate the differences observed in BzaC activity, optimisation trials of *El*BzaCT (the lowest conversion (20 %) were carried out, using protein that had not undergone size exclusion chromatography. These experiments explored various conditions that may be limiting BzaC activity such as requirement for a cation (Mg^{2+}) (76), buffering capacity, salt concentration, pH, protein aggregation, hydroxo[5(6)-OHBza]Cba and product inhibition. Reactions were set up as above, with the respective changes, incubated for 16-24 hr at 30 or 37 °C and analysed by HPLC (method 3).

To assess the requirement for a cation ion, reactions with 10 or 0 mM $MgCl_2$ were utilised. For buffering capacity and pH reactions with 50 vs 20 mM Tris-HCl pH 8 or 20 mM HEPES at pH 7.2, 7.6 and 8 were assayed. In addition, the cobalt - adenosyl upper ligand bond can be cleaved by light producing hydroxo[5(6)-OHBza]Cba, which can occur during reaction preparation. This substrate reacts much more poorly (discussed in section 3.2.5) and therefore could be an inhibitor of the reaction. To test for non-adenosylated substrate inhibition and product inhibition, reactions were spiked with varying concentrations of potential inhibitor (0, 4.2, 8.4 µM hydroxo[5(6)-OHBza]Cba and 0, 2, 4, 10 µM Ado[5(6)-OMeBza]Cba). In addition, the impact of salt concentration was tested using 100 vs 500 mM NaCl and disulphide bond mediated protein aggregation or active site cysteine residue oxidation was assessed by addition of β-mercaptoethanol (0, 0.025%). However, none of the optimisation attempts produced significant increases to the percentage product formed, showing there must be other factors impacting activity.

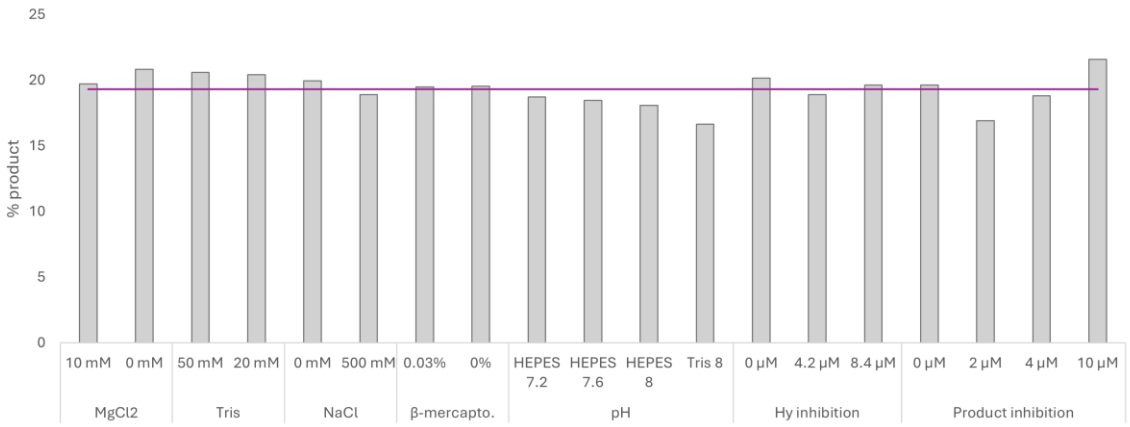
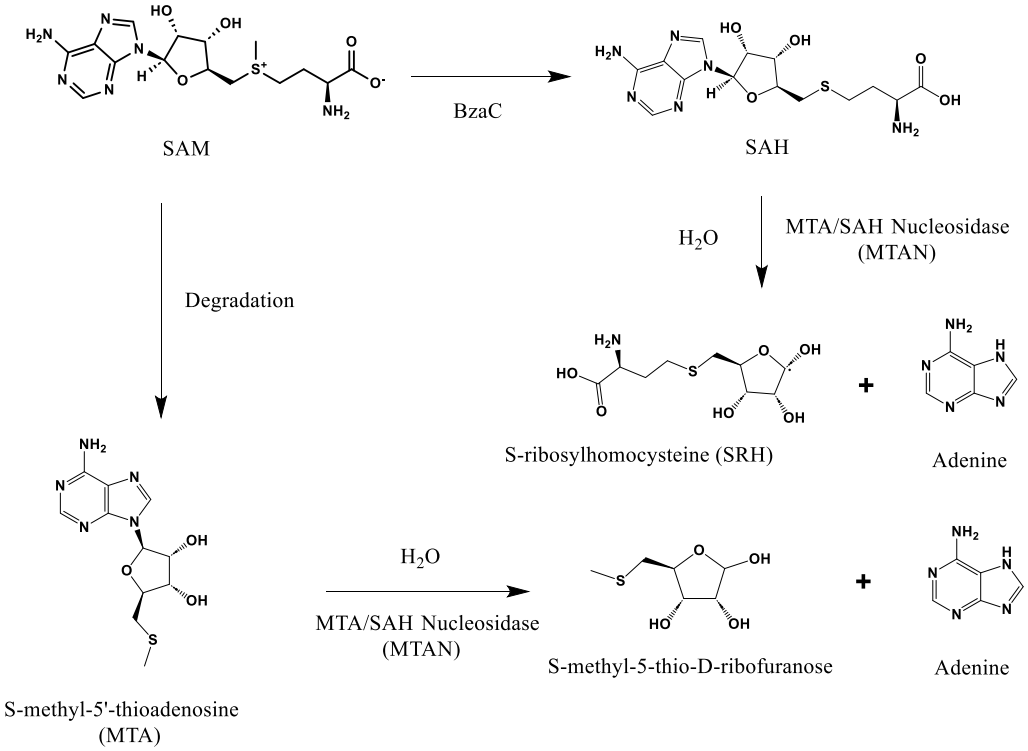


Figure 3.11 - *El*BzaCT reaction optimisation trials.

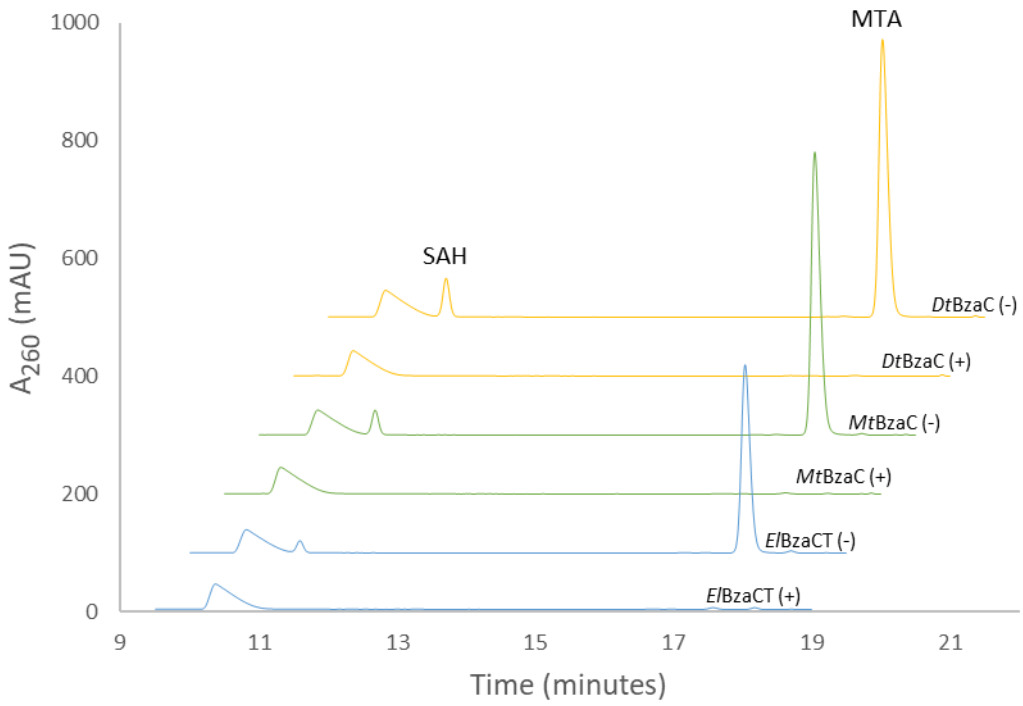
The bar chart represents the percentage conversion of substrate to product for each reaction. The purple line shows the average % conversion across the conditions.

A product of the SAM dependent methylation is S-adenosyl-L-homocysteine (SAH), which is known to provide negative feedback via inhibition of methyltransferases. Furthermore, SAM is a relatively unstable molecule in solution and can degrade into 5-methylthioadenosine (MTA), another known inhibitor of methyltransferases (77). An *E. coli* enzyme called methylthioadenosine nucleosidase (*Ec*MTAN) can hydrolyse both MTA and SAH (**Figure 3.12**, **A**) and was demonstrated to increase the activity of *Mt*BzaC with 5-OHBza (55). Therefore, to investigate if MTA or SAH are inhibiting the BzaC's with a cobamide substrate, reactions (*Dt*, *Mt* and *El*BzaCT) were supplemented with 50 µM *Ec*MTAN and assayed. HPLC analysis at 260 nm (method 2) identified that *Ec*MTAN was able to degrade all the MTA and SAH present in the BzaC reactions (**Figure 3.12**, **B**). Furthermore, removal of SAH and MTA increased both *Dt*BzaC and *Mt*BzaC activity by 8.6 and 15.6 %, respectively, allowing *Dt*BzaC to convert 100 % of the substrate. However, it did not impact the *El*BzaCT reaction. This therefore shows that SAH and/or MTA are inhibitors of BzaC activity, however, it did not allow complete conversion for *Mt*BzaC and had no effect on *El*BzaCT, suggesting that there may be additional factors affecting these enzymes.

(A)


SAM
BzaC
SAH
H2O
MTA/SAH Nucleosidase
(MTAN)
Degradation
S-ribosylhomocysteine (SRH)
Adenine
S-methyl-5'-thioadenosine
(MTA)
H2O
MTA/SAH Nucleosidase
(MTAN)
S-methyl-5-thio-D-ribofuranose
Adenine

(B)


MTA
SAH
DtBzaC (-)
DtBzaC (+)
MtBzaC (-)
MtBzaC (+)
ElBzaCT (-)
ElBzaCT (+)
A260 (mAU)
Time (minutes)
1000
800
600
400
200
0
9
11
13
15
17
19
21

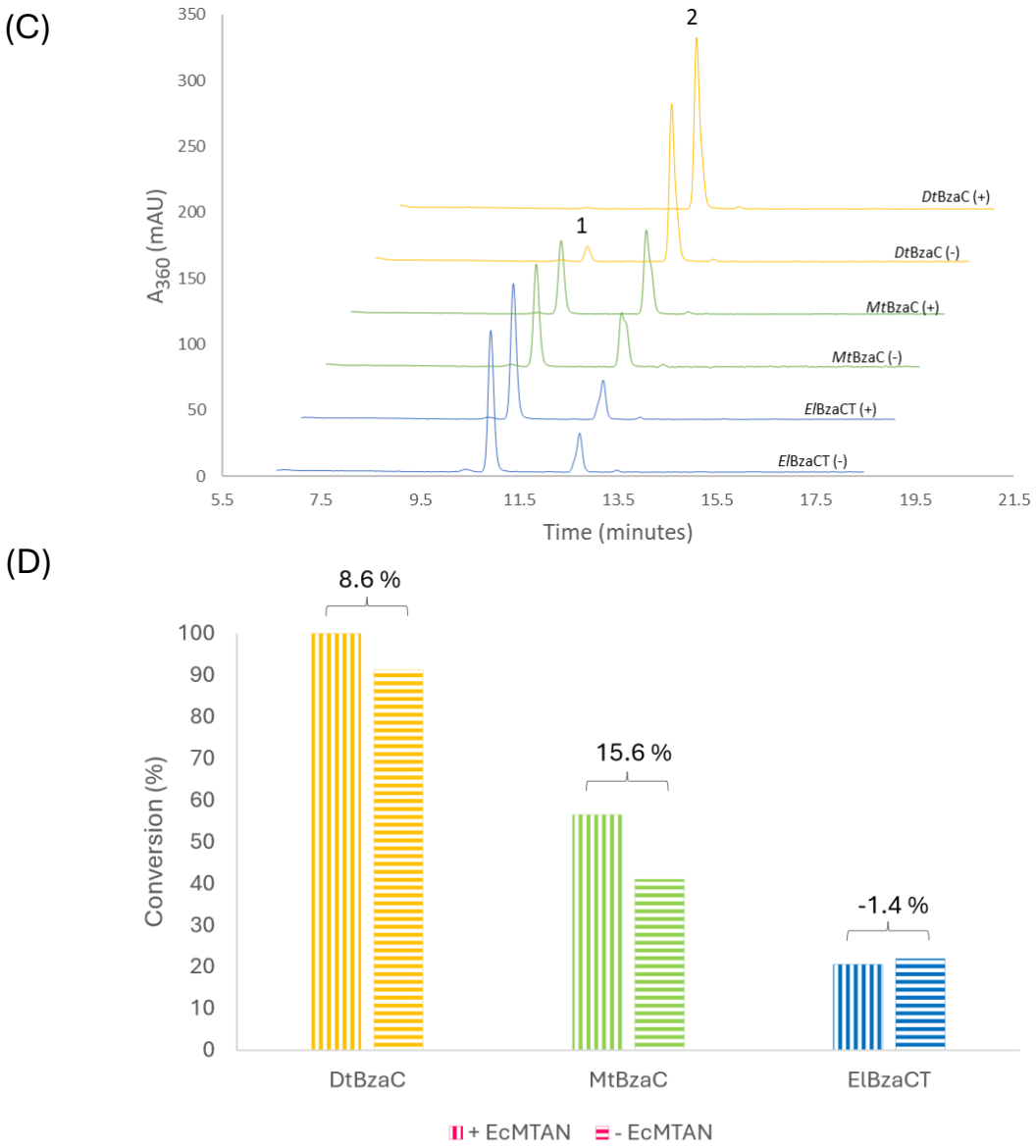


Figure 3.12 - Release of inhibition by *Ec*MTAN.

(A) Schematic of the production of SAH and MTA in a BzaC reaction and their subsequent degradation by *Ec*MTAN. **(B)** HPLC chromatogram at 260 nm representing the degradation of SAH and MTA in BzaC reactions supplemented with *Ec*MTAN. **(C)** HPLC chromatogram at 360 nm showing the impact of *Ec*MTAN on BzaC catalysed conversion of substrate to product. **(D)** Bar chart showing the % conversion of adenosylated substrate to product in the absence (-) and presence (+) of *Ec*MTAN. % difference indicated above each respective bar.

3.2.5 Influence of the cobamide upper ligand on BzaC activity

Next, we wanted to investigate the influence of the cobamide upper ligand on BzaC activity, to understand more about the biochemistry of the reaction. The upper ligand is important as it may be involved directly in protein binding and it influences the base on / base off equilibrium of the cobamide, which in turn can impact binding/ activity. To do this two additional upper ligand forms of [5(6)-OHBza]Cba were produced, Hydroxo (OH) and Cyano (CN). OH[5(6)-OHBza]Cba was generated by exposing Ado[5(6)-OHBza]Cba to direct light for 30-60 minutes, conversion was analysed by HPLC and LC-MS. CN[5(6)-OHBza]Cba was produced by the addition of potassium cyanide (KCN), monitoring the conversion to di-cyano cobamide via UV-VIS. CN[5(6)-OHBza]Cba was then purified using RP18 chromatography and dried via vacuum centrifugation. Resuspended CN[5(6)-OHBza]Cba was then quantified via UV-VIS spectroscopy. The shift in absorption maxima from 367.5 to 361 demonstrates the natural conversion of di-cyano to mono-cyano in solution, and loss of a peak at 260 nm shows the removal of the cleaved adenosyl group my RP18 chromatography (**Figure 3.13**). Therefore, all subsequent assays containing CN[5(6)-OHBza]Cba are in mono-cyano form.

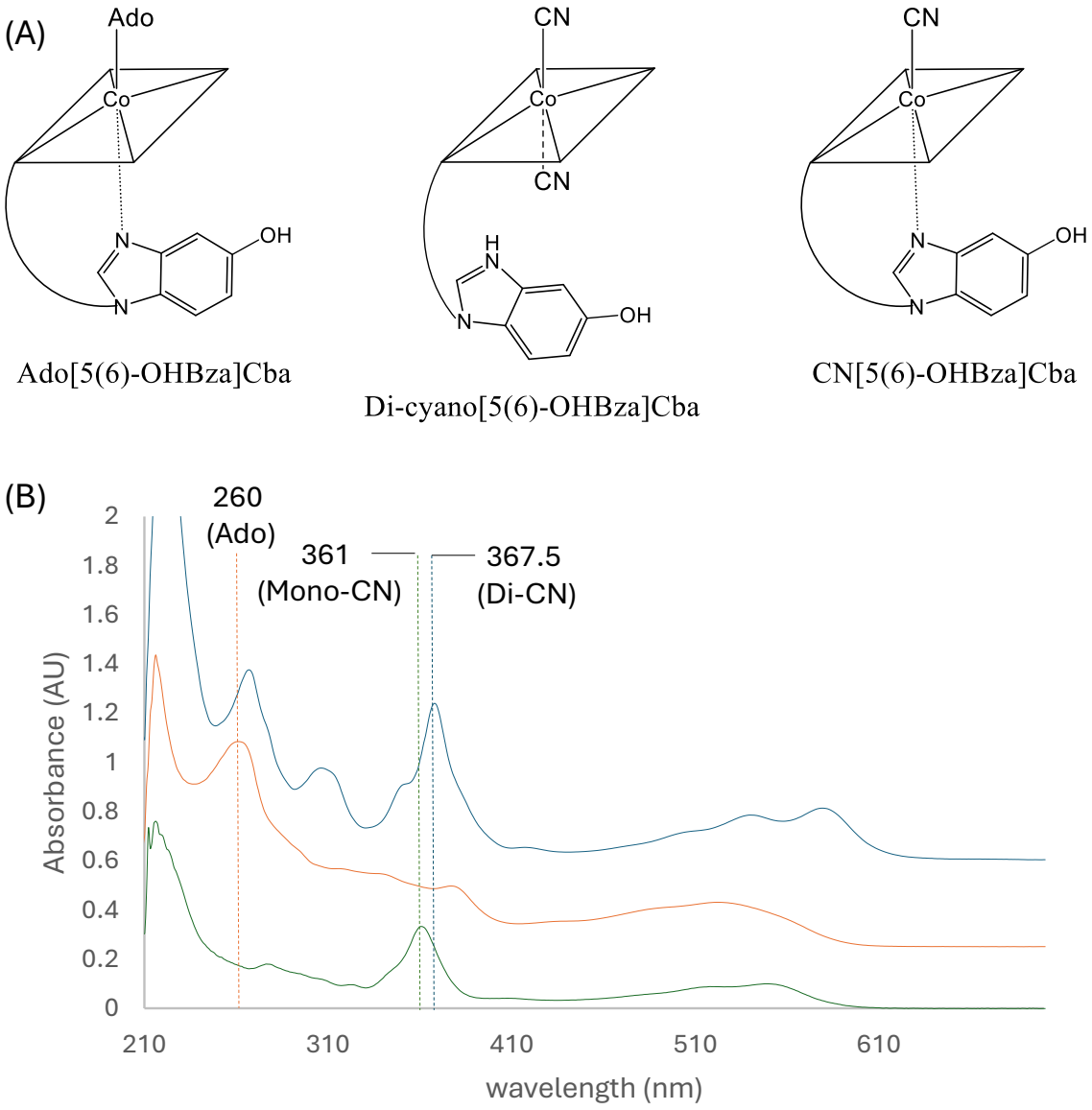


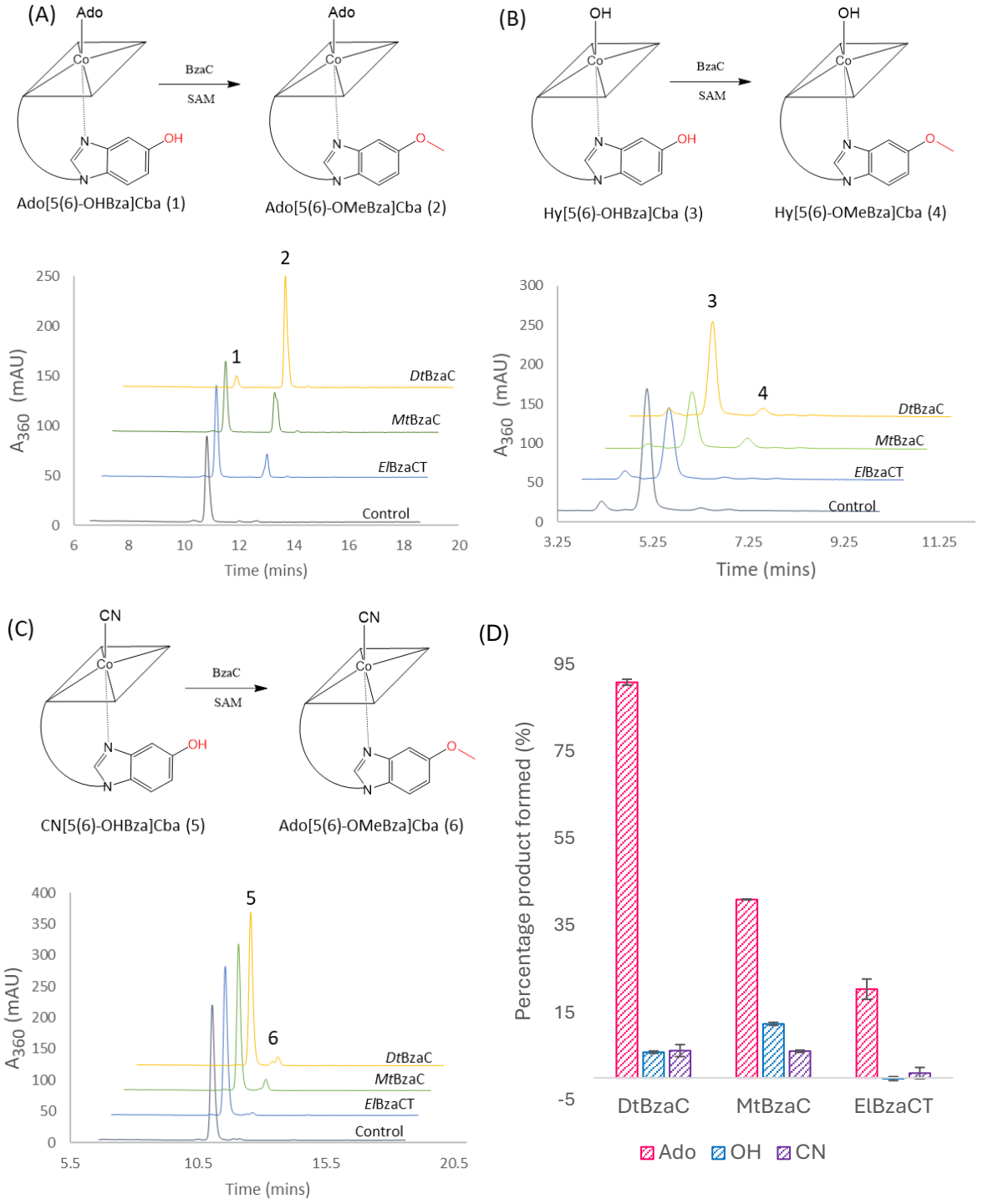
Figure 3.13 - Conversion of adenosylated substrate to cyano.

(A) Chemical structures of mono and di-cyano[5(6)-OHBza]Cba. **(B)** Stacked UV-VIS spectra of Ado[5(6)-OHBza]Cba, di-cyano[5(6)-OHBza]Cba and mono-cyano (CN)[5(6)-OHBza]Cba.

BzaC reactions containing 10 µM BzaC (*Dt*, *Mt* and *El*BzaCT), 20 µM [5(6)-OHBza]Cba (ado, hydroxo and cyano), 1 mM SAM, with and without 50 µM *Ec*MTAN in 20 mM Tris-HCl pH 8, 100 mM NaCl were set up and incubated at 37 °C for 17 hr and 30 minutes.

Reactions were then quenched and analysed by HPLC (method 1) and LC-MS (LC method 9) identifying peaks with increased retention times for each substrate assayed, to varying extents. The expected product was observed again for each BzaC reaction provided Ado[5(6)-OHBza]Cba (m/z values of 791.34 (total cobamide) and 361.09 (lower ligand) corresponding to Ado[5(6)-OMeBza]Cba). MS and MS/MS analysis of the reactions containing hydroxo and CN[5(6)-OHBza]Cba identified the products to be hydroxo and CN[5(6)-OMeBza]Cba with m/z values of 665.79 ($[M]^{2+}$)and 679.29 ($[M+H]^{2+}$) (expected = 665.775 and 679.275), respectively, and lower ligands of 361 ([M+H]).

Each BzaC had a greater activity with Ado[5(6)-OHBza]Cba than either hydroxo or cyano (**Figure 3.14**), with both hydroxo and cyano having activities of 0-12 and 1-6 % respectively with no *Ec*MTAN and 0-14 and 2-16 % respectively with *Ec*MTAN. This supports the hypothesis that a cobamide substrate is acted upon by the enzymes of the *bza* operon as B_{12} is always biosynthesised with adenosyl as the upper ligand.


(A)
Ado
Co
N
N
OH
BzaC
SAM
Ado
Co
N
N
O
Ado[5(6)-OHBza]Cba (1)
Ado[5(6)-OMeBza]Cba (2)
(B)
OH
Co
N
N
OH
BzaC
SAM
OH
Co
N
N
O
Hy[5(6)-OHBza]Cba (3)
Hy[5(6)-OMeBza]Cba (4)
2
1
DtBzaC
MtBzaC
ElBzaCT
Control
A360 (mAU)
Time (mins)
3
4
DtBzaC
MtBzaC
ElBzaCT
Control
A360 (mAU)
Time (mins)
(C)
CN
Co
N
N
OH
BzaC
SAM
CN
Co
N
N
O
CN[5(6)-OHBza]Cba (5)
Ado[5(6)-OMeBza]Cba (6)
5
6
DtBzaC
MtBzaC
ElBzaCT
Control
A360 (mAU)
Time (mins)
(D)
Percentage product formed (%)
DtBzaC
MtBzaC
ElBzaCT
Ado
OH
CN

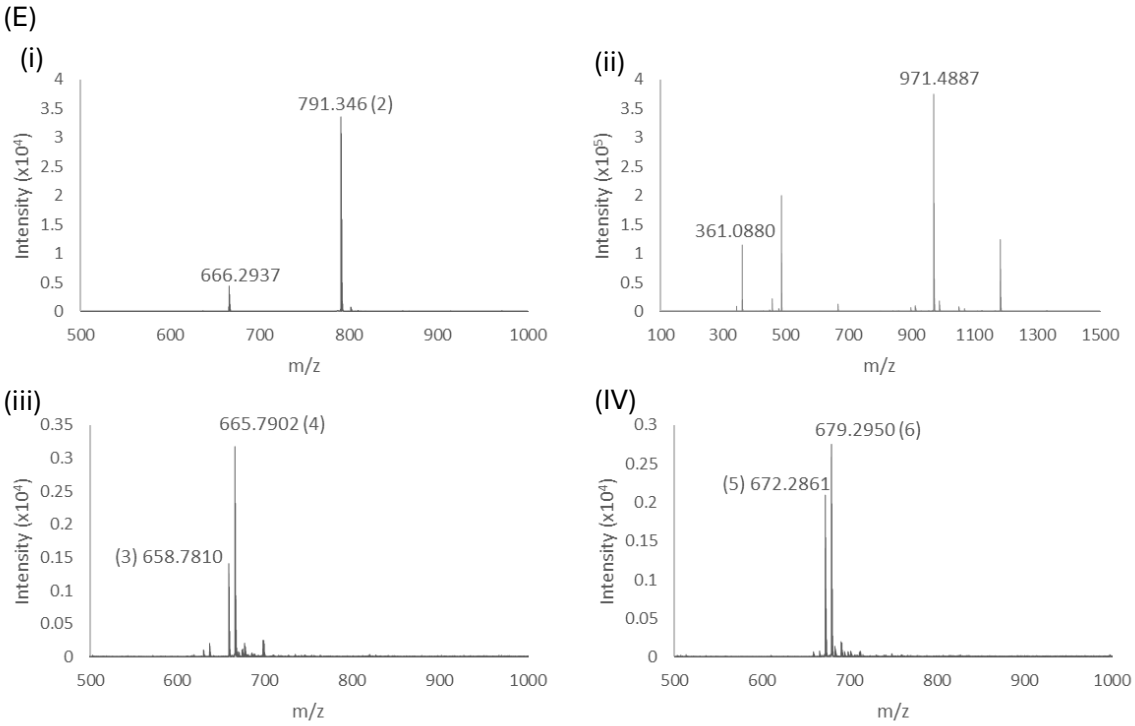


Figure 3.14 - Influence of the upper ligand on BzaC activity.

(A, B, C) HPLC chromatograms at 360 nm and reaction schemes for each BzaC (*Dt*, *Mt* and *El*BzaCT) assayed with each upper ligand variant (Ado, OH and CN[5(6)-OHBza]Cba. **(D)** Percentage conversion of substrate to product in each reaction without *Ec*MTAN, error bars represent standard deviation (n=2). **(E)** MS and MS/MS characterisation of reaction products. Ado[5(6)-OMeBza]Cba MS (i) and MS/MS (ii), OH[5(6)-OMeBza]Cba (iii) and CN[5(6)-OMeBza]Cba (IV).

Furthermore, the data suggest that the adenosyl group may be required for direct binding to the enzyme, as in its absence activity is greatly reduced. In addition, the pKbase-off values for both cyano and hydroxy B_{12} are 0.1 and -2.13, respectively and 3.67 for adenosyl, meaning that cyano and hydroxy are less easily converted to base-off, compared to adenosyl (21, 78, 79). This is due to the increased electron donating

potential of the adenosyl group over cyano and hydroxo, weakening the cobalt-benzimidazole bond, making protonation of the benznidazole and subsequent displacement easier. This suggests that the ability of Ado[5(6)-OHBza]Cba to go base-off more easily may be allowing for the lower ligand to access the BzaC active site in the correct conformation for catalysis, thus increasing activity. It has also been demonstrated that some B_{12} dependent enzymes require a base-off/ His-on mechanism by which the lower ligand is displaced and the cobalt is coordinated by a histidine (His) residue. This data might therefore suggest this is a mechanism by which BzaC could bind its substrate (22).

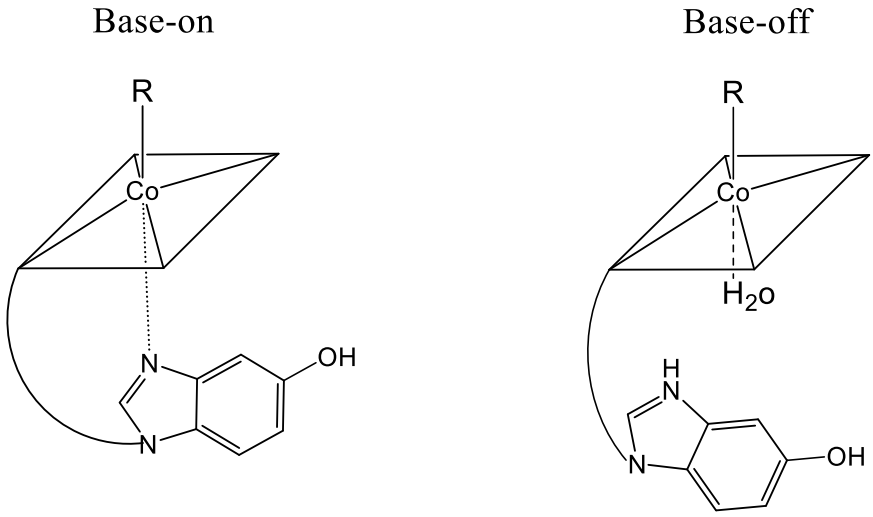


Figure 3.15 - Base-on vs Base-off [5(6)-OHBza]Cba.

3.2.6 Production of alkyl cobamides via BzaC reaction manipulation

Now that the basic biochemistry of the BzaC catalysed reaction has been established we set out to further explore the enzymes substrate scope and to produce novel cobamide analogues that could be used for conjugation. These analogues are of particular interest as they could be used for tracking B_{12} uptake (80) and drug delivery

systems for both eukaryotes and prokaryotes (81) (for example, B_{12} uptake is increased in cancer tumours (82, 83)). Conjugated cobamides have been produced before via chemical and biochemical routes attaching 'cargo' to the ribose ring of the nucleotide loop, cobalt, corrin ring side chain and C5 position of the corrin ring. However, each attachment position has its limitations such as molecules attached to the ribose can easily be lost during remodelling processes within certain organisms or positions may impact cobamide uptake (84).

One biochemical method described in the literature to produce an alkene analogue is by the incorporation of alkyl groups via SAM dependent methyltransferases using modified SAM analogues such as allyl-SAM. This method was utilised by Lawrence at al. 2018 (80) to generate a C5 corrin allyl analogue (C5-allyl-precorrin-8) by the methyltransferase activity of CobL and allyl-SAM. This molecule was then converted to C5-allyl-Hby by CobH, B, and Q, which was then able to have the cobalt inserted and chemically attached to cysteamine, via the allyl group, which the free amine could then be attached to a fluorophore, producing C5-fluorphore-cobyric acid. These conjugates were then used for studying cobalamin uptake by different organisms, such as *E. coli* and *Mycobacterium tuberculosis* (*M. tuberculosis*), whereas only ribose-fluorophore conjugates (described in previous literature) were observed to be taken up and retained in *Caenorhabditis elegans* (*C. elegans*) – supporting that the position of the attachment and structure of the cobamide can be important for uptake. The ribose-probe was utilised in *C. elegans* to identify the regions of the worm in which B_{12} was localised, demonstrating its ability to work as a probe for understanding uptake and trafficking mechanisms.

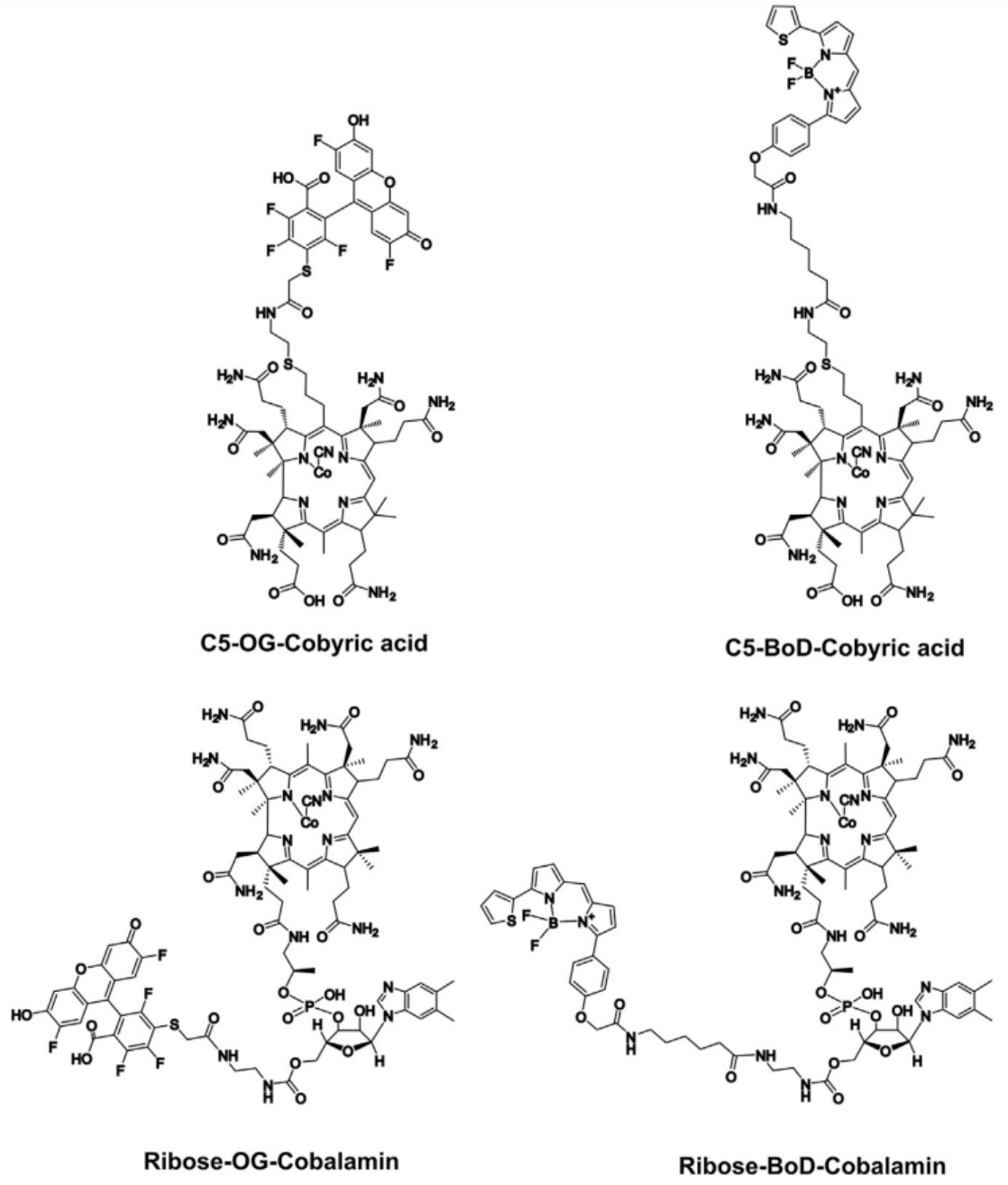


Figure 3.16 - Vitamin B_{12} fluorophore conjugates.

Four conjugates used for studying cobamide uptake systems in bacteria, worms and plants. Taken from (80).

As we have a methyltransferase (BzaC) acting on the benzimidazole of a cobamide we wanted to test the use of alkyl/alkyne-SAM analogues for cobamide analogue production with an alternative site for attachment. These therefore may be recognised and trafficked differently within organisms due to the different attachment / orientation of a fluorophore.

Firstly, reactions containing 180 µM allyl-SAM or SAM were set up with 5.5 µM *Dt*BzaC (non-gel filtrated) and 20 µM Ado[5(6)-OHBza]Cba in 20 mM Tris-HCl pH 8,100 mM NaCl and incubated at 30 °C for 18 hours and 30 minutes. Reactions were then quenched with 1 % acetic acid and analysed via HPLC (method 3) and LC-MS (LC method 11). A peak with an increased retention time compared to Ado[5(6)-OMeBza]Cba was observed with a m/z of 804.35 ($[M+H]^{2+}$)(expected = 804.33). This is a m/z increase of 13 which corresponds to the $[M]^{2+}$ change of the addition of two carbons (24) and two hydrogens (2) present in the allyl group ($-CH_2-HC=CH_2$) compared to a methyl group (CH_3) (Figure 3.17). However, the percentage conversion was lower compared to standard SAM, showing the allyl-SAM is not accepted as well as a cofactor.

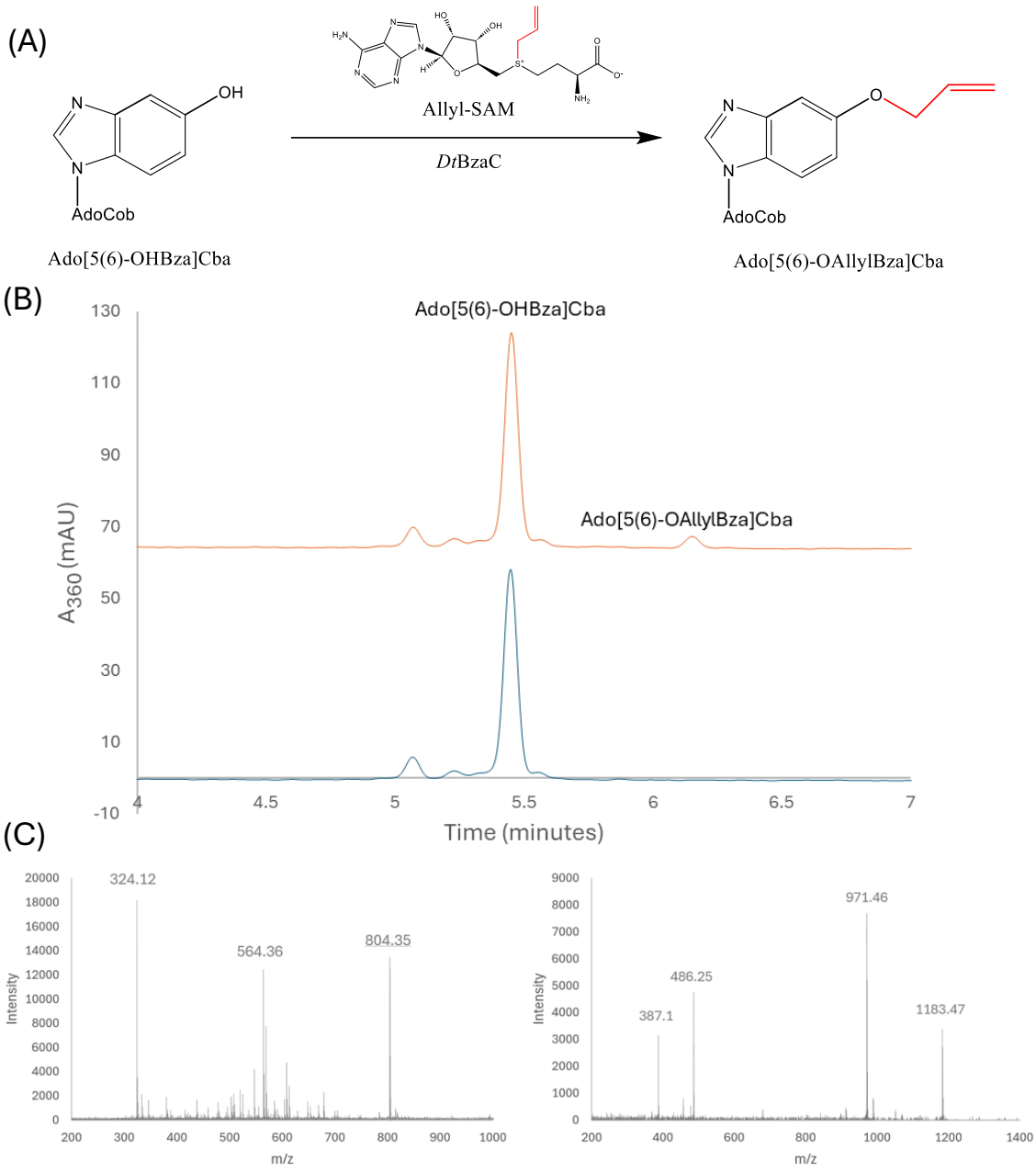


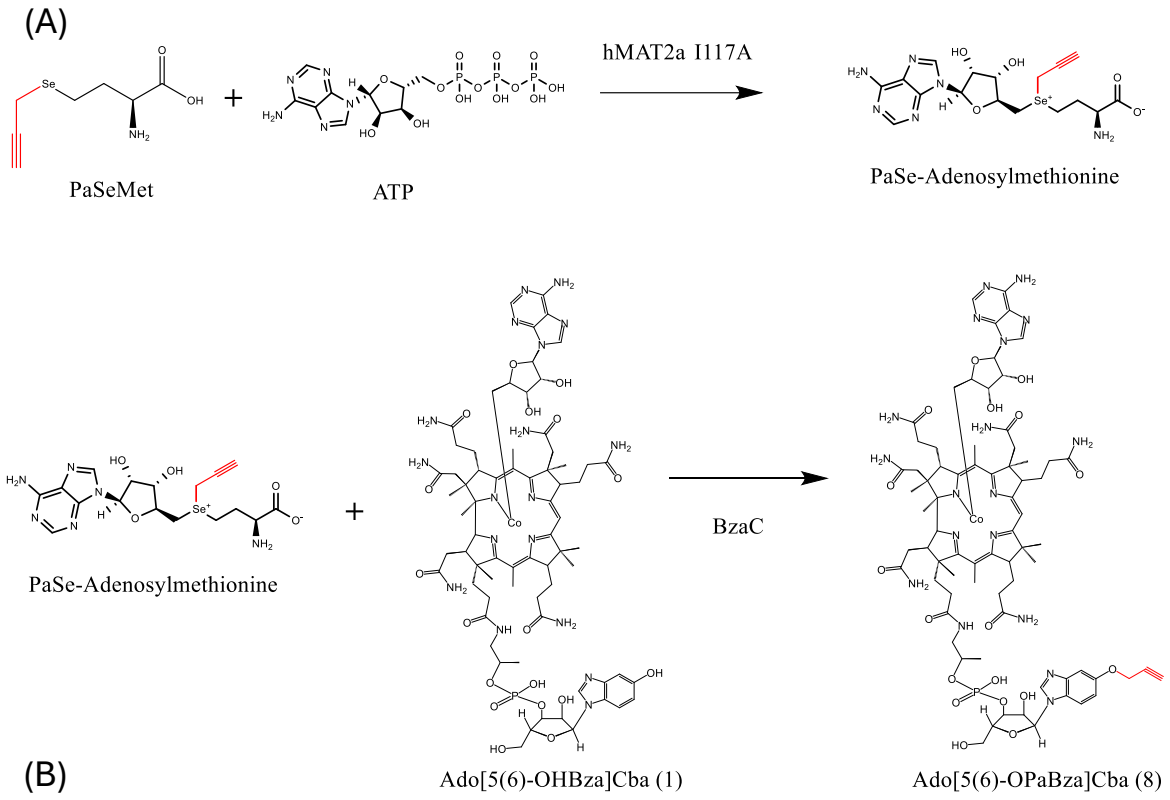
Figure 3.17 - Biosynthesis of Ado[5(6)-OAllyl]Cba using Allyl-SAM as a cofactor. **(A)** A schematic representing the *Dt*BzaC catalysed Allyl group transfer from Ally-SAM to Ado[5(6)-OHBza]Cba to produce Ado[5(6)-OAllylBza]Cba. **(B)** HPLC chromatogram at 360 nm of a reaction containing all components (orange) and a reaction without *Dt*BzaC (blue). **(C, D)** MS and MS/MS spectra of the reaction product.

Next, to produce analogues applicable for simpler downstream 'click' reactions we wanted to make a propargyl-cobamide which contains a triple carbon bond rather than a double carbon bond ($CH{\equiv}C{-}CH_2{-}$). These 'click' reactions are favourable due to the simplicity and efficiency of the methods, making them more practical for completion under standard laboratory conditions.

