Synthesis of Cobalamin Analogues Using Enzymatic and Chemical Modification Methods, and Subsequent Identification of Cobalamin Localisation in a Variety of Organisms

A thesis submitted to the University of Kent for the degree of PhD in the Faculty of Sciences.

2017

Emi H. Nemoto-Smith

Declaration

Name: Emi H. Nemoto-Smith

Degree: PhD Biochemistry

Title: Synthesis of Cobalamin Analogues Using Enzymatic and Chemical Modification Methods, and Subsequent Identification of Cobalamin Localisation in a Variety of Organisms.

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.

Abstract

Cobalamin, also known as vitamin B_{12}, is an essential nutrient for many different organisms including mammals, fish, birds, nematodes, and a variety of bacteria. However, cobalamin is only synthesised by a few bacteria and archaea. Organisms that cannot synthesise cobalamin *de novo* must obtain it from their diet. In humans, the cobalamin uptake mechanism has been studied in detail, but in many organisms, such as *Caenorhabditis elegans*, no method of transport has been defined, and their need for cobalamin is recognised by a cobalamin deficiency phenotype.

Corrin ring modified fluorescent analogues of cobyric acid and ribose conjugated fluorescent analogues of cobalamin were synthesised in order to follow the uptake and localisation of these corrinoids in a variety of organisms. Both the C5 corrin-ring modified and the ribose conjugated analogues were absorbed by *Salmonella enterica*, using the B_{12} uptake system (Btu) and could be converted into active coenzyme forms. The imaging of these fluorescent analogues enabled the identification of the coelomocytes in *C. elegans* as a possible storage cell for cobalamin. However, the C5 cobyric acid analogue was not recognised which suggests that the *C. elegans* cobalamin transport mechanism is specific for complete corrinoid molecules. *Lepidium sativum*, garden cress, was shown to take up both cobalamin analogues from the roots and store it in the vacuoles of the cotyledons in seedlings, even though plants have no cobalamin requirement. In contrast, *Arabidopsis thaliana* did not transport any of the cobalamin analogues.

Cobalamin deficiency has been implicated in impeding disease progression in a number of diseases, such as tuberculosis. The *Mycobacterium tuberculosis* cobalamin uptake protein, BacA, has only recently been identified, and there is still much to learn about the relationship between *M. tuberculosis* and cobalamin. Incubations of a cobalamin dependent strain of *M. tuberculosis*, *ΔmetE*, with a selection of cobalamin biosynthesis intermediates showed that cobyric acid is the earliest intermediate to be taken up and converted into the cofactor form. The C5 corrin ring modified cobyric acid fluorescent analogue is also capable of rescuing this *ΔmetE* strain, and is taken up faster than the ribose conjugated cobalamin analogue. Overall, the research outlined in this thesis demonstrates that fluorescent corrinoid analogues can be used to follow the journey of cobalamin in a broad range of different organisms and systems.

Table of Contents

Section	**Chapter 3**	
3	Synthesising C5-Cobalamin Analogues Using a Combination of Chemical and Native Biological Methods	67

Section	**Chapter 4**	
4	Characterising molecular recognition of cobalamin analogues in a variety of organisms	92

Figures

Tables

Graphs

Abbreviations

Å	Ångstrom
A_{280}	Absorbance at 280 nm
ACN	Aconitase
ADC	Albumin dextrose catalase
AdoCbl	Adenosylcobalamin
ALA	5-Aminolevulinic acid
APS	Ammonium persulphate
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
B_{12}	Cobalamin, Vitamin B_{12}
BOB_{12}	Ribose linked BODIPY® TR-X cobalamin
BODIPY®	Boron-dipyrromethene
BSA	Bovine serum albumin
CA	Cobyric acid
Cbl	Cobalamin
CDI	1,1-carbonyl diimidazole
CIT	Citrate synthase
Da	Dalton
DEAE	Diethylaminoethyl Sepharose
dH_2O	Distilled water
DIC	Differential interference contrast
DIPEA	N, N-di-isopropylethylamine
DMB	Dimethyl benzimidazole
DMF	Dimethylformamide
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
dNDPs	Deoxyribonucleotide diphosphates
dNTPs	Deoxyribonucleotide triphosphates
DTT	Dithiothreitol
ε	Extinction coefficient
EDTA	Ethylene-diamine-tetra-acetic acid
EIC	Extracted ion chromatogram
FAD	Flavin adenine dinucleotide
FPLC	Fast protein liquid chromatography
FUM	Fumerase
GDP	Guanosine diphosphate
GMP	Guanosine monophosphate
GTP	Guanosine triphosphate
HBA	Hydrogenobyrinic acid
HBAD	Hydrogenobyrinic acid-*a*, *c*-diamide
HBAH	Hydrogenobyrinic acid-*a*, *b*, *c*, *d*, *e*, *g*-hexa-amide
HC	Haptocorrin
HPLC-MS	High performance liquid chromatography-Mass spectrometry

ICD	Isocitrate dehydrogenase
IF	Intrinsic factor
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Luria-Bertani medium
MALDI-TOF	Matrix assisted laser desorption/ ionisation-Time of flight
MCD	Methylcitrate dehydratase
MCL	Methylcitrate lyase
MCM	Methylmalonyl coenzyme-A mutase
MCS	Methylcitrate synthase
MDH/MQO	Malate dehydrogenase/ Malate quinine oxidoreductase
MeCbl	Methylcobalamin
MEZ	Malic enzyme
MLS	Malate synthase
MMCE	Methylmalonyl-CoA epimerase
MS	Methionine synthase
MS-MS	Tandem mass spectrometry
m/z	Mass to charge ratio
NAD^+	β-Nicotinamide adenine dinucleotide
NADH	β-Nicotinamide adenine dinucleotide hydrate (reduced)
NADPH	β-Nicotinamide adenine dinucleotide 3' phosphate
NGM	Nematode growth medium
NMR	Nuclear magnetic resonance
SAH	*S*-adenosyl-L- homocysteine
SAM	*S*-adenosyl-L- methionine
SDH	Succinate dehydrogenase
SDS-PAGE	Sodium dodecyl sulphate-Poly acrylamide gel electrophoresis
OD	Optical density
OG_{12}	Ribose linked Oregon green® 514 cobalamin
PC7	Precorrin-7
PCA	Pyruvate casrboxylase
PCC	Propionyl-CoA carboxylase
PCK	Pyruvate carboxykinase
PCR	Polymerase chain reaction
PDHC	Pyruvate dehydrogenase
PEG	Polyethylene glycol
PYK	Pyruvate kinase
RBS	Ribosome binding site
RNA	Ribonucleic acid
rNDPs	Ribonucleotide diphosphates
RNR	Ribonucleotide reductase
rpm	Revolutions per minute
SCS	Succinate synthase
TAE	Tris-Acetate-EDTA
TCII	Transcobalamin II

TEMED	N, N, N', N'-tetramethylenediamine
TFA	Trifluoroacetic acid
Tris	Tris(hydroxymethyl) aminomethane
TR-X	Texas Red
Uro'III	Uroporphyrinogen-III
UV	Ultra violet
v/v	Volume/ volume
VA044	2, 2'-Azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride
w/w	Weight/ weight

Acknowledgements

I have had the greatest of pleasures in meeting so many wonderful people throughout the course of my PhD and I would like to thank all of them. However, in the interests of space (and possibly word count), I will limit myself to a select few to whom I owe a deep gratitude.

Firstly, to the Warren Lab eternal. I could not have dreamt a more fabulous group of people to share this experience with.

To Professor Martin Warren not only for welcoming me into your lab for summer placements and encouraging me to pursue my PhD, but also for introducing me to 'walking hockey', better known as golf. I have thoroughly enjoyed my entire experience.

A huge thank you to Dr. Andrew Lawrence for teaching me so much about chemistry as well as being a fantastic friend. I cannot think of a better start to a PhD than coming up with ludicrous nicknames for the intermediates and proteins we synthesised and used.

To Dr. Evelyne '*Dea DNA*' Deery, your DNA manipulation prowess is truly legendary! Thank you for not only putting up with me for my whole PhD, but also through two summer placements. I don't know how you withstood with so much of my craziness, even I need a break from me sometimes.

Thank you to Dr. Simon Moore for introducing me to 'my' glovebox, for always being at my shoulder 'just-in-case', and for being more excited than I was when I synthesised cobalt precorrin-6b for the first time.

I would also like to thank Dr. Jenny Tullet for donating nematodes to me and showing me how to use them in my experiments.

Thank you to Kevin Howland for your endless help when the HPLC-MS threw strops, as well as regaling me on how to use the flashy new MALDI machine.

To Ian Brown thank you for all your help with using all of the different microscopes, and also for your insights into plant anatomy.

A transatlantic thank you to Dr. Clifton Barry III for the opportunity to join his wonderful lab, and to Dr. Helena Boshoff who helped through my foray into the world of *Mycobacterium tuberculosis*, even after the cherry blossoms beat the boks! And to Theresa Marsh who got me through all the paperwork.

There are not enough words to describe how much I owe to Mum and Dad who put up with their 'boomerang child', and helped me organise my six month stint in the USA. Thank you for coming with me and making a 'holiday' out of the seemingly never-ending search for accommodation!

Naturally, a big hug to Yuri, my big little sister, the only person I know who will be as outraged as I am when someone doesn't know a completely obscure Eurovision reference.

最後に、おじいちゃんとおばあちゃんへ、いつも遠い日本から応援してくれて本当にありがとうございます。心から感謝しています。

Now soars the russet pheasant like a flare,
And like a flame you glimmer unto me.

Now slinks the cat like a silent shadow,
And your midnight coat shines on to me.

Now flies the silent meteor on, and leaves
A shining furrow, as thy thoughts in me.

-Modified from 'The Princess (Part 7)'
By Alfred, Lord Tennyson

Thank you everyone, I am forever grateful.

Chapter 1

An introduction to cobalamin biosynthesis, uptake, and the chemical synthesis of analogues

1.1 Cobalamin as a vitamin

Cobalamin (vitamin B_{12}) is an essential molecule for many organisms across a wide variety of phyla. Although numerous organisms require cobalamin, only some bacteria and archaea can synthesise it (Clardy et al. 2011; Raux et al. 2000). This means that the majority of organisms, including humans, must obtain a regular source of cobalamin through their diet (Clardy et al. 2011; Raux et al. 2000). This is not difficult, as very small quantities of cobalamin are needed for maintaining normal function, in humans the daily requirement for cobalamin is 1µg per day (Martens et al. 2002; Kuzminski et al. 2016). Cobalamin is indispensable for proper brain function and is also involved in cell maintenance, DNA synthesis, neural development and upkeep as well as being involved in red blood cell production, amongst other uses (Battaglia-Hsu et al. 2009; Ames 2001; Fenech 2001). Some cobalamin deficiencies are caused by dietary restriction e.g. veganism, and others by a lack or mutation in the uptake and transport system. A deficiency in cobalamin can lead to pernicious anaemia, hyper-homocysteinemia and many other ailments (Selhub 1999; Marsh 1999). Patients with pernicious anaemia must take cobalamin supplements or have regular intravenous injections to circumvent the lack of transport. There are two types of pernicious anaemia, A, an auto immune disease which targets a cobalamin transporter, intrinsic factor (IF), or B, a non-autoimmune gastritis which is usually associated with *Helicobacter pylori* infection (Toh et al. 1997). If cobalamin deficiency is left untreated patients may develop depression, neurological damage, heart and organ failure, leading to death (Nielsen et al. 2012). Excess cobalamin is stored in the liver (Allen 2008). This store of cobalamin is thought to be the reason why cobalamin deficiency manifests after a significant time gap (sometimes years) from when the cobalamin absorption process breaks down (Allen 2008).