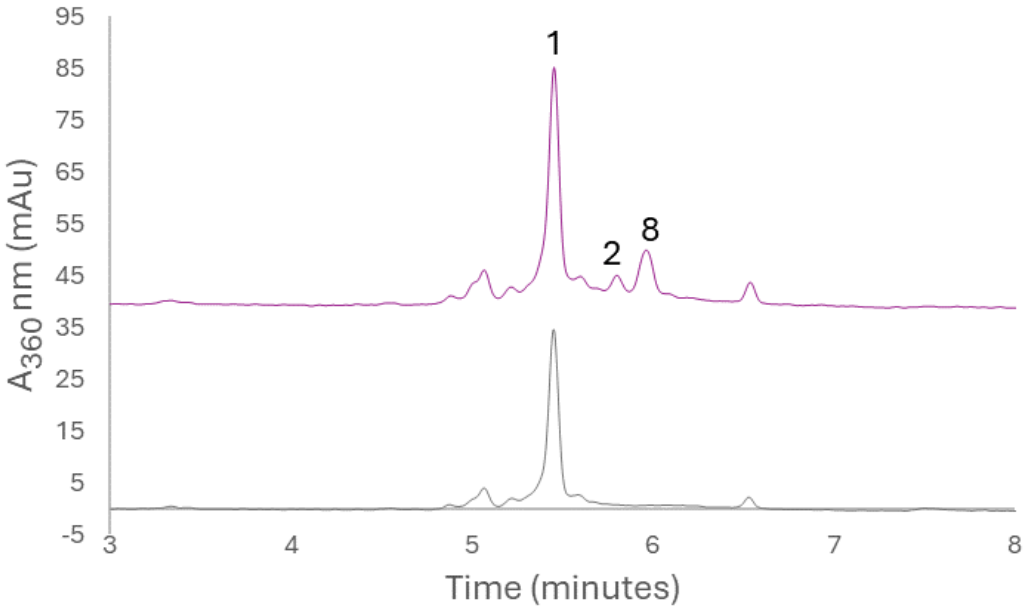
The process for this was more complicated as propargyl-seleno-adenosyl-L-methionine (PaSe-AdoMet) was biosynthesised in situ with the BzaC transformation, due to its short half-life. To do this we used an engineered human methionine adenosyltransferase II-α (hMAT2A) enzyme with a I117A mutation (hMAT2A-I117A) which produces a modified active site able to better accept bulkier alkyne-SAM analogues. This enzyme can catalyse the transfer of an adenosyl group from adenosine triphosphate (ATP) to L-methionine or propargyl-seleno-methionine (PaSeMet) to produce SAM or PaSe-AdoMet.

Reactions containing 2 mM PaSeMet or L-methionine, 2 mM ATP, 50 mM KCL, 10 mM $MgCL_2$, 22 µM hMAT2A-I117A, 10 µM *Dt*BzaC (non-gel filtrated) and 20 µM Ado[5(6)-OHBza]Cba were set up in 20 mM Tris-HCl pH 8, 100 mM NaCl and incubated for 68 hours at room temperature (RT). Reactions were then quenched with 1 % acetic acid glacial and analysed by HPLC (method 3) and LC-MS/MS (method 11). In the reactions supplied with L-methionine, Ado[5(6)-OMeBza]Cba was identified as the product of the reaction, showing that the use of hMAT2a-I117A was able to reconstitute BzaC activity in the absence of purified SAM. This activity is absent in reactions lacking hMAT2a-I117A. Moreover, in reactions provided PaSeMet an additional peak was present with a m/z of 803.3408 ($[M+H]^{2+}$)(expected = 803.325) and a lower ligand m/z

of 385.08 ([M+H])(expected = 385.07), demonstrating the production of the propargyl-cobamide (adenosyl 5(6)-O-propargybenzimidazolylcobamide (Ado[5(6)-OPaBza]Cba)).



(B)



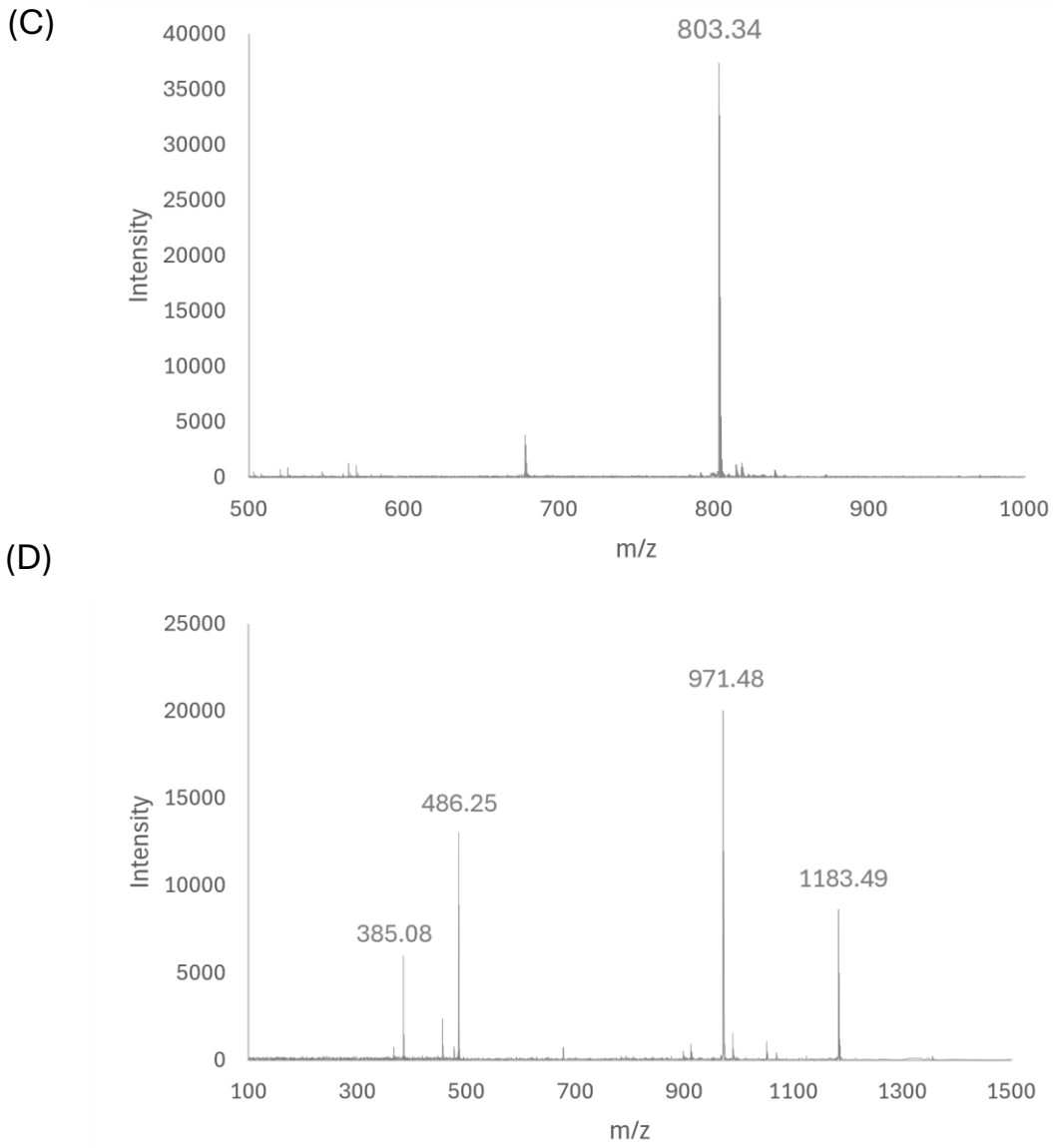


Figure 3.18 - **Ado[5(6)-OPaBza]Cba biosynthesis.**

(A) Schematic representation of the in-situ biosynthesis of PaSe-AdoMet from PaSeMet and ATP catalysed by hMAT2a-I117A and BzaC catalysed alkyne transfer of the propargyl group from PaSe-AdoMet to Ado[5(6)-OHBza]Cba to produce Ado[5(6)-OPaBza]Cba. **(B)** HPLC chromatogram at 360 nm demonstrating the BzaC dependent production of Ado[5(6)-OPaBza]Cba (8) and a small amount of Ado[5(6)-OMeBza]Cba (2) from Ado[5(6)-OHBza]Cba (1). **(C, D)** MS and MS/MS spectra of Ado[5(6)-OPaBza]Cba.

3.2.7 Ado[5(6)-OPaBza]Cba – fluorophore conjugation

Now that the propargyl-cobamide had been produced we wanted to purify it and carry out a click reaction to an azide containing fluorophore. To do this Ado[5(6)-OPaBza]Cba containing samples were purified on an RP18 column, to remove buffer and additional reaction components and dried via vacuum centrifugation. The pellet was resuspended and the Ado[5(6)-OPaBza]Cba was purified by semi-preparative HPLC (method 3). The resulting fraction was then dried and resuspended in 20 mM Tris-HCl pH 8,100 mM NaCl.

To carry out the conjugation reaction to an azide containing fluorophore (Tetramethylrhodamine 5-Carboxamido-(6-Azidohexanyl)), 5-isomer (TAMRA azide)) Copper-Catalysed Azide-Alkyne Cycloaddition (CuAAC) 'click' chemistry was utilised. This method was chosen as it can be carried out efficiently in aqueous conditions at room temperature, making it ideal for biological molecules. A reaction containing Ado[5(6)-OPaBza]Cba, 10 µM TAMRA azide, 0.1 mM $CuSO_4$, 0.5 mM THPTA, 2.5 mM Sodium ascorbate was carried out in 0.1 M Potassium Phosphate buffer pH 7 for 1 hr at RT. Reactions were quenched via addition of 50 mM EDTA pH 7.4 and analysed by LC-MS (LC method 10).

The analysis showed two overlapping peaks with an increased retention time compared to the Ado[5(6)-OPaBza]Cba, indicating increased hydrophobicity, which is expected with the conjugation to the hydrophobic fluorophore. Furthermore, the peaks indicated the presence of two conjugate species (Ado and OH) with the m/z values of 1080.4709 (Ado[5(6)-O-Triazole-TAMRA-Bza]Cba ($[M]^{2+}$), expected m/z = 1080.46) and 720.99 ($[M+2]^{3+}$) expected m/z = 720.97) and 954.9172 (OH[5(6)-O- Triazole-TAMRA-

Bza]Cba ($[M]^{2+}$), expected m/z of 955.41). In addition, an additional 995.4 species was identified, which could potentially be the hydroxo species + SO_4.

Furthermore, the expected m/z for the lower ligand plus + TAMRA ([5(6)-O-Triazole-TAMRA]Ribotide)($[M]^{+}$) of 939.34 was observed during fragmentation of Ado[5(6)-O-Triazole-TAMRA-Bza]Cba and the 995.4 species. However, the fragmentation pattern looked different than what is expected for a standard cobamide and did not fully align with simulated ChemDraw fragmentation patterns. For example, the 720.99 ion presented the expected adenosyl corrin fragment of 971.4560 but presented a lower ligand fragment of 364.17 and a small amount of 727.33. In addition, the 1080.47 ion produced a decreased amount of the 971.5 peak and a large presence of a 955.4 peak. Moreover, the second most abundant species was 727.3 which does not correspond to any simulated fragment. Furthermore, no fragmentation was observed for the expected OH species, only for the putative hydroxo + SO_4 species. However, it contained a 971.5 peak, suggesting the presence of an adenosyl group which is not expected.

Together this suggests that the [5(6)-O-Triazole-TAMRA-Bza]Cba conjugate had been produced, however, to confirm this more accurately the reaction should be scaled up to allow for purification of the conjugate and analysis by NMR and fluorescence excitation / emission. Further investigation into the fragmentation analysis on a higher concentration, purified sample may also unravel the patterns observed.

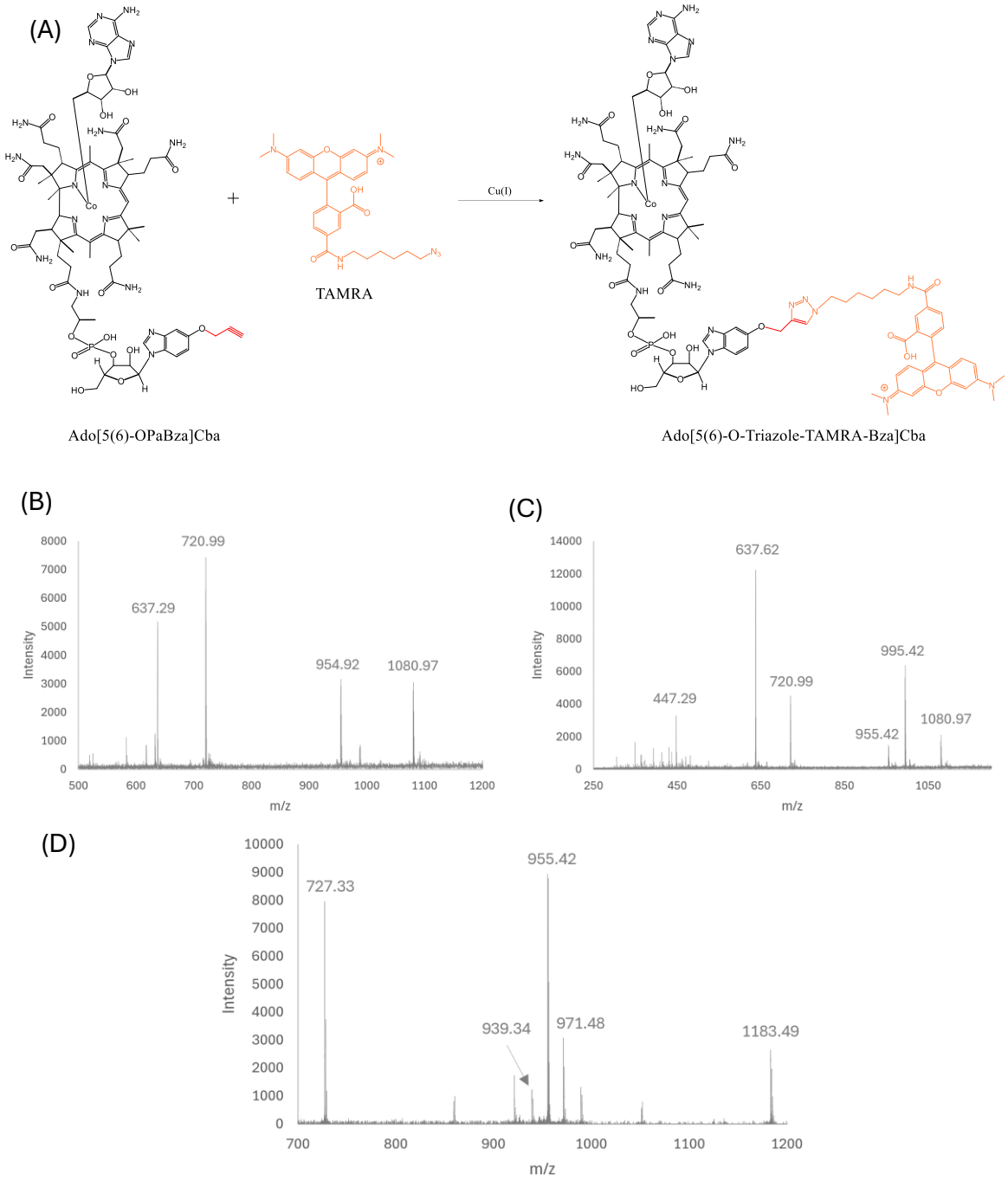


Figure 3.19 - CuAAC conjugation of Ado[5(6)-OPaBza]Cba to TAMRA azide.

(A) Schematic representation of the CuAAC reaction carried out to produce the Ado[5(6)-O-Triazole-TAMRA-Bza]Cba. **(B, C)** MS spectra of the two identified product peaks, with peak B being the most abundant. **(D)** MS/MS spectra of Ado[5(6)-O-Triazole-TAMRA-Bza]Cba 1080.97 species.

3.3 Discussion

The work carried out in this chapter has demonstrated the ability of BzaC to catalyse SAM dependent methyl-transfer to a cobamide substrate, with *Dt*BzaC being able convert 100 % of the substrate to product. Furthermore, the preferred cobamide upper ligand is the adenosyl group, which is the upper ligand incorporated during the biosynthesis of cobalamin, thus supporting the role of BzaC in cobamide biosynthesis. In addition, the presence of later enzymes in the operon (e.g. BzaD, X and E) did not seem to impact the activity of the BzaCs as *Dt*BzaC performed the best and *El*BzaCT performed the worst, however this could be due to *El*BzaCT being a synthetic construct, limiting its activity. Together, this suggests that the preferred substrates for the later enzymes of the *bza* operon (C, D and E) may be cobamides, rather than the previously suggested ribosides. Interestingly, the absence of a known B_{12} binding domain in BzaC suggests a novel mode of binding, most likely in a base off conformation.

Due to these results, I propose a revised pathway scheme starting with the previously demonstrated conversion of AIR to 5-OHBza by BzaAB/F, followed by its activation by CobT to produce 5-OHBza-RP. Next, 5-OHBza-RP needs to be attached at the lower ligand position, to be acted upon by BzaC. This could procced by two routes; 5-OHBza-RP can be de-phosphorlated by CobC and then attached to AdoCbi-GDP by CobS (supported by Maggio Hall LA et al. 1999 (85)) or 5-OHBza-RP can be first attached to AdoCbi-GDP by CobS and subsequently de-phosphorylated by CobC (supported by Zayas C.L. et al. 2007 (86)). BzaC then catalyses the O-methylation of Ado[5(6)-

OHBza]Cba to Ado[5(6)-OMeBza]Cba. The activities of BzaD and E, if present, then convert Ado[5(6)-OMeBza]Cba to Ado[DMB]Cba (adenosylcobalamin) (**Figure 3.20**).

In addition, the only identified inhibitors of BzaC activity were SAH and MTA, due to increased activity in reactions containing *Ec*MTAN. However, this did not completely release the inhibition of *Mt*BzaC (16 % increase to 40 % conversion) and had no impact on *El*BzaCT. This suggests there may be different levels of co-factor inhibition across the BzaCs or there are other factors impacting their activities. It could be that *Dt*BzaC has a tighter affinity for its substrate and therefore converts more substrate in the time period analysed, or *Dt*BzaC may be more stable than *Mt* and *El*BzaCT, thus being able to be active for a longer period of time. Furthermore, the substrate preparation may contain two isomers [5- or 6-OHBza]Cba, due to some CobT/U's being able to produce both isomers of benzimidazoles and ED674 containing two CobT/U's for base activation. Therefore, each BzaC may have different levels of regioselectivity towards either isomer, impacting total activity in a mixed isomer assay. This is discussed in more depth in the next chapter.

Furthermore, the work demonstrated the ability to manipulate the BzaC reaction to make alkyl/alkyne-cobamides which can be used for conjugation to cargo. Future work could explore improving the yield of Ado[5(6)-OPaBza]Cba and Ado[5(6)-O-Triazole-TAMRA-Bza]Cba to allow subsequent purification via BtuF/RP18 chromatography and experimentation. The cobamide-TAMRA conjugate could then be tested against the previously mentioned cobamide-fluorophore conjugates for B_{12} uptake studies to see if the position of fluorophore provides any advantages for uptake or affects its distribution in model organisms.

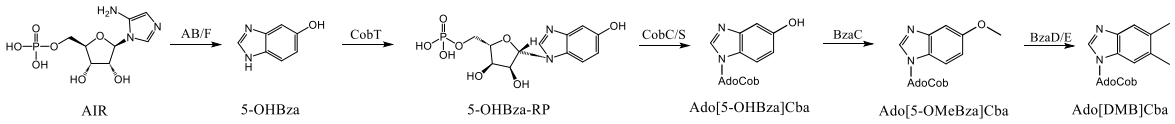


Figure 3.20 - Hypothesised biosynthetic pathway for the anaerobic biosynthesis of Ado[DMB]Cba (adenosylcobalamin) from AIR and AdoCbi-GDP.

Chapter 4: BzaC ligand affinity, kinetics and regioselectivity

4.1 Introduction

Now that the core biochemistry of the BzaC reaction has been established, we wanted to investigate the co-factor/inhibitor and substrate binding affinities, kinetics of the reaction and the potential regiospecificity of the enzyme. These are key areas of research as it will allow for a deeper understanding of mechanisms of inhibition, provide constants to allow for comparison to other known methyltransferases (especially those involved in cobalamin biosynthesis) and provide insight into factors effecting BzaC activity to allow for optimisation to improve product yield to allow for easier downstream characterisation and applications. This information is also useful for future pathway heterologous expression engineering as it will provide information on the optimal concentration of co-factor and substrates for enzyme reaction rates and identify key rate limiting and regulatory processes that could be bypassed for improved pathway flux.

Previous work in Mathur et al. 2020 (55) probed the binding affinities of SAM and SAH to *Mt*BzaC using the quenching of *Mt*BzaC intrinsic tryptophan fluorescence upon ligand binding. This ligand binding induced change in signal was then plotted and fitted to a binding model to obtain K_d values for the interaction. This identified a SAM K_d value of 505.5 µM and SAH of 404.7 µM (**Figure 4.1**). However, no substrate binding or reaction kinetics were explored.

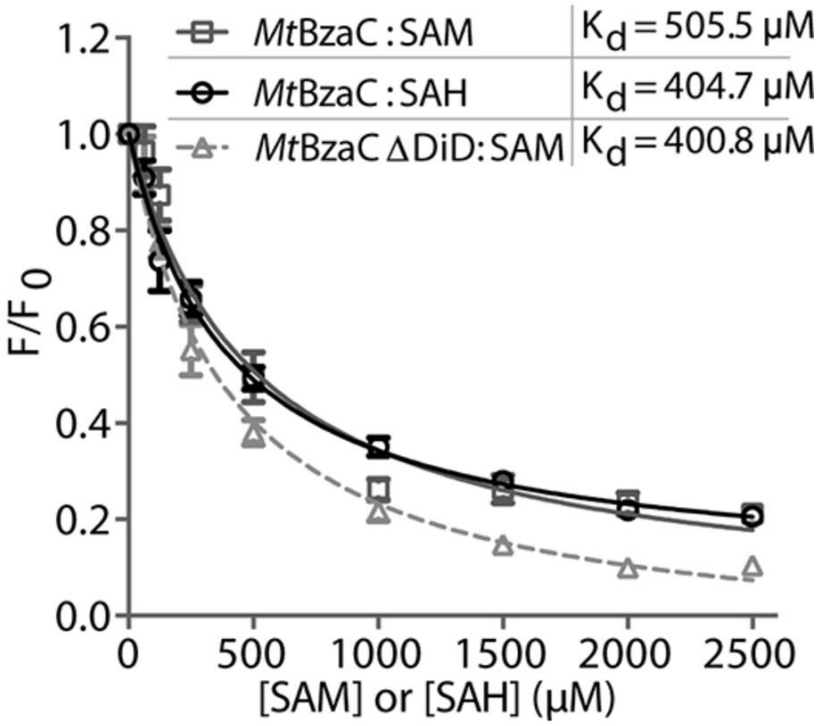


Figure 4.1 - Mathur et al 2020 *Mt*BzaC SAM and SAH affinity determination.

4.1.1 Chapter aims

As there is little previous literature exploring the binding affinities, reaction kinetics and regiospecificity of BzaC enzymes we needed to identify techniques applicable for monitoring BzaC-ligand interactions, product formation over time and isomer levels in both substrates and products. To do this we explored three different methods for determining BzaC-ligand binding affinities; fluorimetry, microscale thermophoresis (MST) and differential scanning fluorimetry (DSF). We tested two methods for analysing the kinetics of the BzaC reactions, a continuous and stopped assay. Additionally, we explored the isomer composition of our substrate preparation via NMR and reaction products via an alternative small-scale method.

4.2 Results

4.2.1 Ligand binding

As we wanted to obtain K_d values for SAH, SAM and MTA for different BzaC enzymes to probe differences in inhibition profiles and to allow for comparison to other methyltransferases in the literature we thought we would first attempt this using a similar technique to Mathur et al. 2020 – fluorimetry.

4.2.1.1 Fluorimetry

Fluorimetry is a technique by which ligand affinities can be determined by the principal of fluorescence changes upon ligand binding. For some proteins, their intrinsic tryptophan fluorescence will be changed upon ligand binding due to conformational changes leading to reduced emission of aromatic residues, especially tryptophan. *Dt*BzaC intrinsic fluorescence was tested in a fluorimeter (rather than Mathur et al. 2020 which used a plate reader), using 1 mL of 10 µM protein, an excitation wavelength of 280 nm and recording emission from 300-400 nm. This produced a good signal of 400 AU with a maxima of ~347 nm, showing the protein was applicable for an intrinsic fluorescence-based assay. Next, a titration of SAM from 0.1 – 2.46 mM was completed, recording emission at 347 nm after each SAM addition. This produced a [SAM] dependent decrease in fluorescence, following a standard binding response curve when ΔF was plotted against [SAM] and nonlinear regression curve fitting with a total

binding model ($y = (B_{max} * x) / (K_d + x) + NS * x$) was carried out using the python package SciPy.

However, due to the absorption maxima of SAM being at 260 nm and the excitation wavelength being 280 nm, it allowed for the possibility that SAM could absorb a portion of the excitation energy, reducing the emission signal independently of *Dt*BzaC binding. To test for the presence of this 'inner filter effect', an ATP titration from 0.2-4.74 mM was carried out as it has the same absorption spectra as SAM but is not expected to bind *Dt*BzaC. A similar [ATP] dependent decrease in signal was observed, with only a small amount of signal from [SAM] binding *Dt*BzaC shown above the inner filter effect (**Figure 4.2**).

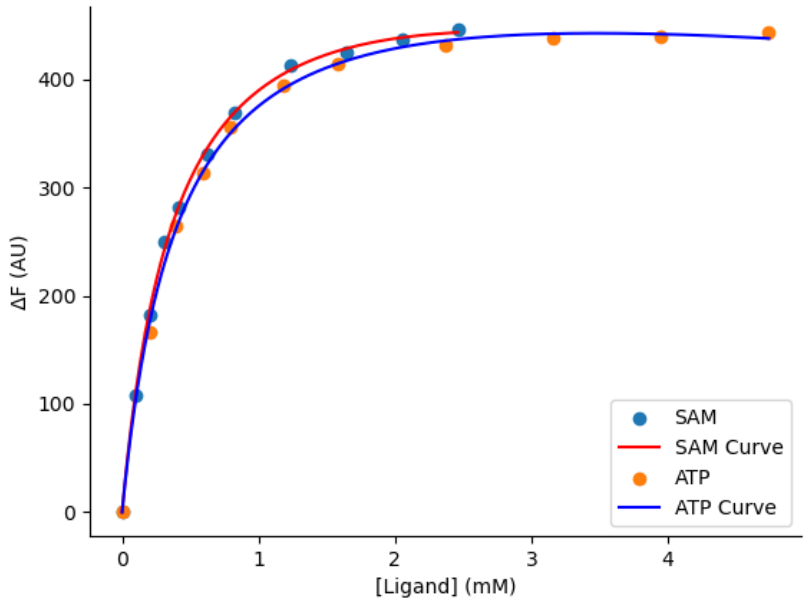


Figure 4.2 - SAM inner filter effect.

A graph representing ΔF against [ligand] for *Dt*BzaC with SAM and ATP with nonlinear regression curves produced using SciPy, fitting data to a total binding model ($y = (B_{max} * x) / (K_d + x) + NS * x$).

To obtain the signal produced from SAM binding, the nonlinear regression curve points from [ATP] were subtracted from [SAM]. Plotting these values and fitting to the total binding model obtained a K_d value of 243 µM (95 % CI: 241-244) (**Figure 4.3**), which is ~ 2 x tighter than the value obtain by Mathur et al. 2020 for *Mt*BzaC (505.5 µM). However, the curve did not fit the expected trend for SAM binding, as after ~ 750 µM the change in fluorescence (ΔF) decreased, rather than plateaued (**Figure 4.3**). This is most likely due to inaccuracies caused by the large inner filter effect observed, as the amount of inner filter effect surpassed the ΔF caused by ligand binding, rather than [SAM] dependent inhibition of binding, as this was not observed in later binding or kinetics experiments (see below).

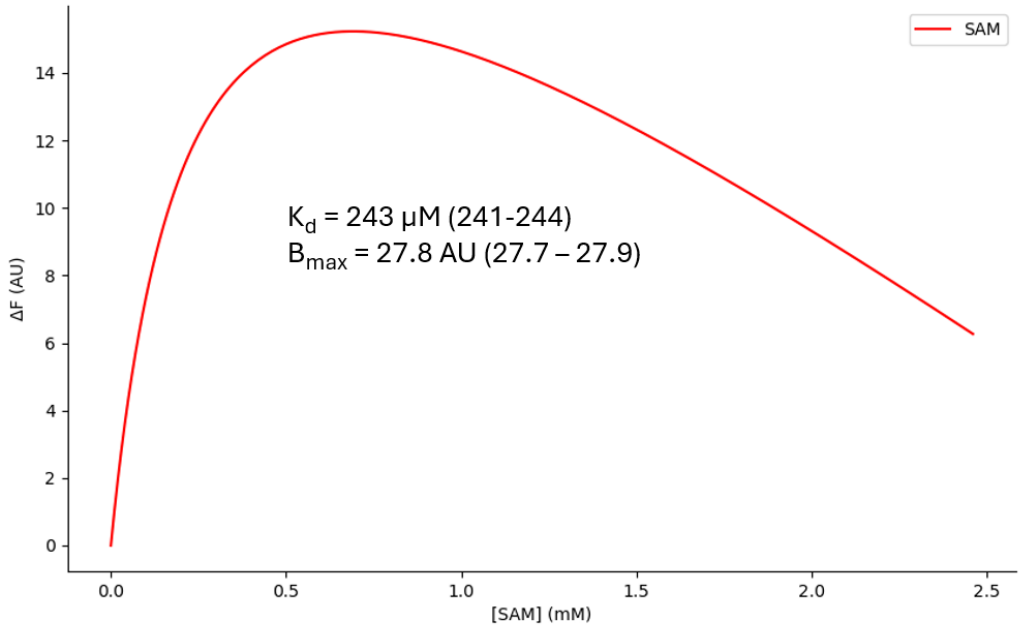


Figure 4.3 - *Dt*BzaC SAM binding accounting for inner filter effect.

Extracted ΔF due to SAM binding induced fluorescence quenching. K_d and B_{max} values obtained by fitting data to the total binding model (y = (B_{max} * x) / (K_d + x) + NS * x).

4.2.1.2 Microscale thermophoresis (MST)

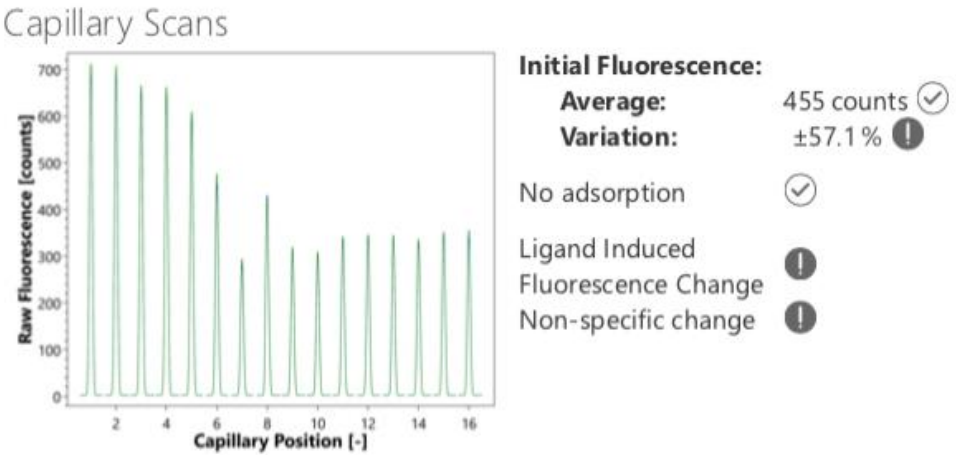
As the data produced from the fluorimetry may not be an accurate representation of nucleotide binding due to the large inner filter effect observed we turned to the use of MST. This technique obtains K_d values from the principal of ligand binding induced changes in the rate of thermophoresis (movement across a thermal gradient). This movement is assessed via the use of fluorescence emission. As the protein (attached to a fluorophore) moves across its gradient, it moves out of the path of the excitation light, reducing the signal. Therefore, the ligand concentration dependent difference in the rate of loss of fluorescence can be monitored, which can be used to determine a K_d value for the system. Furthermore, the protein was labelled with a fluorescent dye (Red-Tris-NTA 2nd generation) which interacts with *Dt*BzaC's His-tag and has an excitation wavelength of 650 and emission at 670 nm, meaning there should be no inner filter effect from any nucleotide or cobamide ligands absorbing any of the excitation or emission energy.

Firstly, the affinity of the dye for the protein of interest needed to be determined, as this affinity will influence the required experimental protocol for studying the binding interaction between the protein-dye complex and ligands of interest. *Dt*BzaC (non-gel filtrated) was diluted to 4 µM in phosphate buffered saline with 0.05 % tween 20 (PBS-T) and a 16-point serial dilution of the protein was performed (4 µM – 0.1 nM). The dye (25 µM final concentration) was then added to each sample and incubated at RT for 30 minutes. Samples were then transferred to the MST capillaries and analysed in a NanoTemper monolith instrument. An initial fluorescence recording of each condition before a thermal gradient was applied showed a [*Dt*BzaC] dependent change in

fluorescence (57 %), which indicates a change in fluorescence upon binding *Dt*BzaC, which should not occur (10 % deviation acceptable).

To investigate this further an Ethylenediaminetetraacetic acid (EDTA) test was performed by addition of 16.6 mM EDTA pH 7.4 to each condition. This chelates the Ni(II) ions from the Tris-NTA dye causing dissociation of the dye from the protein, which helps to determine if the changes in fluorescence are due to protein or non-specific interactions. The experiment showed that addition of EDTA reduced the variation from 36.6 to 17.5 %, demonstrating the presence of both protein and non-specific interactions influencing the initial fluorescence signal. This therefore meant that the method, in its current state of optimisation was not applicable for binding studies. If this study were repeated in the future, purification of the protein into PBS-T, to reduce residual Tris contamination (inhibits reaction) would be optimal. Furthermore, additional purification steps such as size exclusion chromatography and isolation of the dimer species may help reduce aggregation in the sample, which may reduce signal fluctuation. Moreover, a *Dt*BzaC-fluorescent protein fusion could be produced to remove the requirement for the dye or another dye type could be tried (e.g. an N-hydroxysuccinimide (NHS) ester).

(A)



(B)

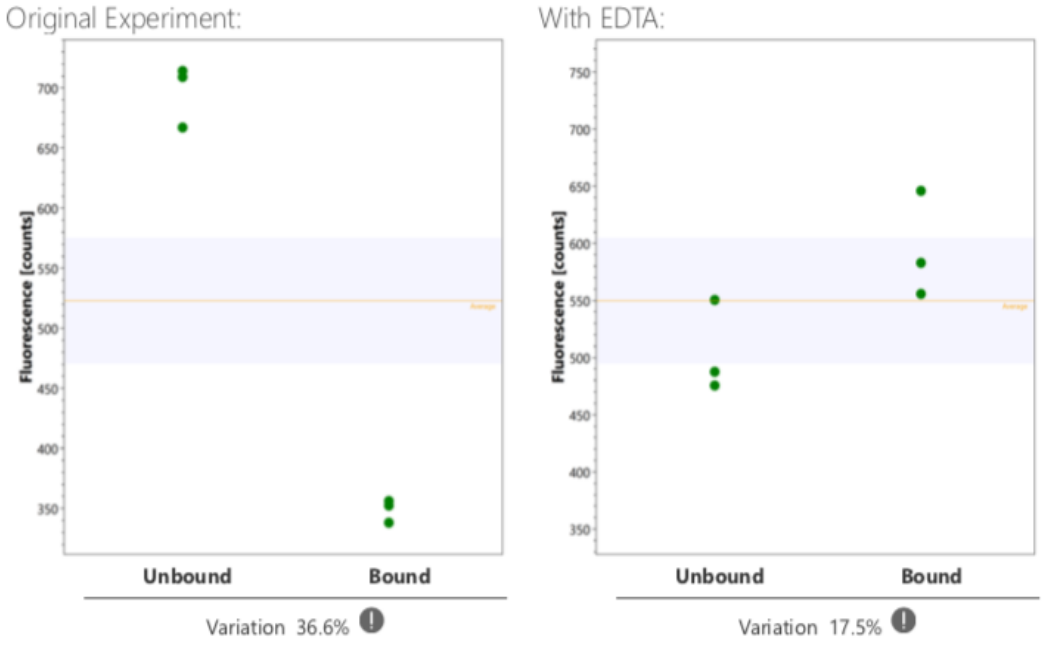


Figure 4.4 - MST troubleshooting.

(A) [*Dt*BzaC] dependent change in initial fluorescence of the Red-Tris-NTA 2nd generation dye. **(B)** Results of the *Dt*BzaC - Red-Tris-NTA 2nd generation dye initial fluorescence experiment in the absence (left) and presence (right) of EDTA

4.2.1.3 Differential Scanning Fluorimetry (DSF)

DSF was tested for its applicability for assessing BzaC ligand binding. This technique relies upon recording the increase in thermal stability observed for a protein when bound to its ligand(s). To do this, a thermal gradient (typically 25-100 °C) is applied to a protein, causing it to unfold. A dye then binds to the hydrophobic core of the protein, inducing a fluorescence signal which correlates to the un-folding of the protein. This can then be plotted and the transition temperature (T_m), the temperature at which 50 % of the protein is un-folded, can be determined. When carried out in a qPCR machine, a 96 well plate can be utilised to carry out this thermal melt process for apo protein and with a series of ligand concentrations, typically showing an increase in the T_m upon ligand-protein complex formation. This can then be utilised to obtain a K_d value for the ligand affinity by fitting the data to binding models.

Firstly, 50 µL reactions, in triplicate, containing 4 µM *Dt*BzaC or *Mt*BzaC, 10 X SYPRO™ orange and a series of SAM, SAH and MTA concentrations in 50 mM Tris-HCL pH 7.5, 100 mM NaCL were set up and incubated at RT for 10 minutes to reach equilibrium. Reactions were then exposed to a 40 – 95 °C gradient over 18.3 minutes recording signal at 558 ± 11 nm using 520 ± 10 nm excitation energy. This generated SAM and SAH concentration dependent increases in T_m (Figure 4.5), which could be used to obtain a K_d value, whereas no significant increase was observed for MTA.

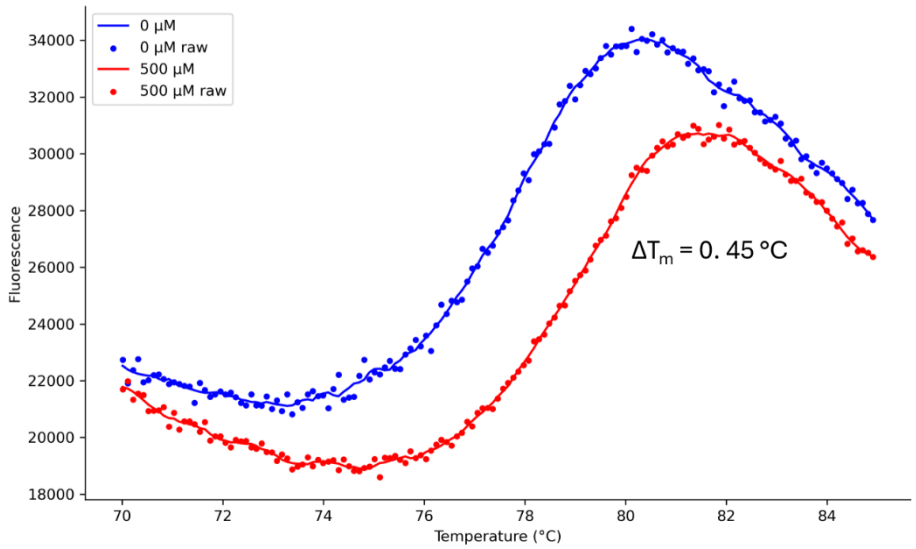


Figure 4.5 - *Dt*BzaC, SAM DSF melt curves.

Thermal unfolding curves of 4 µM Apo *Dt*BzaC (blue) and 4 µM *Dt*BzaC with 500 µM SAM (red), displaying a ΔT_m of 0.45 °C.

However, one limitation of the technique is that to obtain a true K_d value, the system should be at thermal equilibrium, but the ΔT_m used for curve fitting varies in temperature. Therefore, standard nonlinear regression curve fitting of ΔT_m vs [ligand] only produces $K_{d\ apparent}$ values, which tend to be much larger than the actual value (**Figure 4.6**).

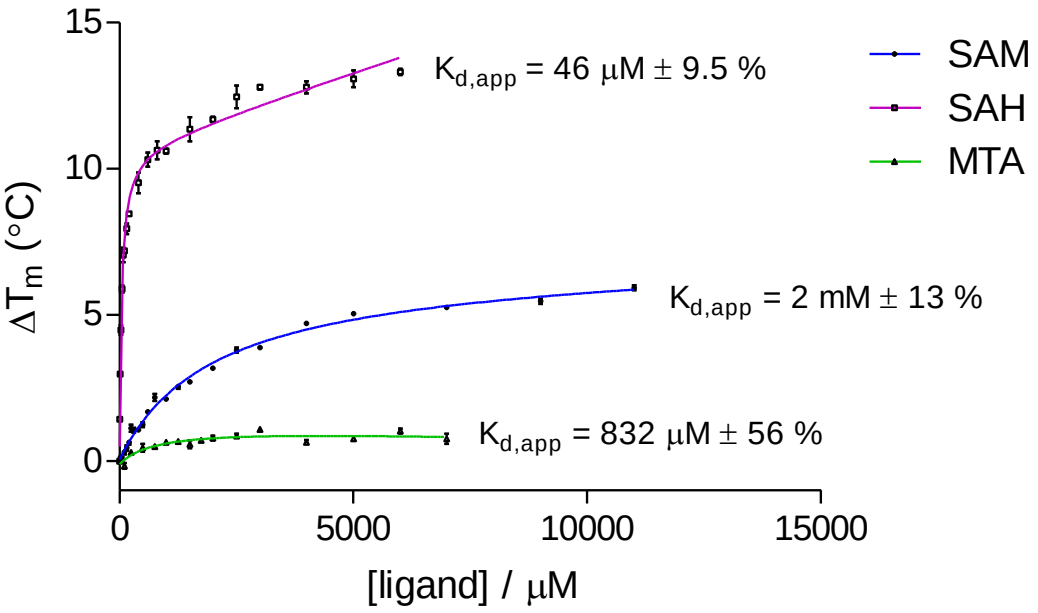


Figure 4.6 - *Dt*BzaC DSF with SAM, SAH and MTA standard nonlinear regression curve fitting for $K_{d\ apparent}$ values.

Error bars represent standard error of the mean (SEM) (n=3).

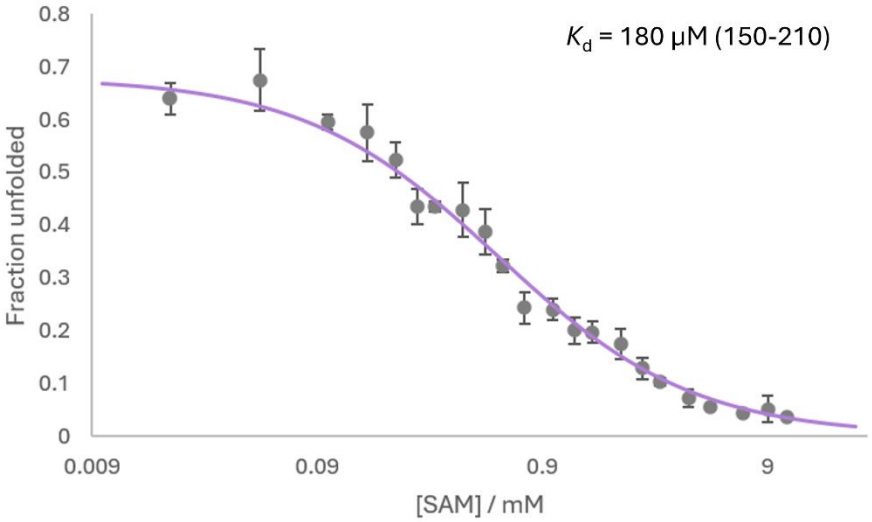
To overcome this limitation the FoldAffinity webserver (60) was used to analyse the data. This uses an isothermal approach utilising the fraction of protein folded at a constant temperature (just above apo protein T_m) at each ligand concentration to calculate K_d values, which correlates much more strongly with other robust experimental techniques such as isothermal titration calorimetry (ITC) or microscale thermophoresis (MST).

Firstly, fluorescence melt curves are generated from the raw data by fitting to a two-state folding model. These are then used to obtain the observed unfolding constant ($K_{u,Obs}$), at a specified temperature, which in turn can be used to calculate the fraction

of protein unfolded at each ligand concentration. This can then be used for isothermal fitting to obtain a K_d value for the system. This produced isothermal curves for *Dt*BzaC (79 °C) and *Mt*BzaC (73 °C) with SAM and SAH but not with MTA, obtaining K_d values of 180 and 82 μM, respectively for SAM and 1 and 3.2 μM for SAH (**Figure 4.7**). This demonstrates that for both *Dt*BzaC and *Mt*BzaC SAH binds much more tightly than SAM (180 x and 21 x, respectively) whereas MTA did not bind at all, thus demonstrating that the actual inhibitor of BzaC activity is SAH rather than MTA.

Furthermore, the SAM affinity is lower than that has been observed for two other cobalamin methyltransferases *Cp*CobM (1.4 μM) (87) and CobA (K_m = 6.3 μM) (88) and higher than CbiL (671 μM) (89) and the previously recorded value for *Mt*BzaC (505.5 μM) (55). If these enzymes were to be used in metabolic engineering experiments in the future, such as pathway heterologous expression, SAM biosynthesis at the mid-high μM range may be required for good BzaC activity. Furthermore, the assay could be used to assess the effects of different mutations on nucleotide binding for enzyme engineering experiments. This could be useful for improving BzaC's ability to utilise non-natural SAM molecules, such as the allyl or propargyl-SAM used in chapter 3.

(A)



(B)

Enzyme	Ligand	K_d (µM)	95 % CI (µM)
*Dt*BzaC	SAM	180	150 - 210
	SAH	1	0.61 - 1.7
	MTA	-	-
*Mt*BzaC	SAM	82	73 - 93
	SAH	3.9	4.3 - 4.4
	MTA	-	-

Figure 4.7 - BzaC nucleotide binding analysis by DSF.

(A) Isothermal curve fitting of *Dt*BzaC with a range of [SAM] (20 µM – 11 mM) at 79 °C.

(B) Table representing the K_d values for *Dt*BzaC and *Mt*BzaC with SAM, SAH and MTA.

In addition, we wanted to test the applicability of the technique for monitoring cobamide binding. Therefore, reactions containing 4 µM *Dt*BzaC, 10 X SYPRO™ orange, 0 - 250 µM Ado[5(6)-OHBza]Cba were set up in 50 mM Tris-HCL pH 7.5, 100 mM NaCl and exposed to a thermal gradient. However, a large decrease in signal was observed with increasing Ado[5(6)-OHBza]Cba concentration, potentially due to an inner filter

effect as there is a concentration dependent loss of initial signal (**Figure 4.8**, **A**). This made ΔT_m calculations unreliable as Boltzmann curve fitting was producing absent or low R^2 values and isothermal fitting was not very accurate and showed no binding.

To overcome this, assays were repeated trying a Nile Red dye; 4 µM *Dt*BzaC, 80 µM Nile Red, 0- 50 µM Ado[5(6)-OHBza]Cba in 50 mM Tris-HCL pH 8, 100 mM NaCL. This dye was chosen due to having an excitation / emission spectrum compatible with using the QuantStudio filters of 580±10 excitation and 623±14 nm emission, which should produce less inner filter effect due to Ado[5(6)-OHBza]Cba. However, a concentration dependent loss of signal was observed. Furthermore, due to there being a large amount of signal, but a reduction in melt curve signal, it could suggest that when the substrate-protein complex forms it is able to withstand the thermal gradient (producing a T_m greater than 95°C), leading to the apparent lack of increase in T_m and a reduction in the melting curve signal, rather than an 'inner filter effect' based loss of signal. This is consistent with the observation that there is not a concentration dependent loss of initial fluorescence (**Figure 4.8, B**). However, it means the technique is not applicable for studying BzaC - cobamide binding.

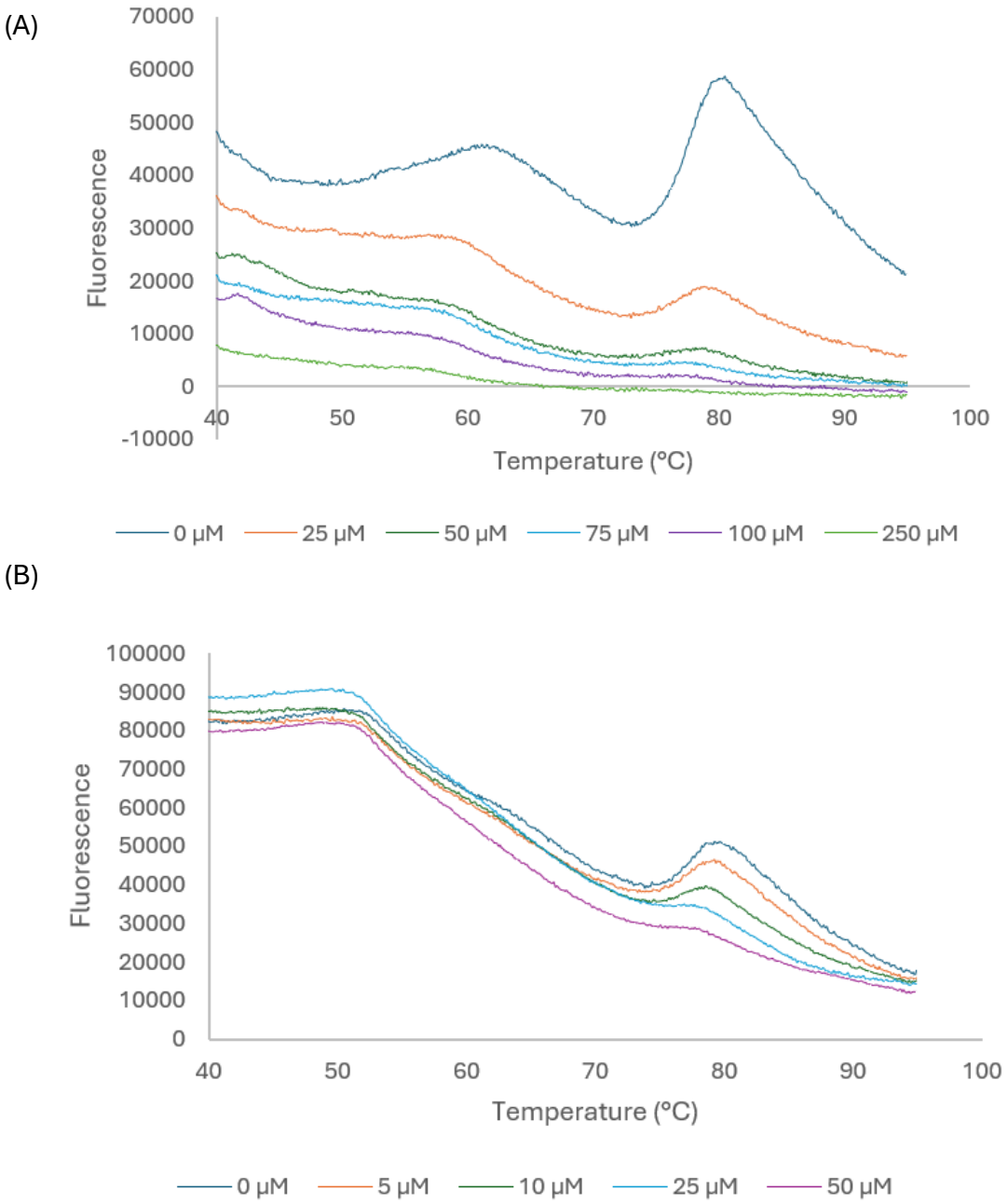


Figure 4.8 - *Dt*BzaC, Ado[5(6)-OHBza]Cba DSF.

(A) SYPRO™ orange *Dt*BzaC, Ado[5(6)-OHBza]Cba assay. **(B)** Nile Red *Dt*BzaC, Ado[5(6)-OHBza]Cba DSF assay.

4.2.2 *Dt*BzaC kinetics

4.2.2.1 Colourimetric, linked assay:

To provide further molecular detail on the reaction catalysed by BzaC we attempted to obtain values for the kinetic parameters K_m and k_{cat} for each substrate. This is important information for future engineering experiments, comparison to other known methyltransferases, studying inhibitors, and comparison between each putative substrate (5-OHBza, 5-OHBza-R and Ado[5(6)-OHBza]Cba. Firstly, we needed to find a suitable experimental set up for recording reaction rates. Due to there being no difference in the UV-VIS spectrum of the substrate and product that could be easily monitored we tried to set up a continuous, coupled colourimetric assay, where BzaC activity was linked to the production of a fluorescent signal (90).

This works by the product of the BzaC reaction, SAH, being hydrolysed by *Ec*MTAN into adenine and S-ribosyl-l-homocysteine. Adenine undergoes two sequential transformations catalysed by Xanthine Oxidase (XOD), to form 8-hydroxyadenine and 2,8-hydroxyadenine, respectively, and hydrogen peroxide (H_2O_2). Horse Radish Peroxidase (HRP) then utilises the H_2O_2 to shuttle electrons onto Amplifu Red, producing the fluorescent molecule Resorufin (**Figure 4.9**). This signal can then be monitored and plotted to calculate the reaction rates. This setup is useful due to the ability to carry out small scale reactions (50-100 µL) in a 96 well plate, recording the data in a plate reader allowing for high throughput analysis of a range of conditions such as multiple substrates, or inhibitors.

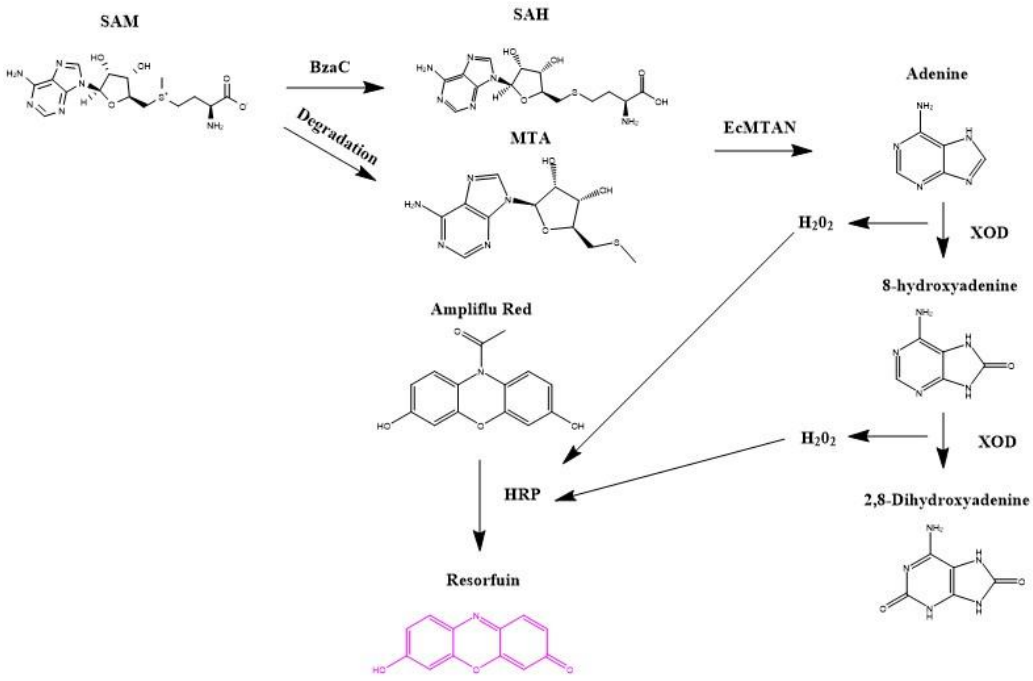


Figure 4.9 - Continuous, colourimetric assay schematic.

Schematic representation of the pathway linking *Dt*BzaC activity (SAH production) to the generation of a fluorescent signal (resorufin).

To test the applicability of the technique for monitoring BzaC, 50 µL reactions (in duplicate) containing 5 µM *Dt*BzaC, 25 µM Ado[5(6)-OHBza]Cba, 50 µM *Ec*MTAN, 0.092 units XOD, 0.022 units HRP, 0.1 mM Amplifu Red, 5 mM $MgCl_2$, and a range of SAM concentrations (0, 50, 125, 250, 500, 1000 µM) in 20 mM Tris-HCL pH 8, 100 mM NaCl were produced. Samples were incubated at 28 °C, for 45 minutes before initiation of reactions via injection of *Dt*BzaC. Signal was then recorded at 544 nm excitation, 620-10 nm emission every 1.5 minutes for 60 minutes. Due to *Ec*MTAN also hydrolysing MTA (produced from SAM degradation) to adenine, reactions lacking *Dt*BzaC at each concentration of SAM were produced to account for this signal in *Dt*BzaC reactions.

This demonstrated a SAM concentration dependent increase in rate over 30 minutes. However, there is a lag in signal production over the first 10 minutes of the reaction, which is likely attributed to by a lag in the reporter system (**Figure 4.10, A**). As we want to obtain initial rates to assess the enzyme kinetics in a non-limited system, we needed to overcome this.

To investigate the source of the lag, reactions with various concentrations of SAH were set up and analysed similarly to above. This demonstrated a SAH concentration dependent increase in signal, however the rate of reaction was significantly greater than with SAM, as expected. This shows that the report system is working faster than BzaC, so should not influence the signal observed. However, at 5 µM SAH, there was a lag in signal production for the first 5 minutes, demonstrating that the lag in the system originates from a low concentration of SAH not generating enough signal over the noise. As there was only 5 µM *Dt*BzaC in the SAM assay, < 5 µM is likely to be generated in the first 10 minutes, generating the lag observed.

To try an overcome this lag, variations in the concentration of the components of the reporter system were attempted. These included increasing / decreasing the concentration of HRP, XOD and *Ec*MTAN, in various combinations, changing the temperature (28 / 37 °C), type and concentration of SAM (di-HCL / p-toluenesulfonate salt, 0.5 / 1 mM), reaction volume (50 / 100 µL) and buffer/pH (20 mM Tris-HCL pH 8, 100 mM NaCL / 50 mM Tris-HCL pH 7.5, 100 mM NaCL) to try and make the reporter system more sensitive to produce stronger signals with lower concentrations of SAH. However, none of these attempts managed to remove the lag in the system.

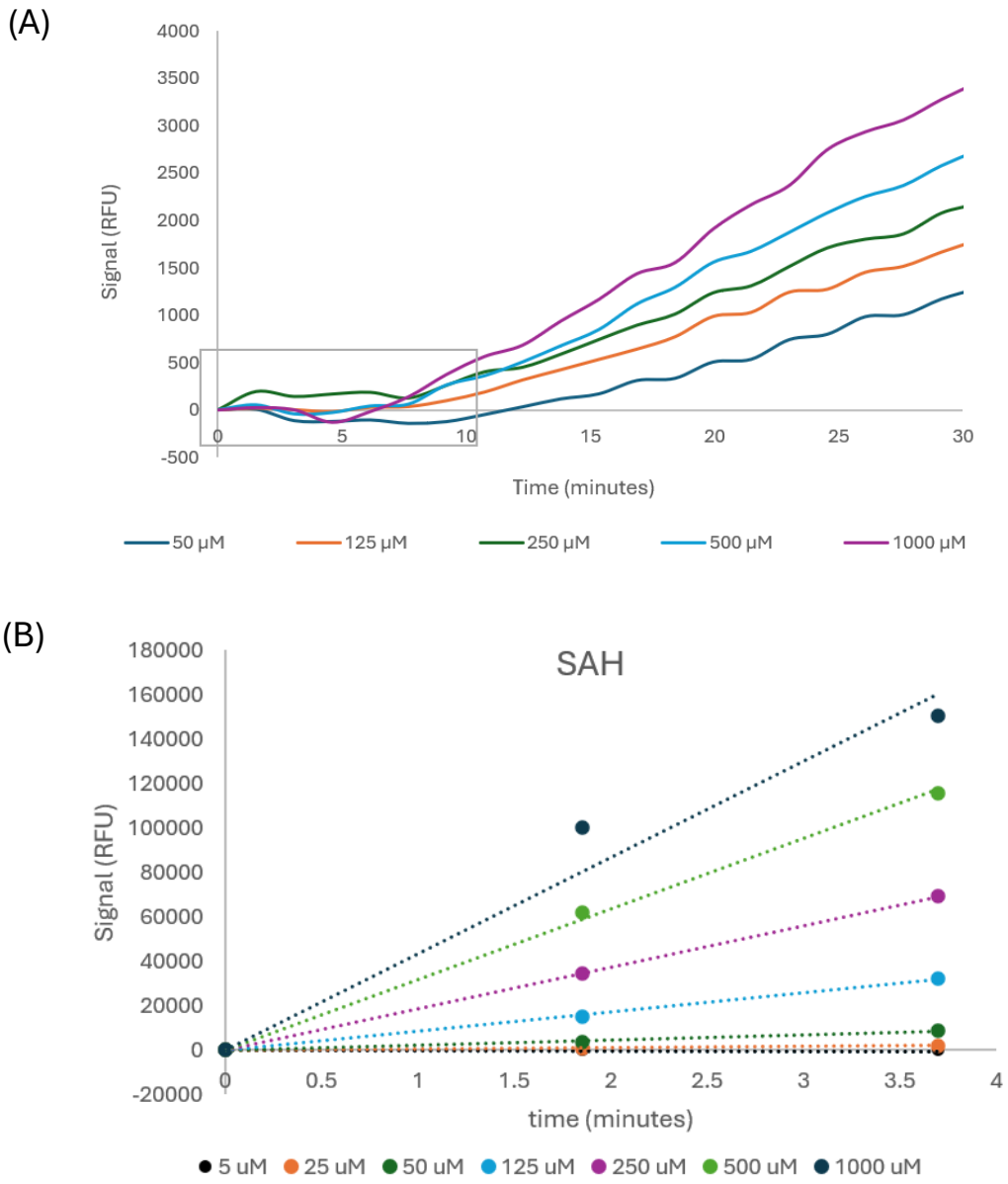


Figure 4.10 - Continuous assay tests.

(A) Graph showing the fluorescent signal over time for 50, 125, 250, 500 and 1000 µM SAM, with the system lag indicated. **(B)** Scatter graph with lines of best fit showing the fluorescent signal over time for 5, 25, 50, 125, 250, 500 and 1000 µM SAH in the absence of BzaC.

4.2.2.2 Stopped, HPLC assay:

Due to the inability to remove the lag in the continuous assay we tried using a stopped assay analysed by HPLC. This works by setting up the reactions and quenching them at various timepoints and quantifying the amount of product formed by HPLC. This technique is less high throughput and requires more reagents, limiting its ease of use. Firstly, to test the feasibility of the technique and the linear reaction time period (non-limiting) 100 µL reactions containing 1 µM *Dt*BzaC, 10 mM SAM, 50 µM Ado[5(6)-OHBza]Cba, 50 µM *Ec*MTAN in 50 mM Tris-HCL pH 8, 100 mM NaCl were set up and stopped at 0, 2.5, 5, 10, 15, 20, 30, 40 and 50 minutes using a final concentration of 2 % acetic acid glacial. Reactions were then analysed by HPLC, and the integrated area of the product peak was plotted against time. This demonstrated an increase in product produced over time, with the linear period of the reaction occurring between 0 and 15 minutes (**Figure 4.11**).

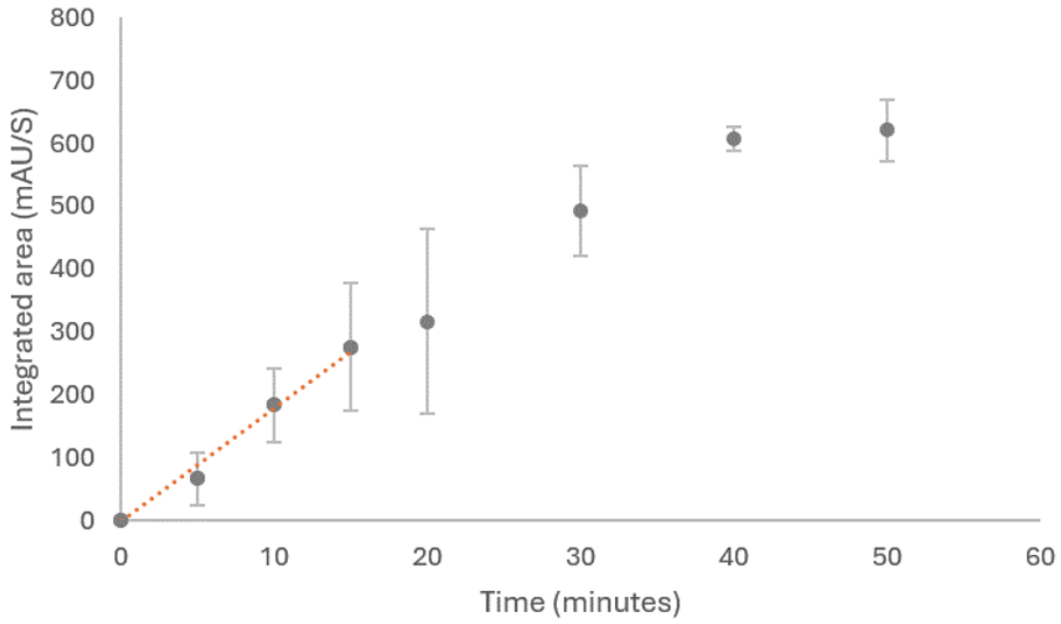


Figure 4.11 - *Dt*BzaC reaction time course.

A scatter graph representing the integrated area of the product peak identified by HPLC over time. Trendline (orange) of the linear reaction time (0-15 minutes) with an R^2 value of 0.9946.

Now that the non-limiting time period had been identified, 100 µL reactions, in duplicate, containing 1 µM *Dt*BzaC, 1 mM SAM, 50 µM EcMTAN, and 0, 12.5, 25, 100, 200 and 400 µM Ado[5(6)-OHBza]Cba in 20 mM Tris-HCL pH 8, 100 mM NaCL were assayed, quenching at 0, 7.5 and 15 minutes. Reactions were then analysed by HPLC (method 4), quantified using a standard curve and the amount of product formed (Ado[5(6)-OMeBza]Cba) was plotted against time. This identified an Ado[5(6)-OHBza]Cba concentration dependent increase in rate of product formation. Furthermore, the rate of product formation (nM/min) was plotted against [Ado[5(6)-OHBza]Cba] (µM) and nonlinear regression curve fitting to a Michaelis Menten enzyme

kinetics equation (Y = V_{max}*X/(K_m + X)) was performed. This identified the K_m to be 11 ± 3.7 µM, V_{max} to be 252.1 ± 15.7 nM/min, k_{cat} of 0.25 min^{-1} and a catalytic efficiency (k_{cat}/K_m) of 0.023 $min^{-1}·µM^{-1}$.

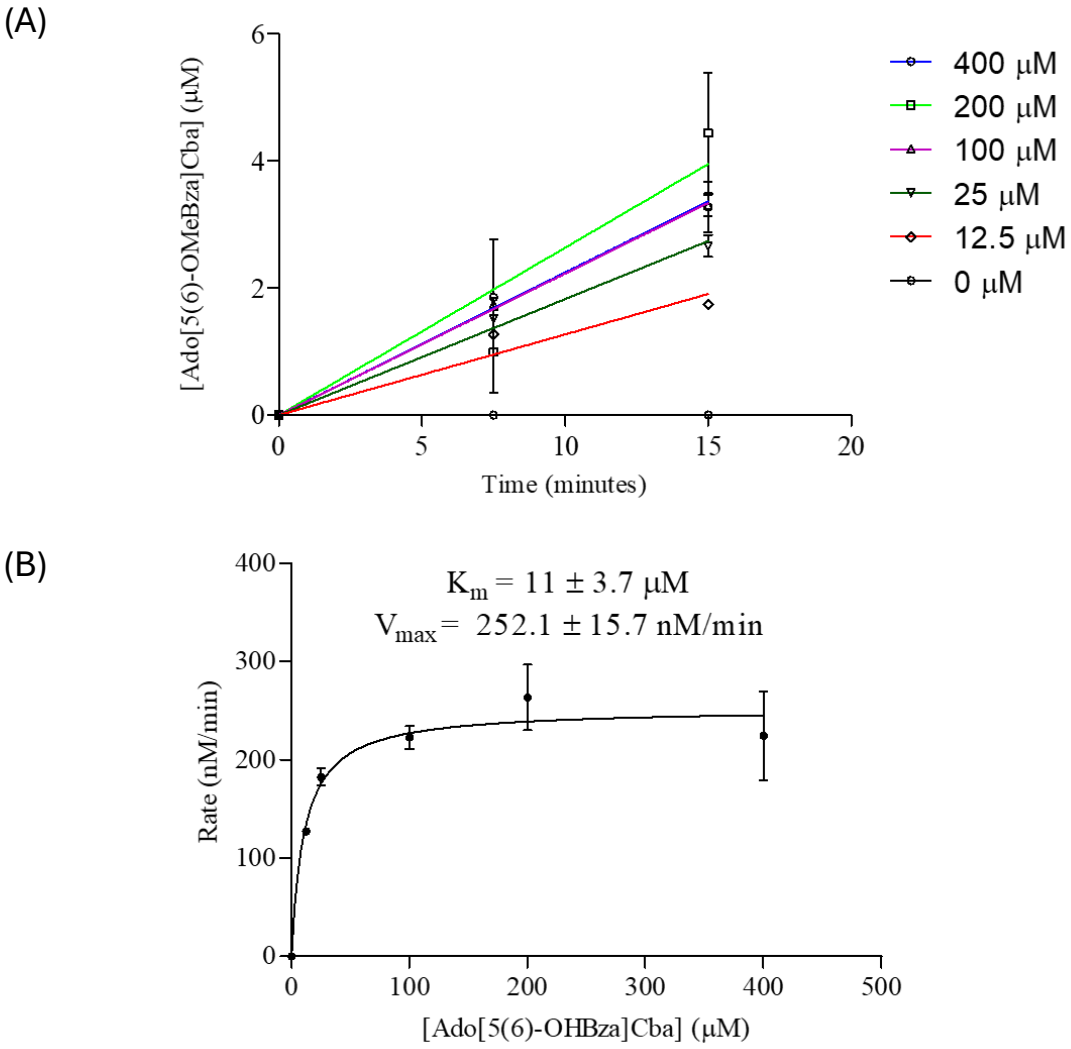


Figure 4.12 - *Dt*BzaC Ado[5(6)-OHBza]Cba kinetics

(A) Scatter graph of [Ado[5(6)-OMeBza]Cba (µM) formed against time (minutes) with error bars representing SEM. **(B)** Nonlinear regression curve of rate (nM/min) against [Ado[5(6)-OHBza]Cba] fitting to the enzyme kinetics model Y = V_{max}*X/(K_m + X) to obtain K_m and V_{max} values.

To study the effect of SAM concentration on the reaction kinetics and to confirm that the 1 mM SAM used for analysing *Dt*BzaC, Ado[5(6)-OHBza]Cba kinetics was not rate-limiting, 100 µL reactions (in duplicate) containing 1 µM *Dt*BzaC, 75 µM Ado[5(6)-OHBza]Cba, 50 µM *Ec*MTAN, and a 6 point serial dilution of SAM from 2400 – 37.5 µM, in 20 mM Tris-HCL pH 8, 100 mM NaCl were assayed, quenching at 0, 7.5 and 15 minutes. Reactions were then analysed by HPLC (method 5), quantified using a standard curve and plotted (**Figure 4.13**). The data were analysed as above determining the K_m to be 12.6 ± 2.3 µM, V_{max} = 257 ± 4.5 nM/min, k_{cat} = 0.26 min^{-1} and a catalytic efficiency of 0.02 $min^{-1} \cdot \mu M^{-1}$.

Together this showed that *Dt*BzaC has a similar affinity for both its substrates, 11 and 12.6 µM for Ado[5(6)-OHBza]Cba and SAM, respectively. Furthermore, they both had the same V_{max} suggesting that neither concentration of substrate (75 µM Ado[5(6)-OHBza]Cba and 1 mM SAM) were limiting reaction rate, which is expected as they are both > 5 X their respective K_ms. Furthermore, these values are like that of CobA (uroporphyrinogen III methyltransferase involved in vitamin B_{12} biosynthesis) SAM = 6.3 µM, uroporphyrinogen III = 1 µM and k_{cat} of 38 h^{-1} (*Dt*BzaC = 15 - 15.6 h^{-1}) (88) and ~ 10-fold tighter than that of DsyB (4-methylthio-2-hydroxybutyrate = 140 µM, SAM = 160 µM) (91) which is the highest homology protein in 3D fold identified by PDBeFold (discussed more in chapter 5). Furthermore, the K_m of 12.6 µM for SAM is > 10-fold tighter affinity than the K_d of 180 µM determined by DSF and correlates more closely with the K_d / K_m values determined for CobM and CobA, which may suggest the DSF values maybe higher than they are, or substrate binding may enhance co-factor binding.

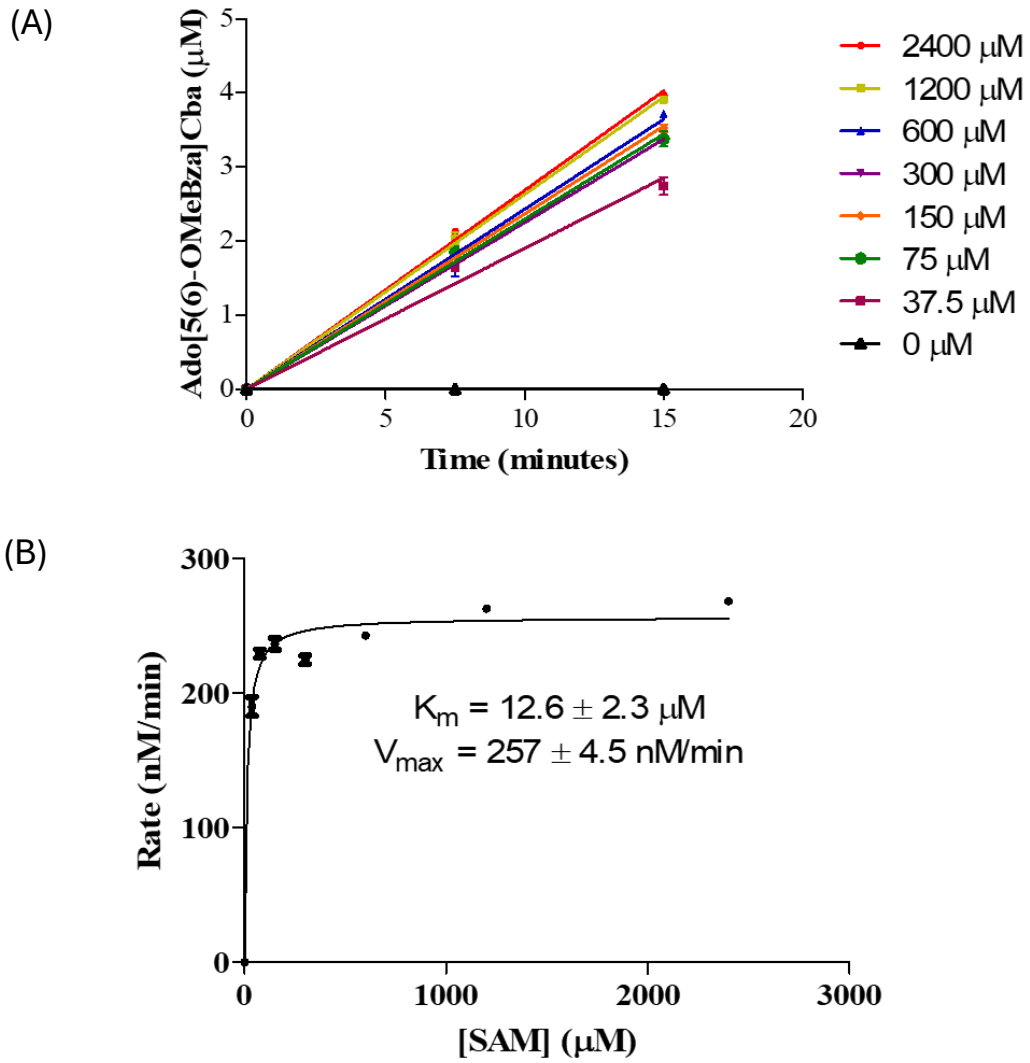


Figure 4.13 - *Dt*BzaC, SAM kinetics.

(A) Scatter graph of [Ado[5(6)-OMeBza]Cba (μM) formed against time (minutes) with error bars representing standard error of the mean (SEM). **(B)** Nonlinear regression curve of rate (nM/min) against [SAM] fitting to the enzyme kinetics model Y = V_{max}*X/(K_m + X) to obtain K_m and V_{max} values.

4.2.3 BzaC regiospecificity

At the end of chapter 3 the possibility of our method of substrate production to make two isomers of cobamides, e.g. Ado[5 and 6-OHBza]Cba was mentioned. This is due to ED674 containing two CobT/U's; *E. coli* CobT (*Ec*CobT, wild type under native promoter) and *Rhodobacter capsulatus* CobU (*Rc*CobU, overexpressed) to allow for the activation of our supplemented benzimidazoles and subsequent attachment to AdoCbi-GDP. Previous literature has demonstrated that *Ec*CobT can make both 5- and 6-OHBza-RP from 5-OHBza (55) showing that the strain could potentially make both isomers. Furthermore, a *Rc*CobU reaction with 5-OHBza and NMN was set up and identified the production of both 5- and 6-OHBza-RP via HPLC (method 6) (**Figure 4.14**), further demonstrating the potential ability of the strain to produce 5- and 6-cobamide isomers, depending on the activity of CobV/S.

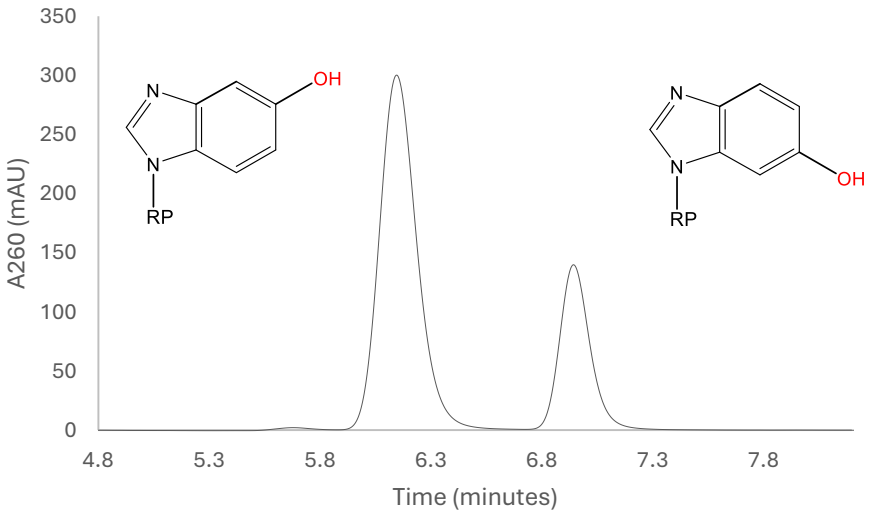
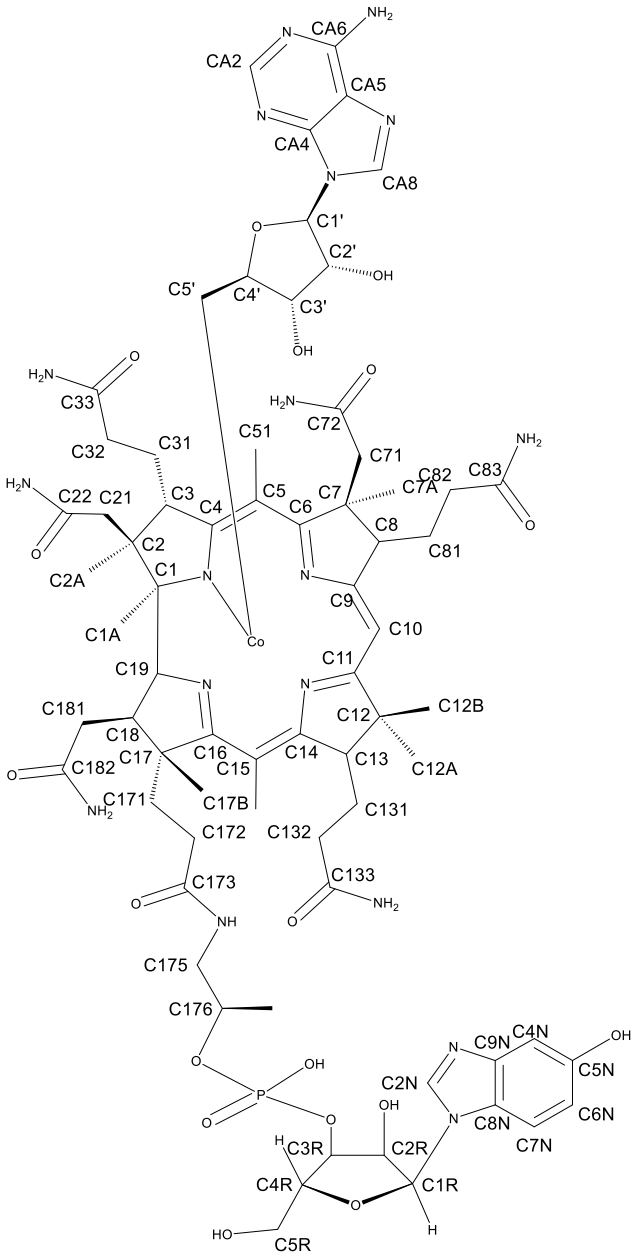


Figure 4.14 - *Rc*CobU assay.

A HPLC chromatogram of a *Rc*CobU reaction with 5-OHBza and NMN producing both 5- and 6-OHBza-RP as products.

To investigate this further we carried out NMR on Ado[5(6)-OHBza]Cba to assess the presence and levels of isomers present in the substrate preparation. To do this ~ 1 mM Ado[5(6)-OHBza]Cba in 10 mM Kpi in D_2O pD 7.4 was subjected to 1D ^{1}H (with Watergate W5 water suppression), ^{1}H-^{13}C HSQC, ^{1}H-^{13}C HMBC, 300ms ROESY and 80ms TOCSY analysis using a Bruker 700 MHz prodigy NMR spectrometer. The Collaborative Computational Project for NMR (CCPN) (70), TopSpin and ACD/Labs software packages were then used to analyse the spectra and make each resonance assignment by Dr Andrew Lawrence and Emily Jones. Based on the position of the protons on the benzimidazole and the carbon/proton shifts the relative abundance of each isomer was estimated via integration of the signals (~ 15 % 5- and 85 % 6-). This was an unexpected result as looking at the *Ec*CobT and *Rc*CobU data it was expected that a ~ 1:1 ratio of isomers would be present, depending on relative expression and activity levels in the strain. However, it suggests that under the conditions in the strain that either the CobT/U's behave differently or CobV/S has a preference for the 6- isomer.

(A)



(B)

	Assignment	δ(^{13}C) [PPM]	δ(^{1}H) [PPM]
Adenosyl	C1'	87.781	5.478
	C2'	72.388	4.438
	C3'	73.255	3.653
	C4'	85.439	2.362
	C5'	23.833 / 23.822 / 23.653	0.458 / 1.372 /1.376
	CA2	152.889	9.138
	CA8	152.889 / 140.664	9.138 / 7.917
	C1A	20.61	0.381
	C2A	16.606	1.247

Corrin ring and aminopropanol linker	C21	43.093 / 43.088	2.357 / 2.305
	C3	55.136	3.994
	C31	25.757	1.901
	C32	34.871	2.427
	C51	15.549	2.337
	C7A	18.472	1.624
	C71	42.221	1.662
	C81	25.944 / 25.944	1.73 / 0.866
	C82	31.117 / 31.71	0.774 / 1.701
	C10	94.881	5.952
	C12A	20.714	1.277
	C12B	31.117	0.774
	C13	52.673	2.834
	C131	27.215 /27.244	1.961 / 2.003
	C132	34.871	2.427
	C151	15.273	2.325
	C17B	16.71	1.276
	C171	31.471 / 31.443	2.341 / 1.963
	C172	31.71	1.701
	C18	39.457	2.561
	C181	31.634 / 31.641	2.555 / 2.435
	C19	73.762	4.167
	C175	44.656 / 44.61	3.056 / 3.44
	C176	72.845	4.244
	C177	18.617	1.123
Ribose	C1R (6-isomer)	86.273	6.097
	C1R (5- isomer)	86.429	6.159
	C2R	68.941	4.194
	C3R	73.005	4.623
	C4R	81.525	4.018
	C5R	60.189 / 59.852	3.789 / 3.62
6-isomer benzimidazole	C2N	141.478	6.903
	C4N	111.729	6.603
	C5N	119.225	6.302
	C7N	96.034	6.699
5-isomer benzimidazole	C2N	142.561	6.949
	C4N	104.1	5.9.19
	C6N	113.225	6.694
	C7N	110.974	7.149

(C)

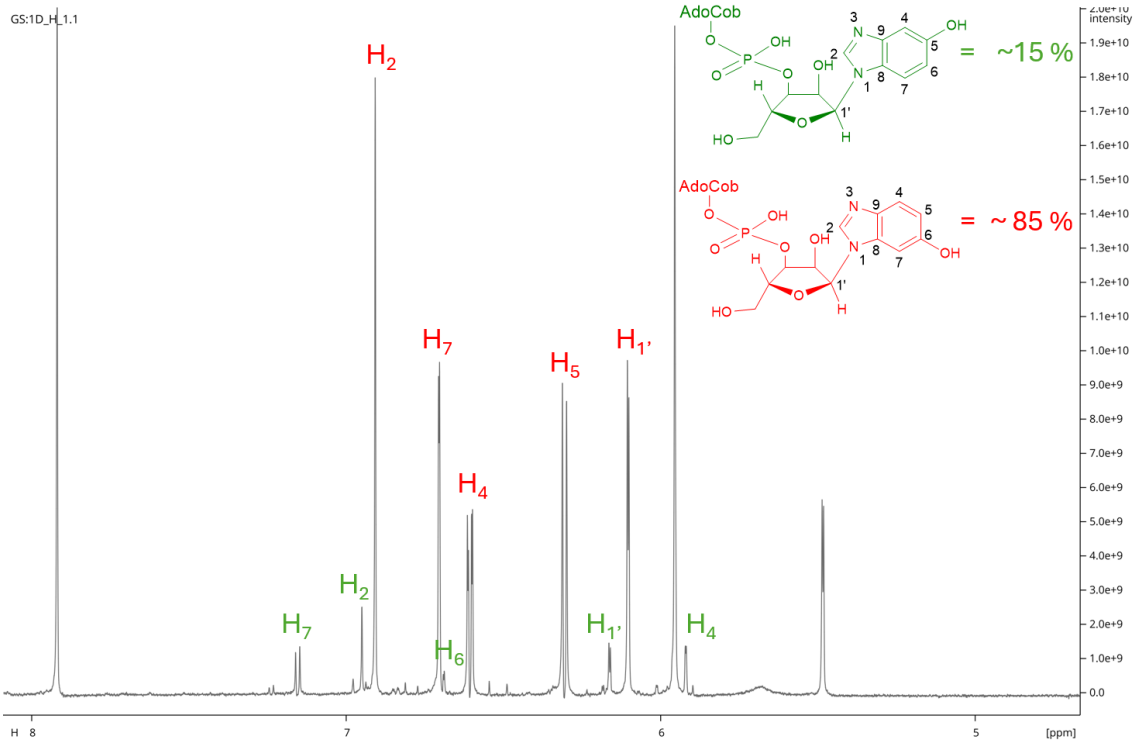


Figure 4.15 - Ado[5(6)-OHBza]Cba NMR analysis.

(A) Ado[5-OHBza]Cba with carbons labelled. **(B)** Table of carbon atom assignments with their respective ^{13}C and ^{1}H shifts (ppm). **(C)** 1D ^{1}H NMR spectra between 5-8 ppm showing the assigned benzimidazole protons for both the 5- and 6- isomers.

Next, we want to investigate potential regioselectivity displayed for each BzaC with the cobamide substrates. To do this we needed to analyse the isomer profiles of the products of BzaC reactions, however due to the large amounts of product required for NMR a smaller scale method was required. As demonstrated by T. Crofts et al. 2014 (92) and Mathur et al. 2020 (55) the benzimidazole ribotide isomers can be separated by HPLC, which cannot be achieved for the cobamides. Therefore, if the nucleotide loop of the cobamide could be cleaved, the resulting benzimidazole ribotide could be purified and isomers could be separated via HPLC. Furthermore, each isomer has a

distinct UV-VIS spectra which makes for simple identification of each respective species.

Previous literature has demonstrated a chemical method of cleaving the nucleotide loop of cobamides utilising alkaline hydrolysis (69). Firstly, the cobamide is exposed to 2.5 M NaOH and heated at 90 °C for 75 minutes, allowing for the alkaline hydrolysis of the 1-amino-propanol 2OP – ribose phosphodiester bond of the nucleotide loop in a nucleophilic acyl substitution reaction. Next, the hydrolysate is cooled and neutralised with HCL. The ribotides are then de-phosphorylated using Alkaline Phosphatase, removing the 3' phosphate to expose the sugars cis diol to allow for purification using boronic acid chromatography (**Figure 4.16**). The resulting ribosides could then be concentrated under air flow or vacuum centrifugation and analysed by HPLC and LC-MS.

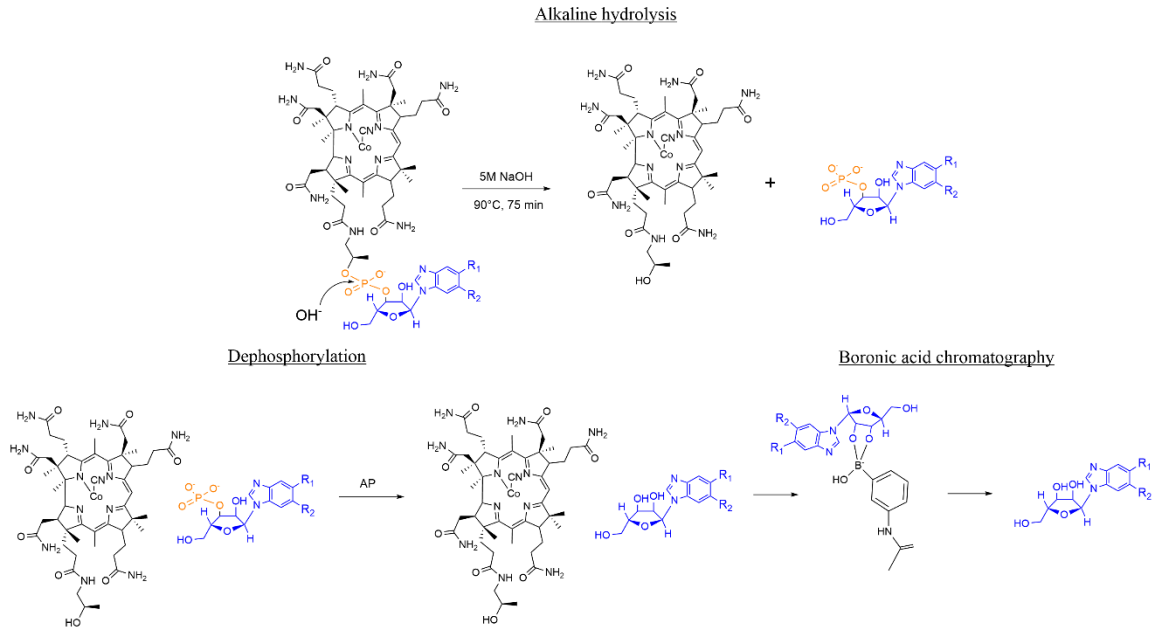


Figure 4.16 - Schematic of the alkaline hydrolysis of the cobamide nucleotide loop and subsequent purification of ribosides.

This technique was then tested using Ado[5(6)-OHBza]Cba. Firstly, 180 µg Ado[5(6)-OHBza]Cba was converted to CN (as the Co-CN bond is less heat liable), purified by RP18 and vacuum dried. CN[5(6)-OHBza]Cba powder was then resuspended in H_2O and subjected to alkaline hydrolysis, the resulting hydrolysate was treated with alkaline phosphatase and the resulting ribosides were purified. Next, the ribosides were separated by HPLC (method 6), identifying two peaks with the respective characteristic UV-VIS spectra for the 5- and 6-OHBza-R. The relative abundance of each species was calculated using the integrated area under the curve, showing 19 % of the preparation was 5- and 81 % was 6- isomer. This was further confirmed by LC-MS which identified both peaks contained a m/z of 267.098 (expected [M+H] = 267.09). These values for the relative abundance of each isomer correlate well with the NMR analysis, demonstrating the applicability of the technique.