1.1.1 Discovery and structure of cobalamin

The discovery of cobalamin started with the identification of pernicious anaemia as an ailment in 1835. Then, in the early 1920s, Whipple observed that feeding raw liver to dogs with anaemia reversed the symptoms (Whipple and Roschelle-Robbins 1925). Around the same time, Minot and Murphy found that patients suffering from pernicious anaemia could be cured by including a whole liver in their diet (Martens et al. 2002). They attributed this cure to an extrinsic factor contained within the liver, but it was not until the late 1940s that the cherry-red coloured cobalamin was isolated from liver and first described as 'vitamin B_{12}' (Banerjee and Ragsdale 2003). Around the same time cobalamin was discovered in many

other foodstuffs including milk, eggs, meat in general, and also in some bacterial cultures. In 1956 Dorothy Hodgkin and her group solved the crystal structure of cobalamin revealing a complex molecule later described as 'nature's most beautiful coenzyme' (Stubbe 1994; Hodgkin et al. 1956).

Cobalamin is a member of the modified tetrapyrrole family. It is unique within this family because of its molecular asymmetry caused by the extrusion of carbon 20 (C20) of the macrocycle, giving rise to a contracted ring system (Figure 1.1.1.1) (Schroeder et al. 2009; Moore and Warren 2012). It is also the only molecule in this family to coordinate a cobalt ion at its centre. The molecule name is generated from this cobalt ion and its vitamin status combining to give: cobal(t) (vital) amin(e), cobalamin. Other modified tetrapyrroles include chlorophyll and haem, which coordinate magnesium and iron in the centre of their macrocycle respectively. The members of the cyclic tetrapyrrole family are known for their distinctive colours: chlorophyll is green, haem is red and cobalamin is cherry red. These colours, and the pivotal functions of these molecules, led to the cyclic tetrapyrrole family being referred to as 'the pigments of life' (Blanche et al. 1995).

The structure of cobalamin is complex, but can be broken down into three parts: the β (upper) cobalt ligand; the α ligand, or nucleotide (lower) loop; and cobinamide (Figure 1.1.1.1). The β ligand changes depending on the cofactor form. In adenosylcobalamin it is an adenosyl group, which is a coenzyme for the cobalamin-dependent isomerases, such as methyl malonyl coenzymeA mutase (MCM) (Banerjee et al. 2009). Exchanging the adenosyl group for a methyl group produces the cofactor form of cobalamin, methylcobalamin, which is used by cobalamin-dependent methyltransferases, as exemplified by methionine synthase, MetH (Banerjee et al. 2009). Cyanocobalamin, the commercial form of vitamin B_{12}, has a cyanide group as the β ligand. The cyanide is not a natural ligand; it is added to aid extraction of cobalamin from bacterial cultures during purification.

The numbering system used for cobalamin is detailed in Figure 1.1.1.1 using adenosylcobalamin as the example. The carbons of the corrin ring are numbered in light purple whilst the sidechains are referred to by letters, a-g (orange), as are the pyrrole rings, A-D (teal). The upper, β ligand is in red and the lower, α ligand is the nucleotide loop composed of the ribophosphate (purple) and dimethyl benzimidazole (DMB) (green). The dotted line illustrates the bond which can form from the DMB to the cobalt ion, known as "base on" binding (Mathews et al. 2007). In some forms of cobalamin, and during binding to some

proteins, this bond is broken (Wuerges et al. 2006). This is known as the "base off" conformation (Wuerges et al. 2006). It is important to note that the carbon atom at position 20 (C20) is not shown in Figure 1.1.1.1 as it has been extruded during the biosynthesis of cobalamin. Prior to extrusion, C20 formed a methylene bridge between C19 and C1 on the left side of the corrin ring, opposite C10. The contracted corrin ring formed by the cyclic tetrapyrrole system tightly coordinates the cobalt ion.
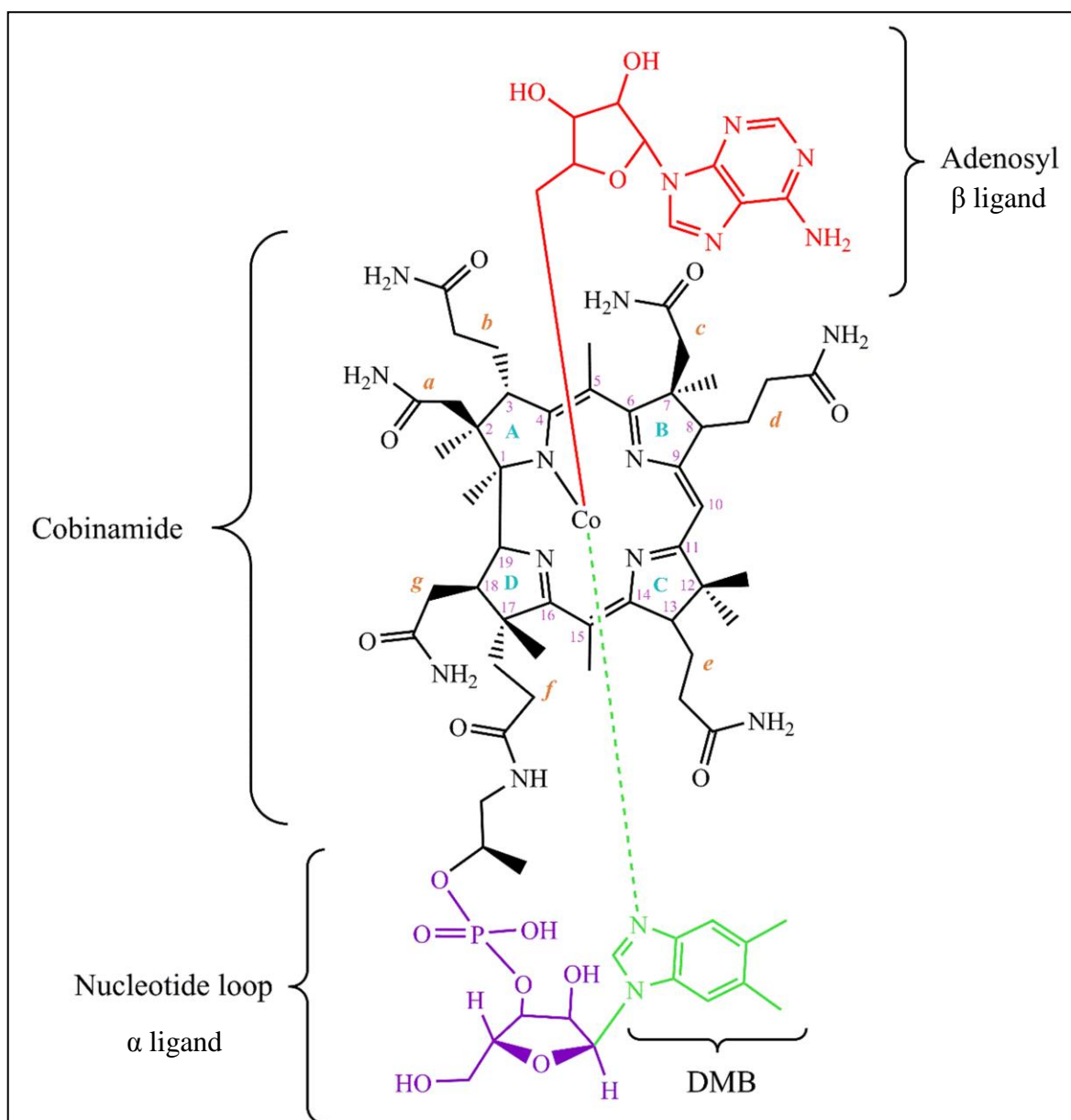
Figure 1.1.1.1: The two dimensional structure of adenosylcobalamin, colour coding the different groups. The corrinoid in black is equivalent to cobinamide, an intermediate of biosynthesis. The carbons of the corrin ring are numbered in light purple, lower case lettered labelling of the side chains are in orange, and capital lettered naming of the pyrrole rings are in teal. The adenosyl group, representing the upper, or β, ligand to the cobalt ion is in red and the lower, α, ligand to the cobalt ion is the nucleotide loop. This is comprised of the ribophosphate (purple) and DMB (dimethyl benzimidazole) (green). The structure is drawn in the "base on" configuration indicated by the dotted line coordinating the cobalt. It is important to note the ring contraction between C1 and C19 indicative of the corrin ring; the excised carbon is referred to as C20 and is only present in the early intermediates.

DMB is the most commonly studied lower base, however, it is not the only form: adenine, phenol, *p*-cresol, and 5-methylbenzimidazole, amongst others, can all be substituted for DMB (Yi et al. 2012). In humans, the DMB form is the only one recognised by the transport mechanism and, consequently, the only one known to work with human cobalamin-dependent proteins (Clardy et al. 2011). Other organisms, such as some cyanobacteria, prefer adenine as the base, known as pseudo-cobalamin, and some monera preferentially synthesise pseudo-cobalamin over cobalamin (Fyfe and Friedmann 1969; Helliwell et al. 2016).

1.2 Cobalamin biosynthesis

There are broadly two routes for cobalamin biosynthesis; the Early insertion (anaerobic) pathway and the Late insertion (aerobic) pathway. The 'Early' and 'Late' refer to the relative stage at which the cobalt is inserted into the macrocycle, whilst 'anaerobic' and 'aerobic' allude to the requirement of molecular oxygen for the ring contraction mechanism.