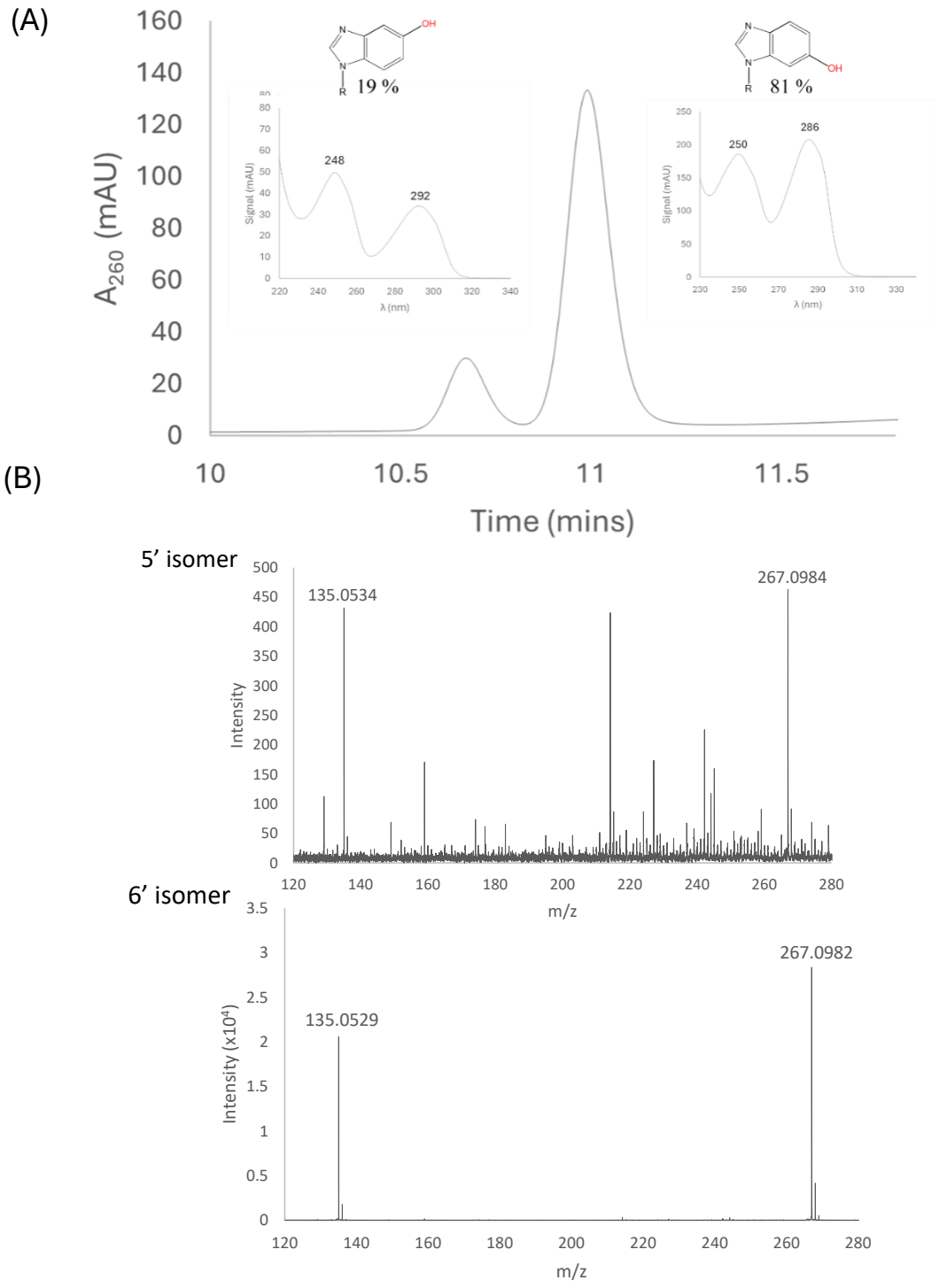


Figure 4.17 - Ado[5(6)-OHBza]Cba isomer profile.

(A) HPLC chromatogram and UV-VIS spectra of 5 and 6-OHBza-R isolated from the alkaline hydrolysis of Ado[5(6)-OHBza]Cba. **(B)** MS spectra of 5- and 6-OHBza-R.

Next, we wanted to use this technique to analyse the isomer profiles of BzaC products, to assess any potential regioselectivity. Firstly, *Mt*BzaC was selected for analysis (~ 40 % conversion in end-point BzaC assays). A larger scale reaction (7.2 mL - to produce enough product to allow for accurate analysis) was set up containing 10 µM *Mt*BzaC, 20 µM Ado[5(6)-OHBza]Cba, 1 mM SAM in 20 mM tris-HCL pH 8, 100 mM NaCl and incubated at 30 °C for 17 hours and 30 minutes. The reaction was then quenched with a final concentration of 1 % glacial acetic acid, and the product was fractionated by HPLC (method 7). The resulting Ado[5(6)-OMeBza]Cba (~ 72 µg) was then converted to CN[5(6)-OMeBza]Cba, purified by RP18 chromatography and vacuum dried. The cobamide was then subjected to alkaline hydrolysis, purified, and analysed by HPLC and LC-MS. This identified two peaks corresponding to 5- and 6-OMeBza-R (m/z = 281.1, expected [M+H] = 281.1) with a relative abundance of 44 % and 56 %, respectively. This demonstrates that *Mt*BzaC does not show complete regiospecificity for one isomer but does suggest that it displays a selectivity for the 5- isomer due to using most of the 5- isomer (88 %) and not much of the available 6- isomer (26 %). This regioselectivity for the less abundant isomer may contribute to why only 40 % of the substrate is converted in the end-point assays, as the weaker activity with the more abundant 6- isomer slows the reaction so only 40 % is converted in the designated time frame or before the protein degrades.

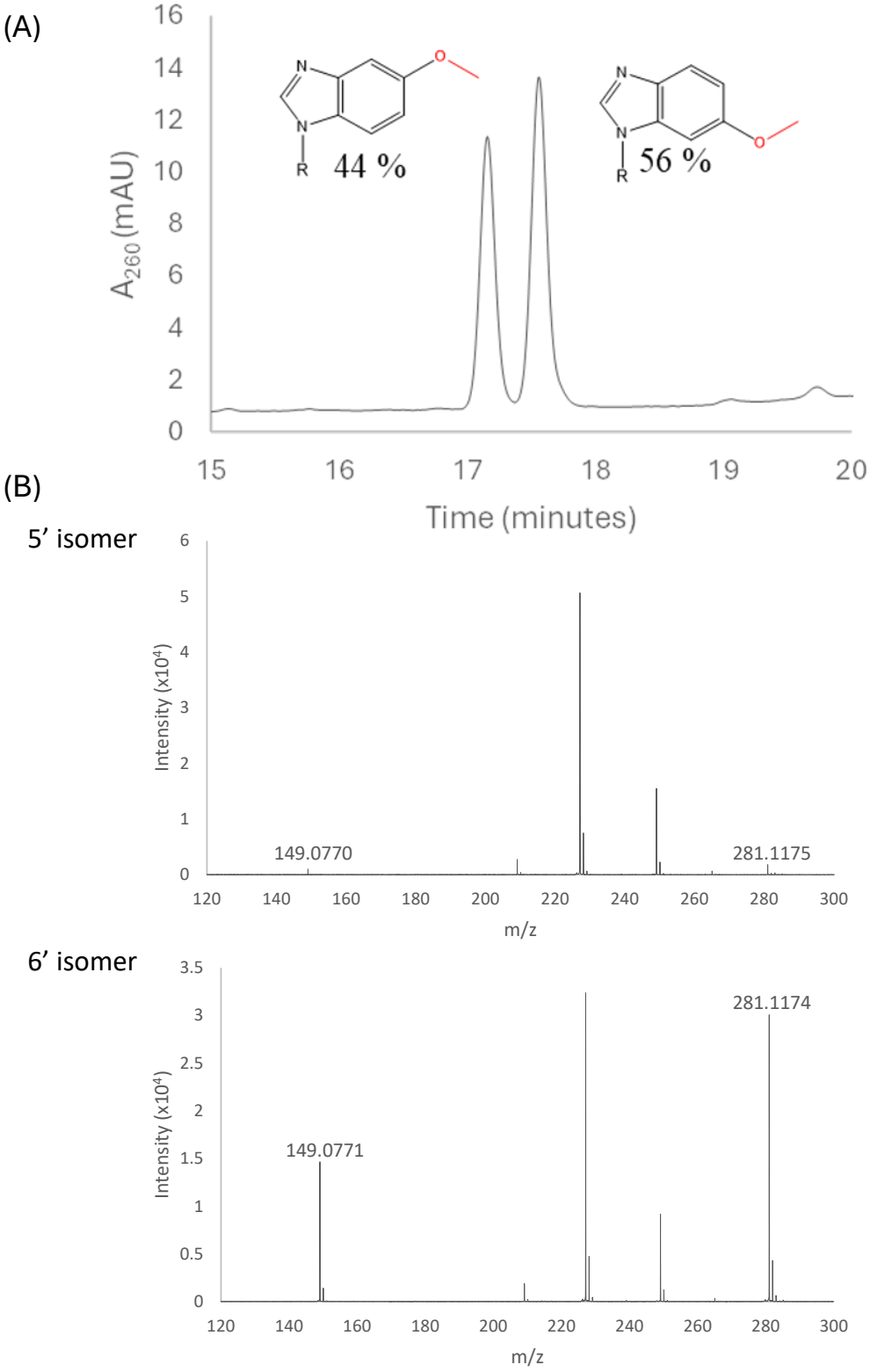


Figure 4.18 - *Mt*BzaC reaction product isomer analysis.

(A) HPLC chromatogram of 5- and 6-OHBza-R resulting from *Mt*BzaC reaction product (Ado[5(6)-OMeBza]Cba) hydrolysis. **(B)** MS spectra of 5- and 6-OHBza-R.

To try and obtain an estimated numerical value for the regioselectivity observed we wanted to use the isomer data and fit it to the *Dt*BzaC enzyme kinetics data. Firstly, to do this the concentration of each isomer at each concentration of substrate used in the kinetics was estimated based off the 19 % 5- and 81 % 6- ratio determined by the hydrolysis assay. For example, 200 µM Ado[5(6)-OHBza]Cba = 40 µM 5- and 160 µM 6- at time t=0. Next nonlinear curve fitting using the excel solver plugin was used to fit the data to a model which represents the relationship between substrate concentrations and overall reaction rates in a competitive system (**Figure 4.19**) allowing estimation of K_m values for each substrate (5- = 1. 9 µM, 6- = 21 µM).

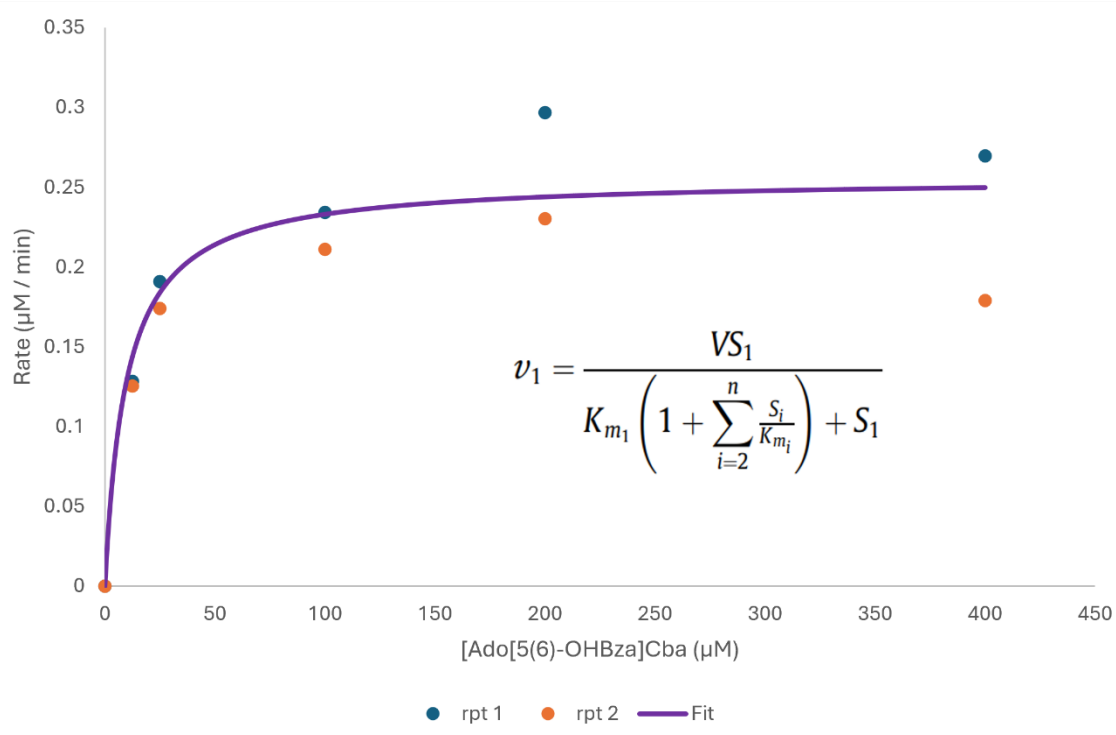


Figure 4.19 - Estimation of substrate isomer affinities.

Scatter graph of rate data obtained from the *Dt*BzaC kinetics experiment and fitted line (estimated rate at each concentration of total substrate (0-400 µM)) with the model used (taken from Schäuble et al 2013 (93)).

The optimised parameters were then used to estimate the reaction rate using each substrate to allow for error, compared to observed total rate, to be calculated (5- = 0.18 μM/min, 6- = 0.061 μM/min). Furthermore, these parameters could then be used to estimate the percentage of product present as each isomer over a series of time points for a given total concentration of substrate. This estimated that in a reaction containing 200 μM Ado[5(6)-OHBza]Cba at 15 minutes 74.6 % would be 5- isomer and 25.4 % would be the 6- isomer (**Figure 4.20**).

To test this a larger scale *Dt*BzaC reaction containing 1 μM *Dt*BzaC, 200 μM Ado[5(6)-OHBza]Cba, 1 mM SAM and 50 μM *Ec*MTAN in 20 mM Tris-HCL pH 8, 100 mM NaCl was produced and incubated at 37 °C for 15 minutes. The reaction was then quenched and subjected to isomer analysis. This identified that 74.6 % of the product was 5- isomer and 25.4 % was the 6-isomer which agrees with the predicted values, thus supporting the K_m values obtained by the model. Overall, this suggests that *Dt*BzaC has ~ 11-fold tighter affinity for the 5- isomer compared to the 6- isomer.

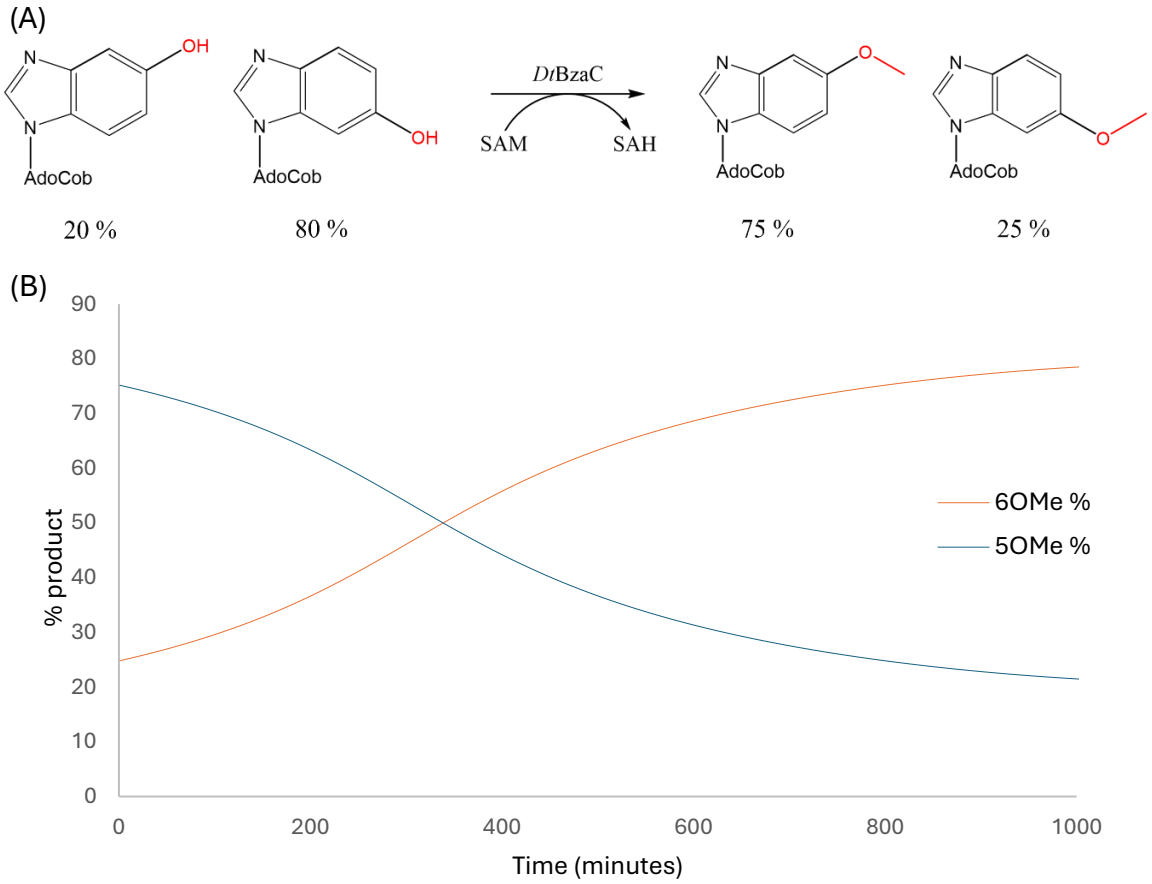


Figure 4.20 - *Dt*BzaC product isomer generation over time.

(A) Schematic representation of the *Dt*BzaC 15-minute reaction with the relative abundance of substrate and product isomers present. **(B)** Graph representing the simulated % of each product isomer present over time.

4.3 Discussion

The data provided in this chapter has explored the specificity of BzaC for a variety of ligands producing both K_d and K_m values to define these affinities. This information and the techniques used to obtain it will be very useful for future engineering efforts, both pathway and enzyme specific engineering. For example, knowing the K_m and k_{cat} values

for each enzyme in the pathway for their substrates will help understand any rate limiting steps and what concentrations of substrates would be required for optimal production.

The DSF data supported the importance of *Ec*MTAN for efficient BzaC activity due to the much tighter affinity for SAH than SAM, therefore incorporation of this in any future engineering efforts would be greatly useful. In addition, the *Dt*BzaC kinetic data aligns well with other known methyltransferases involved in cobamide biosynthesis, supporting BzaC's role in the anaerobic biosynthesis of benzimidazolylcobamides. However, the significant difference between the K_d and K_m values obtained for *Dt*BzaC with SAM (180 vs 11 μM, respectively) questions the accuracy of the K_d values obtained, as the K_m values align with other methyltransferases involved in cobalamin biosynthesis. It either suggests that the DSF assay is not the optimal assay for determining the values, perhaps due it being a linked assay which requires an exogenous dye (SYPRO™ orange) to link the proteins thermal stability to co-factor affinity. Or substrate binding causes active site changes which promotes co-factor binding, which would explain the tighter affinity demonstrated in the kinetics experiment where both substrate and co-factor are present. However, the DSF experiment is still useful for comparing the three nucleosides (SAM, SAH and MTA) as even if the K_d values aren't exactly right, they are relative to each other.

To test the validity of the DSF data a label free approach such as Isothermal Titration Calorimetry (ITC) or Surface Plasmon Resonance (SPR) could be utilised, which are reliable techniques for obtaining ligand affinities that do not have the requirement for

an exogenous dye (which may be limiting the DSF data). However, they were not tested in this thesis as they were not available at the University of Kent, therefore we tried the available in-house techniques.

Furthermore, identifying the best method for BzaC kinetic characterisation, a stopped assay, is useful to allow for future analysis of substrate specificity. Future experiments could carry out analysis of BzaC 5(6)-OHBza and 5(6)-OHBza-R affinity to allow for comparison of K_m, k_{cat} and catalytic efficiency (K_m/k_{cat}) value to determine the preferred substrate for BzaC.

In addition, the isomer analysis identified that the substrate preparation contains both isomer species and the BzaC's prefer the 5- isomer over the 6-. This regiospecificity could contribute towards the differences in percentage conversion of substrate to product identified in chapter 3, for example *Mt*BzaC and *El*BzaCT may be more specific for the less abundant 5- isomer than *Dt*BzaC. To test this further, similar enzyme kinetic experiments and modelling could be carried out to obtain affinities and catalytic efficiencies with each isomer for both *Mt*BzaC and *El*BzaCT or each isomer could be produced and assayed individually. However, to do this, strains of ED674 which are able to make each isomer individually would require developing. This could be achieved by identifying CobT/U's that are selective for specific isomers and then individually incorporating them into the strain, after deleting the other CobT/U's, if required. For example, *Mt*CobT could be used to specifically make 5-OHBza-RP and *Ec*CobT could be used to make just 5-OMeBza-RP (55). However, from published literature CobT/U enzymes that specifically make 6- isomers could not be identified, therefore if these required production a variety of CobT/U's should be identified and

assayed for their regiospecificity for certain substrates, in an attempt to find one with the correct specificity. Furthermore, enzyme engineering attempts could be carried out to modify the active site of the enzymes to modify their specificities.

Chapter 5: *Dt*BzaC structural biology

5.1 Introduction

5.1.1 Structural classification of SAM dependent methyltransferases

SAM dependent enzymes have been present since the last common ancestor, however they have diversified and expanded into five different structural classes. These groups are quite distinct in terms of sequence, structure and mode of SAM binding, however within the groups certain structural motifs have been identified as key for function (74).

5.1.1.1 Class I

This is the largest group of SAM dependent methyltransferases and was first identified by the characterisation of HhaI, a cytosine-5 methyltransferase (94). They are often referred to as the Rossmann-like class due to the alternating alpha helices (α) and beta strands (β) that form the classical seven stranded beta sheet sandwiched between alpha helices. SAM binds between the beta strands 1 and 2 (β1 and β2) coordinated by the glycine rich GxGxG motif between β1 and αA interacting with the adenine moiety and an acidic residue in β1 that forms a bidentate interaction with the ribose hydroxyls (95). The substrate binding and catalytic domain varies drastically, demonstrating the plasticity of the enzymes and diversity of the substrates they can catalyse reactions on (74). The substrate selectivity can sometimes be influenced by additional domains appended to the Rossmann fold domain.

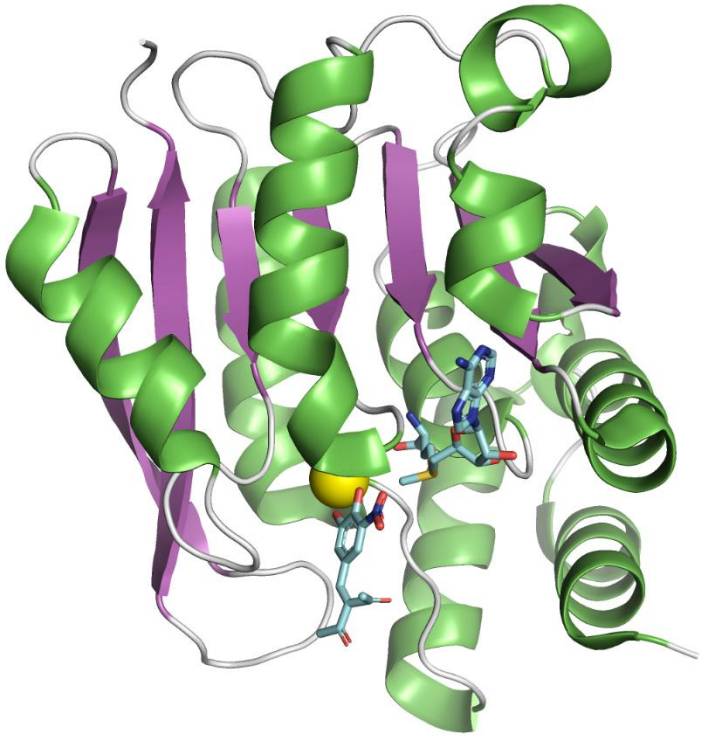


Figure 5.1 - The structure of a class I methyltransferase.

PDB: 6LFE (Rat catechol-O-methyltransferase) (96). The β-sheet is coloured magenta, the α-helices are coloured green, the loops are grey, the substrate (nitecapone) and SAM are coloured primarily blue and the magnesium is coloured yellow.

5.1.1.2 Class II

This class only contains enzymes with one function, which is to regenerate oxidised cobalamin. They form one component of the methionine synthase enzyme, and reactivate any oxidised cob(I)alamin species (cob(II)alamin). To do this flavodoxin firstly reduces cob(II)alamin to cob(I)alamin, which is then brought into proximity to SAM, which then attacked by the cob(I)alamin forming MeCbl. This class is defined by

a long antiparallel β-sheet with a sharp bend near the end and surrounding α-helices at the top and bottom of the sheet making a horse-shoe like topology (97, 98).

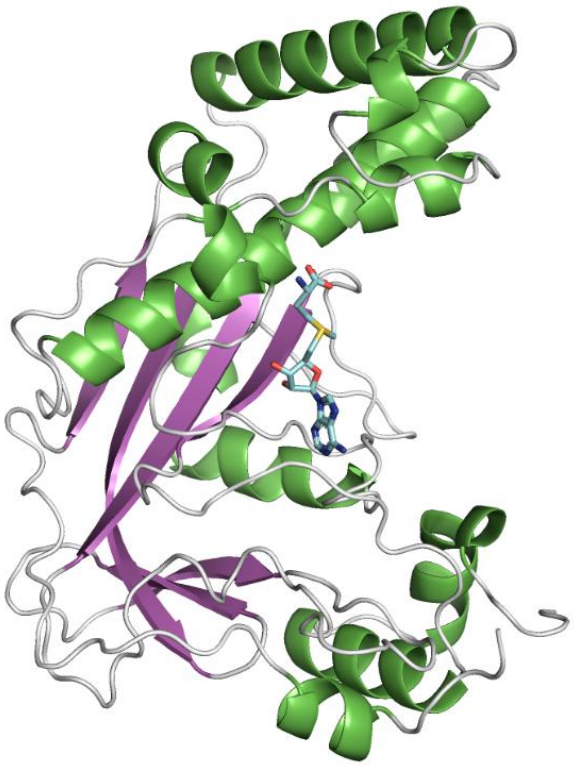


Figure 5.2 - Structure of class II SAM dependent methyltransferases.
PDB: 1MSK activation domain of *E. coli* methionine synthase (98). The β-sheet is coloured magenta, the α-helices are coloured green, loops grey, and the SAM is coloured primarily blue.

5.1.1.3 Class III

This class mainly catalyses reactions with tetrapyrroles substrates, such as the cobalt-precorrin-4 methyltransferase (CbiF) (99), SAM uroporphyrinogen III methyltransferase (UroM) (100) and Siroheme synthase (CysG) (101). These enzymes form dimers with

the active sites formed between the N and C-terminal domains of each monomer. Each monomer has two domains which both contain central β-sheets surrounded by α-helices. However, their topologies are different with the N-terminal domain containing a 3-2-4-1-5 sheet and the C-terminal a 1-2-5-3-4 sheet with a beta turn between strands 4 and 5 producing a mostly antiparallel sheet. Interestingly, unlike the other class, the SAM molecule binds to a pocket on the surface of the protein (99).

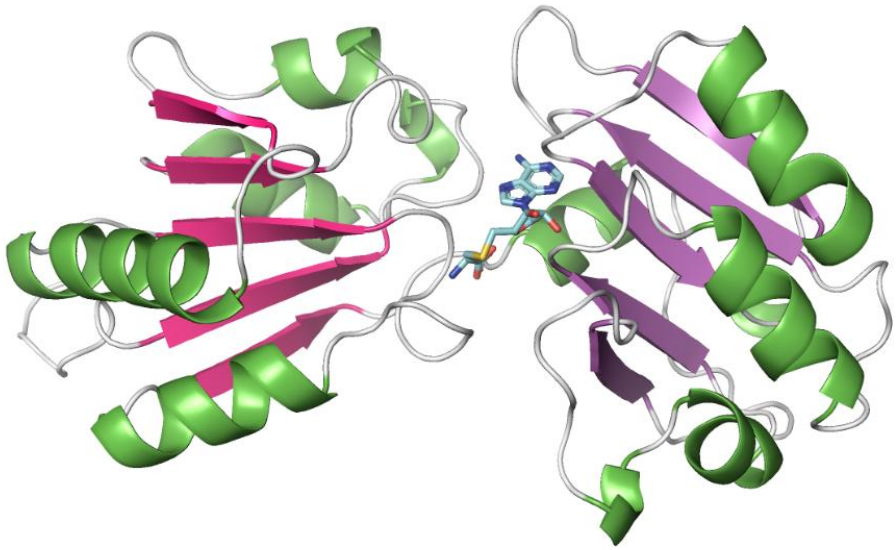


Figure 5.3 - Class III SAM dependent methyltransferase structure.

Crystal structure of CbiF (PDB = 1CBF) (99). The N-terminal β-sheet is coloured pink and the C-terminal magenta, the α-helices are coloured green, loops coloured grey, and the SAH is coloured primarily blue.

5.1.1.4 Class IV

This is the second largest class of SAM dependent methyltransferases and are involved in the post-translational modification of t- and r-RNAs. They are composed of a five/six-stranded parallel β-sheet flanked by alpha helices. Part of the SAM binding domain

contains a rare but characteristic C-terminal 'knot' topology, formed by overlapping loop regions forming a 'knot' conformation (**Figure 5.4**) (74, 102, 103).

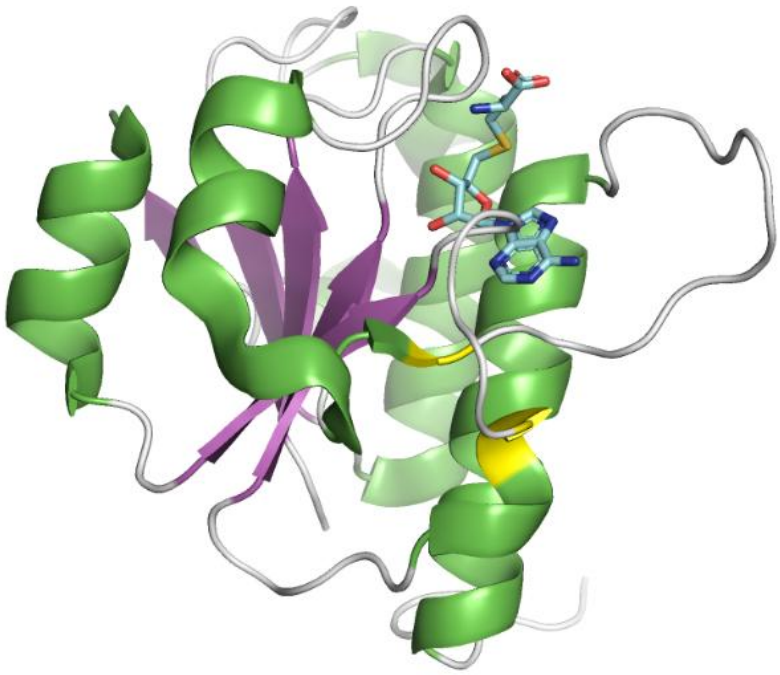


Figure 5.4 - Structure of a class IV SAM dependent methyltransferase. Structure of YibK (a tRNA (cytidine(34)-2'-O)-methyltransferase (103). The β-strands are coloured in magenta, α-helices green, knot structure yellow, loops in grey and SAH in blue.

5.1.1.5 Class V

The final class are the SET (suppressor of variegation, enhancer of zeste and trithorx) domain containing enzymes. They are responsible for histone lysine methylation and other proteins that are important for transcriptional regulation. Their structure typically

consists of four α-helices embedded in a series of twisted β-sheets with a pseudo-knot structure formed by the loop prior to the C-terminal helix (74).

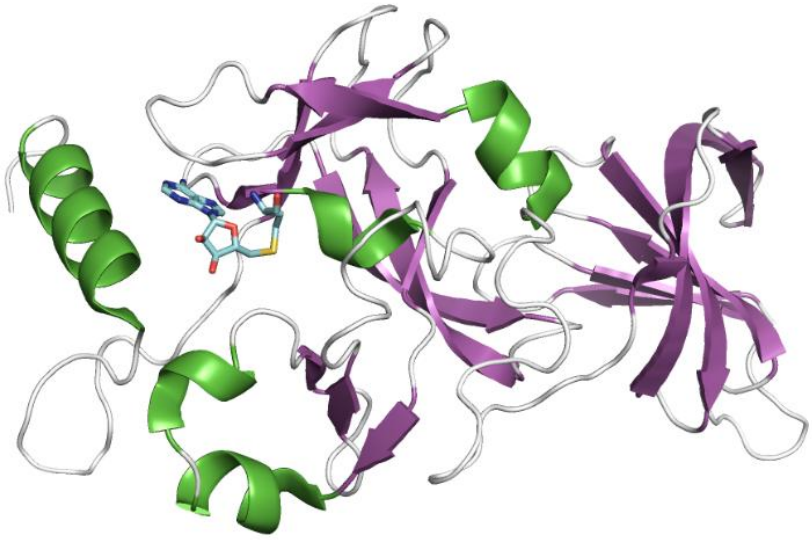


Figure 5.5 - Class IV SAM dependent methyltransferase structure.

Crystal structure of the human histone methyltransferase SET7/9 (PDB = 1O9S) (104). Loops are coloured grey, α-helices green, β-strands magenta, SAH in blue.

5.1.2 Catalytic mechanisms

In addition to classification, structural analysis of these enzymes has played a crucial role in understanding their catalytic mechanisms, which is imperative for understanding their role in biological systems and for utilising them for biotechnological applications.

There are three mechanisms by which methyltransferase catalyse their reactions; 1: the proximity and desolvation mechanism, 2: acid/base mediated methylation and 3: metal-dependent mechanisms (74, 105).

5.1.2.1 Proximity and desolvation

In this mechanism the active site side chains are not directly involved in catalysis, however the pocket brings the substrates together in the optimal orientation to aid catalysis. Furthermore, solvating waters are removed from the area surrounding the acceptor and donor molecules, which increases their electrophilicity and nucleophilicity, respectively.

One example of this is SAMT (SAM:salicylic acid carboxyl methyltransferase) in which the acceptor oxygen on the substrate is brought into 3 Å of the expected position of the SAM methyl group, with no waters are present, allowing for energetically favourable methyltransfer (106).

This is further exemplified by the enzyme DnrK, a carminomycin-4-O-methyltransferase involved in daunorubicin. The crystal structure of DnrK in complex with its product and SAH shows the product oxygen is positioned around 4.1 Å away from the SAH sulfonium atom with the correct orientation to allow for catalysis, furthermore there are no water molecules near the product / SAH interface. There is a tyrosine residue (Tyr142) within 4.1 Å of the product oxygen, which could potentially act as a base for substrate activation. However, mutagenesis of the residue to tryptophan (a non-basic residue) only had a small impact on catalysis, suggesting it is not required for catalysis, thus supporting the proximity and desolvation mechanism (107).

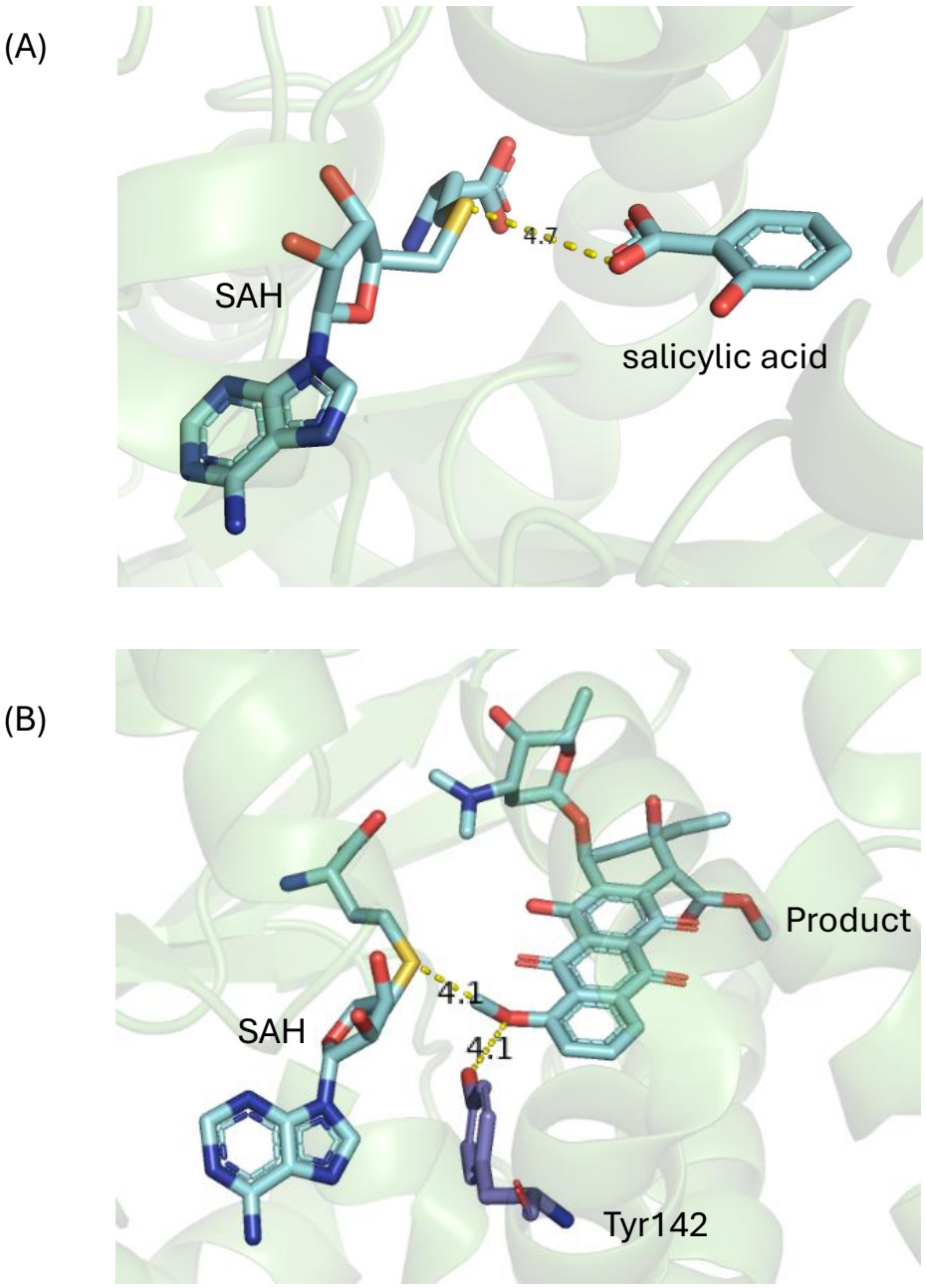


Figure 5.6 - SAMT proximity and desolvation mechanism.

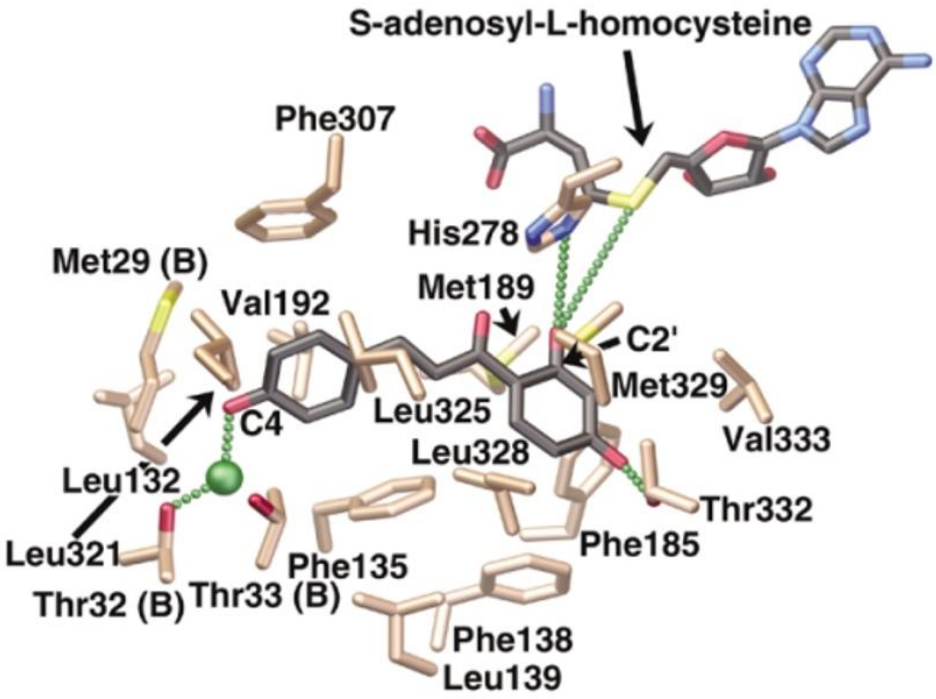
(A) Crystal structure of SAMT with SAH and salicylic acid (PDB ID = 1M6E) (106) with the distance between the acceptor oxygen and SAH sulfonium atom labelled. **(B)** Crystal structure of DnrK (PDB ID = 1TW2) (107) in complex with SAH and product (4-methoxy-e-rhodomycin T) (coloured primarily blue) with Tyr142 coloured primarily purple.

5.1.2.2 Acid / base catalysis

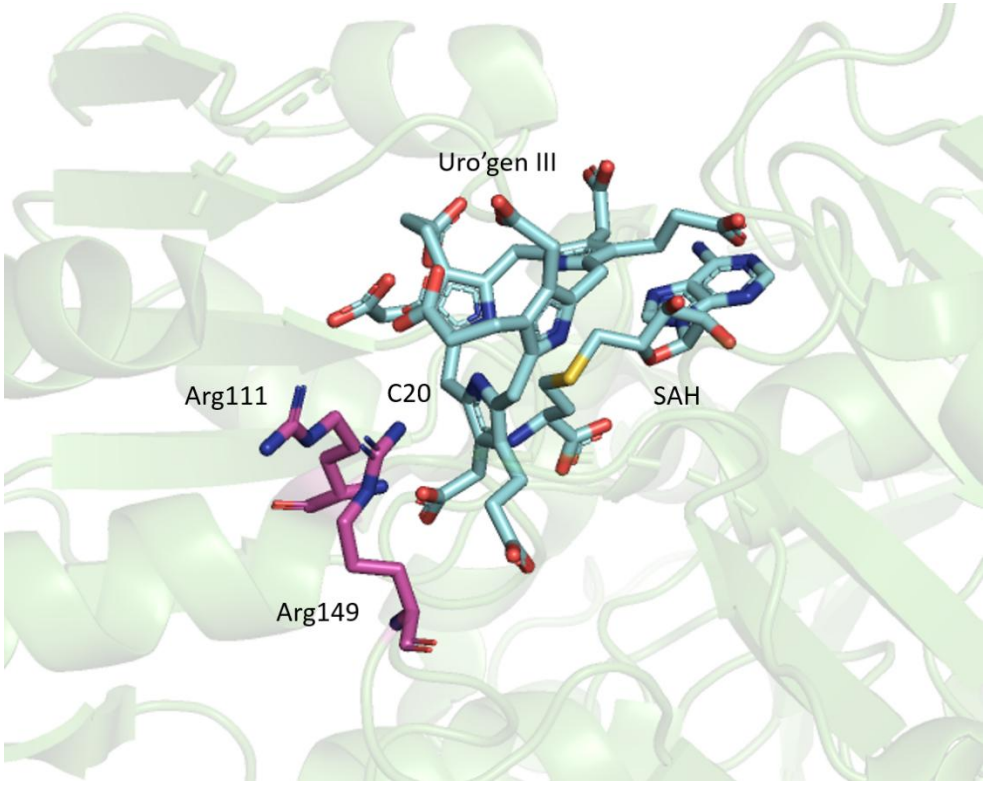
This mechanism relies on active site amino acid side chains that can act as a base, such as histidine and arginine, to deprotonate substrates to increase their nucleophilicity, promoting subsequent attack of SAM's methyl group. This can be exemplified by RebM which catalyses the 4'-oxygen methylation of rebeccamycin via His140 mediated deprotonation of the 4-hydroxyl and subsequent methyltransfer from SAM (108) . This can be further exemplified by Chalcone O-methyltransferase (ChOMT) which catalyses the SAM dependent O-methylation of isoliquiritigenin (4,2',4'-trihydroxychalcone) via the His278 deprotonation of the C2-hydroxyl group which can then attach the SAM methyl group to facilitate methyltransfer. Furthermore, mutation of this residue to Asn, Lys, Leu, Ala or Gln completely abolished activity, further supporting its role in catalysis (109).

In addition, mechanisms by which deprotonation of a group causes the increase in nucleophilicity of a neighbouring atom, leading to its subsequent methylations also exist. For example, NirE catalyses the deprotonation of C-20 of Uro'gen III via Arg111 or Arg149 which causes the subsequent SAM dependent methylation of C-2. This is confirmed by mutation of Arg111 or Arg149 to lysine's decreasing activity (110).

(A)


S-adenosyl-L-homocysteine
Phe307
His278
Met29 (B)
Met189
Val192
C2'
Met329
C4
Leu325
Val333
Leu328
Leu132
Thr332
Phe185
Leu321
Phe135
Thr32 (B)
Thr33 (B)
Phe138
Leu139

(B)


Uro'gen III
Arg111
C20
SAH
Arg149

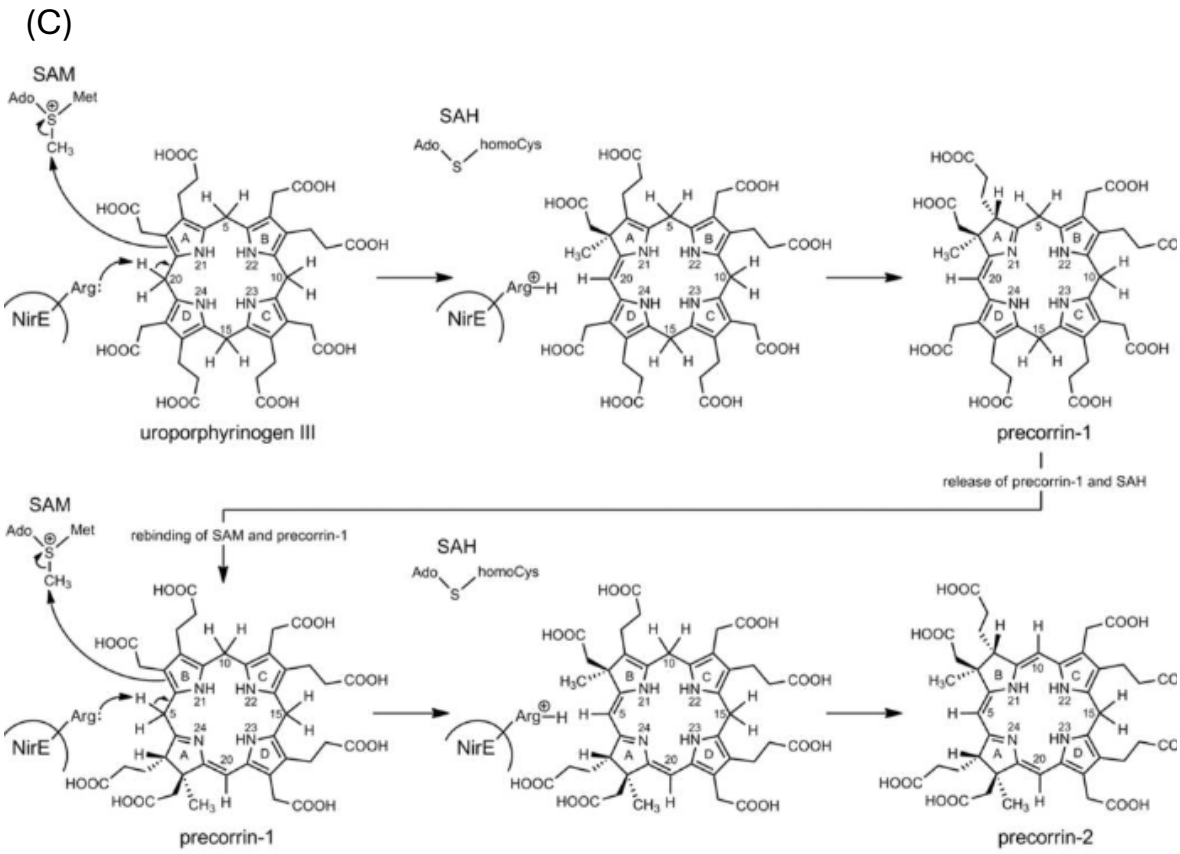


Figure 5.7 - Acid/base catalysis mechanisms.

(A) Active site structure of ChOMT showing the distances (in green) between the catalytic His278, SAH sulfonium atom and the substrate C2' hydroxyl group. Taken from (109). **(B)** Active site of NirE in complex with SAH and Uro'gen III (blue) (PDB ID = 2YBQ) with the catalytic Arg111 and Arg149 residues highlighted in pink. **(C)** Schematic of the NirE mechanism of C20 deprotonation and subsequent C2 methylation by SAM. Taken from (110).

5.1.2.3 Metal-dependent mechanisms

These are almost exclusively observed in plant O-methyltransferases active on substrate phenols and utilise metals for substrate activation. For example, caffeoyl

coenzyme A 3-O-methyltransferase contains an active site Ca^{2+} which can alter the pK_a of the substrate phenol allowing it form an oxyanion (most likely to balance the charge of the microenvironment), increasing its nucleophilicity thus enhancing the methyltransfer from SAM (111).

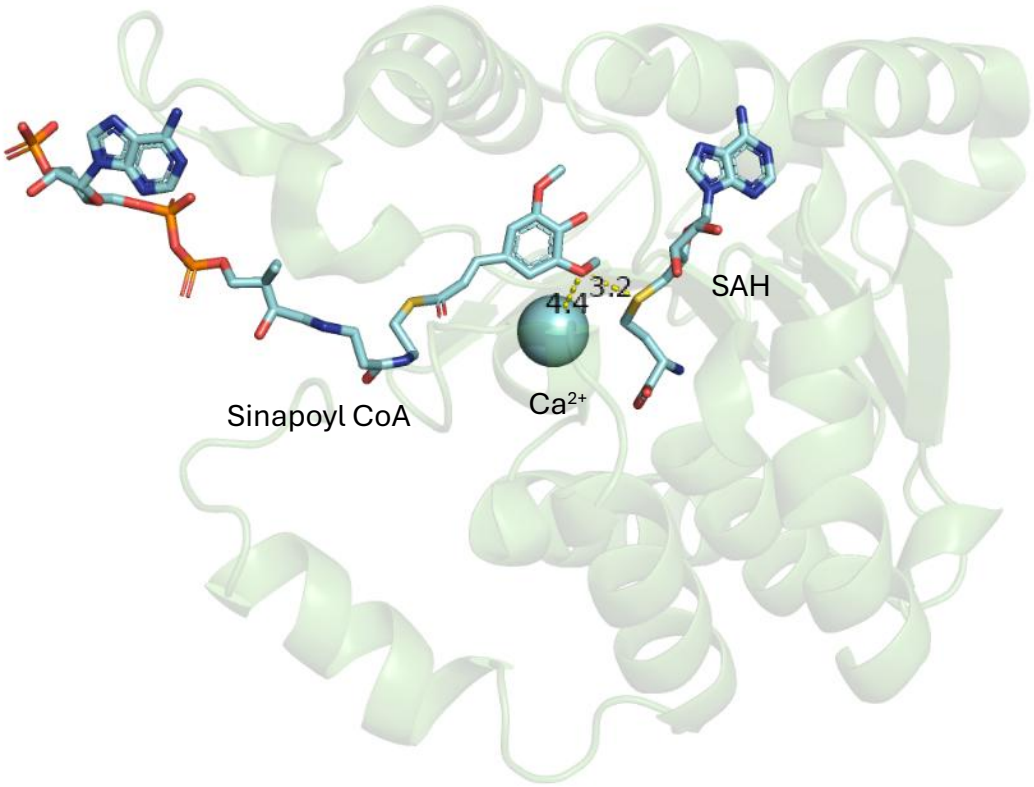


Figure 5.8 - caffeoyl coenzyme A 3/5-O methyltransferase.

Crystal structure of the caffeoyl coenzyme A 3-O methyltransferase active site with product (sinapoyl CoA), SAH and Ca^{2+} (blue) (PDB ID = ISUS) (111). Bond distances between the Ca^{2+} and methyl accepting oxygen (4.4 Å) and SAH and oxygen (3.2 Å) are displayed by the yellow dotted lines.

5.1.3 Chapter aims

As demonstrated by the classification of SAM dependent methyltransferases and elucidation of their mechanisms via crystallography and mutagenesis, obtaining a crystal structure of a BzaC enzyme is critical for understanding how the enzyme binds its substrates, the mechanism of catalysis and how it can be engineered for future applications. This chapter explores the crystallisation, structure determination of *Dt*BzaC, its domain architecture, ligand binding sites and proposes a mechanism for catalysis.

5.2 Results

5.2.1 *Dt*BzaC crystallisation

Since we observed in our DSF experiments that the binding of SAH to *Dt*BzaC results in an increase in protein stability, exemplified by the 13 °C increase in T_m (**Figure 4.6**), we initially attempted to crystalise the *Dt*BzaC-SAH complex. To screen for initial crystallisation conditions, the purified *Dt*BzaC protein (13.6 mg/ml) was incubated with a fivefold excess of SAH and then centrifuged to remove any precipitation.

In collaboration with the Tews lab at the University of Southampton, 400 nL sitting drops with three different protein: precipitant ratios (1:1, 2:1, 1:2) were set up above wells containing 50 µL of each crystallisation condition using the Oryx crystallisation robot. Plates were setup for both the JCSG plus and MORPHEUS screens which were then incubated at 21 °C and in a vibration free incubator and crystal formation was

monitored via microscopy. Multiple crystals suitable for diffraction were observed across a number of conditions, these were picked, transferred to a cryoprotectant and flash frozen in liquid nitrogen. Crystals were then tested for diffraction at the European Synchrotron Radiation Facility (ESRF) by Jack Stubbs. Full datasets were collected for six crystals, with the best diffracting to 2.17 Å.

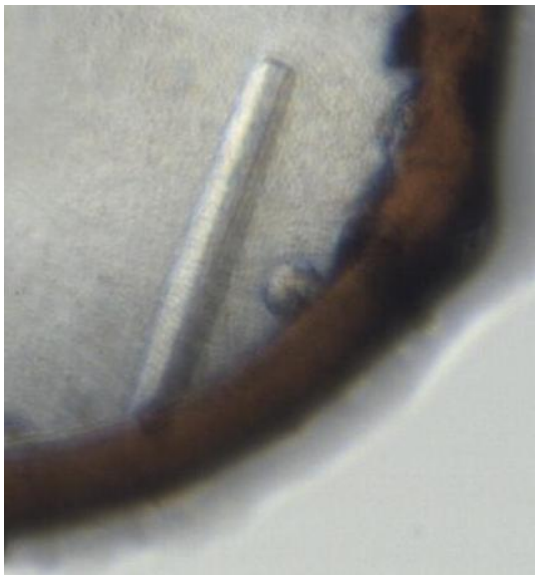


Figure 5.9 - *Dt*BzaC crystal.

Image of the *Dt*BzaC crystal formed in 0.1M Sodium acetate pH 4.6, PEG 4K 8% which diffracted to 2.17 Å. The diffraction data were then used to solve the structure by molecular replacement.

5.2.2 Data processing and refinement

The collected diffraction data were auto processed on the beamline. The space group was estimated through systematic absences by the Pointless programme and the data was scaled using Aimless. The space group was determined to be P 2 2 2 with unit cell

dimensions of a=49.40, b=99.38, c=151.50, alpha=90.00, beta=90.00, gamma=90.00. The structure was solved by molecular replacement using Phenix, which was given the AlphaFold 2 prediction of *Dt*BzaC as the search model. The molecular replacement was performed by Ivo Tews who provided the initial model for refinement. The initial solution shows that the asymmetric unit cell contains two molecules of *Dt*BzaC. We then refined the structure by performing multiple cycles of manual model building in Coot (61) followed by Refmac (63) in the CCP4 software suite (62). Mostly, Ramachandran and rotamer errors were corrected by manually altering the orientation of residues to better fit the electron density (**Figure 5.10**). Furthermore, empty density, determined by the difference map, was fitted with appropriate ligands, water molecules or residues, especially a flexible loop region that was built in chain A. This produced a model with an R-factor of 0.187 and R_{free} value of 0.245 and 0.1 % Ramachandran outliers which is comparable to most PDB entries and shows the model is a good representation of the data (**Table 5.2**).

Wavelength	0.87313
Resolution	47.22 - 2.17
R_{merge} (within I+/I-)	0.112
Mean((I)/sd(I))	7.2
Unique reflections	39332
% completeness	97.7
Multiplicity	3.8
Mid slope anomalous probability	1.011

Table 5.1 - Crystallographic data collection statistics.

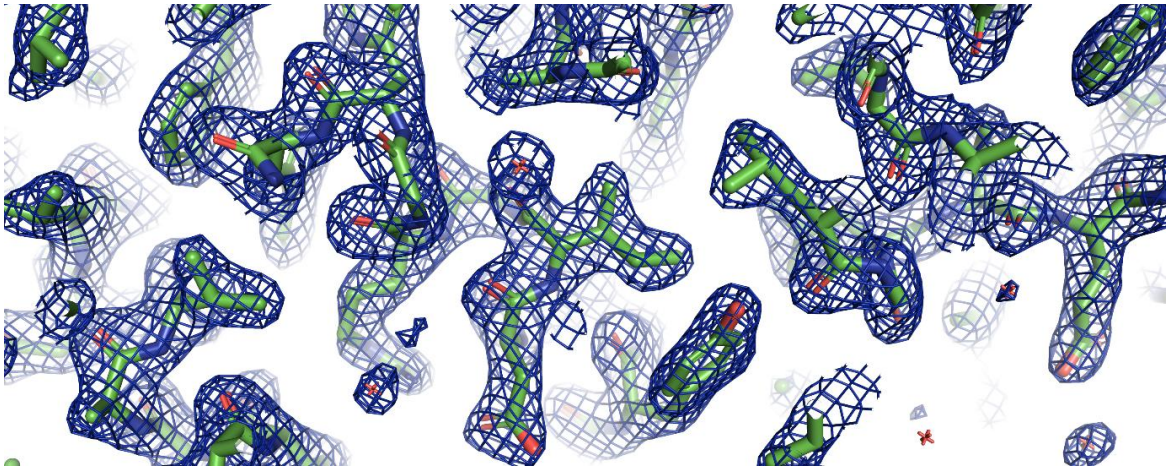


Figure 5.10 - *Dt*BzaC model overlayed with m2Fo-DFc electron density, centred around residues 17 and 18 of chain A.

Image produced by producing the m2Fo-DFc electron density map from the CCP4 Refmac mtz file using FFT (Fast Fourier Transform) in Phenix (112) and overlaying with the Refmac pdb file for the *Dt*BzaC dimer in PyMOL.

Wavelength	0.873
Space group	P 2_1 2_1 2_1
Unit cell	a=49.41, b=99.38, c=151.50, alpha=90.00, beta=90.00, gamma=90.00
Resolution range	47.22 - 2.17
Total reflections	39230
Unique reflections	39230
Wilson B-factor	31.73
Reflections in refinement	39229
Reflections in free set	1928
R_{work}	0.187
R_{free}	0.245
FSC average	0.9677
RMSD bonds	0.0112
RMSD angles	2.187
Ramachandran favoured (%)	98.2
Ramachandran allowed (%)	1.7
Ramachandran outliers (%)	0.1
Rotamer outliers	1.2
Clash score	1.6
MolProbity score	0.97

Table 5.2 - *Dt*BzaC refinement statistics.

Produced from the CCP4 Refmac cycle used to produce the final model.

5.2.3 Structure overview

Visualisation in PyMOL identified the structure to be a homodimer (chain A and chain B) with a molecule of SAH bound to each subunit, thus identifying the active sites (**Figure 5.11, A**). Furthermore, topology analysis using the PDBsum webserver identified that each monomer contains an N-terminal domain involved in dimerisation and a C-terminal methyltransferase domain, which aligns with the previously predicted domain structure (**Figure 5.11, B**).

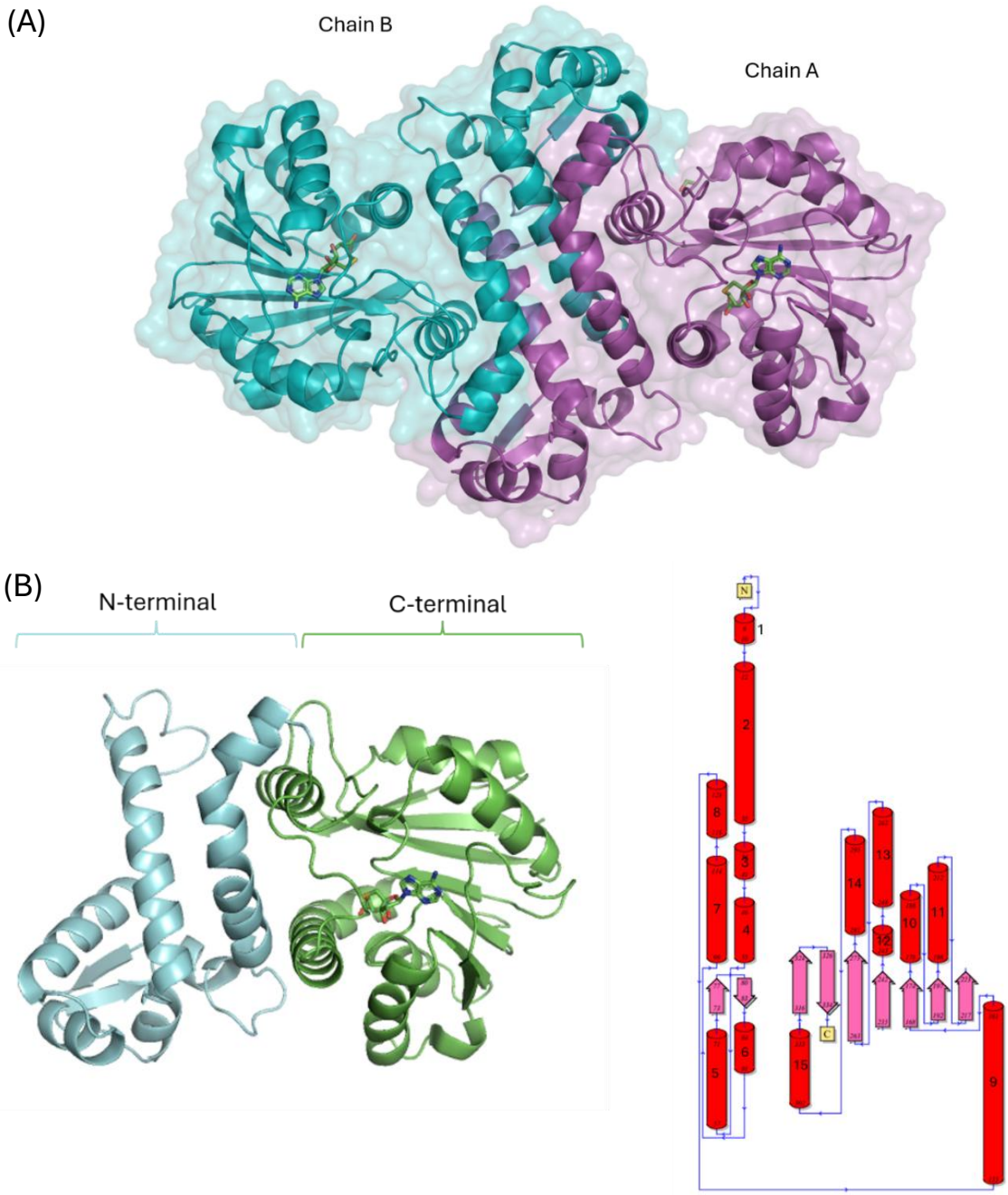


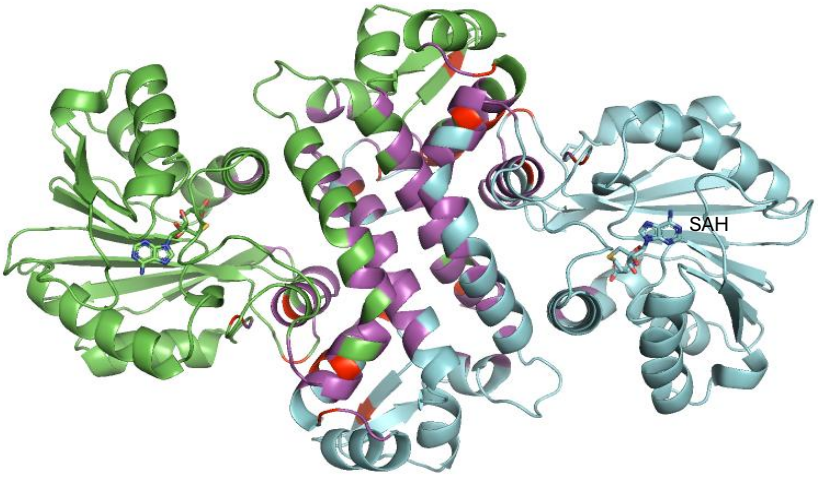
Figure 5.11 - *Dt*BzaC structure and topology.

(A) *Dt*BzaC dimer structure with chain A coloured magenta and chain B coloured cyan. **(B)** Domain architecture of *Dt*BzaC monomer (N-terminal coloured cyan and C-terminal coloured green) and topology (helices in red and beta strands in pink).

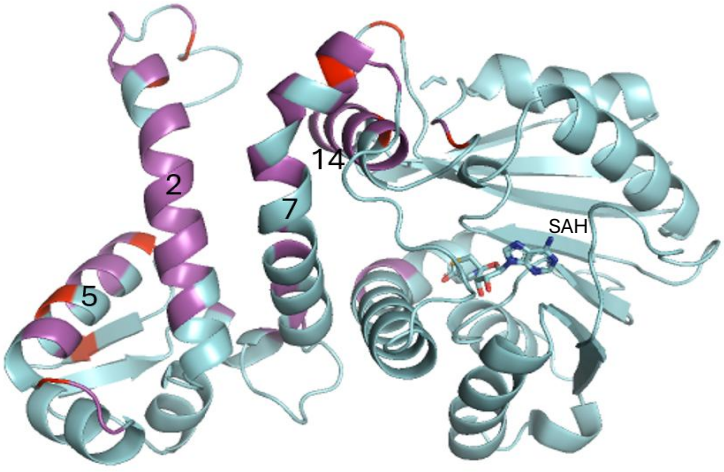
5.2.4 Analysis of the dimerisation domain

The dimerisation domain of the *Dt*BzaC homodimer was investigated using PDBsum (65) and PyMOL (64). This demonstrated the presence of 3 salt bridges, 14 hydrogen bonds and 357 non-bonded contacts between 61 residues of each monomer, mostly occurring between the N-terminal domains, especially between α2 of each chain. However, there are residues outside of the bioinformatically identified N-terminal dimerisation domain that are involved in dimerisation, such as those in C-terminal α14 (**Figure 5.12**).

(A)



(B)



(C)

Chain	A	B
No. interface residues	61	
No. salt bridges	3	
No. disulphide bonds	0	
No. hydrogen bonds	14	
No. non-bonded contacts	357	
total contacts	374	

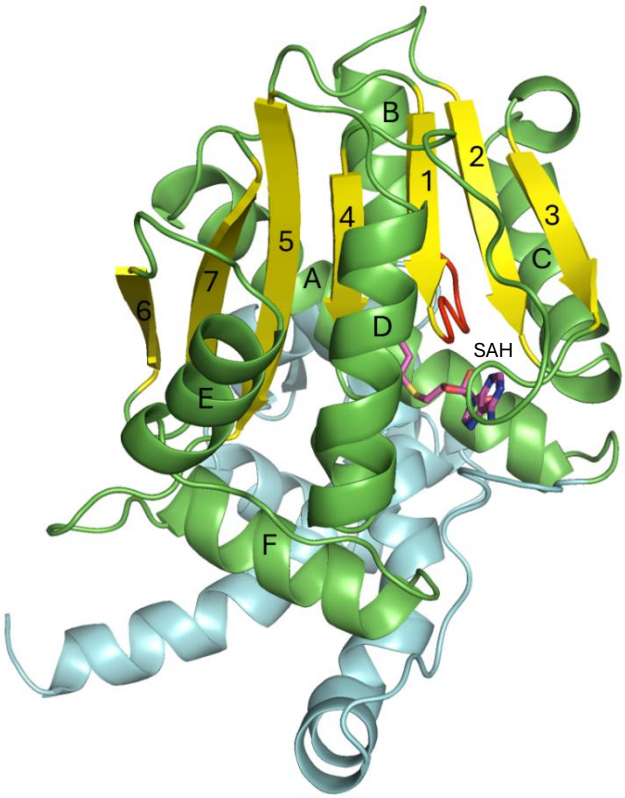
Figure 5.12 - *Dt*BzaC dimerisation domain analysis.

(A) *Dt*BzaC dimer dimerisation domain interface. **(B)** *Dt*BzaC monomer dimerisation domain interface. Residues involved in non-contact interactions are coloured magenta and residues involved in non-contact interactions plus hydrogen bonds and salt bridges are coloured red. **(C)** Table showing the number of residues and bonds involved in dimerisation.

5.2.5 Methyltransferase domain Rossman fold analysis

Analysis of the C-terminal domain structure identified a classical Rossman fold which is present in class 1 SAM dependent methyltransferases. This is composed of seven parallel beta strands, six of which are in the same orientation (towards the dimerisation domain) and one of which (7) in the opposing orientation, all sandwiched between six alpha helices (**Figure 5.13, A**). SAH binds under β1 and β2 with the characteristic GxGxG motif (red) between αB and β1.

(A)



(B)

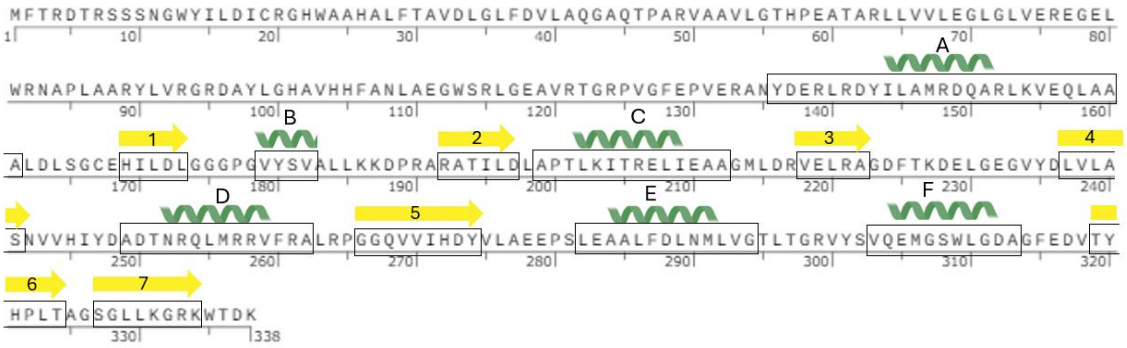
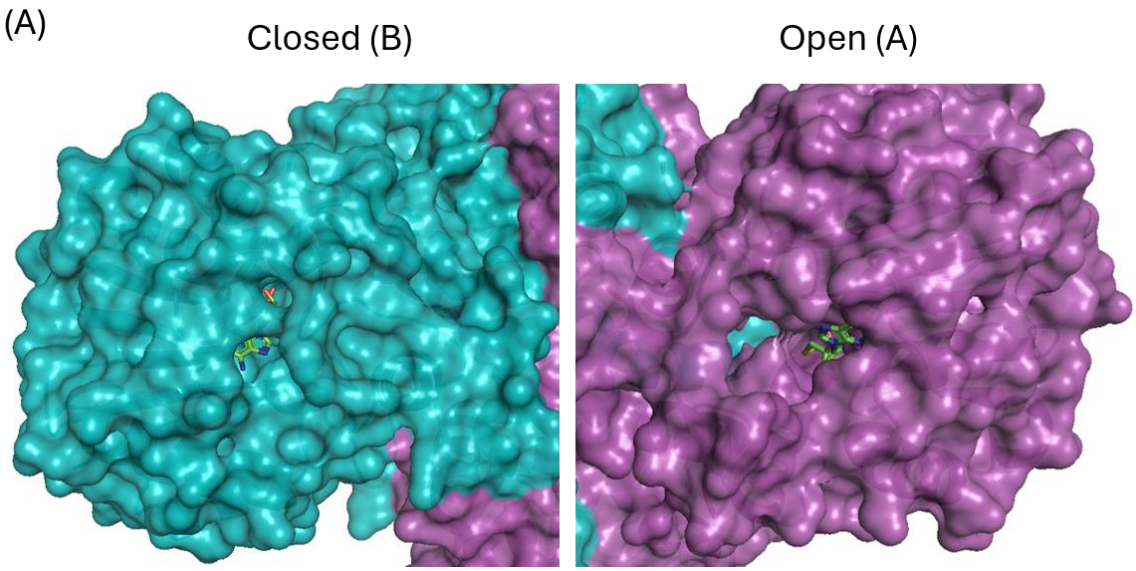


Figure 5.13 - *Dt*BzaC methyltransferase domain analysis.

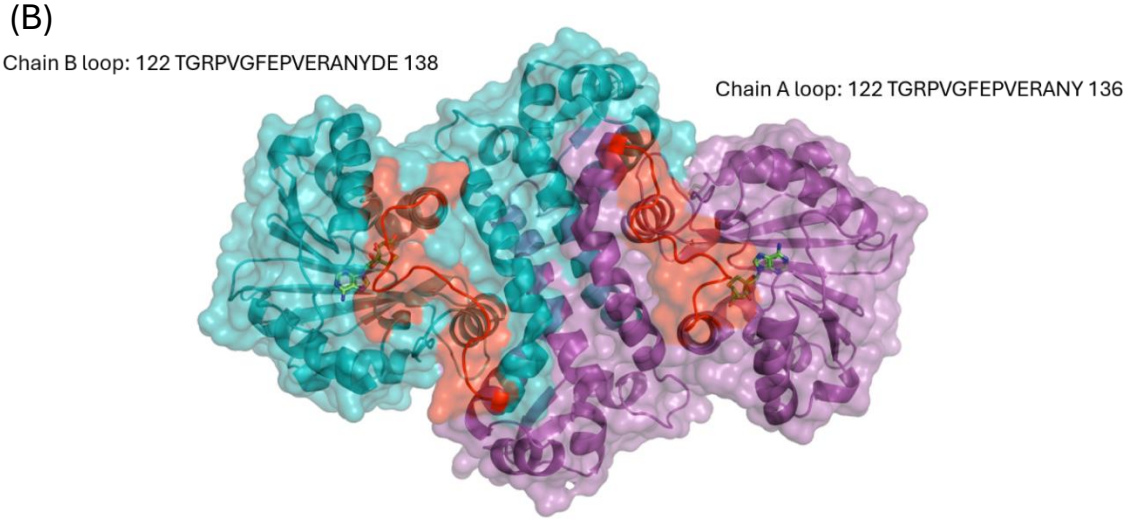
(A) Annotated chain B methyltransferase domain. Alpha helices (green) are alphabetically labelled based on their position in the primary amino acid sequence, beta strands (yellow) are numbered according to their sequence position. The dimerisation domain is coloured cyan and the GxGxG motif is coloured red. **(B)** Annotated *Dt*BzaC amino acid sequence according to the highlighted features of the methyltransferase domain.

5.2.6 Entry to the active site is mediated through a flexible loop

Comparison of the two subunits in the asymmetric unit has identified that the protein can adopt 'open' and 'closed' active site conformations which is mediated by 15-17 amino acid flexible loops at the side of the active sites (**Figure 5.14**). This is shown by the chain A active site being accessible and the chain B active site being inaccessible for freely diffusing molecules. Due to both conformations being present with SAH bound, it appears not to be a co-factor induced conformational change, however it may be a regulatory mechanism for mediating substrate entry and product release. Furthermore, the loop region sequences are not conserved across similar methyltransferases identified by PDBeFold (73), even within BzaCs (**Figure 5.14, C**).



(B)



(C)

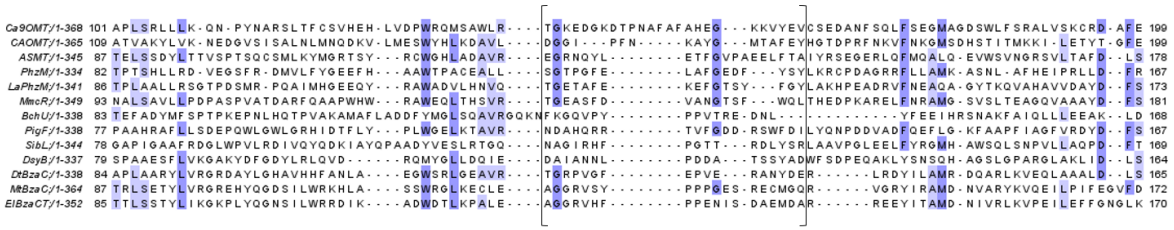


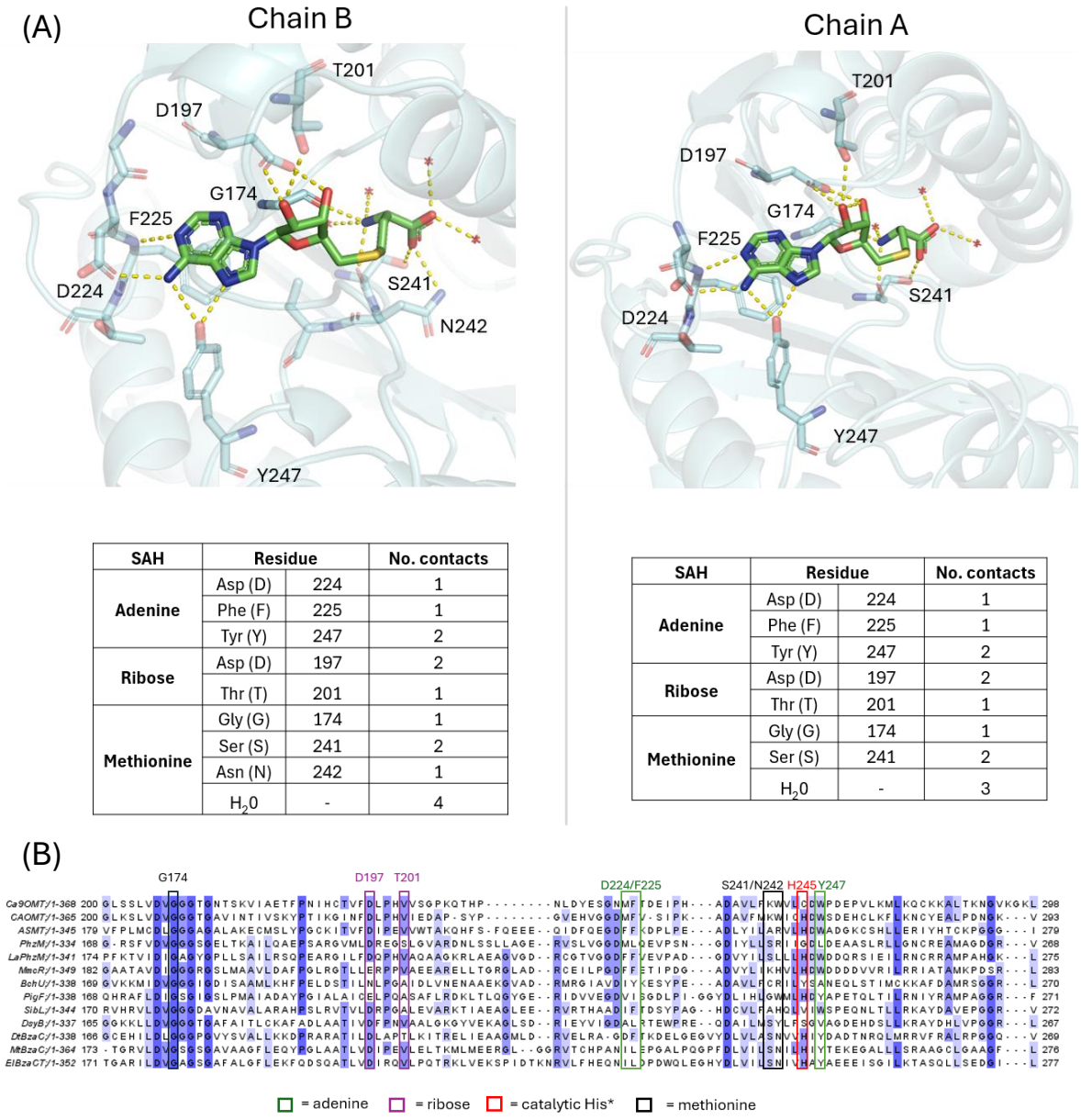
Figure 5.14 - *Dt*BzaC flexible loop mediating active site entry.

(A) PyMOL produced image of the 'closed' conformation of chain B's active site and the 'open' conformation of chain A's. **(B)** PyMOL produced image of the active site flexible loops (red) of each chain. **(C)** Sequence alignment of the top ten characterised proteins with structural homology to *Dt*BzaC and the two other BzaCs, with the *Dt*BzaC flexible loop region within the square brackets ([ ]).

5.2.7 Co-factor binding domain analysis

*Dt*BzaC co-crystallisation with a five-fold excess of SAH also allowed for the analysis of the co-factor (SAM) binding domain. This analysis was carried out in PyMOL by visualising all polar contacts between SAH and *Dt*BzaC / waters. This identified 13 polar contacts in chain A (open) and 15 in chain B (closed) with the two differences

being an extra contact between Asn242 and a water with the methionine moiety of SAH, suggesting that the closed conformation increases the affinity of the co-factor binding. In both active sites the initial glycine (174) of the GxGxG motif interacts with the methionine moiety of SAH which is a classical binding event observed in class 1 SAM dependent methyltransferases and is conserved across all three BzaCs and the top 10 structurally similar proteins identified by PDBeFold. Moreover, each chain produces two interactions between Ser241 (fairly conserved 6/13) and multiple waters with the methionine moiety. Furthermore, each chain forms the expected bidentate interactions between the acidic Asp197 of β2 and the ribose hydroxyl groups. This acidic residue is well conserved across the alignment with ten being Asp, two being Glu (acidic). In addition, Thr201 (fairly conserved 8/13) also interacts with a ribose hydroxyl. Moreover, interactions occur between Asp224, Phe225, and Tyr247 with the adenine moiety with lower conservation (0-7/13).



SAH	Residue		No. contacts
Adenine	Asp (D)	224	1
	Phe (F)	225	1
	Tyr (Y)	247	2
Ribose	Asp (D)	197	2
	Thr (T)	201	1
Methionine	Gly (G)	174	1
	Ser (S)	241	2
	Asn (N)	242	1
	H_20	-	4

SAH	Residue		No. contacts
Adenine	Asp (D)	224	1
	Phe (F)	225	1
	Tyr (Y)	247	2
Ribose	Asp (D)	197	2
	Thr (T)	201	1
Methionine	Gly (G)	174	1
	Ser (S)	241	2
	H_20	-	3

Figure 5.15 - *Dt*BzaC SAH binding domain analysis.

(A) Schematics of the SAH – *Dt*BzaC/H_2O binding interactions occurring in both chain A and B (generated in PyMOL) with respective summary tables. **(B)** Sequence alignment of the three BzaCs and the top 10 highest similarity 3D structures to *Dt*BzaC identified by PDBeFold, with *Dt*BzaC-SAH interacting residues highlighted.

5.2.8 Molecular docking of Ado[5-OHBza]Cba into the *Dt*BzaC-SAH model

We wanted to provide insight into how *Dt*BzaC binds its cobamide substrate (Ado[5(6)-OHBza]Cba) as it does not have a characterised cobamide binding domain. To do this a 3D model of the substrate was produced by taking adenosylcobalamin (in a base on conformation) from the crystal structure of *Thermotoga maritima* Ribonucleotide Reductase, NrdJ, in complex with dTTP and adenosylcobalamin (3O0N) (66) and modifying the benzimidazole to 5-OHBza in PyMOL. The *Dt*BzaC (with SAH bound) and substrates structures were then prepared for docking in Autodock tools (67). This involved deleting the waters and adding hydrogens and charges to *Dt*BzaC and saving both substrate and protein in pdqt format. Furthermore, a grid box (selecting the region of protein used for docking) was set around chain A's 'open' active site. Docking was then carried out using Autodock Vina (68, 113) with exhaustiveness set to 8.

The output docking poses were then assessed in PyMOL which identified two poses, with good scores, which were most viable. Both of which present a method of cobamide binding to the active site with the lower ligand in a base-off conformation and interactions between the cobamide and active site residues orienting the base into the cleft. This aligns with the results of chapter three were removing the adenosyl group reduced BzaC activity, potentially due to the ability of the adenosylated cobamide to go base-off easier.

A histidine residue (His245) was identified above the SAH sulfonium and between the substrate benzimidazole and SAH. This histidine is fairly conserved across the highest homology structures identified by PDBeFold (5/10) and seen in all three BzaCs. Moreover, it has been suggested from structural analysis of human ASMT (114) that this

residue (His255 in ASMT) is involved in substrate hydroxyl deprotonation during catalysis. Therefore, this suggests that *Dt*BzaC's His245 may be acting as a base in catalysis.

However, both poses have the putative catalytic histidine 245 to far away from the substrate hydroxyl (pose 1 = 6.1 Å, pose 2 = 7.8 Å) and the substrate hydroxyl at a distance too far away from the SAH sulphur (pose 1 = 8.1 Å, pose 2 = 8.4 Å) to allow for catalysis. This suggests a conformational change would be required as the SAM methyl needs to be within ~ 3 Å of the substrate hydroxyl.

Despite this, each pose has merit for different reasons. Firstly pose 1 has the 5-hydroxy group in the correct orientation to allow for catalysis, potentially reducing the complexity of the conformational change required for catalysis. However, there is a hydrogen bond between the oxygen of the substrate hydroxyl and the asparagine (Asn) 290, which will reduce the nucleophilicity of the oxygen. But a conformational change bringing the substrate and co-factor together could remove this hydrogen bond. Furthermore, pose 2 has a bond between the adenosyl group and the protein, which also aligns with the biochemical data showing the adenosylated substrate is preferred as this could increase the affinity between the protein and substrate compared to the cyano and hydroxocobamide.

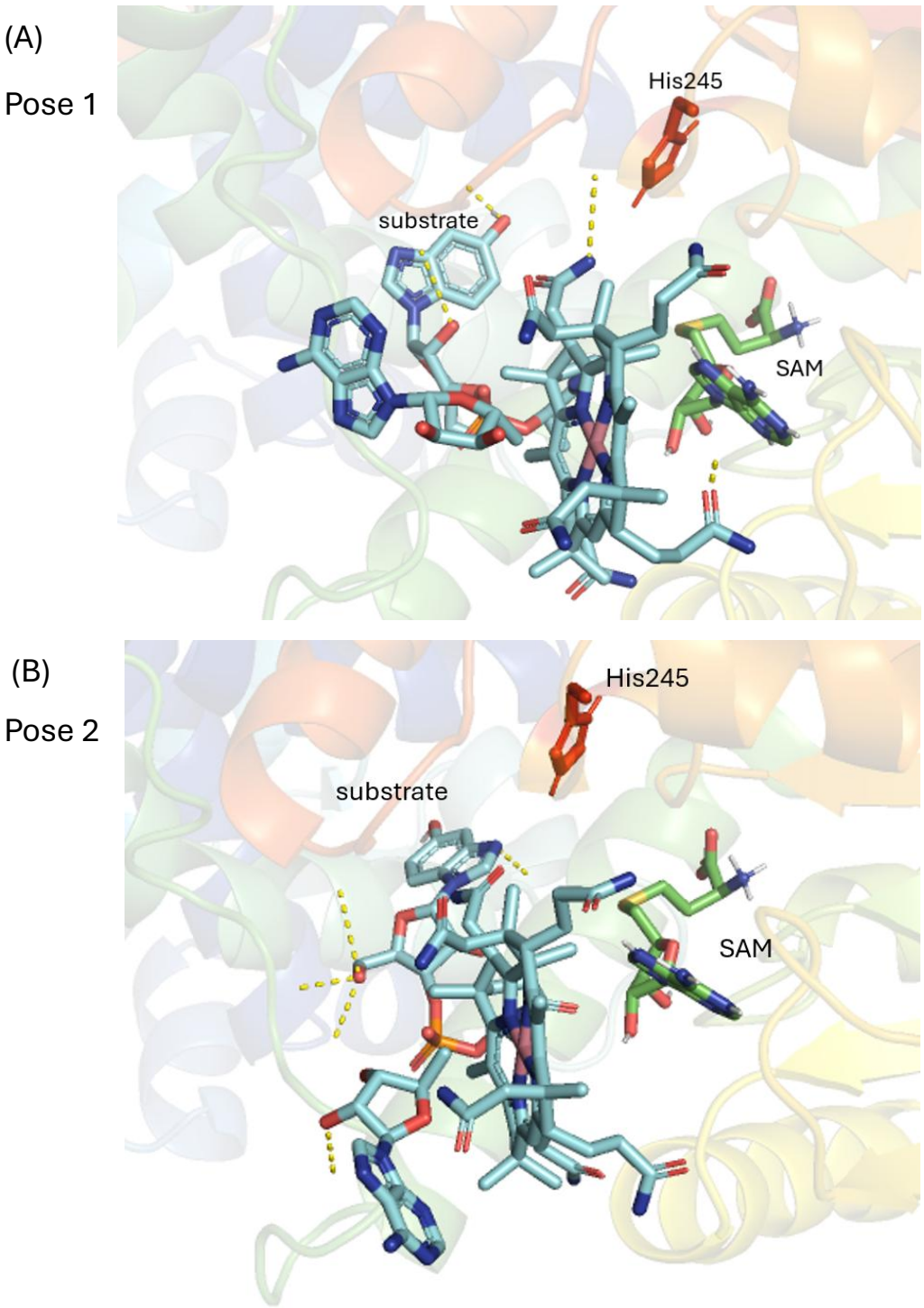


Figure 5.16 - *Dt*BzaC-SAH, Ado[5(6)-OHBza]Cba molecular docking.

(A, B) Poses 1 and 2 generated by docking of the *Dt*BzaC-SAH complex (chain A active site) with Ado[5-OHBza]Cba and visualised in PyMOL.

However, when comparing both docking poses to previously solved structures with both substrates bound and similarity to *Dt*BzaC; ASMT (4A6E) (114) and DSYB (7WDW) (91) it revealed that the docking poses both had the 5-OHBza in the opposite orientation in the active site. With 5-OHBza facing towards the 'back' (inner wall region) of the cleft and ASMT and DSYB substrates facing towards the entrance of the active site (**Figure 5.17**). Furthermore, the DYSB substrate (MTHB) was found at a distance of 6 Å from SAH, placing it in a small channel at the relative 'back' of the active site, suggesting the structure might have captured it entering the active site through this channel. Similarly, there is a channel that leads from the 'back' of the active site to the SAH in *Dt*BzaC which may allow for entrance of the 5-OHBza group in a cobamide base off orientation, however docking around this cleft produced no valid poses.

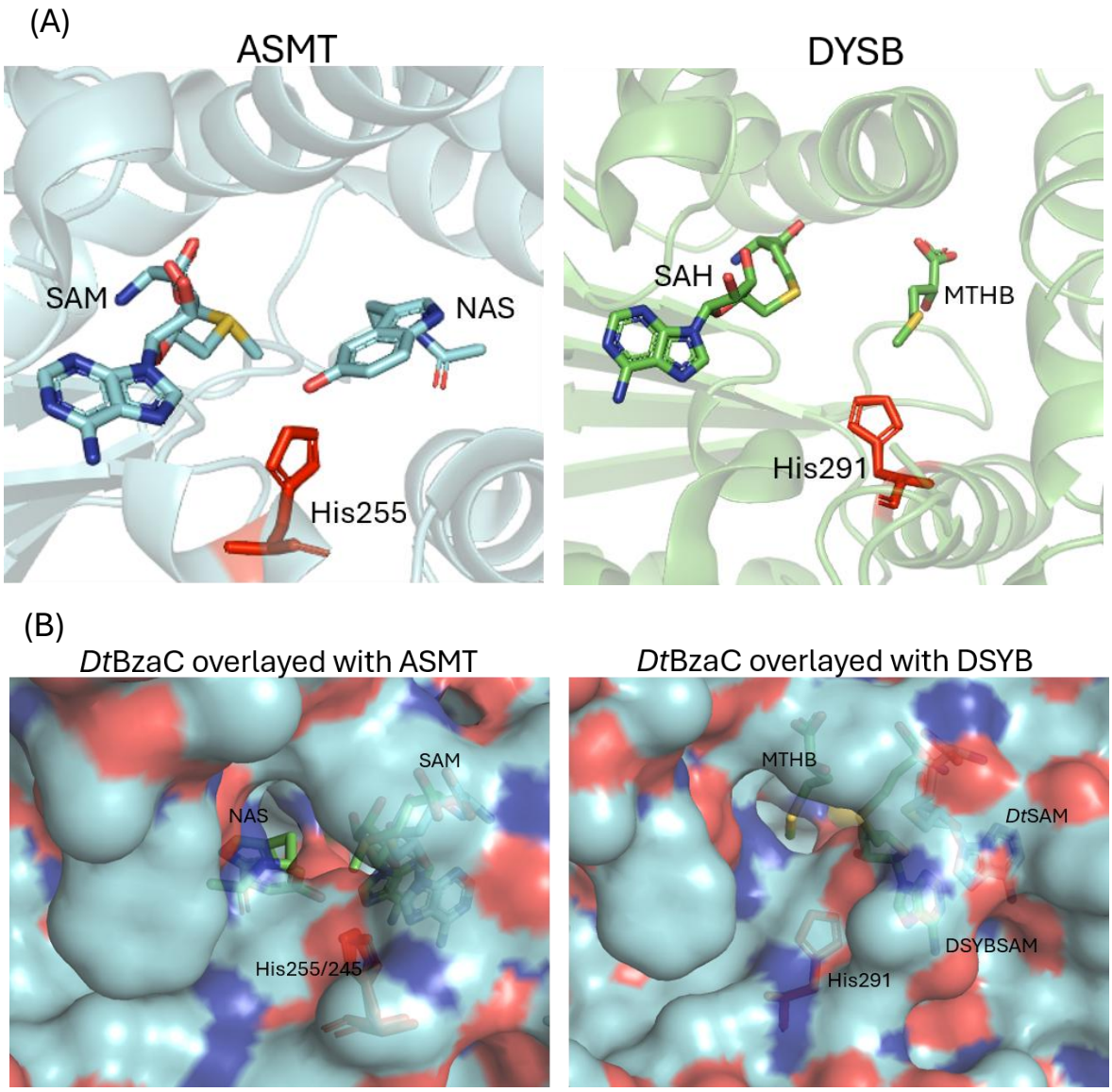


Figure 5.17 - *Dt*BzaC-SAH overlayed with ASMT and DSYB-ligand structures.

(A) Active sites of both ASMT and DYSB with SAM/SAH and NAS/MTHB. **(B)** ASMT and DSYB structures with *Dt*BzaC-SAH overlayed demonstrating the accessibility of the active site in the orientation shown for NAS and MTHB.

5.2.9 Putative mechanism of catalysis

Due to the presence of the potentially catalytic histidine residue in the active site and the absence of waters around this histidine and the SAH sulfonium, it suggests that the enzyme may utilise both proximity and desolvation as well as acid/base chemistry to deprotonate the hydroxyl group of the benzimidazole, increasing its nucleophilicity. This oxygen can then carry out an S_N2-like nucleophilic substitution with SAM's methyl group, producing Ado[5(6)-OMeBza]Cba and SAH (**Figure 5.18**). To test this and investigate if any other residues are required for catalysis, active site mutations (e.g. His 245 to Trp) could be carried out and their impact on the enzyme's activity could be tested. This would differentiate between just proximity and desolvation and acid/base catalysis, as demonstrated by the investigation of DnrK which identified the potential base Tyr 142 wasn't required for catalysis (107).