1.2.1 The Late insertion Pathway

The Late insertion pathway was the first of the two biosynthetic pathways to be solved and has been studied for many years (Raux et al. 1996; Maggio-Hall and Escalante-Semerena 1999) (Figure 1.2.1.1). Uroporphyrinogen III (uro'gen III), the progenitor of all modified tetrapyrroles, is converted to precorrin-2 by methylation at positions C2 and C7 using S-adenosyl-methionine (SAM) as the methyl donor. The digit succeeding the 'precorrin' refers to the number of SAM derived methyl groups bound directly to the corrin ring (including the C20 methyl group even after extrusion). Precorrin-2 is methylated at C20 by CobI, using a SAM donated methyl group, to produce precorrin-3A. This is the first committed step in the biosynthesis of cobalamin, and is a key requirement for the subsequent extrusion of C20 (Warren et al. 2002). The next enzymes, CobG and CobJ, catalyse the extrusion of C20. These reactions not only characterise cobalamin as having a unique asymmetrical macrocycle, but also define the Late insertion pathway as an "aerobic" route. This is because CobG is a mono-oxygenase, and transfers a hydroxyl group derived from molecular oxygen to the C20 position to form a tertiary alcohol. The acetate sidechain, *a*, subsequently forms a γ lactone ring, to produce precorrin-3B (Schroeder et al. 2009). CobJ then promotes ring contraction by catalysing a masked pinacol rearrangement, as well as transferring a methyl group to C17 using SAM as the donor molecule (Schroeder et al. 2009). The product of these reactions is precorrin-4. CobM and CobF both transfer SAM derived methyl groups to C11 and C1 respectively, which synthesises precorrin-6A. CobK reduces the double bond between C18

and 19 to form precorrin-6B using reduced nicotinamide adenine dinucleotide phosphate (NADPH) as the electron donor. Following this, CobL catalyses the transfer of methyl groups to C5 and C15, as well as the decarboxylation of C12. These are the last methyl transfer reactions and are both SAM dependent. The C11 methyl group is migrated to C12 by CobH to produce hydrogenobyrinic acid (HBA). CobH is the last enzyme to alter any of the chemical groups directly attached to the corrin ring. All of the subsequent reactions affect the sidechains or are related to the cobalt ion, either its insertion or the nature of the β ligand.

CobB catalyses the amidation of sidechains *a* and *c*, using glutamine as the amide donor. Next, the cobalt chelation complex, CobNST, inserts cobalt into the macrocycle. This is a slow and very metabolically expensive reaction, potentially using 15 molecules of adenosine triphosphate (ATP) to insert one cobalt ion (Heldt et al. 2005). This central cobalt is reduced and an adenosyl group is added as the β ligand by CobR and CobO, respectively. In the majority of organisms which utilise the Late insertion pathway, these two reactions must be complete before CobQ will proceed to amidate sidechains, *b*, *d*, *e*, and *g*, producing adenosyl cobyric acid (Warren et al. 2002). The final enzymes, CobD, P and V, catalyse the incorporation of the lower loop onto sidechain *f* of adenosyl cobyric acid. CobD attaches the aminopropanol phosphate to the propionic acid sidechain forming adenosyl-cobinamide. Succeeding this, CobP converts cobinamide into adenosyl-GDP-cobinamide, firstly by phosphorylating the hydroxyl group of the aminopropanol, using ATP as the donor molecule, and then, transferring a GMP moiety from GTP onto the phosphate group (Cohen 2014). CobV transfers α-ribazole phosphate (or α-ribazole) onto the first phosphate group to produce cobalamin-5'-phosphate. This 5'-phosphate is subsequently removed to form adenosylcobalamin (Cohen 2014).

The nomenclature used in Figure 1.2.1.1 is specific for *Pseudomonas denitrificans* and may be different in other organisms. However, in general the genes of the Late insertion pathway are prefixed '*cob*'.

Figure 1.2.1.1: The Late insertion pathway. This is an abridged representation of cobalamin biosynthesis, starting from uroporphorinogen III and showing the modifications (red) catalysed by each of the pathway enzymes (blue) to produce adenosyl cob(III)yric acid. The enzymes and cofactors required to convert adenosyl cob(III)yric acid into adenosylcobalamin are represented by the 5 arrows.

1.2.2 The Early insertion pathway

     The Early insertion pathway has only been elucidated in the last few years (Moore and Warren 2012; Frank et al. 2005) (Figure 1.2.2.1). This was partly because the intermediates produced are oxygen sensitive and very unstable, due to the early incorporation of the cobalt ion into the macrocycle (Moore, Lawrence, et al. 2013). This pathway also begins with the conversion of uro'gen III into precorrin-2. The C14-15 bond of precorrin-2 is oxidised by SirC, using nicotinamide adenine dinucleotide (NAD$^+$) as a cofactor, to produce factor-II. Factor-II, also known as sirohydrochlorin, is the oxidised form of precorrin-2. The numbers after the name of the intermediate are in Roman numerals and again represent the number of methyl groups that have been added to the macrocycle.

Figure 1.2.2.1: The Early insertion pathway. This abridged representation of the Early insertion pathway begins with the conversion of precorrin-2 to factor-II, the first committed step in cobalamin biosynthesis (Frank et al. 2005), up to cob(II)yrinic acid. From here cobalamin *de novo* synthesis is the same as the Late insertion pathway. The changes in structure are shown in red, whilst the enzymes are in black, with their cofactors in pink.

     Cobalt is inserted into factor-II by CbiX, a cobaltochelatase, to form cobalt factor-II. Unlike the Late insertion chelatase this reaction does not require ATP or oxygen, but, apart

from this, little is known about the cobalt chelation reaction (Frank et al. 2005; Moore and Warren 2012). Cobalt factor-II is subsequently methylated at C20 and C17 by CbiL and CbiH, respectively, both in a SAM-dependent manner. CbiH also catalyses the C20 extrusion, ring contraction reaction, by promoting the formation of a δ-lactone between rings A and D, and it is thought that the 4Fe-4S centre of CbiH is involved in the reduction of cobalt factor-III to cobalt precorrin-4, but little is known about the mechanism of this reduction reaction (Moore, Biedendieck, et al. 2013). The next enzyme, CbiF, is another SAM dependent methyltransferase, methylating C11 to produce cobalt precorrin-5a. CbiG is an enzyme unique to the early insertion pathway (Moore, Lawrence, et al. 2013). It is thought to open the δ-lactone ring, leaving C1 free to accept the methyl group transferred by CbiD. CbiD is a protein with no known similarity to any other characterised enzyme (Moore, Lawrence, et al. 2013). Despite it being a SAM-dependent methyltransferase it does not have the characteristic canonical GXGXG sequence in the N-terminus of the protein (Roessner et al. 2005). However, it has been shown that CbiD methylates the C1 position in a SAM-dependent manner (Moore, Lawrence, et al. 2013). CbiJ reduces the double bond between C18 and C19 using NADPH as a cofactor to form cobalt precorrin-6b. The last two SAM mediated methyl transfers are catalysed by CbiET. This is a fused protein in *Bacillus megaterium* but exists as two separate proteins in *Methanobacterium thermoautotrophicum*: CbiE catalyses the methylation of C5 and CbiT the methylation of C15 and the decarboxylation of C12. CbiC subsequently catalyses the migration of the C11 methyl group to C12 to form cobyrinic acid. CbiA promotes the amidation of sidechains *a* and *c* to make cobyrinic acid-*a, c*-diamide, which then follows the same pathway as in the Late insertion pathway to produce cobalamin (Moore and Warren 2012).

1.3 Cobalamin-dependent enzymes

Cobalamin is classified as a vitamin in humans because deficiency causes detrimental effects to life. These include demyelinating nerves, gastrointestinal dysfunction, hyper-segmentation of neutrophils, and potentially dementia, although the connection with the latter is inconclusive (Nielsen et al. 2012; Andrès et al. 2004; Werder 2010). Humans have two cobalamin-dependent enzymes: methyl malonyl coenzymeA mutase (MCM), and methionine synthase (encoded by *metH* in *Escherichia coli* and *Mycobacterium tuberculosis*, and by *cblG* in *Homo sapiens*) (Nielsen et al. 2012; Willard et al. 1978; Banerjee and Ragsdale 2003; Gopinath, Moosa, et al. 2013; Banerjee et al. 2009; Dobson et al. 2002; Gonzalez et al. 1992). MCM requires adenosylcobalamin to catalyse the isomerisation of methylmalonyl coenzyme

A to succinyl coenzymeA in mitochondria (Clardy et al. 2011; Brown 2005; Andrès et al. 2004). The cobalamin-dependent methionine synthase is dependent on the methylcofactor to methylate L-homocysteine to produce L-methionine in the cytosol (Brown 2005; Stubbe 1994). In some organisms there is a methylcobalamin independent methionine synthase (encoded by *metE* in *E. coli* and *M. tuberculosis*, and by *cblE* in *H. sapiens* (Gonzalez et al. 1992; Gopinath, Moosa, et al. 2013; Banerjee et al. 2009; Dobson et al. 2002)) which is active when cobalamin is not available (Figure 1.3.1), but it is transcriptionally terminated in the presence of cobalamin due to the existence of a cobalamin riboswitch upstream of the *metE* gene (Warner et al. 2007). It has been shown in *E. coli* that MetH has a catalytic turnover of 1500 molecules per minute and is over 100 times more active than MetE (12.3 molecules per minute) revealing a clear metabolic advantage of using the cobalamin-dependent enzyme when possible (Gonzalez et al. 1992; Jeter et al. 1984). From now on the cobalamin-dependent methionine synthases will be referred to as MS in humans, MetH in all other organisms, and the independent equivalent as MetE.

Figure 1.3.1: **The two different mechanisms of methionine synthesis.** A schematic representation of the reactions catalysed by the cobalamin-dependent, MetH, and independent, MetE, enzymes. Changes in structure are coloured in red.
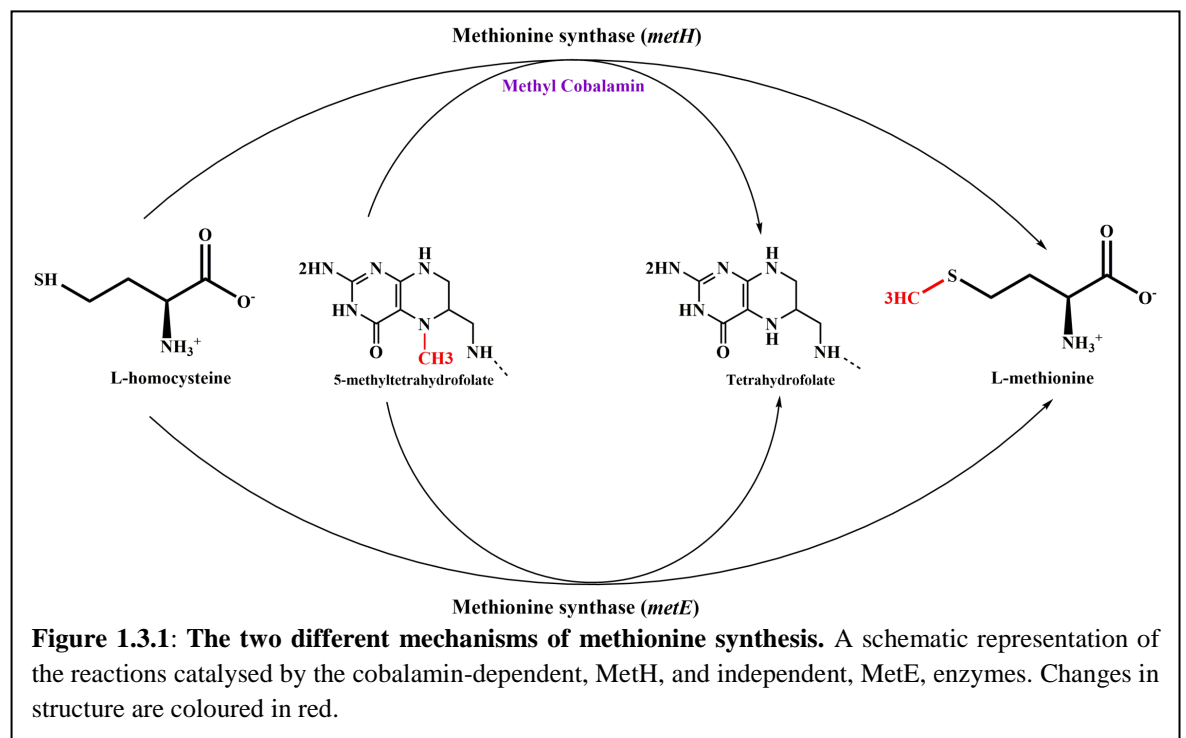
In many bacteria, cobalamin-associated genes, for instance those encoding cobalamin biosynthetic enzymes or cobalamin-dependent enzymes, are regulated, in part, by upstream riboswitches (Rodionov et al. 2003). Riboswitches are short regions of RNA (in bacteria) preceding, or within, the regulated gene (Warner et al. 2007; Vitreschak et al. 2003). They can either regulate at the transcriptional or translational level and both are dependent upon ligand