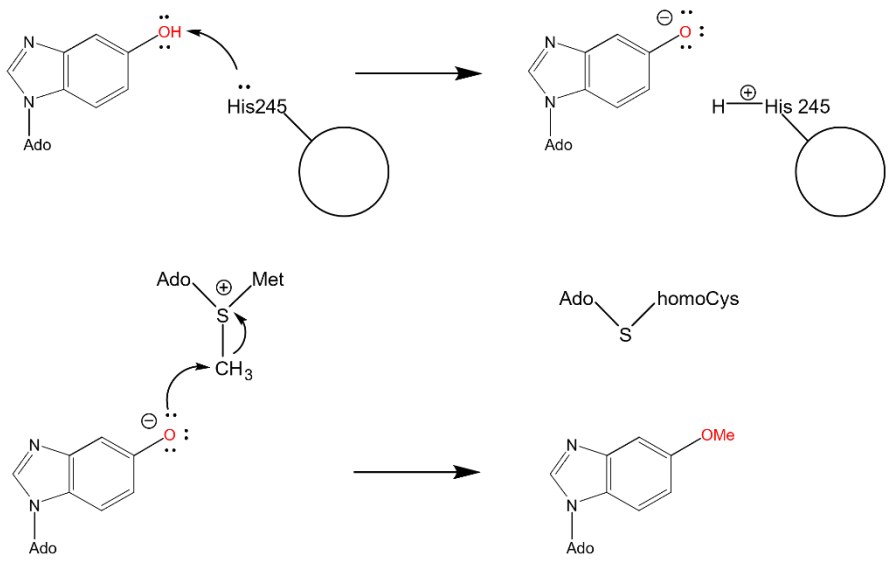


Figure 5.18 - Putative BzaC catalytic mechanism.

His245 may be able to facilitate SAM dependent methyltransfer via deprotonation of the substrate hydroxyl group.

5.3 Discussion

X-ray crystallography is a powerful tool for the investigation of protein structure and function. Utilisation of this tool provided insight into *Dt*BzaC's structure and how it catalyses SAM dependent methyl-transfer to a cobamide substrate. Together the data in this chapter supports BzaC's classification as a class 1 SAM dependent methyltransferase due to its classical seven β-stranded Rossman fold, and SAH binding domain with the conserved GxGxG motif and acidic Asp (D) 197 of β2.

The docking carried out provided two potential modes of cobamide binding, both in a base off conformation which supports the previous analysis of the cobamide upper ligand influence on the base-on / base-off equilibrium. However, both poses presented the substrate hydroxyl in a position too far away from the SAH sulphur to allow for catalysis, either presenting a limitation in the models or suggesting that a conformational change would be required for catalysis.

In addition, comparison to two characterised methyltransferases with high structural similarity (ASMT and DSYB) identified that their substrates are in an alternative orientation, with the hydroxyl groups facing the 'front' of the *Dt*BzaC active site, potentially entering from entrances at the 'back' of the active site. However, docking poses simulating this in *Dt*BzaC were not identified.

This therefore suggests that *Dt*BzaC either has a different mode of substrate binding, which could be expected due to the large difference in substrate size, or the ligand binding models are incorrect and the substrate binds at the 'back' side of the active site and the base enters via a small channel in a base-off conformation. Interestingly,

this channel isn't present in the active site in the 'closed conformation' therefore the flexible loop may cause a conformational change at the 'back' of the active site also, additionally mediating any potential entry from that side of the protein also.

Moreover, a histidine (245) residue located in between the substrate hydroxyl and the SAH was identified, suggesting it may be involved in the mechanism of catalysis by acting as a base to deprotonate the substrate hydroxyl increasing its nucleophilicity, thus promoting methyltransfer.

To be able to draw stronger conclusions about substrate binding and catalysis I suggest that future efforts would focus on obtaining an Apo *Dt*BzaC structure and a *Dt*BzaC structure with SAM or SAH bound with Ado[5(6)-OHBza]Cba, as this would give a more definitive model for how the enzyme binds its substrate and provide insight into any conformational changes that might occur upon substrate and cofactor binding.

Furthermore, mutagenesis of the putative catalytic histidine (His245) and analysis by an endpoint or kinetic based assay would be good to assess its involvement in catalysis. Moreover, obtaining crystal structures of *El*BzaCT and *Mt*BzaC would also shed light on any structural features that may contribute towards the differences in activity observed, which could be useful information for future enzyme engineering efforts to increase enzyme activity or relax/tighten regiospecificity.

Chapter 6:

Characterisation of *Dt*BzaE, a hypothesised B_{12} dependent radical SAM enzyme

6.1 Introduction

6.1.1 Radical SAM enzymes

There are another class of SAM dependent enzymes which use radical chemistry rather than the previously described polar mechanisms. These are called the radical SAM enzymes, with currently ~600,000 members spanning the kingdoms of life (115). These enzymes utilise $[4Fe\text{-}4S]^{+}$ clusters to reductively cleave the C5'-sulfonium bond of SAM, producing a 5'-dexoyadenosine radical (Ado·). This radical, in most cases, then abstracts a hydrogen atom from a substrate unactivated C-H bond, initiating a large variety of radical reactions, such as rearrangements, methylation, or carbon-carbon bond formation between unactivated sp^2 and sp^3-hybridised carbon centres, forming the product (116, 117).

Due to the generation of an adenosyl radical, but the smaller and less energetically intensive route to the biosynthesis of SAM, radical SAM enzymes have been referred to as the poor man's adenosylcobalamin. However, radical SAM enzymes are much more widespread in nature. The enzymes of this family share very limited sequence similarity with the most characteristic being the CX_3CX_2C motif, of which the three cysteine residues coordinate three of the irons of the [4Fe-4S] cluster, with the final iron coordinated by the SAM methionine moiety (118). They also typically contain a full (β/α)8 or partial (β/α)6 triose phosphate isomerase (TIM) barrel fold (115).

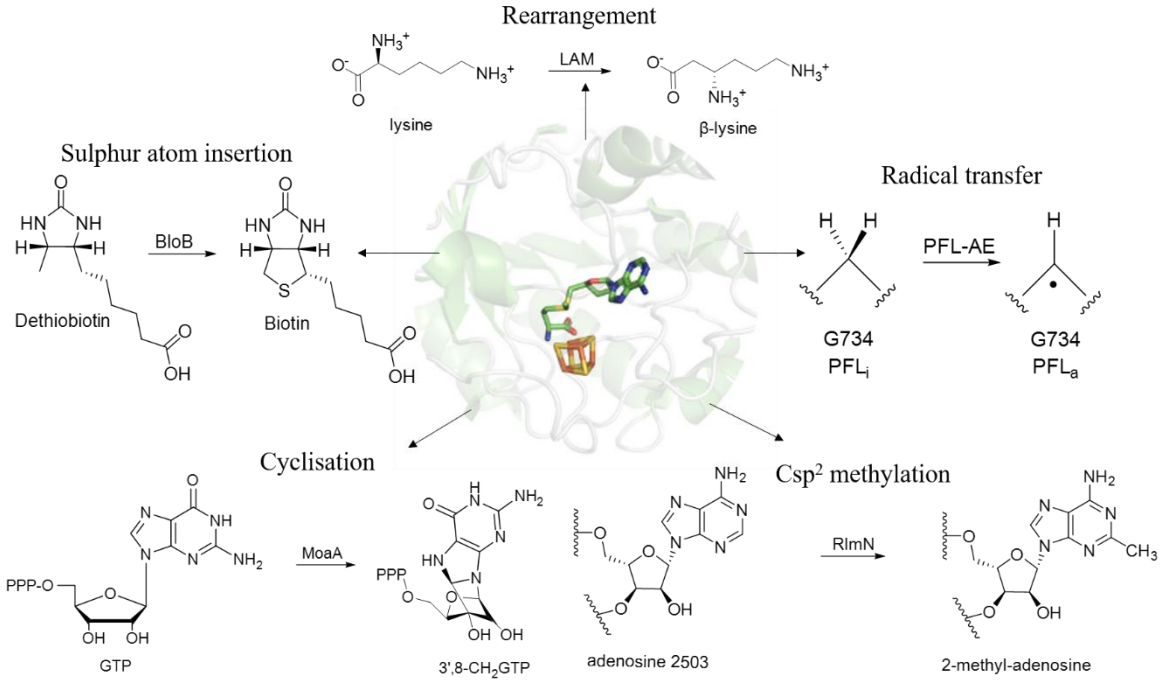


Figure 6.1 - Radical SAM enzyme catalysed reactions.

Schematic representation of the diversity of reactions that radical SAM enzymes can catalyse using [4Fe-4S] clusters and SAM (119–123). The central protein is HydE (a radical SAM enzyme PDB ID = 3IIZ) (123) demonstrating an active site conformation with the iron sulphur cluster coloured yellow and orange and the SAM primarily in green.

6.1.1.1 Generalised mechanism

The enzyme mechanism starts with a $[4Fe\text{-}4S]^{2+}$ cluster utilising SAM to chelate one of the irons. This cluster is then reduced by a single electron (e^-) to produce the activated $[4Fe\text{-}4S]^{+}$ cluster. This reduction in biological systems is usually requires flavodoxin or other single electron donors, in vitro strong reductants such as dithionite or titanium citrate are utilised. The reduced cluster then transfers a single electron to the bound

SAM, homolytically cleaving the S-C(5') bond producing $[4Fe\text{-}4S]^{2+}$-methionine and Ado·. This radical then abstracts a hydrogen from the substrate, producing a substrate radical and AdoH. In some cases, this substrate radical is the end product as in the case of pyruvate formate lyase activating enzyme (PFL-AE), where the product is a glycl radical on PFL. However, usually the substrate radical is just an intermediate which then undergoes a radical based transformation (117) (**Figure 6.2**). In many cases the SAM is used up in the reaction, however in a few cases it has been observed that the product radical can abstract a hydrogen from AdoH, thus regenerating the Ado· which can recombine with methionine to form SAM (118).

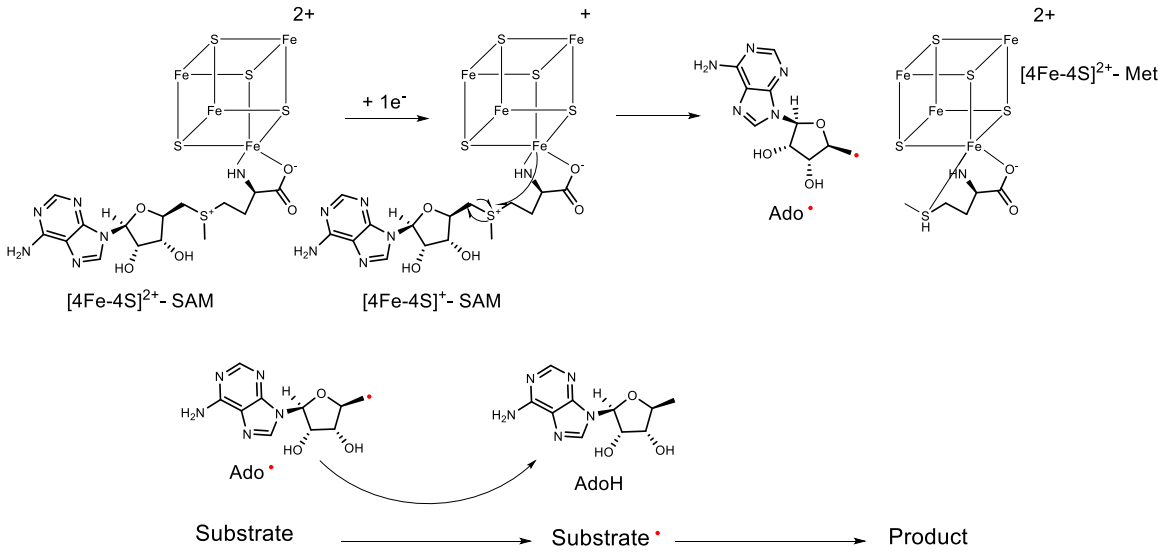


Figure 6.2 - Generalised mechanism of a radical SAM enzyme catalysed reaction.

6.1.2 Cobalamin dependent radical SAM enzymes

There is another class of these radical SAM enzymes which employ cobalamin as an additional co-factor to catalyse their reactions. These were first identified alongside

the other radical SAM enzymes by Sofia et al. 2001 (53), due to having an additional domain with similarity to methionine synthase and malonyl CoA mutase, which are known to bind cobalamin. In this original paper only 25 sequences were identified, including enzymes involved in various biosynthetic pathways such as bacteriochlorophyll (BchE) and oxetanocin (OxsB). However, since then they have grown into a superfamily in their own right with around 200,000 annotated members (115).

Due to challenges associated with studying these enzymes such as solubility, substrate availability and the requirement for strict anaerobic conditions, not too much biochemical detail is known about their mechanisms (54). However, they catalyse a large variety of transformations, usually using radical SAM chemistry and MeCbl to methylate unactivated carbon centres. Other members have been shown to catalyse complex radical rearrangement reactions e.g. OxsB (124) and in one case it has been observed to not use radical chemistry at all (TsrM) (125).

The first two B_{12} dependent radical SAM enzymes to be biochemically characterised were PhpK (126) and TsrM which are P- and a C(sp^2)-methyltransferases involved in bialaphos (a herbicide) and thiostrepton A (an antibiotic). However, since then most of the enzymes of this class that have been studied have been shown to methylate C(sp^3)-hybridised carbons utilising SAM and MeCbl (e.g. Fom3 (127), GenK (128) and Mmp10) (129). This is of high importance as they are thus far the only known enzymes that are able to catalyse these challenging reactions. In addition, it has been shown that some members of the family such as TokK (130) and CysS (131) can catalyse serial methylation of the same carbon centre, producing alkane chains e.g. *t*-butyl groups.

Moreover, some have also been shown to be able to catalyse thioether bond formation (ThnL) (132) and ring contraction (OxsB), utilising OHCbl (hydroxocobalamin).

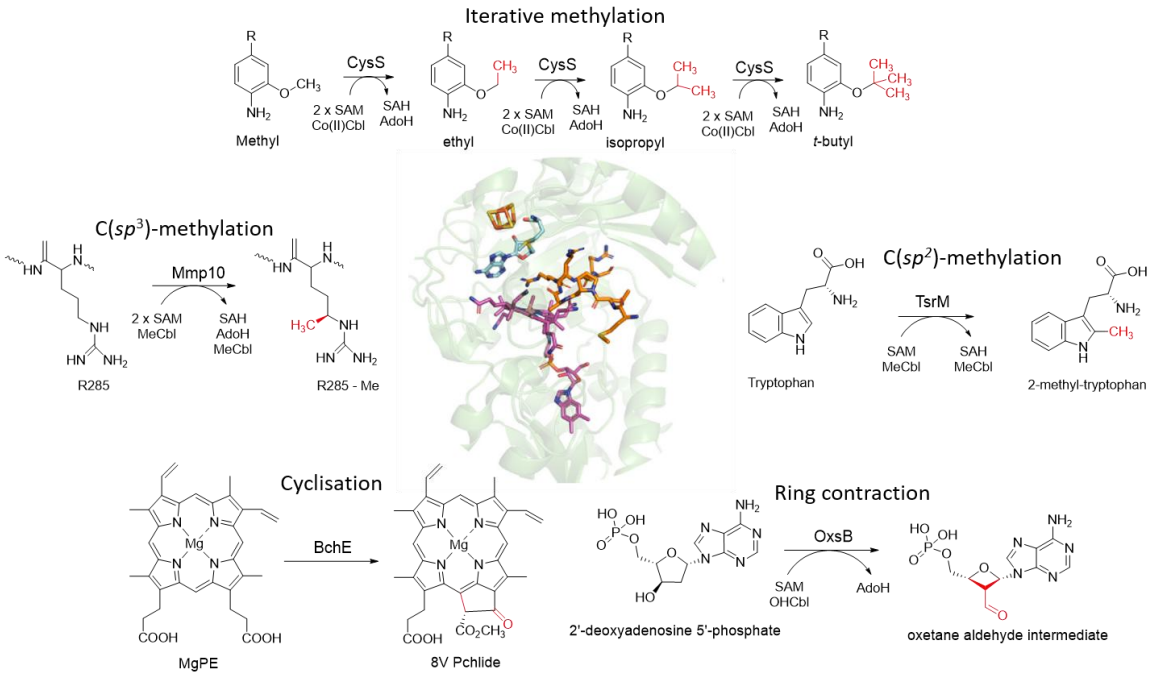


Figure 6.3 - Reactions catalysed by B_{12} dependent radical SAM enzymes.

Example active site of a B_{12} dependent radical SAM enzyme (Mmp10 PDB ID = 7QBS) (129) with the [4Fe-4S] cluster represented in yellow, SAM in blue, methylcobalamin (base-off) in pink, substrate (methyl-coenzyme M reductase peptide chain) in orange. Example reactions demonstrating iterative methylation (131), C(sp^3)-methylation (129), cyclisation (133), ring contraction (124) and C(sp^2)-methylation (125) shown.

6.1.3 *Dt*BzaE

The *Dtbza* operon hosts a gene encoding a BzaE2 enzyme, named 2 due to having only 32.3% identity to *El*BzaE and also 21.3 % homology to *El*BzaD (calculated using

emboss-needle pairwise alignment). Domain architecture analysis using InterPro scan identified a B_{12} binding domain and a Radical SAM domain, suggesting it may belong to the B_{12} dependent radical SAM family of enzymes. However, due to the homology to both *El*BzaD and E, its unsure what reaction it may catalyse.

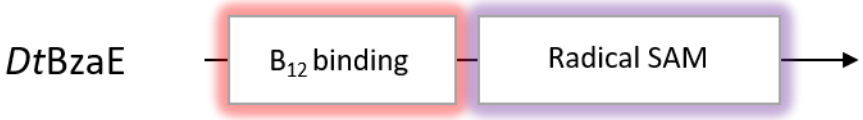


Figure 6.4 - *Dt*BzaE domain architecture.

6.1.4 Chapter aims

Due to the lack of understanding of how B_{12} dependent radical SAM enzymes work, finding enzymes of this family capable of study is important for shedding more light on the family. In addition, due to the difference in the *Dtbza* operon structure compared to *E. limosum* and the lack of characterisation of *D. thermosubterraneus* cobamide production it is unknown what benzimidazolylcobamide it will make. It may present an alternative biosynthetic route for producing adenosylcobalamin or it may produce another benzimidazolylcobamide. Therefore, studying *Dt*BzaE in vitro might provide more detail on how the family of proteins operate and also provide insight into the final product of its pathway.

6.2 Results

6.2.1 Expression and purification of *Dt*BzaE

Firstly, the pET28-*DtbzaE* plasmid (provided by Dr Yamini Mathur) was transformed into BL21 Star (DE3)pLysS-btuB competent cells. 1 L LB cultures of transformed cells were then grown to OD_{600} of 0.6, induced with 400 µM IPTG and incubated overnight at 19°C, 160 rpm (or cultures grown in 2YT and supplemented with 25 µM Fe(III)Cl_2 and 150 µM L-cysteine at induction). Cultures were then harvested and pellets resuspended in buffer. Due to the presence of the iron sulphur cluster, which is highly oxygen sensitive, all steps from this point on containing *Dt*BzaE were carried out in an anaerobic chamber.

Resuspended pellets were sonicated at 60 % amplitude 10 sec on and 50 sec off until a homogenous cell lysate was obtained. The lysates were then transferred to anaerobic centrifuge bottles and centrifuged at 18, 000 rpm, 4-8°C for 20 minutes. *Dt*BzaE was then purified from the lysates via IMAC, and buffer exchanged via PD-10 chromatography. The presence of the predicted [4F-4S] cluster was then assessed via UV-VIS spectroscopy, which showed a large peak at 280 nm a very small bump at ~ 420 nm, demonstrating a small amount of intact cluster and the need for chemical reconstitution. Furthermore, the protein was diluted to a 280 nm absorbance of ~ 1 and quantified by UV-VIS spectroscopy.

6.2.2 Iron sulphur cluster reconstitution

To reconstitute the iron sulphur cluster, the protein cysteines were reduced with 10 mM DTT and sequentially supplemented with iron in the form of Fe(II) citrate and sulphur via lithium sulphide. The reconstituted protein was the purified via PD-10 chromatography and analysed via UV-VIS spectroscopy. This demonstrated an increase in the ~ 300-600 nm signal ($A_{280/300}$ pre reconstitution = 4.7, post reconstitution = 2.6) demonstrating successful reconstitution. The protein concentration was recalculated at this point to account for any loss during reconstitution and re-purification.

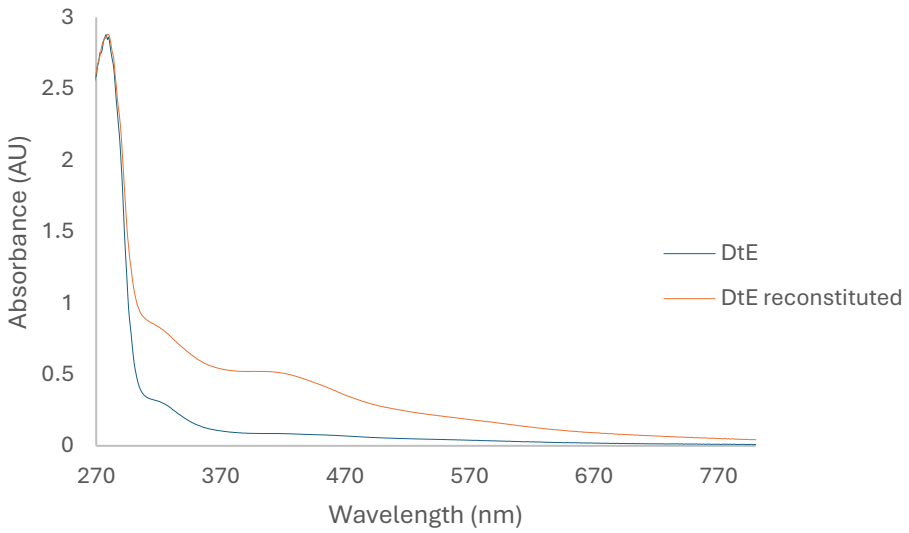


Figure 6.5 - *Dt*BzaE iron sulphur reconstitution.

UV-VIS spectra of *Dt*BzaE pre reconstitution (blue) and post reconstitution (orange).

6.2.3 Purity assessment

To assess the purity of the protein at each stage of purification and 12 % SDS-PAGE was carried out with samples of each 2 mL elution (EB), post first PD10 column and 5 µg of reconstituted protein (Rc). This demonstrated that a protein of the correct molecular weight (expected = 61 kDa) was purified and still intact post reconstitution. The relative level of purity was not high, but due to not having easily accessible anaerobic gel filtration it was accepted and taken forward for further experiments. Furthermore, the presence of mainly lower MW bands in the reconstituted and post PD-10 samples suggest that the contaminants may mainly be degradation products.

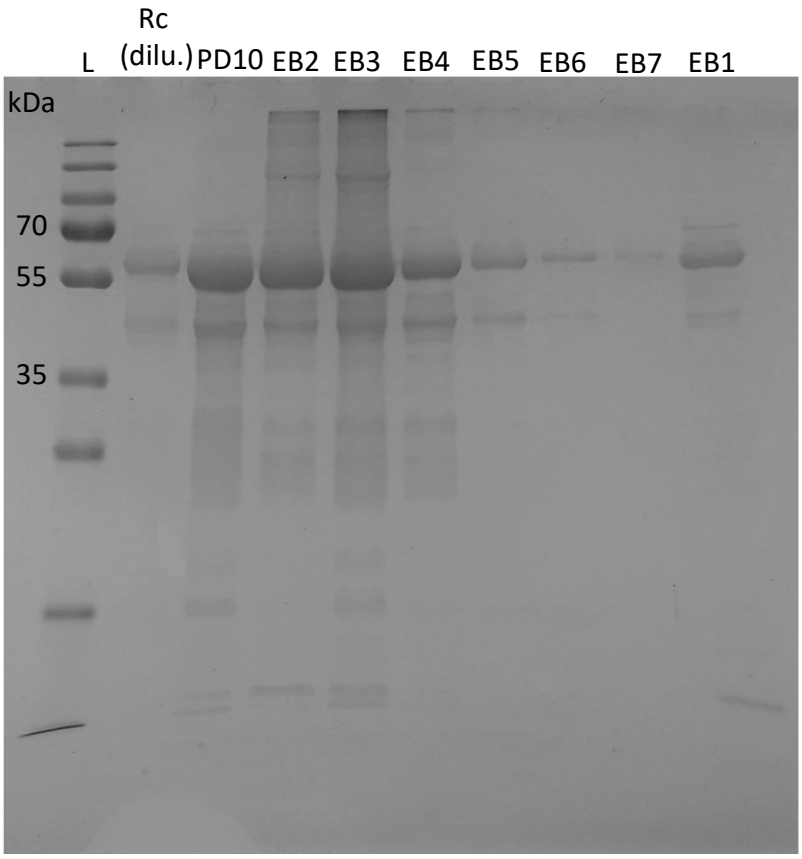


Figure 6.6 - *Dt*BzaE SDS-PAGE.

An image of a 12 % SDS-PAGE of protein samples collected at different stages of purification. Rc (dilu.) = diluted reconstituted protein, PD10 = post PD10 and EB = each elution fraction.

6.2.4 Ado[5(6)-OMeBza]Cba production by guided biosynthesis

As the product of *Dt*BzaC was Ado[5(6)-OMeBza]Cba and the function of *Dt*BzaX (DUF domain containing protein) was unknown we first wanted to test if *Dt*BzaE was able to catalyse any reaction with Ado[5(6)-OMeBza]Cba. To do this we needed to generate suitable quantities of it, thus we turned to the method of guided biosynthesis. Firstly, production was optimised using the method followed in chapter three, however this

time using 0, 2.5, 5, 10, 20 and 40 mg/L 5-methoxybenzimidazole (5-OMeBza). This identified that the optimal concentration for supplementation was 20 mg/L. Production was then scaled up and purity was assessed as before, showing a purity at 360 nm of ~ 94% (**Supplementary figure 5**). Furthermore, substrate was subjected to isomer analysis we developed in chapter 4, which demonstrated that the samples contain 65 % 5- and 35 % 6- isomer.

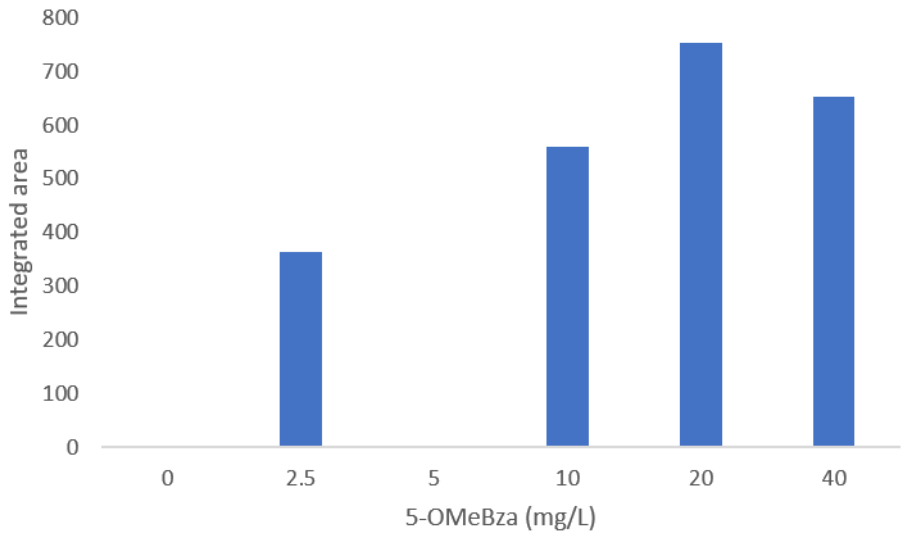


Figure 6.7 - Optimisation of the production of Ado[5(6)-OMeBza]Cba.
Integrated area under the curve of hydroxy[5(6)-OMeBza]Cba produced from the supplementation of ED674 cultures with 0, 2.5, 5, 10, 20 and 40 mg/L 5-OMeBza.

6.2.5 *Dt*BzaE activity assays

To test the activity of *Dt*BzaE a reaction containing 10 µM reconstituted *Dt*BzaE, 6 mM dithiothreitol (DTT), 20 µM Ado[5(6)-OMeBza]Cba, 1 mM SAM, 100 µM hydroxy or methyl-cobalamin in 20 mM Tris-HCL pH 8, 100 mM NaCl (and controls lacking each

component individually) were set up and incubated at 25°C for 18 hours. Reactions were then quenched using a final concentration of 1 % glacial acetic acid and analysed by HPLC (method 4) and LC-MS (LC method 9). This showed the broadening of a peak at ~ 20 minutes with the addition of a small shoulder peak in the reactions containing enzyme but lacking in each control apart from those without hydroxy or methyl-cobalamin (**Figure 6.8, A**). This thus demonstrated enzymatic product formation. In addition, activity in the absence of a cobalamin co-factor (hydroxy or methyl) suggests the enzyme may be a radical SAM enzyme rather than a B_{12} dependent radical SAM, or the substrate may also act as the cobamide cofactor. Furthermore, HPLC analysis using method 2, identified the presence of both 5-deoxyadenosine (5'-dA / Ado) and SAH demonstrating the enzyme must require a 5'-dA radical (5'-dA· / Ado·). However, the SAH is present in a no enzyme control, were as 5'-dA is not, suggesting the SAH produced is just from the DTT in the reaction and not from the enzymatic reaction.

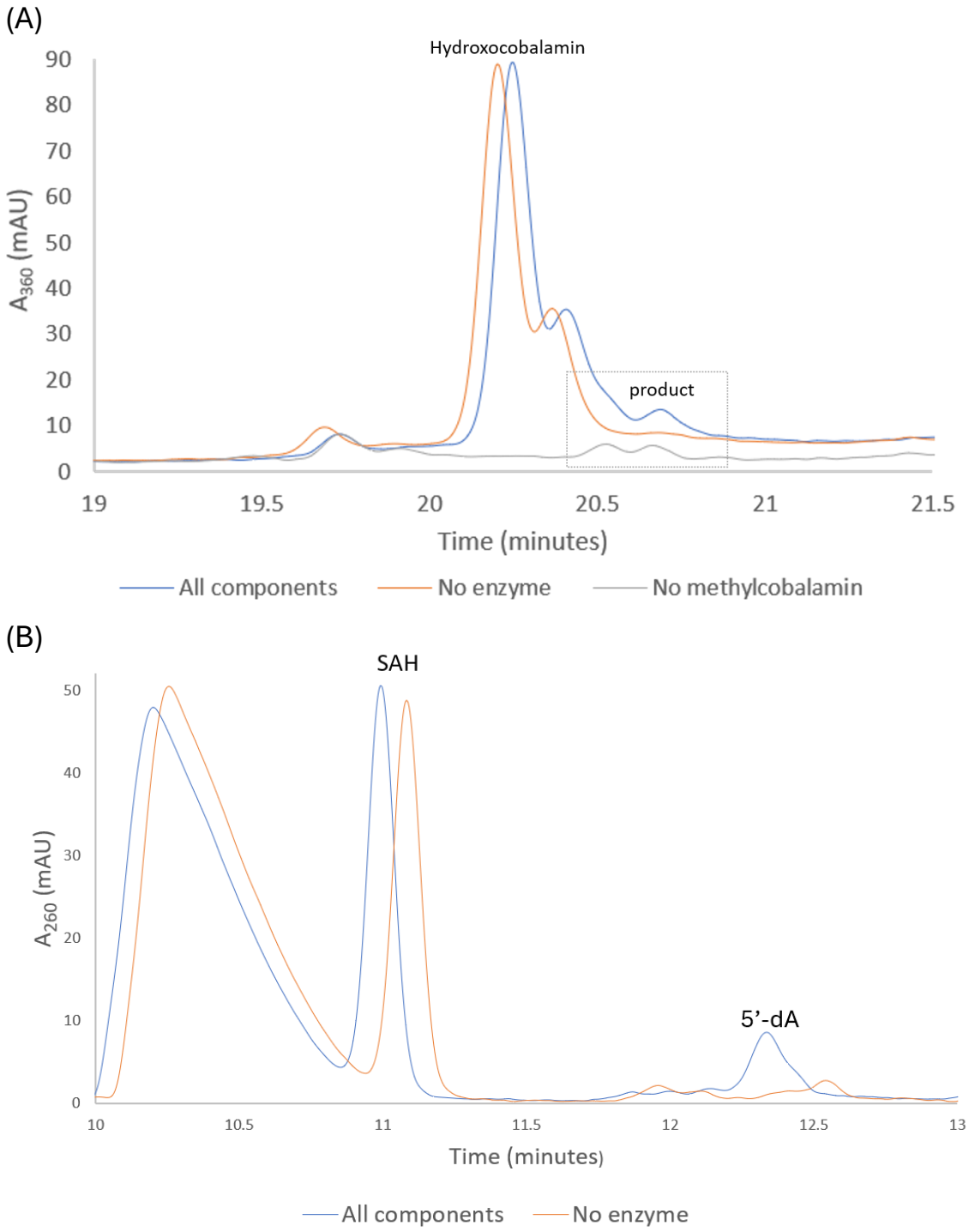


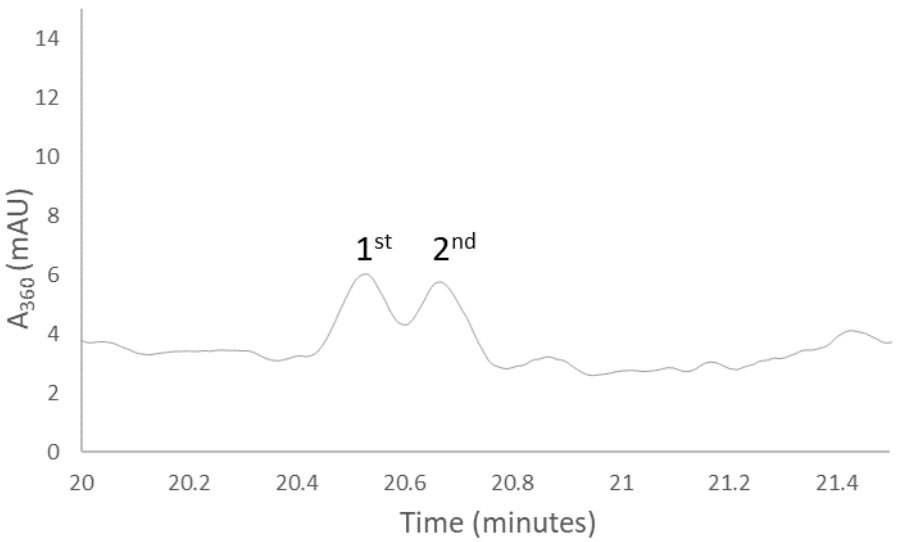
Figure 6.8 - *Dt*BzaE activity assays.

(A) HPLC trace at 360 nm of *Dt*BzaE reactions either with all components, or without enzyme or methylcobalamin. **(B)** HPLC trace at 260 nm of *Dt*BzaE reactions with and without enzyme.

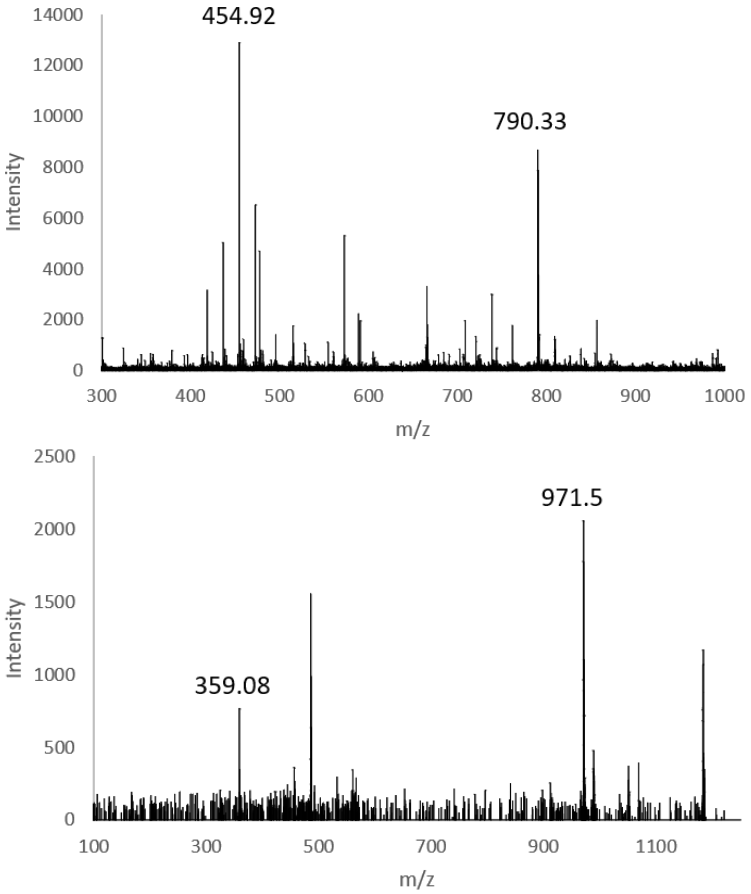
Furthermore, LC-MS analysis of the corresponding product peaks shown in **Figure 6.8**, A showed the additional shoulder contained a m/z of 791. 3, which is the same as the substrate mass, suggesting the enzyme catalyses a rearrangement reaction. Furthermore, when LC-MS analysis was applied to the reactions lacking a cobalamin co-factor it made identification of the peaks easier due to the absence of the hydroxocobalamin (either present due to addition to the reaction or from light hydrolysis of methylcobalamin). This identified that the first peak contained a m/z of 790.33 (correlating to the expected $[M+H]^{2+}$ of adenosylcobalamin) and the second peak contained the m/z of 791.3.

It is unlikely that *Dt*BzaE produced DMB from 5-OMeBza, however peaks with similar retention times to this putative adenosylcobalamin are only present in reactions containing all components or the one lacking a cobalamin co-factor. Running an adenosylcobalamin standard will help to identify this product, as the reduced polarity (exemplified by its reduced retention times) is unexpected for adenosylcobalamin, suggesting that it may be another cobamide with the same mass. Furthermore, carrying out reactions containing no cobalamin co-factor and controls lacking the other components individually may help with identifying if this is an enzymatic product, as there will be no cobalamin co-factor present to interfere with the LC-MS analysis.

(A)



(B) 1st peak MS and MS/MS



(C) 2nd peak MS and MS/MS

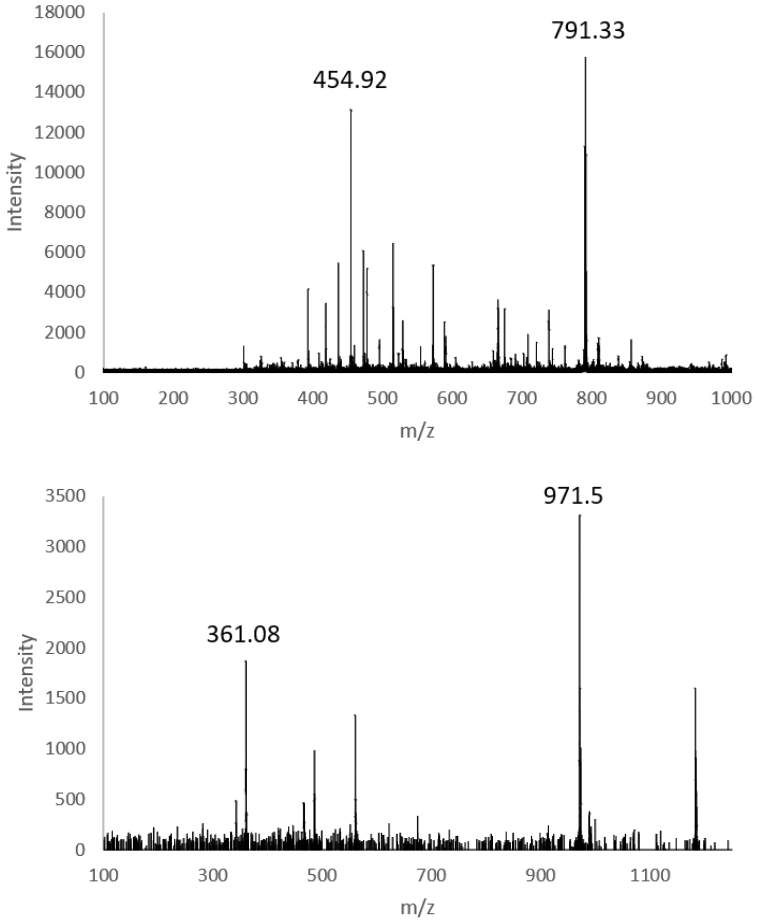


Figure 6.9 - *Dt*BzaE products MS and MS/MS analysis.

(A) HPLC chromatogram of a *Dt*BzaE reaction in the absence of a cobalamin co-factor. **(B)** MS analysis of the 1st product peak and MS/MS fragmentation of the 790.33 ion. **(C)** MS analysis of the 2nd product peak and MS/MS fragmentation of the 791.33 ion (putative rearrangement product).

There are three feasible options for the rearrangement product formed; 1: *Dt*BzaE catalyses the migration of the 5-OMe methyl group to the benzimidazole 6- carbon, leaving a hydroxyl group at 5- producing Ado[5-OH, 6-MeBza]Cba or 2: leaving a keto group at the 5- position producing Ado[5-keto, 6-MeBza]Cba or 3: rearranges the methoxy group to produce a hydroxymethyl (CH_2-OH) group at the 5- position

producing Ado[5-CH_2-OHBza]Cba (**Figure 6.10**). The third option has been suggested as a potential product for another BzaE in unpublished work by Olivier Berteau's group (our collaborators).

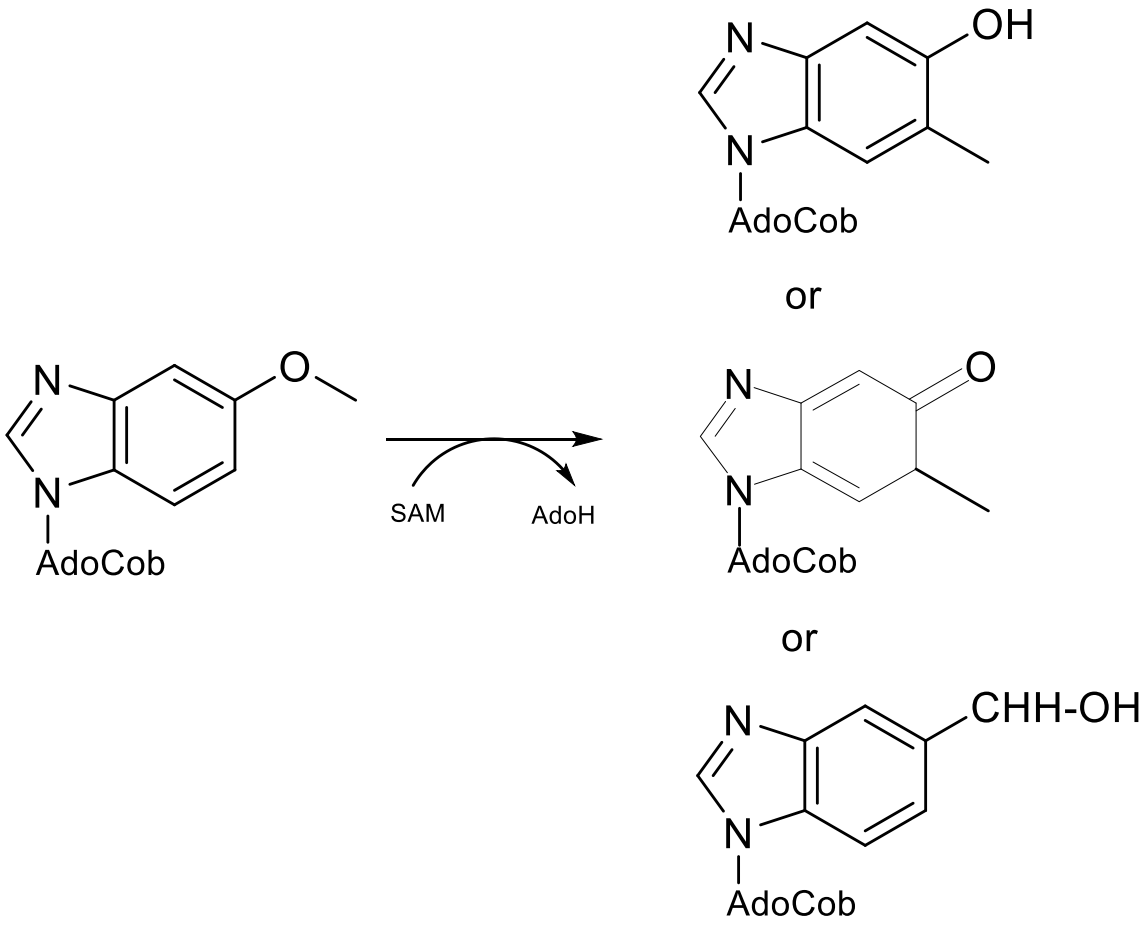


Figure 6.10 - Schematic representation of the *Dt*BzaE reaction, with the three potential products.

Both Ado[5(6)-CH_2-OHBza]Cba (**Supplementary figure 6**) and Ado[5-OH, 6-MeBza]Cba (**Supplementary figure 7**) were then produced via our method of guided biosynthesis using the commercially available 5-CH_2-OHBza and 5-OH, 6-MeBza synthesised by Aniket Vartak from the Hazara group. These were then run as standards on the LC-MS and it was observed that Ado[5(6)-CH_2-OHBza]Cba aligns closely with

the *Dt*BzaE rearrangement product peak (**Figure 6.11**) suggesting that this is the likely product. However, to elucidate the correct structure of the product, it would require isolation and NMR or crystallisation. To allow for suitable amounts of product formation for characterisation the reaction requires optimisation to increase its yield.

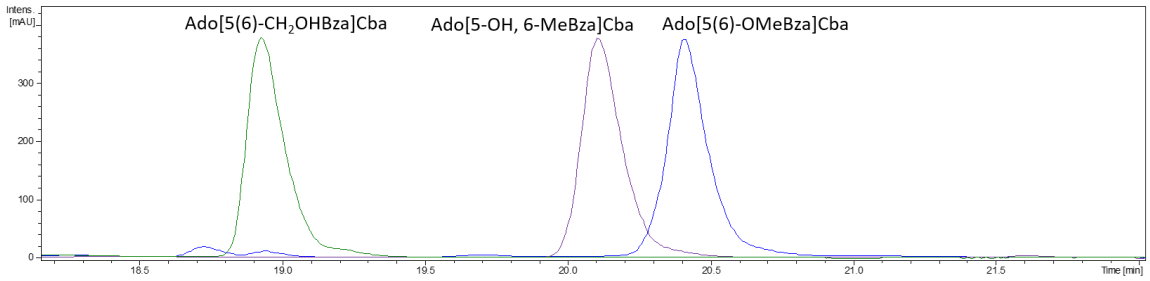


Figure 6.11 - Potential *Dt*BzaE product standards.

HPLC chromatogram of a *Dt*BzaE reaction (blue) overlayed with an Ado[5(6)-CH_2-OHBza]Cba (green) and Ado[5-OH, 6-MeBza]Cba (purple) standard.