binding. In the case of *metE*, adenosylcobalamin acts as the ligand. When adenosylcobalamin is bound to the 'B$_{12}$ element', the cobalamin regulated riboswitch, the RNA folds and forms a ligand-stabilised pseudoknot (Vitreschak et al. 2003; Rodionov et al. 2003). This interacts with the antisequestor or antiterminator regions (which are sometimes part of the pseudoknot itself) and instigates the formation of a ribosome sequester hairpin or terminator in translational or transcriptional regulation, respectively (Vitreschak et al. 2003). For the *metE* gene this equates to a decrease in transcription, resulting in lower protein production in the presence of adenosylcobalamin.

1.4 Transport of cobalamin

As mentioned previously there are many different organisms that require cobalamin, but only some bacteria and archaea synthesise it *de novo*. Therefore, cobalamin requiring organisms, including many bacteria, algae, nematodes, birds, fish and mammals, must acquire cobalamin from cobalamin producing organisms (Grossman 2016; Helliwell et al. 2016; Greibe, Fedosov, and Nexo 2012; Greibe, Fedosov, Sorensen, et al. 2012; Bito et al. 2013; Nielsen et al. 2012). Different organisms have different methods of transporting cobalamin, of which the bacterial uptake systems are the most extensively studied.

M. tuberculosis has a single known transporter, BacA, which was identified in 2013 (Gopinath, Venclovas, et al. 2013). This spans the single membrane of *M. tuberculosis*, transporting cobalamin from the waxy lipid rich coat of the bacterium exterior, into the cytoplasm (Figure 1.4.1) (Bansal-mutalik and Nikaido 2014; Gopinath, Venclovas, et al. 2013). BacA is an ATP binding cassette (ABC) protein, predicted to be a homodimer. Although most ABC importers work cooperatively with a high affinity substrate binding protein, no such protein has been identified in *M. tuberculosis* (Gopinath, Venclovas, et al. 2013).

Figure 1.4.1: **The structure of the *Mycobacterium tuberculosis* cell wall.** The single, plasma membrane is covered in lipoarabinomannans (LAM), lipomannans (LM), and phosphatidylinositol (PIM). The presence of these is what causes *M. tuberculosis* to be impervious to Gram staining. Instead the bacterium is classified as acid-fast. This figure was reproduced with permission from Park and Bendelac 2000.

In gram-negative bacteria import requires a few more proteins. In *E. coli*, the Ton-B dependent BtuB is the outer membrane transporter which imports cobalamin to the periplasm. BtuF then tightly binds it and trafficks it to the BtuC/D complex on the inner membrane. BtuF has an extremely high affinity for cobalamin (15 nM) whereas BtuC/D alone does not bind free cobalamin well (Lewinson et al. 2010). However, upon the association of BtuF to BtuC/D, BtuC/D causes BtuF to release the bound cobalamin which is internalised (Lewinson et al. 2010). Using ATP, BtuF then dissociates from BtuC/D allowing BtuF to bind free cobalamin and for the process to be repeated (Lewinson et al. 2010) (Figure 1.4.2). BtuD is an ABC protein, like BacA, whereas BtuC is a transmembrane protein (Cadieux et al. 2002).

Figure 1.4.2: **The mechanism of cobalamin import in *E. coli*.**This mechanism proposed in (and reproduced with permission from) Lewinson et al. 2010 shows BtuF, the protein on top, binding cobalamin, the black 8 shape, and trafficking it to the BtuC/D complex. Cobalamin is then transported through BtuC, transmembrane section, and BtuD, ATP binding protein at the bottom. Upon the hydrolysis of ATP to ADP, BtuF is realeased from the complex and available for periplasmic cobalamin binding.

In algae only one protein, CBA1, has been implicated in cobalamin uptake, although the role of this protein and the mechanism of uptake is unknown (Helliwell et al. 2016). Plants have no designated cobalamin uptake mechanism as they do not need it. However, it has been shown that some plants, e.g. soya bean, can take up cobalamin from their environment (Watanabe et al. 2013).

C. elegans has a cobalamin deficiency phenotype, which indicates that it requires cobalamin (Bito et al. 2013). The origin of the cobalamin is likely to be the bacterial food source, but the uptake system is currently not known.

Zebrafish (*Danio rerio*) have one identified cobalamin binding protein which is excreted in large quantities into the ambient water to sequester cobalamin (Greibe, Fedosov, and Nexo 2012). Similarly, rainbow trout (*Oncorhynchus mykiss*) also have one known extracellular cobalamin binding protein, which is distributed through various tissues in the fish particularly in the kidney and roe (Greibe, Fedosov, Sorensen, et al. 2012). Both of the proteins identified in these fish behave like intermediaries of the three known extracellular mammalian cobalamin binding proteins (see Section 1.4.1) (Greibe, Fedosov, and Nexo 2012; Greibe, Fedosov, Sorensen, et al. 2012).

1.4.1 Cobalamin absorption in *Homo sapiens*

The human cobalamin uptake system is more complex than those of the organisms discussed previosly. Humans can absorb cobalamin at different points during digestion (Nielsen et al. 2012) (Figure 1.4.1.1). Haptocorrin (HC), the initial cobalamin binding protein (also called R-protein or transcobalamin I) is in many bodily fluids including plasma and saliva. Once HC is bound to cobalamin it 'escorts' it through the upper gastrointestinal tract. This is thought to prevent the hydrolysis of cobalamin by stomach acid (Nielsen et al. 2012).

Figure 1.4.1.1: **Overview of uptake and transport of cobalamin in humans.** In the upper gastrointestinal tract, cobalamin is released from food and is bound by HC. The HC is degraded in the duodenum and cobalamin is captured by IF. Further down the gut, in the terminal ileum, IF-bound cobalamin is endocytosed by cubam. Inside these enterocytes, IF is degraded and cobalamin is released to plasma by MRP1. In plasma, cobalamin is bound to HC or TCII (Nielsen et al. 2012). This figure was reproduced with permission from Nielsen et al. 2012.

In the duodenum the HC is degraded by pancreatic enzymes and the cobalamin passes onto intrinsic factor (IF). IF serves as the first quality control point as it specifically binds cobalamin (with DMB as the lower base), whereas HC will bind cobinamide and other

intermediates too (Nielsen et al. 2012). Once bound the IF-cobalamin complex travels down the gut to the terminal ileum where it is absorbed into cells via the IF-cobalamin receptor, cubam, a complex between cubilin and amnionless (Clardy et al. 2011). IF is then degraded in the cells of the intestinal lining (enterocytes) and cobalamin is excreted by multidrug resistance protein (MRP) 1 into the bloodstream where it is picked up by transcobalamin II (TCII) or HC. Although the HC bound cobalamin accounts for about 80 % of cobalamin in the bloodstream, only the TCII-cobalamin complex is internalised via the transcobalamin receptor (TCR) (Banerjee et al. 2009). This is most likely because TCII only binds completed cobalamin, ensuring that no incomplete forms of cobalamin are present in the cell. The TCII-cobalamin complex is degraded in the lysosome of cells, and the cobalamin is exported into the cytosol, whereupon it is either converted into a methyl cofactor, or transported into the mitochondria.

1.5 Links between disease and cobalamin

The causes of cobalamin deficiency are widespread. Some causes are intrinsic, such as a lack of a cobalamin transporter, some are self-imposed lifestyle choices, and others are due to infection or disease. Low dietary cobalamin intake resulting from lifestyle choices, such as veganism, can cause deficiency in adults as well as in breastfed infants, and even if the child subsequently adopts an omnivorous diet they are at greater risk of developing deficiency. These children may have stunted growth due to their childhood deficiency (Allen 2008).

Polymorphisms in the cobalamin uptake proteins e.g. TCII, or an autoimmune response to a transport protein e.g. IF (pernicious anaemia), result in malabsorption either from the bloodstream into cells or from the intestine into the enterocytes, respectively. Consequently cobalamin is depleted or absent from the bloodstream and deficiency occurs. There are also several mutations which obstruct the conversion of cobalamin into either one or both of the cofactor forms (Figure 1.5.1). Depletions in methionine synthase (MS) activity can cause homocystinuria, whilst reduced MCM activity can cause methylmalonic aciduria. Disruptions in MS and MCM can result in both disorders (Froese and Gravel 2010).

Figure 1.5.1: Eight defects in intracellular processing of cobalamin in humans. The complementation groups, shown in blue, are positioned at sites of metabolic blocks (red). Cobalamin intermediates are also in red. Cbl-R denotes a cobalamin with an undefined group at the β position on the cobalt ion, whilst Me and Ado refer to methyl and adenosyl, respectively. Three complementation groups, *cblF, cblC* and *cblD*, correspond to blocks in steps that are common to the synthesis of both cofactors with resulting deficiency of MS and MCM activities. Patients from these groups have combined homocystinuria and methylmalonic aciduria. Three groups, *cblD* variant 1, *cblE* and *cblG*, have blocks in the cytosolic pathway leading to methyl-cobalamin (MeCbl) synthesis or apo-MS, and result in deficient MS activity and homocystinuria. The final groups, *cblD* variant 2, *cblA*, *cblB* and *mut*, affect steps occurring in the mitochondrion leading to adenosyl-cobalamin (AdoCbl) synthesis or apo-MCM and result in deficient MCM activity and methylmalonic aciduria (Froese and Gravel 2010). This figure was reproduced with permission from Froese and Gravel 2010.

Many parasites, such as *Diphyllobothrium latum*, a fish tapeworm, starve the host of cobalamin by absorbing it directly from the hosts' gut. *H. pylori*, is a gram-negative spiral bacterium, which infiltrates the interface of the gastric epithelial cells and mucosa resulting in gastritis and gastric atrophy (Allen 2008). This reduces IF production causing cobalamin malabsorption from the gut into the blood, leading to deficiency (Allen 2008).

The levels of cobalamin, particularly methylcobalamin decrease with age, especially in the frontal cortex, and are depleted in autistic and schizophrenic patients as well (Zhang et al. 2016). Cobalamin deficiency has also been implicated in increasing the risk of cognitive

impairment disorders such as Alzheimer's or dementia (Spence 2016). There are also suggestions that cobalamin deficiency due to diet increases the likelihood of *M. tuberculosis* infection (Chanarin and Stephenson 1988; Bakhshi et al. 2010). Conversely, *M. tuberculosis* infection has been cited to cause pernicious anaemia and a paper published in the early 1930s proposes antagonism between the two (Ramagopalan et al. 2013; Barron 1933). It has been reported that the expression of the adenosylcobalamin requiring ribonucleotide reductase (type II) is upregulated in *M. tuberculosis* during latent infection (see Section 1.7) (Gopinath, Moosa, et al. 2013; Boshoff and Barry 2005). Therefore, *M. tuberculosis* may have a high demand for cobalamin during disease progression.

There is also evidence to suggest patients suffering from pernicious anaemia and chronic myeloid leukaemia (CML) exhibit retarded CML progression. Providing such patients with cobalamin hastened CML progression whilst withholding cobalamin treatment improved their condition (Corcino et al. 1971; Mclean et al. 1997). Patients with breast cancer have been reported to become cobalamin deficient, and cancer cells were shown to exhibit an unusually high requirement for cobalamin in order to support DNA synthesis prior to cell division (Lee and Grissom 2009; Hogenkamp et al. 1999). In addition, some cancer cells (human colorectal adenocarcinoma, for example) have a decreased proliferation rate when deprived of cobalamin (Lai et al. 2011).

1.6 Cobalamin in therapeutics

Since the initial identification of cobalamin as an essential molecule, the biosynthesis and chemical synthesis of cobalamin has been researched extensively resulting in the elucidation of the two biosynthesis pathways (Moore and Warren 2012; Raux et al. 2000; Eschenmoser 1988; Heldt et al. 2005; Raux et al. 1996). This has spawned a variety of cobalamin related research including the assembly of different cobalamins (Hazra et al. 2013), the structure of cobalamin used by different organisms (Helliwell et al. 2016; Greibe, Fedosov, and Nexo 2012; Greibe, Fedosov, Sorensen, et al. 2012), the chemistry of cobalamin (Dereven'kov et al. 2016; Widner et al. 2016; Shell and Lawrence 2015), and the relationship between cobalamin and disease (Zhang et al. 2016; Young et al. 2015). Through these more specific research areas, the window to the world of cobalamin modification and potential therapeutics has opened (Shell and Lawrence 2015; Brown 2005; Clardy et al. 2011).

Cobalamin is an ideal molecule for therapeutics. It is water soluble, has no known toxicity, has known links to many diseases, is indispensable to most organisms, and therefore,

it is unlikely that mutational arrest of cobalamin uptake will occur (Waibel et al. 2008). The idea of using cobalamin as a therapeutic is not new. Radioactive [57]Co-cobalamin was shown to accumulate in transplanted mice tumours nearly 50 years ago (Flodh and Ullberg 1968). In the 1980s, an anilide of cobalamin was shown to successfully treat acute myelogenous leukaemia, albeit in one patient (Herbert 1983). The anilide analogues were produced by chemical hydrolysis of the amide groups on the sidechains of cobalamin, to form acids, followed by esterification with a aniline, $C_6H_5NH_2$, an amino group attached to a phenyl group (E. L. Smith et al. 1955; Anton et al. 1980; Robinson 1966). Since then research into the production of cobalamin analogues as therapeutics increased dramatically. Chemical modifications of cobalamin were used to determine the possible points of conjugation, or substitution, and how these affected the binding to TCII and other cobalamin transport proteins (Figure 1.6.1) (Pathare et al. 1996). When these analogues were tested in binding assays, the results showed that modifications to the cobalt ion, 5-*O*-ribose moiety, and the *e* sidechain did not significantly decrease binding to TCII. The *b* sidechain conjugates decreased binding to an intermediate degree, whereas alterations to the *c* and *d* sidechains of ring B dramatically decreased binding (Pathare et al. 1996). Further research conducted by the same group revealed that conjugates on the *e* sidechain may be taken into the murine lymphoma BW5147 cells but do not support proliferation (Mclean et al. 1997). Previous reports have shown that *e* analogues cause reduction in MCM and MS activities, and increase methylmalonic acid and homocysteine levels, which suggest that cobalamin metabolism would be a good target for anti-proliferative drugs, and cobalamin analogues could be used to block native cobalamin metabolism (Kolhouse et al. 1993; Stabler et al. 1991; Mclean et al. 1997).

Figure 1.6.1: Possible sites of conjugation or substitution on cobalamin. This schematic was derived from Pathare et al. 1996. It is important to note that there are no sites shown which modify the corrin macrocycle directly.
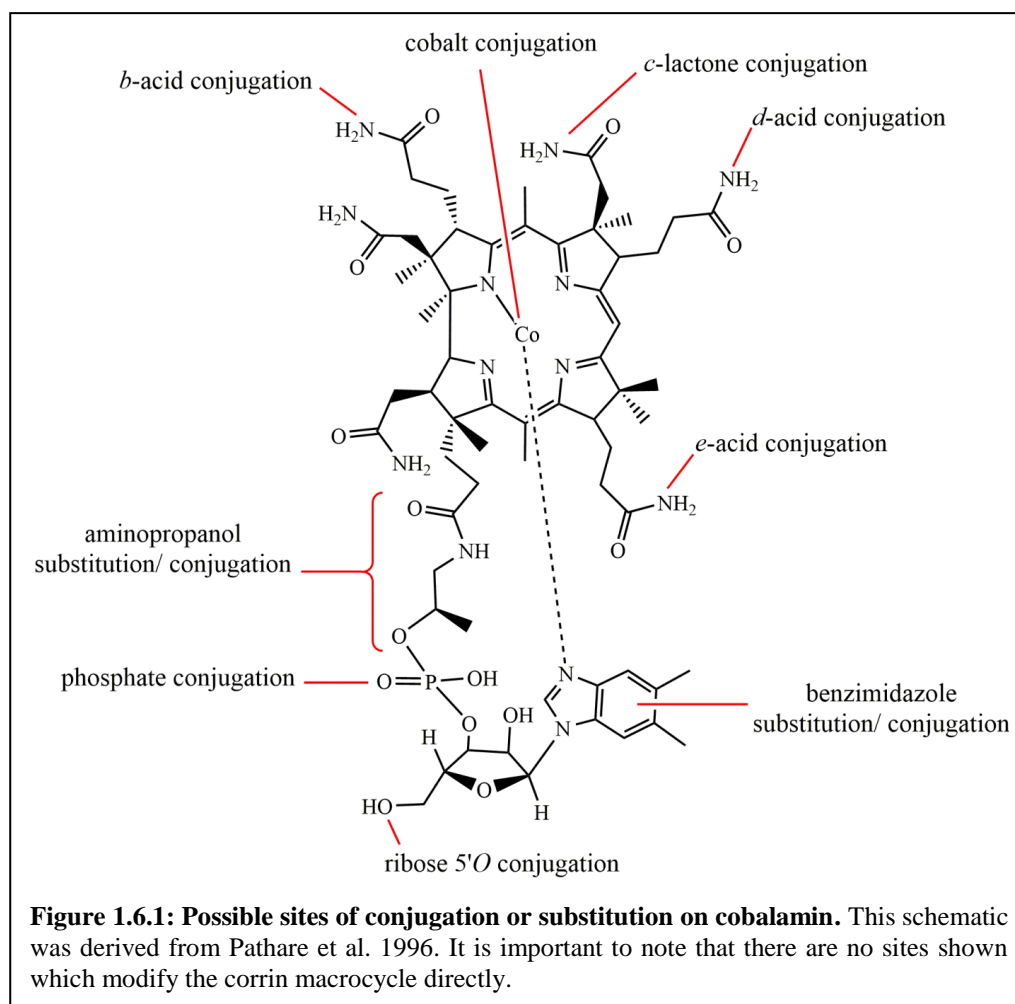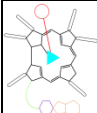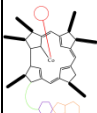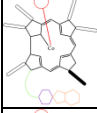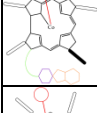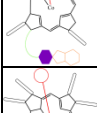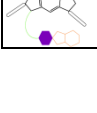
One Swiss study used a radioactive technetium (^{99m}Tc) in an organometallic tracer conjugated to the *b* sidechain to target tumours in murine models (Waibel et al. 2008). Earlier attempts at such experiments used ^{57}Co or ^{111}In to label tumours but, although successful, these also resulted in high accumulation of radioactivity in normal tissue (Flodh and Ullberg 1968). However, in the Swiss study they used the discovery that tumours expressed high levels of cytoplasmic and membrane bound HC to synthesise analogues which did not bind TCII but did bind HC (Kim et al. 1993). TCII is usually the protein which facilitates cellular internalisation of cobalamin, but using this TCII non-binding analogue they found it targeted to tumours, with very little non-specific accumulation (Waibel et al. 2008). These analogues were used in a clinical trial to locate tumours in cancer patients which showed that tumour visualisation is possible, although the sensitivity is low and non-specific accumulation occurs (Sah et al. 2014).
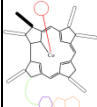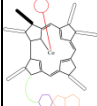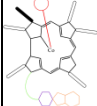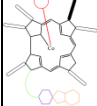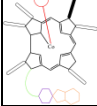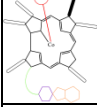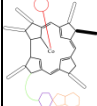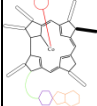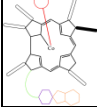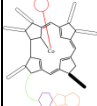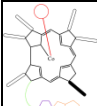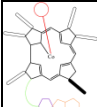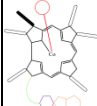
A different study attempted to use fluorophores conjugated to the 5-*O*-ribose moiety on the lower loop to identify rapidly dividing cancer cells, with a view to aiding surgeons to discern cancerous tissue from healthy tissue (Lee and Grissom 2009). Stable cobalamin
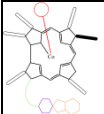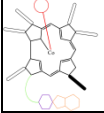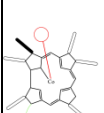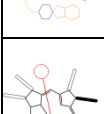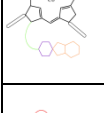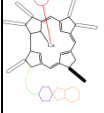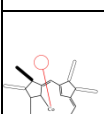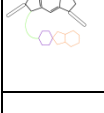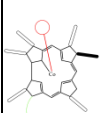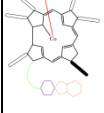
conjugates were produced, however, binding kinetics were not tested for the human cobalamin transport proteins. Previous studies have attached a fluorophore to the β ligand of cobalt before but these analogues suffer from photochemical instability due to the low dissociation energy of the Co-C bond, which falls within the range of visible photons (Smeltzer et al. 2001; Martin and Finke 1992).
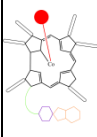
A recent study also used the β ligand of cobalt but this time to conjugated drugs using a light labile bond to trigger the release of the drug to target tumour cells (Shell and Lawrence 2015). Lipidated fluorophore antennae were paired with β ligand drug conjugated cobalamins and loaded onto erythrocytes, which were subsequently incubated with target adherent HeLa cell lines. These were illuminated at the appropriate wavelength causing the drug to be released and migrate into the HeLa cells (W. J. Smith et al. 2014). This has only been shown in culture and, as with all photo-dynamic treatments, it is limited by the distance of light penetration. A summary of all the various cobalamin analogues that have been made is given in Table 1.6.2.