6.2.6 Reaction optimisation

Due to the low level of conversion of substrate to product it suggests that either the reaction is not very favourable, or the reaction set up was not optimal. Furthermore, optimisation of product formation would make the identification of the product(s) much easier and more reliable. To try and optimise the reaction conditions a concentration gradient of DTT was applied to test whether a more optimal reducing agent concentration is required. This identified that 6 mM was better than the lower concentrations either suggesting a higher concentration is required or a single electron donor such as titanium (III) citrate (TiC), dithionite or a flavodoxin system is required for optimal reduction of the iron sulphur cluster (**Figure 6.12**).

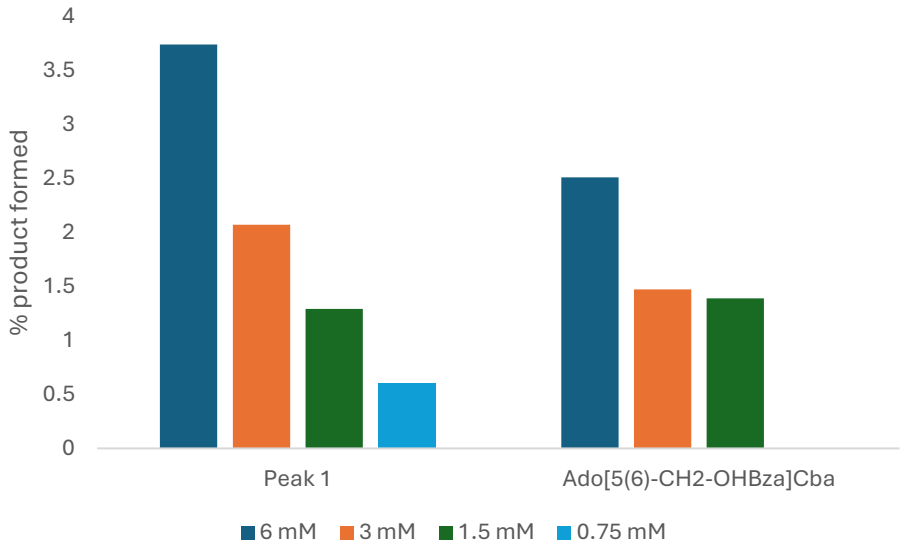


Figure 6.12 - *Dt*BzaE [DTT] optimisation.

A bar chart representing the percentage product formed (peak 1 or Ado[5(6)-CH_2-OHBza]Cba) at four different concentrations of DTT (6, 3, 1.5 and 0.75 mM).

Additionally, we tested the impact of the source of sulphur in the reconstitution method on iron sulphur cluster formation and the enzyme reactivity. To do this purified *Dt*BzaE was split into two equal parts and the reconstitution method was followed simultaneously, however 1 part received lithium sulphide (LiS) and the other sodium sulphide (NaS). This produced protein with similar iron sulphur cluster levels ($A_{280/300}$ LiS = 3.13 and NaS = 3.08) and similar reaction levels (**Figure 6.13**).

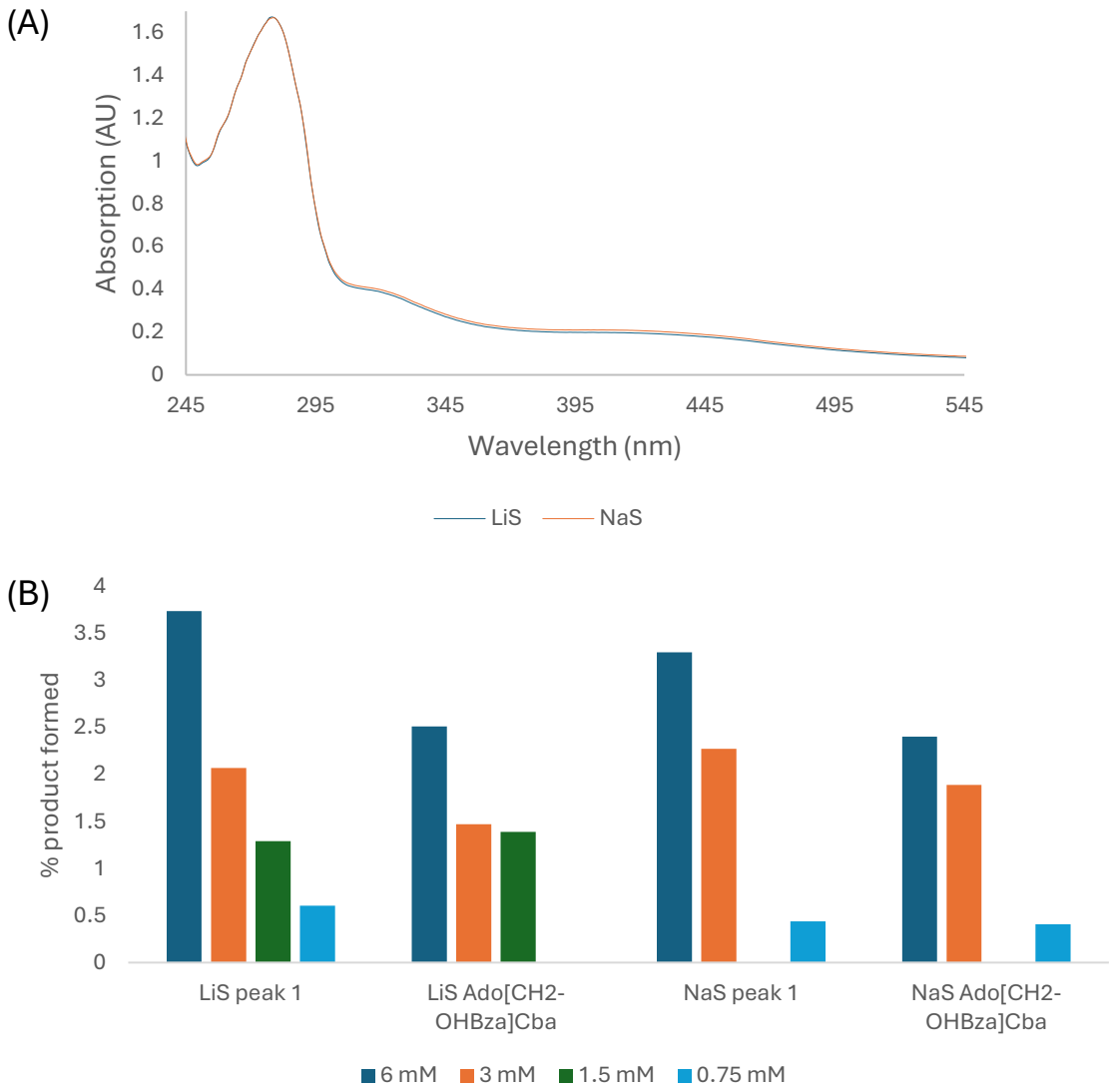


Figure 6.13 - LiS vs NaS as the source of sulphur for iron sulphur cluster reconstitution.

(A) UV-VIS spectra of *Dt*BzaE reconstituted with LiS or NaS, absorption corrected to be equivalent at 280 nm. **(B)** A bar chart representing the % product formed (either peak 1 or Ado[5(6)-CH_2-OHBza]Cba) for *Dt*BzaE reconstituted with LiS or NaS at each concentration of DTT (6, 3, 1.5 and 0.75 mM).

6.2.7 *Dt*BzaX

As previously shown in **Figure 3.6**, the *Dtbz*a operon contains genes encoding BzaA, B, C, X and E enzymes. The homology of BzaA and B to ThiC suggest that they should convert AIR to 5-OHBza. The enzymology thus far in this thesis then suggests that *Dt*BzaC should produce Ado[5(6)-OMeBza]Cba from Ado[5(6)-OHBza]Cba. This can then be acted upon by *Dt*BzaE to produce Ado[5(6)-CH_2-OHBza]Cba (and potentially another product). The only remaining enzyme in the pathway is the DUF 2284 domain (iron sulphur cluster containing) containing protein *Dt*BzaX.

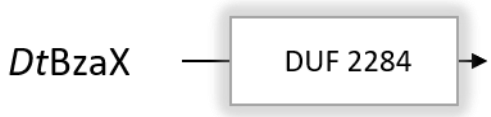
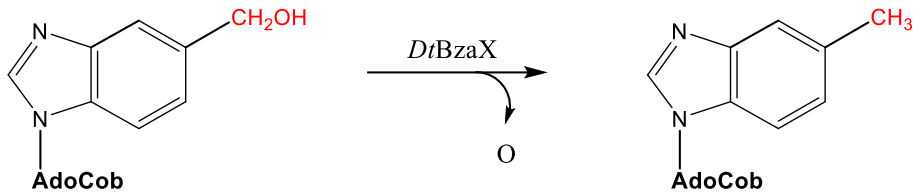


Figure 6.14 - *Dt*BzaX domain architecture.

Work carried out by Amira Abood expressed and purified this recombinant protein anaerobically, reconstituted its iron sulphur cluster(s) and carried out enzyme assays to characterise its reactivity. This identified that upon incubation with Ado[5(6)-CH_2-OHBza]Cba an additional peak with an increased retention time was identified. Mass spectrometry analysis of this peak identified a molecule with a m/z of 783.5 and MS/MS fragmentation identified a lower ligand mass of 345.1. This suggests that the product is Ado[5(6)-MeBza]Cba which has an expected mass of 783.325 ($[M+H]^{2+}$) and lower ligand of 345.08 ($[M+H]^{+}$). All together suggesting that the pathway produces Ado[5(6)-

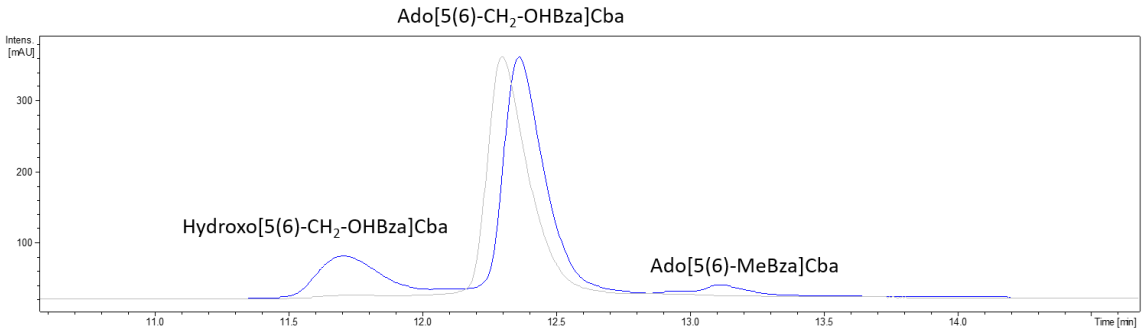
MeBza]Cba as its end product rather than the adenosylcobalamin produced by *E. limosum*.

(A)

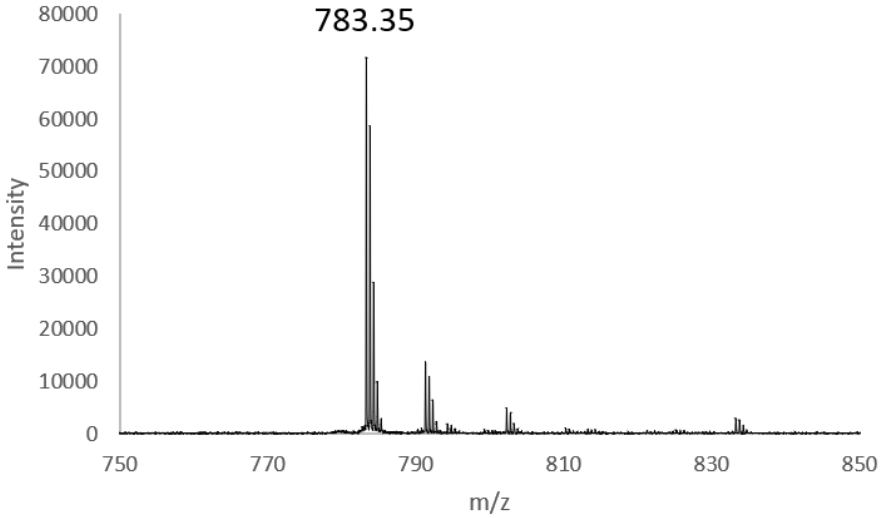


Ado[5(6)-CH_2OHBza]Cba Ado[5(6)-MeBza]Cba

(B)



(C)



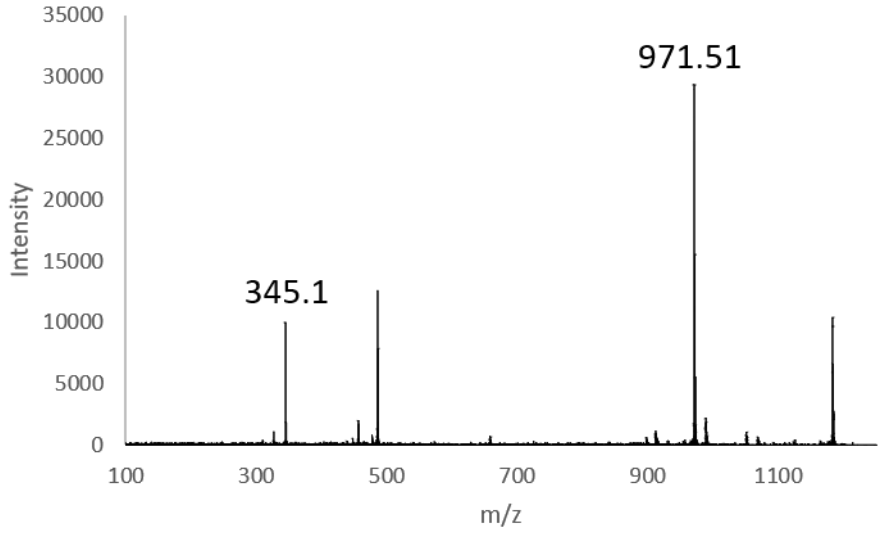


Figure 6.15 - LC-MS analysis of a *Dt*BzaX reaction with Ado[5(6)-CH_2-OHBza]Cba.

(A) Reaction scheme. **(B)** HPLC chromatogram. Blue trace = with *Dt*BzaX, grey trace = without *Dt*BzaX. **(C)** MS and MS/MS analysis.

6.3 Discussion

The work presented in this chapter present some preliminary findings characterising *Dt*BzaE shedding light into its catalytic function and role within the biosynthetic pathway. The main findings are that *Dt*BzaE could be a radical SAM enzyme (produces 5'-dA) rather than a B_{12} dependent radical SAM enzyme, suggesting the characterised B_{12} binding domain is for substrate rather than co-factor, or the substrate also acts as a cofactor. Furthermore, it catalyses a radical rearrangement reaction as a product with the same mass but with a reduced retention time was identified. This produced 3 potential products: Ado[5-OH, 6-MeBza]Cba, Ado[5-Keto, 6-MeBza]Cba or Ado[5(6)-CH_2-OHBza]Cba. However, the product has a similar retention time to Ado[5(6)-

CH_2OHBza]Cba suggesting that this is the likely product, however further analytical chemistry is required to confirm this.

Moreover, another product was identified with a m/z value correlating with the expected m/z of adenosylcobalamin ($[M+H]^{2+}$), however this product was not expected. Furthermore, it has a reduced polarity which is not expected for adenosylcobalamin, suggesting that it may be another product. Further controls and analysis are required to determine what the product is and if it is enzymatic.

Due to the low activity observed (maximally ~3.5 % for Ado[5(6)-CH_2-OHBza]Cba) it suggests that the reaction is not optimised. Furthermore, as the optimisation of DTT and sulphide source didn't have an impact on *Dt*BzaE activity, I would suggest future experiments test the applicability of a signal electron donor reducing agent for improving reactivity.

In addition, the *Dt*BzaX studies suggest that it is able to catalyse the reduction of Ado[5(6)-CH2-OH]Cba (removal of an oxygen) to produce Ado[5(6)-MeBza]Cba in the presence of the reducing agent DTT. Together with the characterisation of *Dt*BzaC and E this suggests that the pathway starts with the production of 5-OHBza from AIR. This is then activated (by CobT/U) and attached to cobinamide producing Ado[5(6)-OHBza]Cba. *Dt*BzaC is then able to catalyse the SAM dependent O-methylation of the 5(6)-hydroxy group to produce Ado[5(6)-OMeBza]Cba. *Dt*BzaE is then capable of catalysing the rearrangement of the methoxy group to produce the hydroxymethyl group of Ado[5(6)-CH_2OHBza]Cba. *Dt*BzaX is then able to catalyse the reduction of the hydroxymethyl group, leaving Ado[5(6)-MeBza]Cba.

Future work should be carried out to explore the mechanism of catalysis for both *Dt*BzaE and X to provide an increased understanding of how they catalyse their interesting reactions. In addition, to confirm the pathway operates in the fashion described, reactions containing *Dt*BzaC, X and BzaE should be carried out and the products characterised to see if Ado[5(6)-MeBza]Cba is produced. In addition, culturing the host bacterium and identifying the type of cobamide it synthesises may help confirm the results of the chapter.

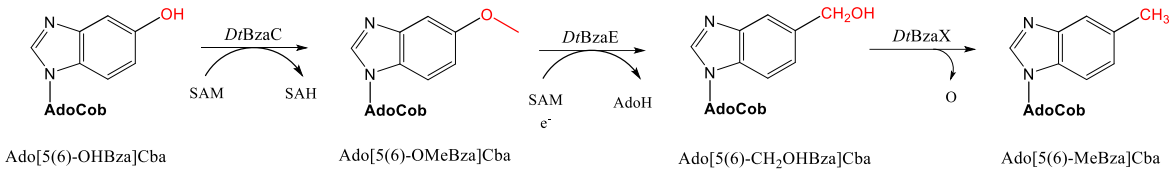


Figure 6.16 - Potential products of the *Dtbza* operon.

General discussion

The work carried out in this thesis set out to investigate and address some open questions regarding the anaerobic biosynthesis of vitamin B_{12}'s α-ligands. This is an important area of research due to being one of the larger gaps left unsolved in the biosynthesis of vitamin B_{12} and contains the potential to provide novel fundamental biochemical insight into how enzymes and biosynthetic pathways work.

The first, and possibly one of the most important, results of the work carried out in this thesis was the identification that BzaC enzymes are capable of catalysing the SAM dependent O-methylation of a cobamide substrate, namely Ado[5(6)-OHBza]Cba. Furthermore, it appears that this substrate is preferred over the free base (5-OHBza) and the riboside intermediates 5- and 6-OHBza-R, due to the low percentage 5-10 % conversions observed in Mathur et al. 2020 and the ability of *Dt*BzaC to catalyse 100 % conversion with a cobamide substrate, with the lowest being around 20 % by *El*BzaCT. In addition, when the adenosyl group, the group natively present during cobalamin biosynthesis, was replaced with either a hydroxyl or cyano group a lower activity was observed which further supports the preferred substrate in the biosynthetic pathway to be an adenosylated cobamide. Moreover, the *Dt*BzaC kinetics analysis identified that it has a similar affinity for its substrate (Ado[5(6)-OHBza]Cba) and SAM as other characterised SAM dependent methyltransferases involved in cobalamin biosynthesis (CobA and CobM) which further supports the argument.

The experiments carried out in this thesis are not completely comparable to Mathur et al. 2020's due to difference in assay set up, therefore, to fully confirm BzaC's preference for the cobamide substrate the riboside intermediates could be produced and the kinetics of the reactions could be monitored in the same set up as utilised in chapter four of this thesis. This would obtain comparable K_m, k_{cat} and catalytic efficiency values which could be compared to Ado[5(6)-OHBza]Cba to be conclusive.

Furthermore, this finding is very useful for coming to study the more complex putative B_{12} dependent radical SAM enzymes identified later in the biosynthetic pathways, BzaD, BzaX and BzaE. Due to having to carry out assays in an anaerobic chamber and issues with production and solubility of this class of enzyme, these enzymes already present difficulties for study. Therefore, by presenting them with a more preferred substrate, larger amounts of product can be produced, which makes characterisation of their mechanisms, pathway intermediates etc significantly easier and produces a better representation of what occurs with the cell. Moreover, this class of enzyme are often found within natural product biosynthetic pathways, which can sometimes make identification and isolation of their substrates difficult. Due to the work carried out by Dr Evelyne Deery making the engineered strain ED674 capable of producing various benzimidazolylcobamides at good yields, it makes characterisation of putative B_{12} dependent radical SAM enzymes in Bza and Bza-like pathways achievable due to the ease of producing various potential substrates and products.

The identification that a cobamide is likely the preferred substrate of BzaC also speaks to the timing of attachment of the ribotide intermediate to the cobinamide in the anaerobic pathways. It suggests that 5-OHBza produced from the action of BzaAB/F is

activated by CobT/U and is then attached to cobinamide to produce the cobamide which is then acted upon later enzymes in the pathway (**Figure 3.20**).

Furthermore, *Dt*BzaC is quite promiscuous with its ability to utilise SAM analogues (Allyl-SAM and Propargyl- SeMet) to produce cobamide analogues harbouring these functional groups at the 5- / 6- oxygen. The ability of these analogues to then be attached to azide containing cargo using click chemistry presents a novel route for the production of 'clickable' cobamide analogues. Future experiments should explore reaction scalability and their applicability in cell systems to track uptake / deliver cargo to assess potential applications.

The work presented in chapter four provided insight into the enzymes mechanism of inhibition, regioselectivity and provided further experimental support for the validity of the cobamide substrate. Firstly, the significantly tighter affinity of SAH over SAM and lack of binding observed for MTA showed that the true inhibitor of BzaC activity is SAH. Interestingly, removal of SAH by the activity of *Ec*MTAN only increased the lower conversion of *Mt*BzaC by 15.6 % and had no influence on the *El*BzaCT reaction, thus suggesting there are other factors limiting their activity. This could potentially be explained by the results of the regioselectivity analysis developed in chapter 4. This method identified that our way of biosynthesising cobamides produced both 5- and 6-isomers and that *Dt* and *Mt*BzaC both show a preference for the 5- isomer of Ado[5(6)-OHBza]Cba, thus suggesting that *Mt*BzaC and *El*BzaCT may shower a weaker affinity for the 6- isomer, thus reducing their activity compared to *Dt*BzaC.

Moreover, obtaining the crystal structure of *Dt*BzaC provided great insight into how it operates as a methyltransferase but also left some unaddressed questions. Firstly, it

showed it belongs to class I of the SAM dependent methyltransferase family and also presents a very classical mode of co-factor binding. However, due to not obtaining a crystal structure of the protein bound to both substrate and cofactor, its cobamide binding domain, and any substrate binding induced conformational changes could not be determined, leaving uncertainty with potential catalytic mechanisms.

Obtaining a crystal structure with the cobamide bound would provide an understanding of how this protein, which contains no bioinformatically identified cobamide binding domain, binds its substrate. Molecular docking was utilised to try and address some of these questions and two reasonable poses were produced, however due to too large distances between the 5-OH group and SAH, it would suggest these binding modes aren't quite correct or a conformational change is required for catalysis. Therefore, a substrate bound crystal would be fantastic to investigate this further.

Furthermore, a substrate/cofactor co-crystal may provide insight into the mechanism of catalysis as substrate binding may or may not induce a conformational change bringing the substrate and co-factor closer together with the potential catalytic base (His245), compared to the docking model. However, to confirm if this residue is involved in catalysis mutation of the residue to a non-basic amino acid and subsequent activity assays should be completed.

The work carried out exploring *Dt*BzaE and *Dt*BzaX also provided biochemical evidence for a Bza-like pathway producing a non-DMB benzimidazolylcobamide (Ado[5(6)-MeBza]Cba). Furthermore, it shows a derivation from the original biosynthetic scheme described by Amrita et al. 2015 with the production of Ado[5(6)-CH_2OHBza]Cba from

Ado[5(6)-OMeBza]Cba rather than the expected Ado[5-OMe, 6-MeBza]Cba and then subsequent removal of the oxygen by *Dt*BzaX which was suggested to be carried out by a BzaE enzyme.

All together this allows us to produce a potential biosynthetic route in which 5-OHBza is activated by CobT/U to produce 5-OHBza-RP which can be utalised by CobV for the displacement of GDP at the lower ligand position of Adenosyl-GDP-Cobinamide producing Ado[5(6)-OHBza]Cba. BzaC can then catalyse the O-methylation of the substrate Ado[5(6)-OHBza]Cba hydroxyl group to produce Ado[5(6)-OMeBza]Cba. In the case of the *Dt*Bza pathway, *Dt*BzaE can then catalyse the rearrangement of the Ado[5(6)-OMeBza]Cba methoxy group to produce Ado[5(6)-CH_2OH]Bza. *Dt*BzaX can then catalyse the reduction of the hydroxymethyl group to produce Ado[5(6)-MeBza]Cba. In the case of the *E. limosum* pathway another biosynthetic scheme must be followed which results in the production of Adenosylcobalamin (**Figure 7.1**).

The findings of this work also suggest that the original *E. limosum* Bza pathway may not operate in the originally suggested way, as the *El*BzaC (DUF) domain may act like *Dt*BzaX (due to being the same class of DUF (DUF2284)) which is supported by Dr Amira Abood's work which demonstrated that the full length *El*BzaC (including DUF domain) did not increase *El*BzaC's activity compared to the *El*BzaCT (data not shown) suggesting it does not participate in the O-methylation reaction. Furthermore, either *El*BzaD or *El*BzaE may catalyse a similar rearrangement reaction to *Dt*BzaE, thus providing a different chemical route to DMB.

Future experiments characterising a variety of BzaD, X and E enzymes from a variety of different *bza* and *bza*-like operons could provide significant biochemical insight into a

variety of chemical routes for the functionalisation of benzimidazolylcobamides, potential novel biotransformation's and insight into how radical SAM enzymes and iron sulphur cluster containing protein's function.

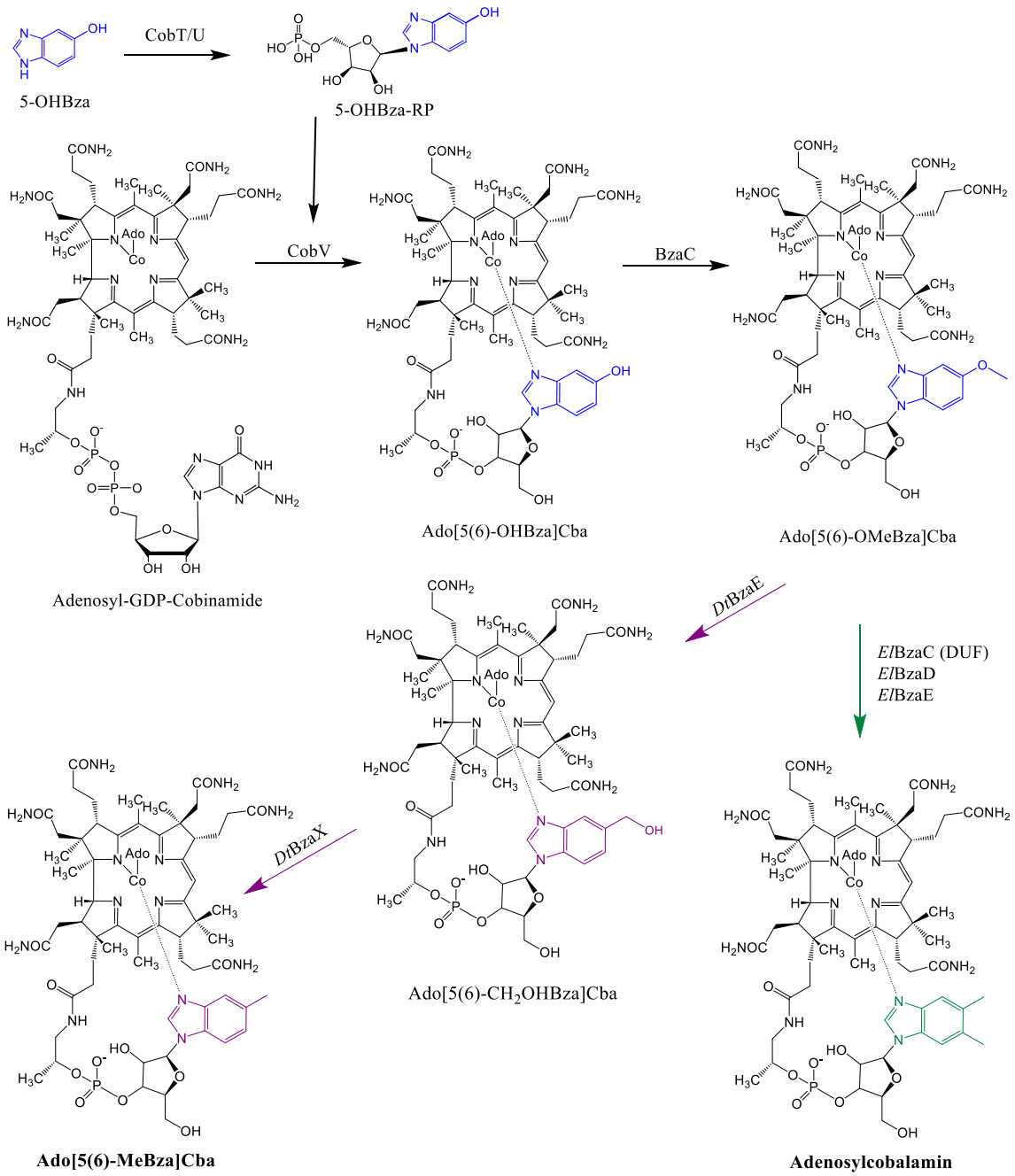


Figure 7.1 - Final steps of cobamide biosynthesis

Schematic representation of the revised potential final steps of cobamide biosynthesis. Shared benzimidazole intermediates are highlighted in blue, *Dt*Bza pathway benzimidazole intermediates are highlighted in purple and *El*Bza pathway benzimidazole highlighted in green. Final pathway product names are in bold.

References

1. Bryant, D. A., Hunter, C. N., and Warren, M. J. (2020) Biosynthesis of the modified tetrapyrroles—the pigments of life. *Journal of Biological Chemistry*. **295**, 6888–6925

2. Battersby, A. R. (2000) Tetrapyrroles: The pigments of life. *Nat Prod Rep*. **17**, 507–526

3. Wilks, A., and Heinzl, G. (2014) Heme oxygenation and the widening paradigm of heme degradation. *Arch Biochem Biophys*. **544**, 87–95

4. Glazer, A. N. (1989) Light guides. Directional energy transfer in a photosynthetic antenna. *J Biol Chem*. **264**, 1–4

5. Lokstein, H., Renger, G., and Götze, J. P. (2021) Photosynthetic Light-Harvesting (Antenna) Complexes-Structures and Functions. *Molecules*. **26**, 3378

6. Standfuss, J., Van Scheltinga, A. C. T., Lamborghini, M., and Kühlbrandt, W. (2005) Mechanisms of photoprotection and nonphotochemical quenching in pea light-harvesting complex at 2.5 Å resolution. *EMBO Journal*. **24**, 919–928

7. Duerringt, M., Schmidt, G. B., and Huber, R. (1991) Isolation, crystallization, crystal structure analysis and refinement of constitutive C-phycocyanin from the chromatically adapting cyanobacterium *Fremyella diplosiphon* at 1.66 Å resolution. *Journal of Molecular biology*. **217**, 577-592

8. Perutz, M. F. (1970) Stereochemistry of Cooperative Effects in Haemoglobin: Haem–Haem Interaction and the Problem of Allostery. *Nature*. **228**, 726–734

9. Park, S. Y., Yokoyama, T., Shibayama, N., Shiro, Y., and Tame, J. R. H. (2006) 1.25 Å Resolution Crystal Structures of Human Haemoglobin in the Oxy, Deoxy and Carbonmonoxy Forms. *J Mol Biol*. **360**, 690–701

10. Birukou, I., Schweers, R. L., and Olson, J. S. (2010) Distal Histidine Stabilizes Bound O_2 and Acts as a Gate for Ligand Entry in Both Subunits of Adult Human Hemoglobin. *Journal of Biological Chemistry*. **285**, 8840–8854

11. Gendron, A., and Allen, K. D. (2022) Overview of Diverse Methyl/Alkyl-Coenzyme M Reductases and Considerations for Their Potential Heterologous Expression. *Front Microbiol*. **13**, 867342

12. Nayak, D. D., Mahanta, N., Mitchell, D. A., and Metcalf, W. W. (2017) Post-translational thioamidation of methyl-coenzyme M reductase, a key enzyme in methanogenic and methanotrophic Archaea. **6**:e29218

13. Crane, B. R., Siegel, L. M., and Getzoff, E. D. (1995) Sulfite Reductase Structure at 1.6 Å: Evolution and Catalysis for Reduction of Inorganic Anions. *Science*. **270**, 59–67

14. Robscheit-Robbins, F. S., and Whipple, G. H. (1925) Blood Regeneration in Severe Anemia: II. Favorable Influence of Liver, Heart and Skeletal Muscle in Diet. *American journal of physiology*. **72**, 408-418.

15. Musser, J. H., Minot, G. R., Murphy, W. P., and Huntington, P (1926). Treatment of Pernicious Anemia by a Special Diet. *Journal of the American Medical Association*. **87**, 470-476

16. Hodgkin, D. C., Kamper, J., MacKay, M., Pickworth, J., Trueblood, K. N., and White, J. G. (1956) Structure of Vitamin B_{12}. *Nature*. **178**, 64–66

17. Smith, A. D., Warren, M. J., and Refsum, H. (2018) Vitamin B_{12}. *Adv Food Nutr Res*. **83**, 215–279

18. Hazra, A. B., Tran, J. L. A., Crofts, T. S., and Taga, M. E. (2013) Analysis of Substrate Specificity in CobT Homologs Reveals Widespread Preference for DMB, the Lower Axial Ligand of Vitamin B_{12}. *Chem Biol*. **20**, 1275–1285

19. Osman, D., Cooke, A., Young, T. R., Deery, E., Robinson, N. J., and Warren, M. J. (2021) The requirement for cobalt in vitamin B_{12}: A paradigm for protein metalation. *Biochim Biophys Acta Mol Cell Res*. **1868**, 118896

20. Giedyk, M., Goliszewska, K., and Gryko, D. (2015) Vitamin B_{12} catalysed reactions. *Chemical Society Reviews*. **44**, 3391-3404

21. Matthews, R. G., Koutmos, M., and Datta, S. (2008) Cobalamin-dependent and cobamide-dependent methyltransferases. *Curr Opin Struct Biol*. **18**, 658–666

22. Gruber, K., Puffer, B., and Kräutler, B. (2011) Vitamin B_{12}-derivatives—enzyme cofactors and ligands of proteins and nucleic acids. *Chem Soc Rev*. **40**, 4346–4363

23. Marsh EN, Meléndez GD. (2012) Adenosylcobalamin enzymes: Theory and experiment begin to converge. *Bone*. **23**, 1–7

24. Gruber, K., and Kratky, C. (2002) Coenzyme B_{12} dependent glutamate mutase. *Curr Opin Chem Biol*. **6**, 598–603

25. Kräutler, B. (2005) Vitamin B_{12}: chemistry and biochemistry. *Biochemical Society Transactions*. **33**, 806–810

26. Toraya, T. (2003) Radical Catalysis in Coenzyme B_{12}-Dependent Isomerization (Eliminating) Reactions. *Chem Rev*. **103**, 2095–2127

27. O'Brien, R. J., Fox, J. A., Kopczynski, M. G., and Babior, B. M. (1985) The mechanism of action of ethanolamine ammonia-lyase, an adenosylcobalamin-dependent enzyme. *Journal of Biological Chemistry*. **260**, 16131–16136

28. Berkovitch, F., Behshad, E., Tang, K.-H., Enns, E. A., Frey, P. A., and Drennan, C. L. (2004) A locking mechanism preventing radical damage in the absence of substrate, as revealed by the x-ray structure of lysine 5,6-aminomutase. *Proc Natl Acad Sci U S A*. **101**, 15870-15875

29. Jost, M., Simpson, J. H., and Drennan, C. L. (2015) The Transcription Factor CarH Safeguards Use of Adenosylcobalamin as a Light Sensor by Altering the Photolysis Products. *Biochemistry*. **54**, 3231–3234

30. Ortiz-Guerrero, J. M., Polanco, M. C., Murillo, F. J., Padmanabhan, S., and Elías-Arnanz, M. (2011) Light-dependent gene regulation by a coenzyme B_{12}-based photoreceptor. *Proc Natl Acad Sci U S A*. **108**, 7565–7570

31. Hunter, G. A., and Ferreira, G. C. (2009) 5-aminolevulinate synthase: catalysis of the first step of heme biosynthesis. *Cell Mol Biol (Noisy-le-grand)*. **55**,102–110.

32. Kikuchi, G., Kumar, A., Talmage, P., and Shemin, D. (1958) The Enzymatic Synthesis of δ-Aminolevulinic Acid. *J Biol Chem*. **233**, 1214–1219

33. Huang, D. D., and Wang, W. Y. (1986) Chlorophyll Biosynthesis in *Chlamydomonas* Starts with the Formation of Glutamyl-tRNA. *Journal of Biological Chemistry*. **261**, 13451–13455

34. Moser, J., Lorenz, S., Hubschwerlen, C., Rompf, A., and Jahn, D. (1999) *Methanopyrus kandleri* Glutamyl-tRNA Reductase. *Journal of Biological Chemistry*. **274**, 30679–30685

35. Gamini Kannangara, C., Gough, S. P., Bruyant, P., Kenneth Hoober, J., Kahn, A., and von Wettstein, D. (1988) $tRNA^{Glu}$ as a cofactor in δ-aminolevulinate biosynthesis: steps that regulate chlorophyll synthesis. *Trends Biochem Sci*. **13**, 139–143

36. Nandi, D. L., Baker-Cohen, K. F., and Shemin, D. (1968) δ-Aminolevulinic Acid Dehydratase of *Rhodopseudomonas spheroides*. *Journal of Biological Chemistry*. **243**, 1224–1230

37. Louie, G. V, Brownlie, P. D., Lambert, R., Cooper, J. B., Blundell, T. L., Wood, S. P., Warren, M. J., Woodcock, S. C., and Jordan, P. M. (1992) Structure of porphobilinogen deaminase reveals a flexible multidomain polymerase with a single catalytic site. *Nature*. **359**, 33–39

38. Fan, C., and Bobik, T. A. (2008) The PduX Enzyme of *Salmonella enterica* Is an L-Threonine Kinase Used for Coenzyme B_{12} synthesis. *Journal of Biological Chemistry*. **283**, 11322–11329

39. Tavares, N. K., VanDrisse, C. M., and Escalante-Semerena, J. C. (2018) *Rhodobacterales* use a unique L-threonine kinase for the assembly of the nucleotide loop of coenzyme B_{12}. *Mol Microbiol*. **110**, 239–261

40. Blanche, F., Cameron, B., Crouzet, J., Debussche, L., Thibaut, D., Vuilhorgne, M., Leeper, F. J., and Battersby, A. R. Vitamin B_{12}: How the Problem of Its Biosynthesis Was Solved. *Angew. Chem. Int. Ed. Engl*. **34**, 383-411

41. Escalante-Semerena, J. C., and Warren, M. J. (2008) Biosynthesis and Use of Cobalamin (B_{12}). *EcoSal Plus*. **3**, 10.1128/ecosalplus.3.6.3.8

42. Brushaber, K. R., O'toole, G. A., and Escalante-Semerena, J. C. (1998) CobD, a Novel Enzyme with L-Threonine-O-3-phosphate Decarboxylase Activity, Is Responsible for the Synthesis of (R)-1-Amino-2-propanol O-2-Phosphate, a Proposed New Intermediate in Cobalamin Biosynthesis in *Salmonella typhimurium* LT2*. *Journal of Biological Chemistry*. **273**, 2684-2691

43. Blanche, F., Debussche, L., Famechon, A., Thibaut, D., Cameron, B., and Crouzet, J. (1991) A Bifunctional Protein from *Pseudomonas denitrificans* Carries Cobinamide Kinase and Cobinamide Phosphate Guanylyltransferase Activities. Journal of Bacteriology. **173**, 6052-6057.

44. O'Toole, G. A., and Escalante-Semerena, J. C. (1995) Purification and Characterization of the Bifunctional CobU Enzyme of *Salmonella typhimurium* LT2. *Journal of Biological Chemistry*. **270**, 23560–23569

45. Woodson, J. D., Peck, R. F., Krebs, M. P., and Escalante-Semerena, J. C. (2003) The *cobY* Gene of the Archaeon *Halobacterium sp.* Strain NRC-1 Is Required for De Novo Cobamide Synthesis. *Journal of Bacteriology*. **185**, 311–316

46. Trzebiatowski, J. R., O'toole, G. A., and Escalante-Semerena, J. C. (1994) The *cobT* Gene of *Salmonella typhimurium* Encodes the NaMN: 5,6-Dimethylbenzimidazole Phosphoribosyltransferase Responsible for the Synthesis of N^1-(5-Phospho-α-D-Ribosyl)-5,6-Dimethylbenzimidazole, an Intermediate in the Synthesis of the Nucleotide Loop of Cobalamin. *Journal of Bacteriology*. **176**, 3568-3575

47. Cameron, B., Blanche, F., Rouyez, M.-C., Bisch, D., Famechon, A., Couder, M., Cauchois, L., Thibaut, D., Debussche, L., and Crouzet1, J. (1991) Genetic Analysis, Nucleotide Sequence, and Products of Two *Pseudomonas denitrificans cob* Genes Encoding Nicotinate-Nucleotide:Dimethylbenzimidazole Phosphoribosyltransferase and Cobalamin (5'-Phosphate) Synthase. *Journal of Bacteriology*. **173**, 6066-6073

48. O'Toole, G. A., Trzebiatowski, J. R., and Escalante-Semerena, J. C. (1994) The *cobC* Gene of *Salmonella typhimurium* Codes for a Novel Phosphatase Involved in the Assembly of the Nucleotide Loop of Cobalamin. *Journal of Biological Chemistry*. **269**, 26503–26511

49. Taga, M. E., Larsen, N. A., Howard-Jones, A. R., Walsh, C. T., and Walker, G. C. (2007) BluB cannibalizes flavin to form the lower ligand of vitamin B_{12}. *Nature*. **446**, 449–453

50. Hazra, A. B., Han, A. W., Mehta, A. P., Mok, K. C., Osadchiy, V., Begley, T. P., and Taga, M. E. (2015) Anaerobic biosynthesis of the lower ligand of vitamin B_{12}. *Proc Natl Acad Sci U S A*. **112**, 10792–10797

51. Claas, K. R., Parrish, J. R., Maggio-Hall, L. A., and Escalante-Semerena, J. C. (2010) Functional Analysis of the Nicotinate Mononucleotide:5,6-Dimethylbenzimidazole Phosphoribosyltransferase (CobT) Enzyme, Involved in the Late Steps of Coenzyme B_{12} biosynthesis in *Salmonella enterica*. *J Bacteriol*. **192**, 145–154

52. Crofts, T. S., Seth, E. C., Hazra, A. B., and Taga, M. E. (2013) Cobamide Structure Depends on Both Lower Ligand Availability and CobT Substrate Specificity. *Chem Biol*. **20**, 1265–1274

53. Sofia, H. J., Chen, G., Hetzler, B. G., Reyes-Spindola, J. F., and Miller, N. E. (2001) Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods. *Nucleic Acids Res*. **29**, 1097-1106

54. Bridwell-Rabb, J., Li, B., and Drennan, C. L. (2022) Cobalamin-Dependent Radical S-Adenosylmethionine Enzymes: Capitalizing on Old Motifs for New Functions. *ACS Bio and Med Chem Au*. **2**, 173–186

55. Mathur, Y., Sreyas, S., Datar, P. M., Sathian, M. B., and Hazra, A. B. (2020) CobT and BzaC catalyze the regiospecific activation and methylation of the 5-hydroxybenzimidazole lower ligand in anaerobic cobamide biosynthesis. *Journal of Biological Chemistry*. **295**, 10522–10534

56. SnapGene software (www.snapgene.com)

57. Benchling (https://benchling.com)

58. Lee, P. H., Huang, X. X., Teh, B. T., and Ng, L. M. (2019) TSA-CRAFT: A Free Software for Automatic and Robust Thermal Shift Assay Data Analysis. *Slas Discovery*. **24**, 606-612

59. Burastero, O., Niebling, S., Defelipe, L. A., Günther, C., Struve, A., and Garcia Alai, M. M. (2021) eSPC: an online data-analysis platform for molecular biophysics. *Acta Cryst D Struct Biol*. **77**, 1241–1250

60. Niebling, S., Burastero, O., Bürgi, J., Günther, C., Defelipe, L. A., Sander, S., Gattkowski, E., Anjanappa, R., Wilmanns, M., Springer, S., Tidow, H., and García-Alai, M. (2021) FoldAffinity: binding affinities from nDSF experiments. *Sci Rep*. **11**, 9572

61. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K (2010). Features and development of *Coot*. *Acta Cryst D Struct Biol*. **66**, 486-501.