Table 1.6.2: Summary of all the previously synthesised cobalamin analogues described in the text

Position of modification	Modification	Date	Authors
Cobalt ion	Replaced with ^{57}Co, ^{58}Co or ^{60}Co	1968	Flodh, H and Ullberg, S.
Unspecified sidechain	Aniline ($C_6H_5NH_2$)	1956, 1980	Smith, EL et al , Herbert, V
e sidechain	Exchange NH_2 for $N(CH_3)_2$	1991	Stabler, S et al
e sidechain	Exchange NH_2 for OH in cobinamide	1993	Kolhouse, J et al
5'-*O*-ribose	Exchange H for $CO(CH_2)_2CONH(CH_2)_{12}NH_2$	1996	Pathare, PM et al
5'-*O*-ribose	Exchange H for $COCH_2CH_2COOH$	1996	Pathare, PM et al

Position of modification	Modification	Date	Authors
Cobalt β ligand	$CH_2CH_2CH_2NHCO(CH_2)_5NH$-Biotin	1996	Pathare, PM et al
Cobalt β ligand	$CH_2CH_2CH_2NH_2$	1996	Pathare, PM et al
b sidechain	Exchange NH_2 for OH	1996	Pathare, PM et al
b sidechain	Exchange NH_2 for $NH(CH_2)_{12}NH_2$	1996	Pathare, PM et al
b sidechain	Exchange NH_2 for $NH(CH_2)_{12}NH$-Biotin	1996	Pathare, PM et al
c sidechain	*c* lactone	1996	Pathare, PM et al
c sidechain	Exchange NH_2 for $NH(CH_2)_{12}NH_2$	1996	Pathare, PM et al
c sidechain	Exchange NH_2 for $NH(CH_2)_{12}NH$-Biotin	1996	Pathare, PM et al
d sidechain	Exchange NH_2 for OH	1996	Pathare, PM et al
d sidechain	Exchange NH_2 for $NH(CH_2)_{12}NH_2$	1996	Pathare, PM et al
d sidechain	Exchange NH_2 for $NH(CH_2)_{12}NH$-Biotin	1996	Pathare, PM et al
e sidechain	Exchange NH_2 for OH	1991, 1996	Stabler, S et al Pathare, PM et al
e sidechain	Exchange NH_2 for $NH(CH_2)_{12}NH_2$	1996	Pathare, PM et al
e sidechain	Exchange NH_2 for $NH(CH_2)_{12}NH$-Biotin	1996	Pathare, PM et al
b sidechain	Exchange NH_2 for $NH(CH_2)_{12}NH$-*p*-Iodohippurate	1997	M[c]Lean, GR et al

Position of modification		Modification	Date	Authors
	d sidechain	Exchange NH$_2$ for NH(CH$_2$)$_{12}$NH-*p*-Iodohippurate	1997	McLean, GR et al
	e sidechain	Exchange NH$_2$ for NH(CH$_2$)$_{12}$NH-*p*-Iodohippurate	1997	McLean, GR et al
	b sidechain	Dimeric cobalamin linked by exchange of NH$_2$ for NH(CH$_2$)$_{12}$NH$_2$ in both monomers and then linked via an isophthaloyl moiety	1997	McLean, GR et al
	d sidechain	Dimeric cobalamin linked by exchange of NH$_2$ for NH(CH$_2$)$_{12}$NH$_2$ in both monomers and then linked via an isophthaloyl moiety	1997	McLean, GR et al
	e sidechain	Dimeric cobalamin linked by exchange of NH$_2$ for NH(CH$_2$)$_{12}$NH$_2$ in both monomers and then linked via an isophthaloyl moiety	1997	McLean, GR et al
	b sidechain	Dimeric cobalamin linked by exchange of NH$_2$ for NH(CH$_2$)$_{12}$NH$_2$ in both monomers and then linked via an isophthaloyl moiety with *p*-iodobenzoate	1997	McLean, GR et al
	d sidechain	Dimeric cobalamin linked by exchange of NH$_2$ for NH(CH$_2$)$_{12}$NH$_2$ in both monomers and then linked via an isophthaloyl moiety with *p*-iodobenzoate	1997	McLean, GR et al
	e sidechain	Dimeric cobalamin linked by exchange of NH$_2$ for NH(CH$_2$)$_{12}$NH$_2$ in both monomers and then linked via an isophthaloyl moiety with *p*-iodobenzoate	1997	McLean, GR et al
	Cobalt β ligand	3-aminopropyl linked Oregon Green, fluorescein, and naphthofluorescein	2001	Smeltzer, C et al
	b sidechain	Exchange H for [pyridine-2-ylmethyl-amino]-acetic acid [^{99m}Tc(OH$_2$)$_3$(CO)$_3$]	2008	Waibel, R et al
	5'-*O*-ribose	Exchange H for trans-1,4-Diaminocyclohexane with Rhodamine 6G or fluorescein	2009	Lee, M and Grissom, C

Position of modification		Modification	Date	Authors
	Cobalt β ligand	C18 linked methotrexate, colchicine, dexamethasone, Tetramethylrhodamine, or fluorescein	2014	Smith, W et al
	Cobalt β ligand	$CH_2CH_2CH_2COHN$ linked SulphoCy5, AlexaFluor700, Atto725, or DyLight800	2015	Shell, T and Lawrence, D

1.6.1 The need for new cobalamin analogues

All of the potential cobalamin therapeutics listed in Table 1.6.2 have limitations. The sidechain *e* analogues are not specific to certain cell types and rely upon competition against intrinsic cobalamin to instigate a decrease in cell proliferation. The fluorophore conjugates hoping to target rapidly dividing cells have yet to be tested and are likely to photobleach during the proposed surgery. The most useful development is the observation that radioactive cobalamin analogues with no TCII recognition, accumulate in tumour cells by exploiting the potential HC uptake route, thereby increasing specificity to the tumours. However, in clinical trials it was shown to have low sensitivity and non-specific accumulation.

Cobalamin analogues have great potential for therapeutic use not only in tumour identification but also in drug delivery. It is important to retain recognition by human transport proteins for targetting non-tumour highly proliferative cells, but the ^{99m}Tc studies show that only HC recognition is required for tumour targeting (Waibel et al. 2008). This is potentially exploitable as HC recognises cobalamin intermediates prior to complete lower loop assembly, whilst TCII requires the full DMB lower loop to be present (Nielsen et al. 2012). Therefore, providing this uptake mechanism is legitimate, earlier intermediate analogues should be specifically internalised by tumour cells. Bacteria often have the end of the cobalamin biosynthesis pathway so they can complete coenzyme synthesis from earlier cobalamin intermediates. Humans do not have this ability to complete the synthesis of partial corrinoids, which suggests that earlier intermediate analogues can also be used to target bacterial infections, particularly those that cause cobalamin deficiency.

An earlier cobalamin intermediate analogue cannot be modified on the lower loop as the loop is incomplete or not present. Sidechain modifications have mixed recognition, whilst cobalt β ligand analogues are light sensitive and the conjugate could be exchanged *in vivo* by

the native enzymes for a methyl or adenosyl group. Ideally the conjugation should be directly on the corrin ring to limit the impediment of recognition. A ring modified corrin analogue could have a drug, a fluorophore or a radioactive moiety conjugated, but will not have any of the issues of light labile bonds or reduced recognition due to sidechain modifications. Macrocycle altered analogues could pave the way to a new generation of cobalamin therapeutics.

1.7 The enigma of cobalamin and *Mycobacterium tuberculosis*

M. tuberculosis is an obligate human pathogen. The World Health Organisation (WHO) cites tuberculosis (TB) as a major global health problem, alongside the Human Immunodeficiency Virus (HIV). TB is in fact the main cause of death in HIV positive patients. It is predicted that a third of the human population is latently infected with *M. tuberculosis* which accounts for the maintenance of TB in the human population despite a cocktail of drug treatments having been available for a long time (World Health Organisation 2015). Only about 5 % of individuals with latent TB will develop active infection, termed 'reactivation disease', this figure is much higher in immunosuppressed people (e.g. via HIV infection, or by therapeutic administration of tumour necrosis factor α). Latency is typically characterised in two stages of non-replicating persistence (NRP) (Boshoff and Barry 2005): NRP1, recognised by cessation of cell division with oxygen levels of 1 % saturation, and NRP2, which is a more comprehensive arrest of metabolism at oxygen levels of 0.06 % saturation (Boshoff and Barry 2005). *M. tuberculosis* is able to persist at 0.06 % oxygen without supplementation of external terminal electron acceptors, suggesting that an energised membrane is maintained without respiration; however, it is not known how this is achieved.

M. tuberculosis has three cobalamin-dependent enzymes; MCM, MetH and a ribonucleotide reductase (RNR), encoded by *nrdZ* (Gopinath, Venclovas, et al. 2013; Gopinath, Moosa, et al. 2013; Warner et al. 2007). None of the three cobalamin-dependent enzymes in *M. tuberculosis* appear to be essential (Gopinath, Moosa, et al. 2013). RNRs reduce ribonucleotide diphosphates (rNDPs) into deoxyribonucleoside diphosphates (dNDPs), which are then used to make deoxyribonucleoside triphosphates (dNTPs) and can be used in DNA synthesis, hence they are essential during cell replication (Gopinath, Moosa, et al. 2013; Dawes et al. 2003). There are broadly three classes of RNRs, I to III, and they each depend on different metal coenzymes; class I contain diiron-oxygen cluster, class II use adenosyl-cobalamin, and class III have a 4Fe-4S iron-sulphur cluster and require SAM (Kolberg et al.

2004). *M. tuberculosis* has both class I, which are oxygen dependent, and class II RNRs. The *nrdZ* gene is predicted to be a class II ribonucletide reductase and is completely oxygen independent (Dawes et al. 2003). In contrast to class I RNRs, the class II enzymes only have one subunit, as the cleavage of the (Co-C)-bond adenosyl-cobalamin generates the thiol radical instead of the second protein subunit present in class I RNRs (Kolberg et al. 2004; Gruber et al. 2011). In *M. tuberculosis* it is the *nrdEF2* class I ribonucleotide reductase which is essential for growth, and the *nrdZ* is only induced in the dormancy response (Gopinath, Moosa, et al. 2013). It can be concluded that *nrdZ* requirement increases in low oxygen environments in the presence of adenosyl-cobalamin. It is here that *M. tuberculosis* has been postulated to become cobalamin-dependent during infection, due to the limited oxygen causing the other RNRs to be compromised (Gopinath, Moosa, et al. 2013). This makes the exit from dormancy a potential access point for a cobalamin linked therapeutic.

The *M. tuberculosis* methionine synthases, MetE and MetH, both use 5-methylterahydrofolate (N^5-MeTHF) as the methyl donor when converting L-homocysteine into L-methionine (Figure 1.3.1). Unlike MetE, MetH requires methylcobalamin to function, and although the catalytic turnovers have not been calculated in *M. tuberculosis*, they are likely to show a similar benefit to cobalamin dependency as in *E. coli* (Gonzalez et al. 1992).

MCM has no specific cobalamin independent enzyme to compensate its function in cobalamin deficient circumstances, although there is an alternative pathway (Figure 1.7.1). MCM is responsible for converting methylmalonyl-CoA to succinyl-CoA with an adenosyl-cobalamin coenzyme in the final step of the methylmalonyl pathway. This pathway is one of the ways in which *M. tuberculosis* can remove propionate from the cell. Propionate accumulates as the three carbon terminal product of odd and branched chained fatty acids, branched amino acids and cholesterol catabolism. Although propionate is a key precursor in lipid biosynthesis, it is toxic if accumulated. The methylcitrate cycle and the glyoxylate cycle can also stop propionate from accumulating. Interestingly, in *M. tuberculosis* methylcitrate lyase (MCL) is not present and instead isocitrate lyase (ICL) from the glyoxylate cycle compensates for this methylcitrate cycle enzyme (underlined in red in Figure 1.7.1) (Savvi et al. 2008). This means that if ICL is compromised *M. tuberculosis* will become reliant on the methylmalonyl pathway and therefore cobalamin-dependent. Although the existence of these cobalamin-dependent enzymes is acknowledged, very little is understood about their regulation or their influence on viability and pathogenicity of *M. tuberculosis* (Warner et al. 2007).

Figure 1.7.1: **Predicted routes of propionate metabolism in *M. tuberculosis*.** The ICL (isocitrate lyase) is underlined in red to highlight the reliance of both the methyl citrate and the glyoxylate pathways on this enzyme due to the loss of MCL (Methylcitrate lyase) in *M. tuberculosis*. Should ICL be knocked out *M. tuberculosis* propionate metabolism would become adenosyl-cobalmain (Ado-B$_{12}$) dependent due to the reliance of MCM (boxed in orange) on this coenzyme. ACN: Aconitase, CIT: citrate synthase, FUM: Fumerase, ICD: Isocitrate dehydrogenase, MCD: Methylcitrate dehydratase, MCS: Methylcitrate synthase, MDH/MQO: Malate dehydrogenase/ Malate: quinine oxidoreductase, MEZ: Malic enzyme, MLS: Malate synthase, MMCE: Methylmalonyl-CoA epimerase, PCA: Pyruvate casrboxylase, PCC: Propionyl-CoA carboxylase, PCK: Pyruvate carboxykinase, PDHC: pyruvate dehydrogenase, PYK: Pyruvate kinase, SCS: Succinate synthase, SDH: Succinate dehydrogenase. This figure was reproduced

The genome of *M. tuberculosis* has been sequenced (Cole et al. 1998). It reveals that *M. tuberculosis* has all of the genes for the cobalamin biosynthesis pathway except for *cobF*, encoding the C1 deacetylase and methyltransferase, which has led to speculation about whether it can produce cobalamin *de novo* (Savvi et al. 2008). Curiously, *Mycobacterium smegmatis*, which is not a human pathogen but a soil bacterium, has all of the biosynthetic

genes, and does synthesise cobalamin whereas *Mycobacterium leprae* has only the biosynthesis genes to convert cobinamide into cobalamin (Cole et al. 2001; Szumowski et al. 2013; Dawes et al. 2003). Both these organisms have homologues of MetH and MCM (Savvi et al. 2008; Young et al. 2015). *M. leprae* has undergone a severe genome reduction so it is not surprising that it has lost most of the biosynthetic genes, but by keeping the genes to allow the salvage of incomplete corrinoids, *M. leprae* is able to adapt to environmental conditions and to satisfy its cobalamin-dependent enzymes (Savvi et al. 2008).

M. tuberculosis has never been shown to synthesise cobalamin *de novo*, although it could scavenge exogenous corrinoids, and it has known cobalamin-dependent enzymes that are not essential. It is therefore unknown to what extent cobalamin is needed in the lifecycle of *M. tuberculosis*.

1.8 Cobalamin and *Caenorhabditis elegans*

C. elegans is a soil dwelling nematode worm which feeds on microbes, particularly on bacteria. They are an ideal model for humans in many research areas including, genomics, cell biology, neuroscience and aging, and have been shown to exhibit similar cobalamin deficiency phenotypes as humans (Altun and Hall 2009b; Froese and Gravel 2010; Bito et al. 2013). *C. elegans* has both MetH and MCM, but very little is known about the cobalamin transport system (Bito et al. 2013). Like all non-monera, *C. elegans* must obtain cobalamin from its diet; therefore, it must have a transport protein from the gut to the pseudocoelom, however, one has not been identified. A study has been conducted showing that ribose linked dodecylamine derivatives of cobalamin inhibit nematode MCM and MetH, inducing a deficiency phenotype (Bito et al. 2014). These analogues have the same effect on mammalian cell lines but have not been tested *in vivo* in mammals (Bito et al. 2014).

C. elegans are predominantly hermaphrodites, although males do also occur (Altun and Hall 2009b). They have very simple bodies and a complete cell lineage has been established (Brenner 1974; Sulston et al. 1983). They are small (adults are around 1130 µm long), can easily be grown under lab conditions with a lifecycle of about 3 days at 20°C and are transparent throughout their lifecycle (Altun and Hall 2009b).

1.9 Cobalamin, *Lepidium sativum* and *Arabidopsis thaliana*

A. thaliana is the most commonly used model organism for plants owing to its small stature and short lifecycle. Additionally, the whole genome of *A. thaliana* has been sequenced (Boyes et al. 2001). *L. sativum*, garden cress, is an easily obtainable plant which can be grown alongside *A. thaliana* to provide a comparison. Although no plants require or synthesise cobalamin, some have been shown to take it up, which offers a counterpoint to the organisms described above (Watanabe et al. 2013; Helliwell et al. 2011).

1.10 The basis and aims of the investigations presented in this thesis

Work into cobalamin analogue synthesis is not a new area of research. Initial usage was intended to be as a tumour detection mechanism, as studies dating back over 40 years have shown high accumulation of ^{57}Co cobalamin in tumours (Waibel et al. 2008). From this start point the applications have diversified into proliferation inhibition (Mclean et al. 1997), tumour specific localisation (Waibel et al. 2008), and now specific drug/ cargo transport (Clardy et al. 2011). It is known that many forms of cancer cells, particularly solid malignant tumours, increase their requirement of cobalamin and, recently, *M. tuberculosis* infected cells have been implicated for the same reason (Waibel et al. 2008; Brown 2005; Gopinath, Moosa, et al. 2013). Most of the previously synthesised analogues target these compromised cells. However, these analogues revolve around modifications to the cobalt ion ligands, the peripheral sidechains, or to the nucleotide loop, all of which have various shortcomings (Section 1.6). This by no means implies that cobalamin analogues are limited in usage, in fact they have already been successfully applied in treatment (Herbert 1983; Corcino et al. 1971), but rather that the site of modification should ideally be directly on the corrin macrocycle. This will increase specificity and stability of modification without compromising recognition.

The objectives of the investigations detailed forthwith are:

1) To synthesise a corrin modified analogue of cobalamin.

2) To attach a fluorophore or drug at this position

3) To analyse the efficacy of these analogues in different organisms

4) To identify corrinoid specificity and localisation in *M. tuberculosis* and *C. elegans*.

There is much still unknown about cobalamin uptake, localisation and utilisation, but by designing fluorescent analogues that can be followed *in vivo*, the journey and fate of cobalamin within these organisms can be studied in greater detail than formerly possible. A potential outcome of this research could be a corrin conjugated drug transportation mechanism which uses the cobalamin uptake pathway to enter the cell whereupon the drug takes effect in a Trojan horse-like subterfuge.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Chemicals and equipment

Most of the chemicals were bought from Sigma Aldrich, Ltd. or Thermo Fisher Scientific, Inc.. Those that were not are as follows: Isopropyl β-D-1-thiogalactopyranoside (IPTG) and ampicillin from Melford Laboratories, Ltd., VA044 from Wako Pure Chemical Industries, Ltd., chelating sepharose, buffer exchange and empty PD-10 columns from GE Healthcare Lifesciences, Amicon® 5 kDa centrifuge protein concentrators and LiChroprep® RP-18 from Merck Millipore, Merck KGaA, QIAprep® Spin Miniprep Kit, QIAquick® Gel Extraction Kit from QIAGEN GmbH, restriction enzymes from Promega, cloning strains from Novagen, tryptone, yeast extract and bacterial agar from Oxoid, Ltd., Middlebrook 7H9 media and Bovine serum albumin from Becton, Dickinson and Company, Roche FastStart High Fidelity PCR system from Roche Diagnostics GmbH, 4-20 % Mini-PROTEAN® TGX™ precast polyacrylamide gels (USA) and Bradford reagent from Bio-Rad Laboratories, Inc., 4-20 % Novex TBE polyacrylamide gels (UK) were also purchased from Thermos Fisher Scientific, Inc., ATTO 1D dual mini slab gel tank from ATTO corporation, Ace 5 AQ column (2.1 x 150 mm) 5 µm particle size, 100 Å, from Advanced Chromatography Technologies Ltd..

2.1.2 Bacterial strains and plasmids

The bacterial strains were purchased from Novagen, Invitrogen or Promega. All the *Mycobacterium tuberculosis* strains were provided by H. Boshoff and C. Barry III, National Institutes of Health, MD, USA and originally sourced from . Plasmids which were previously constructed are listed alongside the name of the researcher who assembled them.