62. Agirre, J., Atanasova, M., Bagdonas, H., Ballard, C. B., Baslé, A., Beilsten-Edmands, J., Borges, R. J., Brown, D. G., Burgos-Mármol, J. J., Berrisford, J. M., Bond, P. S., Caballero, I., Catapano, L., Chojnowski, G., Cook, A. G., Cowtan, K. D., Croll, T. I., Debreczeni, J., Devenish, N. E., Dodson, E. J., Drevon, T. R., Emsley,

P., Evans, G., Evans, P. R., Fando, M., Foadi, J., Fuentes-Montero, L., Garman, E. F., Gerstel, M., Gildea, R. J., Hatti, K., Hekkelman, M. L., Heuser, P., Hoh, S. W., Hough, M. A., Jenkins, H. T., Jiménez, E., Joosten, R. P., Keegan, R. M., Keep, N., Krissinel, E. B., Kolenko, P., Kovalevskiy, O., Lamzin, V. S., Lawson, D. M., Lebedev, A. A., Leslie, A. G. W., Lohkamp, B., Long, F., Malý, M., McCoy, A. J., McNicholas, S. J., Medina, A., Millán, C., Murray, J. W., Murshudov, G. N., Nicholls, R. A., Noble, M. E. M., Oeffner, R., Pannu, N. S., Parkhurst, J. M., Pearce, N., Pereira, J., Perrakis, A., Powell, H. R., Read, R. J., Rigden, D. J., Rochira, W., Sammito, M., Rodríguez, F. S., Sheldrick, G. M., Shelley, K. L., Simkovic, F., Simpkin, A. J., Skubak, P., Sobolev, E., Steiner, R. A., Stevenson, K., Tews, I., Thomas, J. M. H., Thorn, A., Valls, J. T., Uski, V., Usón, I., Vagin, A., Velankar, S., Vollmar, M., Walden, H., Waterman, D., Wilson, K. S., Winn, M. D., Winter, G., Wojdyr, M., and Yamashita, K. (2023) The *CCP*4 suite: integrative software for macromolecular crystallography. *Acta Cryst D Struct Biol*. **79**, 449–461

63. Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F., and Murshudov, G. N. (2004) *REFMAC5* dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr*. **60**, 2184–2195

64. The PyMOL Molecular Graphics System, Version 3.0 Schrödinger, LLC.

65. Laskowski, R. A., Hutchinson, E. G., Michie, A. D., Wallace, A. C., Jones, M. L., and Thornton, J. M. (1997) PDBsum: A Web-based database of summaries and analyses of all PDB structures. *Trends Biochem Sci*. **22**, 488–490

66. Larsson, K. M., Logan, D. T., and Nordlund, P. (2010) Structural Basis for Adenosylcobalamin Activation in AdoCbl-Dependent Ribonucleotide Reductases. *ACS Chem Biol*. **5**, 933–942

67. Morris, G. M., Ruth, H., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., and Olson, A. J. (2009) AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J Comput Chem*. **30**, 2785–2791

68. Eberhardt, J., Santos-Martins, D., Tillack, A. F., and Forli, S. (2021) AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J Chem Inf Model*. **61**, 3891–3898

69. Malalasekara, L., and Escalante-Semerena, J. C. (2022) A method for the isolation of α-ribazole from vitamin B_{12}, and its enzymatic conversion to α-ribazole 5'-phosphate. *Methods in Enzymology*. **668**, 125–136

70. Skinner, S. P., Fogh, R. H., Boucher, W., Ragan, T. J., Mureddu, L. G., and Vuister, G. W. (2016) CcpNmr AnalysisAssign: a flexible platform for integrated NMR analysis. *J Biomol NMR*. **66**, 111–124

71. Madeira, F., Pearce, M., Tivey, A. R. N., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A., and Lopez, R. (2022) Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res*. **50**, W276–W279

72. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. **25**, 1189–1191

73. E. Krissinel and K. Henrick Protein structure comparison service PDBeFold at European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/ssm)

74. Abdelraheem, E., Thair, B., Varela, R. F., Jockmann, E., Popadić, D., Hailes, H. C., Ward, J. M., Iribarren, A. M., Lewkowicz, E. S., Andexer, J. N., Hagedoorn, P. L., and Hanefeld, U. (2022) Methyltransferases: Functions and Applications. *ChemBioChem*. **23**, e202200212

75. Struck, A. W., Thompson, M. L., Wong, L. S., and Micklefield, J. (2012) S-Adenosyl-Methionine-Dependent Methyltransferases: Highly Versatile Enzymes in Biocatalysis, Biosynthesis and Other Biotechnological Applications. *ChemBioChem*. **13**, 2642–2655

76. Akey, D. L., Li, S., Konwerski, J. R., Confer, L. A., Bernard, S. M., Anzai, Y., Kato, F., Sherman, D. H., and Smith, J. L. (2011) A New Structural Form in the SAM/Metal-Dependent O-Methyltransferase Family: MycE from the Mycinamycin Biosynthetic Pathway. *J Mol Biol*. **413**, 438-450

77. Zhang, J., and Zheng, Y. G. (2016) SAM/SAH Analogs as Versatile Tools for SAM-Dependent Methyltransferases. *ACS Chem Biol*. **11**, 583–597

78. Fleischhacker, A. S., and Matthews, R. G. (2007) Ligand Trans Influence Governs Conformation in Cobalamin-Dependent Methionine Synthase. *Biochemistry*. **46**, 12382-12392

79. Randaccio, L., Geremia, S., Demitri, N., and Wuerges, J. (2010) Vitamin B_{12}: Unique Metalorganic Compounds and the Most Complex Vitamins. *Molecules*. **15**, 3228–3259

80. Lawrence, A. D., Nemoto-Smith, E., Deery, E., Baker, J. A., Schroeder, S., Brown, D. G., Tullet, J. M. A., Howard, M. J., Brown, I. R., Smith, A. G., Boshoff, H. I., Barry, C. E., and Warren, M. J. (2018) Construction of Fluorescent Analogs to Follow the Uptake and Distribution of Cobalamin (Vitamin B_{12}) in Bacteria, Worms, and Plants. *Cell Chem Biol*. **25**, 941-951

81. Zhao, S., Wang, Z. P., Wen, X., Li, S., Wei, G., Guo, J., and He, Y. (2020) Synthesis of Vitamin B_{12}-Antibiotic Conjugates with Greatly Improved Activity against Gram-Negative Bacteria. *Org Lett*. **22**, 6632–6636

82. Lacombe, V., Lenaers, G., and Urbanski, G. (2022) Diagnostic and Therapeutic Perspectives Associated to Cobalamin-Dependent Metabolism and Transcobalamins' Synthesis in Solid Cancers. *Nutrients*. **14**, 2058

83. Waibel, R., Treichler, H., Schaefer, N. G., Van Staveren, D. R., Mundwiler, S., Kunze, S., Küenzi, M., Alberto, R., Nüesch, J., Knuth, A., Moch, H., Schibli, R., and Schubiger, P. A. (2008) New Derivatives of Vitamin B_{12} Show Preferential Targeting of Tumors. *Cancer Res*. **68**, 2904–2911

84. Wierzba, A. J., Maximova, K., Wincenciuk, A., Równicki, M., Wojciechowska, M., Nexø, E., Trylska, J., and Gryko, D. (2018) Does a Conjugation Site Affect Transport of Vitamin B_{12}–Peptide Nucleic Acid Conjugates into Bacterial Cells? *Chemistry – A European Journal*. **24**, 18772–18778

85. Maggio-Hall, L. A., and Escalante-Semerena, J. C. (1999) In vitro synthesis of the nucleotide loop of cobalamin by *Salmonella typhimurium* enzymes. *Proc Natl Acad Sci U S A*. **96**, 11798–11803

86. Zayas, C. L., and Escalante-Semerena, J. C. (2007) Reassessment of the Late Steps of Coenzyme B_{12} Synthesis in *Salmonella enterica*: Evidence that Dephosphorylation of Adenosylcobalamin-5'-Phosphate by the CobC Phosphatase Is the Last Step of the Pathway. *J Bacteriol*. **189**, 2210–2218

87. Peinado, R. dos S., Olivier, D. S., Eberle, R. J., de Moraes, F. R., Amaral, M. S., Arni, R. K., and Coronado, M. A. (2019) Binding studies of a putative *C. pseudotuberculosis* target protein from Vitamin B_{12} Metabolism. *Scientific Reports*. **9**, 6350

88. Blanche, F., Debussche, L., Thibaut, D., Crouzet, J., and Cameron, B. (1989) Purification and Characterization of S-Adenosyl-L-Methionine: Uroporphyrinogen III Methyltransferase from *Pseudomonas denitrificans*. *J Bacteriol*. **171**, 4222-4231

89. Frank, S., Deery, E., Brindley, A. A., Leech, H. K., Lawrence, A., Heathcote, P., Schubert, H. L., Brocklehurst, K., Rigby, S. E. J., Warren, M. J., and Pickersgill, R. W. (2007) Elucidation of Substrate Specificity in the Cobalamin (vitamin B_{12}) Biosynthetic Methyltransferases: Structure and Function of the C20 Methyltransferase (CbiL) from *Methanothermobacter thermautotrophicus*. *Journal of Biological Chemistry*. **282**, 23957–23969

90. Akhtar, M. K., Vijay, D., Umbreen, S., McLean, C. J., Cai, Y., Campopiano, D. J., and Loake, G. J. (2018) Hydrogen Peroxide-Based Fluorometric Assay for Real-Time Monitoring of SAM-Dependent Methyltransferases. *Front Bioeng Biotechnol*. **6**, 146

91. Li, C.-Y., Crack, J. C., Newton-Payne, S., J Murphy, A. R., Chen, X.-L., Pinchbeck, B. J., Zhou, S., Williams, B. T., Peng, M., Zhang, X.-H., Chen, Y., Le Brun, N. E., Todd, J. D., Zhang, Y.-Z., and by Ning-Yi Zhou, E. (2022) Mechanistic insights into the key marine dimethylsulfoniopropionate synthesis enzyme DsyB/DSYB. *mLife*. **1**, 114-130

92. Crofts, T. S., Hazra, A. B., Tran, J. LA, Sokolovskaya, O. M., Osadchiy, V., Ad, O., Pelton, J., Bauer, S., and Taga, M. E. (2014) Regiospecific Formation of Cobamide Isomers Is Directed by CobT. *Biochemistry*. **53**, 7805–7815

93. Schäuble, S., Stavrum, A. K., Puntervoll, P., Schuster, S., and Heiland, I. (2013) Effect of substrate competition in kinetic models of metabolic networks. *FEBS Lett*. **587**, 2818–2824

94. Cheng, X., Kumar, S., Posfai, J., Pflugrath, J. W., Robertst, R. J., and Keck, W. M. (1993) Crystal Structure of the Hhal DNA Methyltransferase Complexed with S-Adenosyl-L-Methionine. *Cell*. **74**, 299-307.

95. Schubert, H. L., Blumenthal, R. M., and Cheng, X. (2003) Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem Sci*. **28**, 329–335

96. Iijima, H., Takebe, K., Suzuki, M., Kobayashi, H., Takamiya, T., Saito, H., Niwa, N., and Kuwada-Kusunose, T. (2020) Crystal Structure of Catechol O-Methyltransferase Complexed with Nitecapone. *Chem Pharm Bull (Tokyo)*. **68**, 447–451

97. Jarrett, J. T., Huang, S., and Matthews, R. G. (1998) Methionine Synthase Exists in Two Distinct Conformations That Differ in Reactivity toward

Methyltetrahydrofolate, Adenosylmethionine, and Flavodoxin. *Biochemistry.* **37**, 5372-5382.

98. Dixon, M. M., Huang, S., Matthews, R. G., and Ludwig, M (1996). The structure of the C-terminal domain of methionine synthase: presenting S-adenosylmethionine for reductive methylation of B_{12}. *Structure*. **4**, 1263-1275

99. Schubert, H. L. , Wilson, K. S. , Raux, E. , Woodcock. S. C. , Warren. J. M. (1998) The X-ray structure of a cobalamin biosynthetic enzyme, cobalt-precorrin-4 methyltransferase. *Nat Struct Biol*. **5**, 585–592

100. Vévodová, J., Graham, R. M., Raux, E., Schubert, H. L., Roper, D. I., Brindley, A. A., Ian Scott, A., Roessner, C. A., Stamford, N. P. J., Elizabeth Stroupe, M., Getzoff, E. D., Warren, M. J., and Wilson, K. S. (2004) Structure/Function Studies on a S-Adenosyl-L-methionine-dependent Uroporphyrinogen III C Methyltransferase (SUMT), a Key Regulatory Enzyme of Tetrapyrrole Biosynthesis. *J Mol Biol*. **344**, 419–433

101. Stroupe, M. E., Leech, H. K., Daniels, D. S., Warren, M. J., and Getzoff, E. D. (2003) CysG structure reveals tetrapyrrole-binding features and novel regulation of siroheme biosynthesis. *Nat Struct Biol*. **10**, 1064–1073

102. Tkaczuk, K. L., Dunin-Horkawicz, S., Purta, E., and Bujnicki, J. M. (2007) Structural and evolutionary bioinformatics of the SPOUT superfamily of methyltransferases. *BMC Bioinformatics*. **8**, 73

103. Lim, K., Zhang, H., Tempczyk, A., Krajewski, W., Bonander, N., Toedt, J., Howard, A., Eisenstein, E., and Herzberg, O. (2003) Structure of the YibK

methyltransferase from *Haemophilus influenzae* (HI0766): A cofactor bound at a site formed by a knot. *Proteins*. **51**, 56–67

104. Xiao, B., Jing, C., Wilson, J. R., Walker, P. A., Vasisht, N., Kelly, G., Howell, S., Taylor, I. A., Blackburn, M. G., and Gamblin, S. J. (2003) Structure and catalytic mechanism of the human histone methyltransferase SET7/9. *Nature*. **421**, 652–656

105. Sun, Q., Huang, M., and Wei, Y. (2021) Diversity of the reaction mechanisms of SAM-dependent enzymes. *Acta Pharm Sin B*. **11**, 632–650

106. Zubieta, C., Ross, J. R., Koscheski, P., Yang, Y., Pichersky, E., and Noel, J. P. (2003) Structural Basis for Substrate Recognition in the Salicylic Acid Carboxyl Methyltransferase Family. *The Plant Cell*. **15**, 1704–1716

107. Jansson, A., Koskiniemi, H., Mäntsälä, P., Niemi, J., and Schneider, G. (2004) Crystal Structure of a Ternary Complex of DnrK, a Methyltransferase in Daunorubicin Biosynthesis, with Bound Products. *Journal of Biological Chemistry*. **279**, 41149–41156

108. Singh, S., McCoy, J. G., Zhang, C., Bingman, C. A., Phillips, G. N., and Thorson, J. S. (2008) Structure and Mechanism of the Rebeccamycin Sugar 4'-*O*-Methyltransferase RebM. *J Biol Chem*. **283**, 22628-22636

109. Zubieta, C., He, X. Z., Dixon, R. A., and Noel, J. P. (2001) Structures of two natural product methyltransferases reveal the basis for substrate specificity in plant O-methyltransferases. *Nature Structural Biology*. **8**, 271–279

110. Storbeck, S., Saha, S., Krausze, J., Klink, B. U., Heinz, D. W., and Layer, G. (2011) Crystal Structure of the Heme d_1 Biosynthesis Enzyme NirE in Complex with Its Substrate Reveals New Insights into the Catalytic Mechanism of S-Adenosyl-L-methionine-dependent Uroporphyrinogen III methyltransferases. *Journal of Biological Chemistry*. **286**, 26754–26767

111. Ferrer, J. L., Zubieta, C., Dixon, R. A., and Noel, J. P. (2005) Crystal Structures of Alfalfa Caffeoyl Coenzyme A 3-*O*-methyltransferase. *Plant Physiol*. **137**, 1009–1017

112. Liebschner, D., Afonine, P. V, Baker, M. L., Bunkó, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L. W., Jain, S., Mccoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V, Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J., and Adams, P. D. (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Cryst D Struct Biol*. **75**, 861–877

113. Trott, O., and Olson, A. J. (2010) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J Comput Chem*. **31**, 455-461

114. Botros, H. G., Legrand, P., Pagan, C., Bondet, V., Weber, P., Ben-Abdallah, M., Lemière, N., Huguet, G., Bellalou, J., Maronde, E., Beguin, P., Haouz, A., Shepard, W., and Bourgeron, T. (2013) Crystal structure and functional mapping of human

ASMT, the last enzyme of the melatonin synthesis pathway. *J Pineal Res*. **54**, 46–57

115. Benjdia, A., and Berteau, O. (2023) B_{12}-dependent radical SAM enzymes: Ever expanding structural and mechanistic diversity. *Curr Opin Struct Biol*. **83**, 102725

116. Nicolet, Y. (2020) Structure–function relationships of radical SAM enzymes. *Nat Catal*. **3**, 337–350

117. Booker, S. J., and Lloyd, C. T. (2022) Twenty Years of Radical SAM! The Genesis of the Superfamily. *ACS Bio and Med Chem Au*. **2**, 538–547

118. Broderick, J. B., Duffus, B. R., Duschene, K. S., and Shepard, E. M. (2014) Radical S - Adenosylmethionine Enzymes. *Chemical Reviews*. **114**, 4229-4317

119. Yan, F., Lamarre, J. M., Röhrich, R., Wiesner, J., Jomaa, H., Markin, A. S., and Fujimori, D. G. (2010) RlmN and Cfr are Radical SAM Enzymes Involved in Methylation of Ribosomal RNA. *J Am Chem Soc*. **132**, 3953–3964

120. Yokoyama, K., Li, D., and Pang, H. (2022) Resolving the Multidecade-Long Mystery in MoaA Radical SAM Enzyme Reveals New Opportunities to Tackle Human Health Problems. *ACS Bio and Med Chem Au*. **2**, 94–108

121. Berkovitch, F., Nicolet, Y., Wan, J. T., Jarrett, J. T., and Drennan, C. L. (2004) Crystal Structure of Biotin Synthase, an S-Adenosylmethionine-Dependent Radical Enzyme. *Science*. **303**,76-79

122. Frey, P. A., and Booker, S. J. (2001) Radical mechanisms of S-adenosylmethionine-dependent enzymes. *Advances in Protein Chemistry*. **58**, 1-45

123. Nicolet, Y., Amara, P., Mouesca, J. M., and Fontecilla-Camps, J. C. (2009) Unexpected electron transfer mechanism upon AdoMet cleavage in radical SAM proteins. *Proc Natl Acad Sci U S A*. **106**, 14867–14871

124. Bridwell-Rabb, J., Zhong, A., Sun, H. G., Drennan, C. L., and Liu, H. W. (2017) A B_{12}-dependent radical SAM enzyme involved in oxetanocin A biosynthesis. *Nature*. **544**, 322–326

125. Benjdia, A., Pierre, S., Gherasim, C., Guillot, A., Carmona, M., Amara, P., Banerjee, R., and Berteau, O. (2015) The thiostrepton A tryptophan methyltransferase TsrM catalyses a cob(II)alamin-dependent methyl transfer reaction. *Nat Commun*. **6,** 8377

126. Werner, W. J., Allen, K. D., Hu, K., Helms, G. L., Chen, B. S., and Wang, S. C. (2011) In Vitro Phosphinate Methylation by PhpK from *Kitasatospora phosalacinea*. *Biochemistry*. **50**, 8986–8988

127. McLaughlin, M. I., Pallitsch, K., Wallner, G., Van Der Donk, W. A., and Hammerschmidt, F. (2021) Overall Retention of Methyl Stereochemistry during B_{12}-Dependent Radical SAM Methyl Transfer in Fosfomycin Biosynthesis. *Biochemistry*. **60**, 1587–1596

128. Kim, H. J., McCarty, R. M., Ogasawara, Y., Liu, Y. N., Mansoorabadi, S. O., Levieux, J., and Liu, H. W. (2013) GenK-Catalyzed C-6' Methylation in the Biosynthesis of

Gentamicin: Isolation and Characterization of a Cobalamin-Dependent Radical SAM Enzyme. *J Am Chem Soc*. **135**, 8093–8096

129. Fyfe, C. D., Bernardo-García, N., Fradale, L., Grimaldi, S., Guillot, A., Brewee, C., Chavas, L. M. G., Legrand, P., Benjdia, A., and Berteau, O. (2022) Crystallographic snapshots of a B_{12}-dependent radical SAM methyltransferase. *Nature*. **602**, 336–342

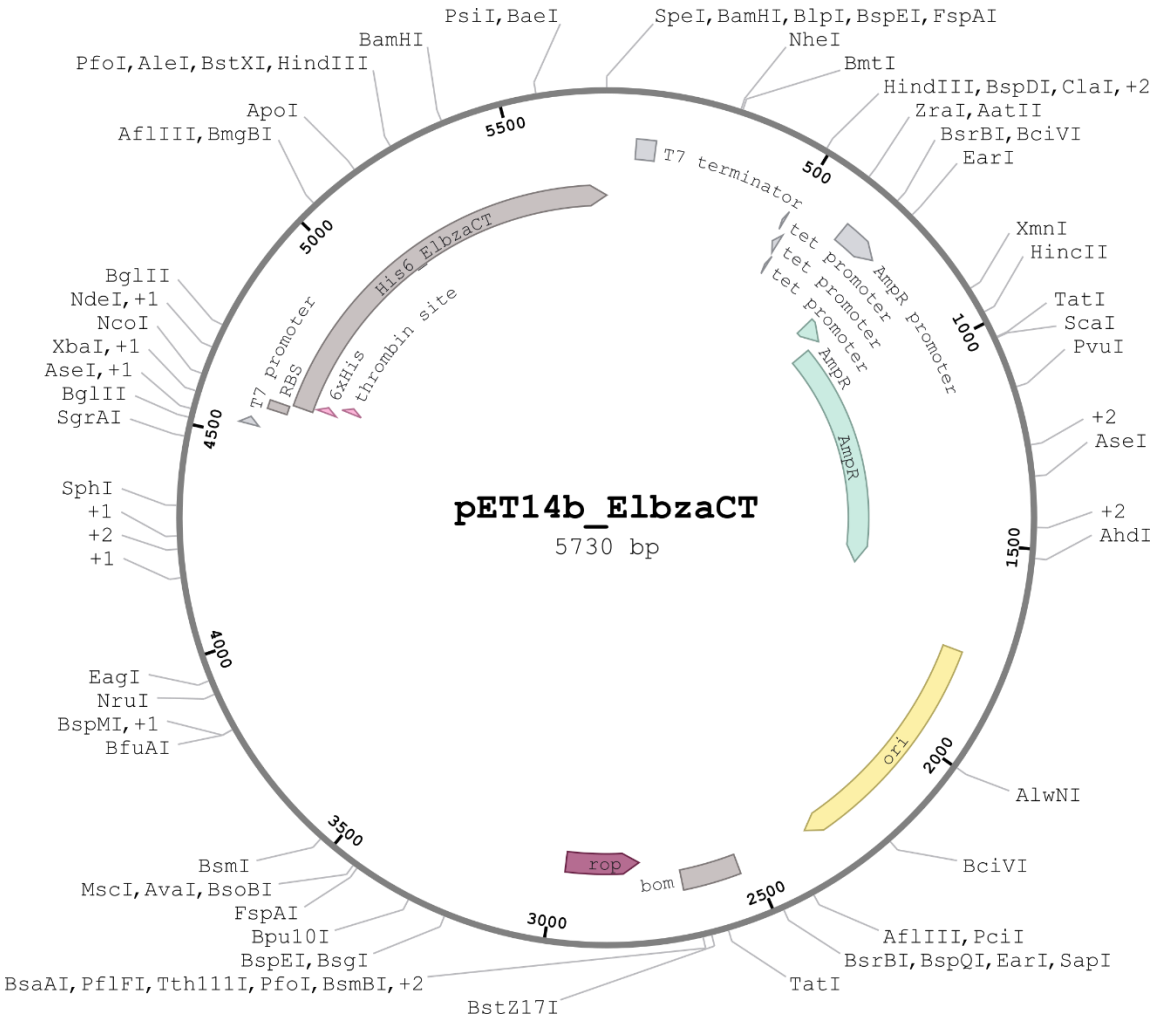
130. Knox, H. L., Sinner, E. K., Townsend, C. A., Boal, A. K., and Booker, S. J. (2022) Structure of a B_{12}-dependent radical SAM enzyme in carbapenem biosynthesis. *Nature*. **602**, 343-348

131. Wang, Y., Schnell, B., Baumann, S., Muller, R., and Begley, T. P. (2017) Biosynthesis of Branched Alkoxy Groups: Iterative Methyl Group Alkylation by a Cobalamin-Dependent Radical SAM enzyme. *J Am Chem Soc*. **139**, 1742–1745

132. Sinner, E. K., Li, R., Marous, D. R., and Townsend, C. A. (2022) ThnL, a B_{12}-dependent radical S-adenosylmethionine enzyme, catalyzes thioether bond formation in carbapenem biosynthesis. *Proc Natl Acad Sci U S A*. **119**, e2206494119

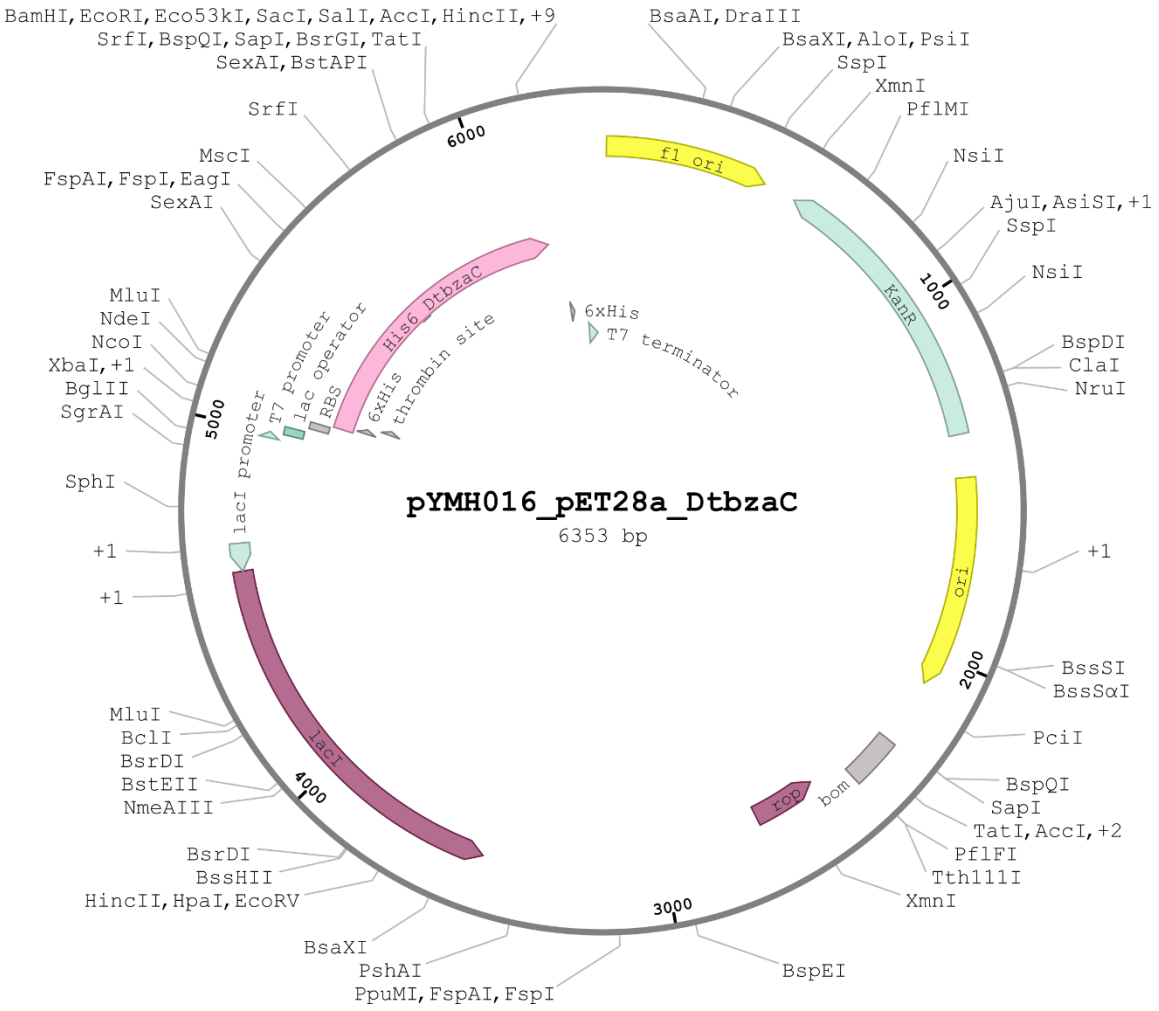
133. Suzuki, J. Y., Bollivar, D. W., and Bauer, C. E. (1997) Genetic analysis of chlorophyll biosynthesis. *Annu Rev Genet*. **31**, 61-89

Appendix



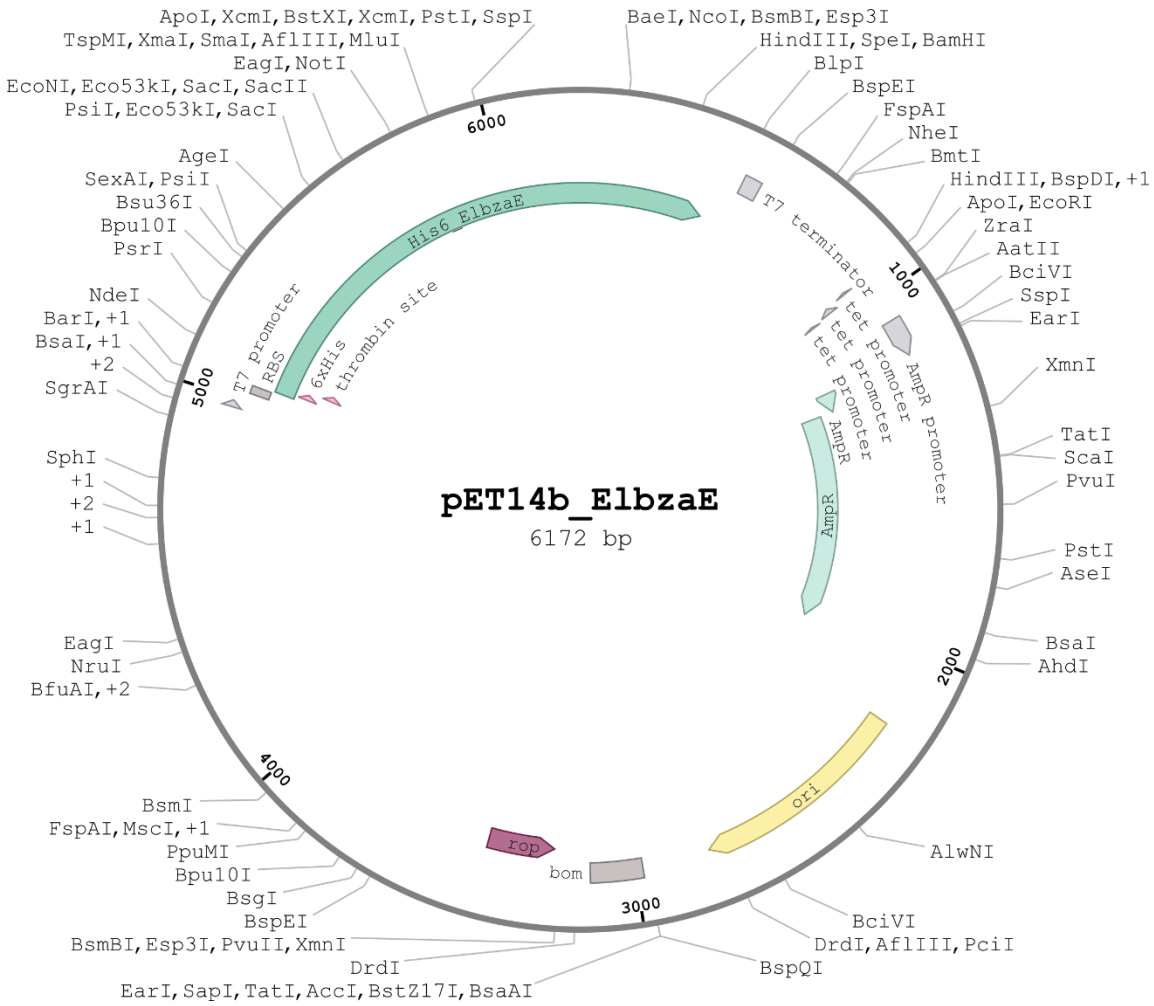
Supplementary figure 1 - pET14b_*ElbzaCT* plasmid map**

Schematic representation of the pET14b_*ElbzaCT* plasmid containing the His_6-*ElbzaCT* gene under control of the T7 promoter system and an ampicillin resistance (AmpR) cassette for selection.



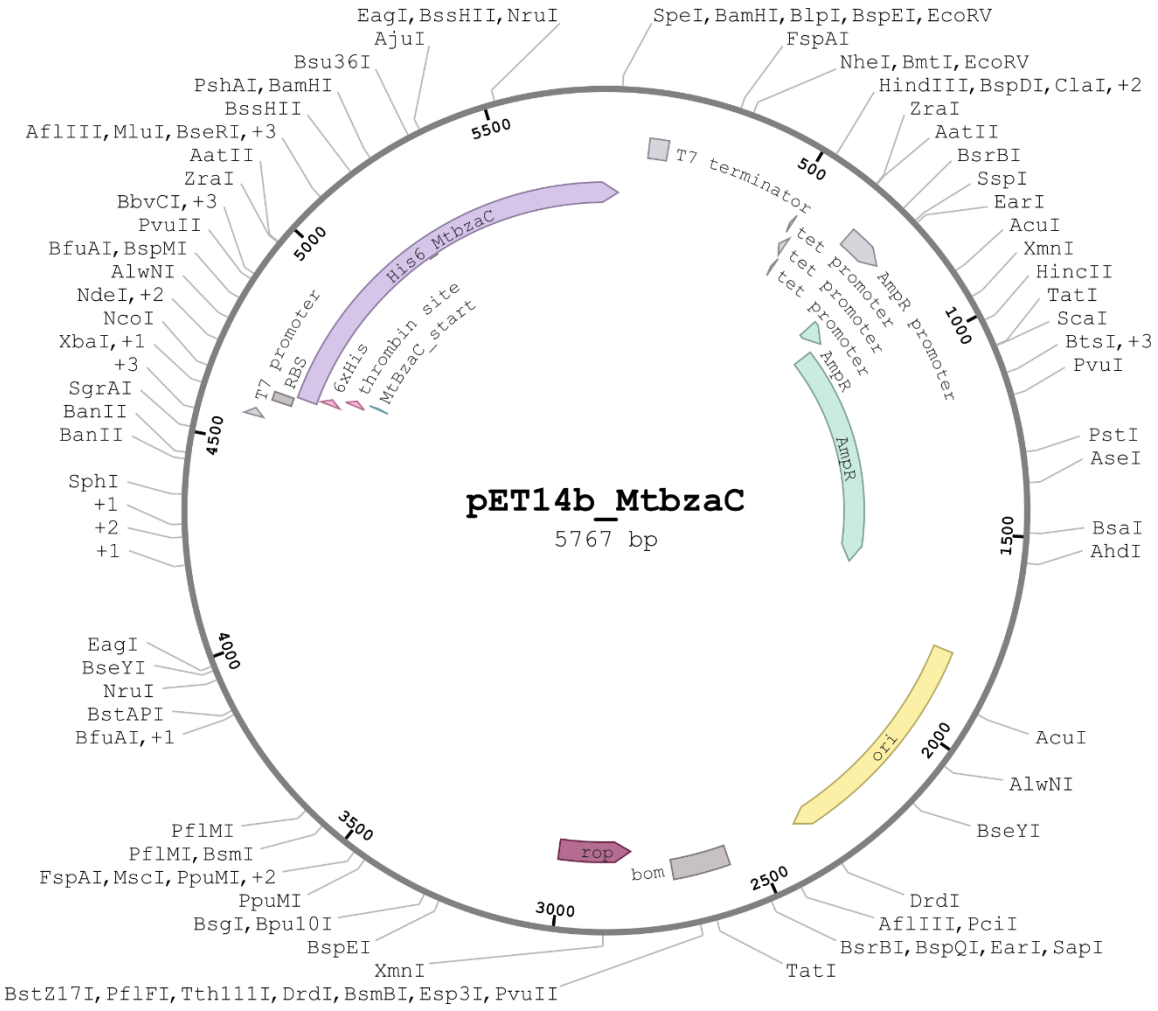
Supplementary figure 2 - pET28_*DtbzaC* plasmid map

Schematic representation of the pET28_*DtbzaC* plasmid containing the His_6-*Dt*bzaC gene under control of the T7 promoter system and a kanamycin resistance cassette (KanR) for selection.



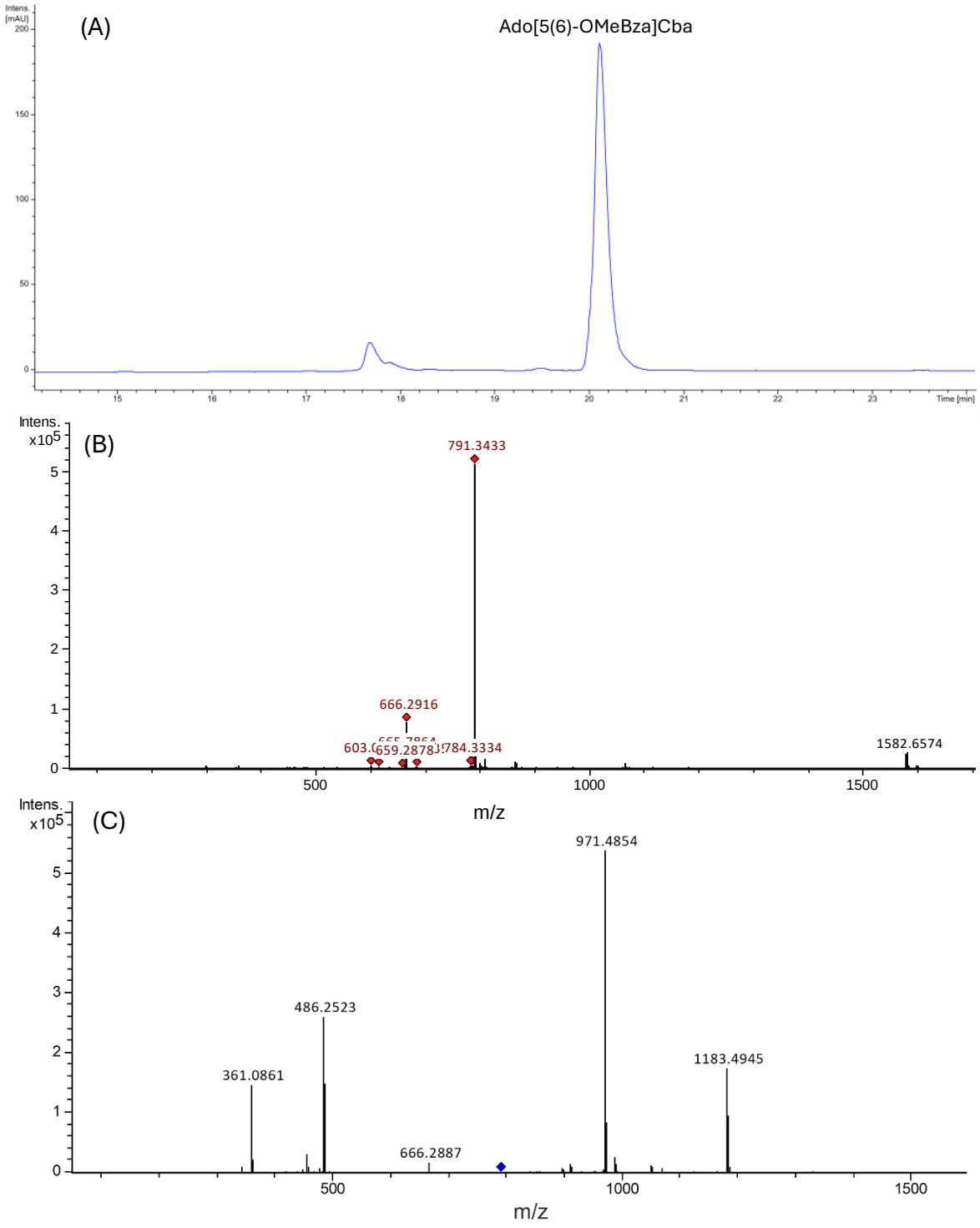
Supplementary figure 3 - pET14b_*ElbzaE* plasmid map

Schematic representation of the pET14b_*ElbzaE* plasmid containing the His_6-*ElbzaE* gene under control of the T7 promoter system and an ampicillin resistance cassette (AmpR) for selection.



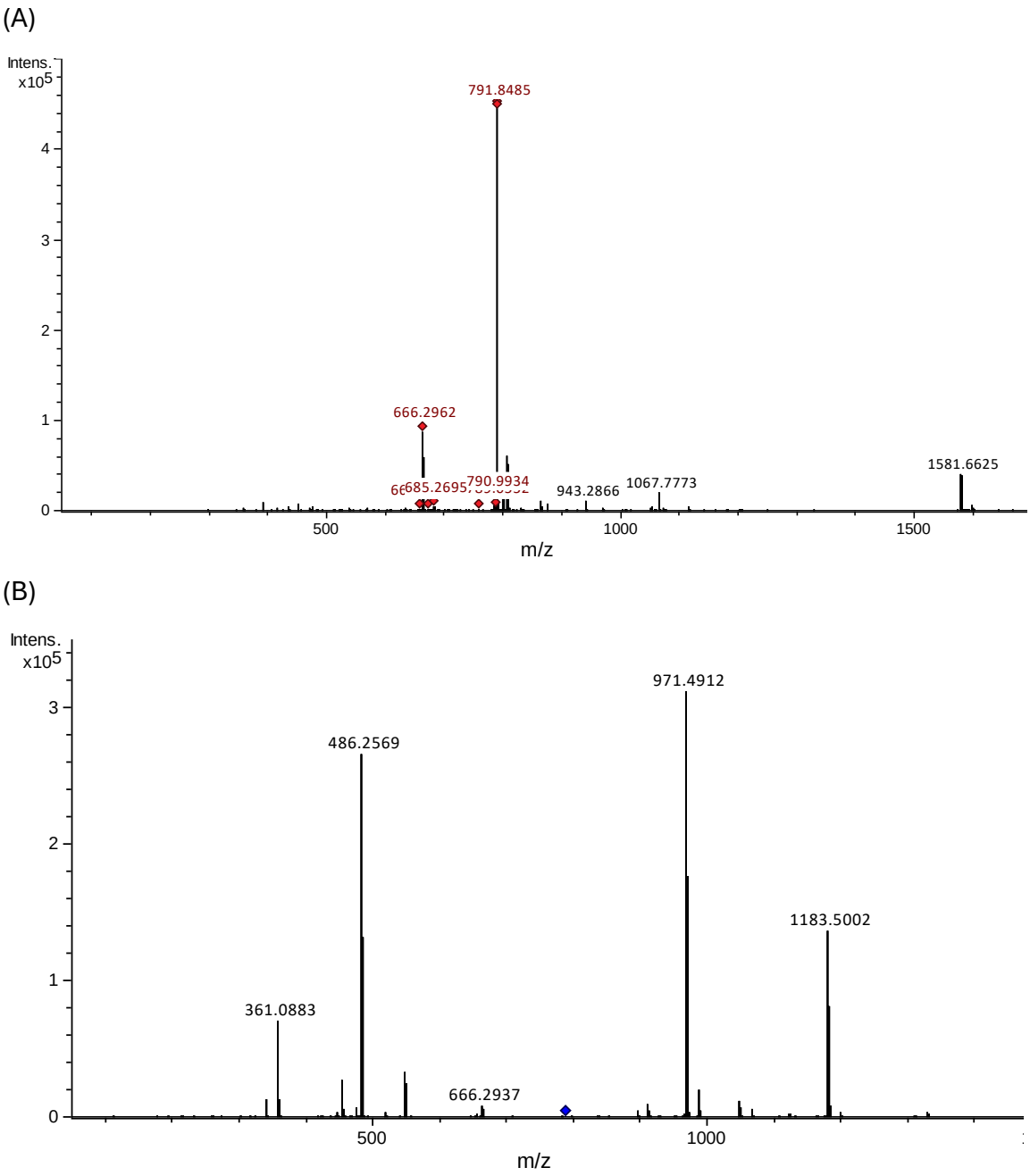
Supplementary figure 4 - pET14b_*Mt*bzaC plasmid map

Schematic representation of the pET14b_*MtbzaC* plasmid containing the His_6-*MtbzaC* gene under control of the T7 promoter system and an ampicillin resistance cassette (AmpR) for selection.



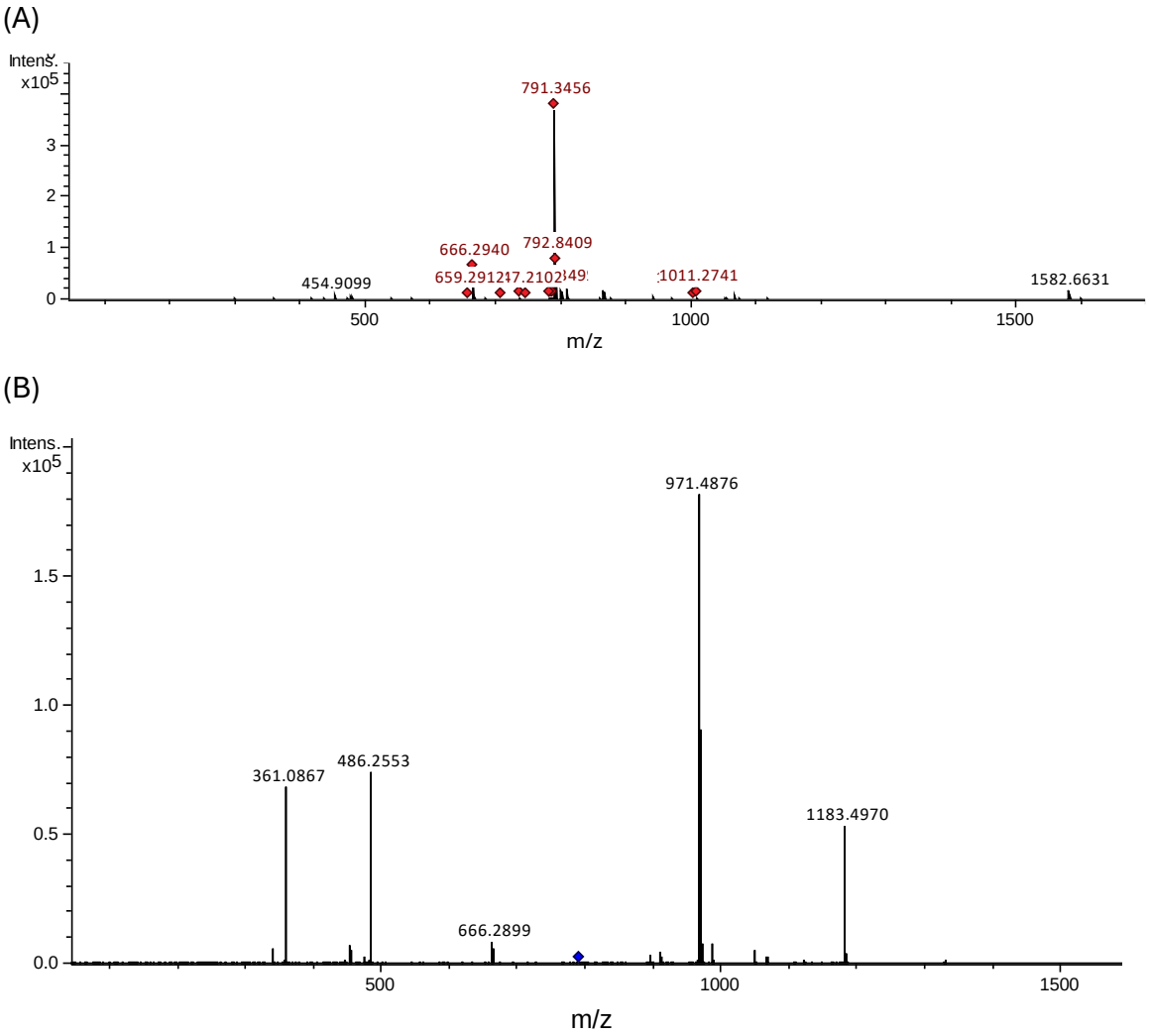
Supplementary figure 5 - Ado[5(6)-OMeBza]Cba

LC-MS analysis of Ado[5(6)-OMeBza]Cba produced by guided biosynthesis. **(A)** HPLC chromatogram. **(B)** MS. **(C)** MS/MS fragmentation.



Supplementary figure 6 - Ado[5(6)-CH_2OHBza]Cba mass spectra

Mass spectra of Ado[5(6)-CH_2OHBza]Cba produced by guided biosynthesis. **(A)** MS, expected m/z = 791.825 ($[M+2H]^{2+}$). **(B)** MS/MS fragmentation, expected m/z of the lower ligand fragment = 361.07 ([M+H]).



Supplementary figure 7 - Ado[5-OH, 6-MeBza]Cba mass spectra

Mass spectra of Ado[5-OH, 6-MeBza]Cba produced by guided biosynthesis. **(A)** MS, expected m/z = 791.325 ($[M+H]^{2+}$). **(B)** MS/MS fragmentation, expected m/z of the lower ligand fragment = 361.07 ([M+H]).