2.1.2.1 Bacterial strains

Table 2.1.2.1.1 Bacterial strains used

Strains	Genotype/ Phenotype	Description	Origin
Escherichia coli			
JM109	*end*A1 *rec*A1 *gyr*A96, *thi hsd*R17 (r_κ-m_κ^+) *rel*A1 *sup*E44 Δ(*lac-pro*AB) [F', *tra*D36 *pro*AB *lac*I^{q}ZΔM15]		Promega

Strains	Genotype/ Phenotype	Description	Origin
BL21(DE3) pLysS	F⁻ *ompT hsdS*$_B$ (r$_B^-$, m$_B^-$) *gal dcm* (DE3) pLysS (Cm^R)		Novagen
BL21(DE3)	F⁻ *ompT hsdS*$_B$ (r$_B^-$, m$_B^-$) *gal dcm* (DE3)		Novagen
Rosetta (DE3)	F⁻ *ompT hsdS*$_B$(r$_B^-$, m$_B^-$) *gal dcm* (DE3) pRARE (Cam^R)		Novagen
OP50		Uracil auxotroph mutant of *E. coli* B (Berkeley) (May et al. 2009)	Dr J. Tullet
OP50 BtuBF		OP50 transformed with pET-BAD-*btuBF*, arabinose promoter	Dr E. Deery
OP50 (-)		OP50 transformed with pET 3a	Dr E. Deery
BL21(DE3) ABCDCX^{S}L		BL21(DE3) transformed with pETcoco-2-*cobA, hemBCD,sirC, cbiX^S, cbiL. hemB, sirC* and *cbiX^{S}L* from *Methanobacterium thermoautotrophicum, cobA* from *Methanosarcina barkeri*, and *hemCD* from *Bacillus megaterium.*	Dr S. Frank with *cbiX^{S}L* added by Dr S. Moore
Bacillus megaterium			
B. meg CbiH^{His}	*Bacillus megaterium* DSM509 (Moore and Warren 2012)	*B. megaterium* DSM509 transformed with P$_{xylA}$-*CbiH*^{His}. *CbiH* from *B. megaterium.*	Dr S. Moore
B. meg CbiF^{His}	*Bacillus megaterium* DSM509 (Moore and Warren 2012)	*B. megaterium* DSM509 transformed with P$_{xylA}$-*cbiF*^{His}. *cbiF* from *B. megaterium.*	Dr S. Moore
B. meg CbiD^{His}	*Bacillus megaterium* DSM509 (Moore and Warren 2012)	*B. megaterium* DSM509 transformed with P$_{xylA}$-*cbiD*^{His}. *cbiD* from *B. megaterium.*	Dr S. Moore
B. meg CbiG^{His}	*Bacillus megaterium* DSM509 (Moore and Warren 2012)	*B. megaterium* DSM509 transformed with P$_{xylA}$-*cbiG*^{His}. *cbiG* from *B. megaterium.*	Dr S. Moore
B. meg CbiJ^{His}	*Bacillus megaterium* DSM509 (Moore and Warren 2012)	*B. megaterium* DSM509 transformed with P$_{xylA}$-*cbiJ*^{His}. *cbiJ* from *B. megaterium.*	Dr S. Moore
Mycobacterium tuberculosis			

Strains	Genotype/ Phenotype	Description	Origin
WT	H37Rv ATCC® 27294	Clinical isolate from 1934.	Dr H. Boshoff
H37Rv	H37RvJO	Clinical isolate from 1998.	Dr H. Boshoff
ΔbacA	H37RvJO with *hyg*-marked *ΔbacA* deletion (Warner et al. 2007).		Dr H. Boshoff
ΔmetE	H37RvJO with *hyg*-marked *ΔmetE* deletion (Warner et al. 2007).		Dr H. Boshoff
ΔmetH	H37RvJO with *ΔmetH* deletion (Warner et al. 2007).		Dr H. Boshoff

2.1.2.2 Plasmids

Table 2.1.2.2.1 Plasmids used

Name	Description	Origin
pET14b	Over-expression N-terminal His-tag fusion protein vector with T7 promoter, Amp^R.	Novagen
pET3a	Over-expression vector with T7 promoter, Amp^R.	Novagen
pLysS	Vector with basal T7 lysozyme expression in λDE3 lysogenic host to suppress the T7 promoter.	Novagen
pETcoco-2	A low copy plasmid maintained in a single copy state which can be switched to a medium copy state by inducing the expression of the *trfA* gene with arabinose.	Novagen
pETcoco-2-ABCDCX^{S}L	Low copy plasmid with *cobA, hemBCD,sirC, cbiX^S, cbiL. hemB, sirC* and *cbiX^{S}L* from *Methanothermus thermoautotrophicus*, *cobA* from *Methanosarcina barkeri*, and *hemCD* from *Bacillus megaterium*.	Dr S. Moore
pET14b-*cbiE*	PCR product of the *cbiE* gene from *Methanothermus thermoautotrophicus* inserted *Nde*I and *Spe*I into pET14b.	Dr S. Moore
pET14b-*cbiT*	PCR product of the *cbiT* gene from *Methanothermus thermoautotrophicus* inserted *Nde*I and *Spe*I into pET14b.	Dr S. Moore
pET14b-*cbiC*	PCR product of the *cbiC* gene from *Salmonella enterica* serovar typhimurium inserted *Nde*I and *Spe*I into pET14b.	Dr S. Moore
pET3a-*cobAIGJFMKL^{C}E*^{His}	PCR product of *Rhodobacter capsulatus* SB1003 for *cobA, I, J, F, M, K, L* and *H, Brucella melitensis* 16M for *cobE*, and *Pseudomonas denitrificans* for *cobG* inserted *Ase*I or *Nde*I at the 5' and *Spe*I at the 3' into pET3a.	Dr E. Deery

Name	Description	Origin
pET14b-*cobL*	PCR product of *Rhodobacter capsulatus* SB1003 for *cobL* inserted *Nde*I at the 5' and *Spe*I at the 3' into pET14b.	Dr E. Deery
pET14b-*cobH*	PCR product of *Rhodobacter capsulatus* SB1003 for *cobH* inserted *Nde*I at the 5' and *Spe*I at the 3' into pET14b.	Dr E. Deery
pET14b-*cobB*	PCR product of *Rhodobacter capsulatus* SB1003 for *cobB* inserted *Nde*I at the 5' and *Spe*I at the 3' into pET14b.	Dr E. Deery
pET14b-*cobQ*	PCR product of *Allochromatium vinosum* DSM 180 for *cobQ* (*Alvin_2223*) inserted *Nde*I at the 5' and *Spe*I at the 3' into pET14b.	This study
pET14b-*btuF*	PCR product of *Escherichia coli* for *btuF* inserted *Nde*I at the 5' and *Spe*I at the 3' into pET14b.	Dr E. Deery
pLysS-*btuB*	PCR product of *Escherichia coli* for *btuB* inserted *Nde*I at the 5' and *Spe*I at the 3' into pLysS.	Dr E. Deery
pET14b-*cobH*(T85A)	PCR product of *Rhodobacter capsulatus* SB1003 for *cobH* with T85A mutant introduced, inserted *Nde*I at the 5' and *Spe*I at the 3' into pET14b.	Dr. A Lawrence

2.1.2.3 PCR of *cobQ* from *Allochromatium vinosum*

Primers, obtained from Invitrogen Life Technologies by Mr J. Baker:

Forward: CTA<u>CATATG</u>ACCGATTCAGCCCCCAC

Reverse: CAT<u>ACTAGT</u>TCAGCGTGCCAGTTCGAG

The restriction sites are underlined: *Nde*I in the forward primer, and *Spe*I in the reverse primer.

2.1.3 Media and solutions for bacterial work

Luria-Bertani (LB) broth: Tryptone   10 g

                             Yeast Extract 5 g

                             NaCl   5 g

This was made up to 1 litre with distilled H_2O and autoclaved.

2YT broth: Tryptone   16 g

                             Yeast Extract 10 g

NaCl   5 g

This was made up to 1 litre with distilled H_2O and autoclaved.

Super optimal broth with catabolite repression (SOC) medium:

Tryptone   2 g

Yeast Extract 0.5 g

NaCl   10 mM (Final concentration)

KCl   2.5 mM (Final concentration)

This was made up to 97.2 mL and autoclaved, and then the following were added:

Mg^{2+} 20 mM (Final concentration)*

Glucose   0.36 % (Final concentration)

*Mg^{2+} was made up of 2 g $MgCl_2\cdot6H_2O$ and 2.5 g $MgSO_4\cdot7H_2O$ made up to 10 mL with dH_2O and filter sterilised.

The two latter additions were filter sterilised before addition to the autoclaved solution.

LB agar: Bacterial agar 15 g

This was added to 1 L LB broth, shown above.

Fermenter medium: YE   40 g

10x M9 salts   400 mL

Glucose   8 g

Glycerol   160 g

1M $MgSO_4$ 8 mL

0.1M $CaCl_2$ 4 mL

This was made up to 4 L with distilled H_2O.

10x M9 salts Na_2HPO_4 60 g

                                  KH_2PO_4 30 g

                                  NH_4Cl 10 g

                                  NaCl   5 g

This was made up to a litre with distilled H_2O and then autoclaved.

7H9-ADC media   Middlebrook 7H9   9.4 g

                                  Glycerol   0.2 %

                                  Tween 80   0.05 %

This was made up to 1800 mL with distilled H_2O and filter sterilised.

ADC   Bovine serum albumin   50 g

                                  Glucose   20 g

                                  NaCl   8.1 g

This was made up to 1 L with distilled H_2O and filter sterilised.

The 1800 mL of 7H9 medium was made up to 2 L with 200 mL of the ADC and mixed thoroughly. The rest of the ADC was stored at 4 °C.

Sauton's media KH_2PO_4 0.5 g

                                  $MgSO_4 \cdot 7H_2O$ 0.5 g

                                  Citric acid   2.0 g

                       Ferric Ammonium Citrate   0.05 g

                                  Glycerol   6 %

                                  Asparagine   4 g

                                  TWEEN80   0.05 %

NaOH pellets were added until the medium was at pH 7.

This was made up to 1 L with distilled H_2O and filter sterilised.

Calcium Chloride solution, 0.1 M: $CaCl_2 \cdot H_2O$ 2.94 g

This was made up to 200 mL with distilled H_2O.

The calcium chloride solution with 15 % glycerol was made in the same way but with 30 mL of glycerol as well, and then made up to 200 mL with distilled H_2O.

Isopropyl β-D-1-thiogalactopyranoside (IPTG), 1 M:

IPTG   2.383 g

This was made up to 10 mL with distilled H_2O and then filter sterilised.

Arabinose, 2 % (w/v)   Arabinose   1 g

This was made up to 50 mL with distilled H_2O and filter sterilised.

Xylose, 5 % (w/v)   Xylose   2.5 g

This was made up to 50 mL with distilled H_2O and then autoclaved.

Glucose, 20 % (w/v)   Glucose   40 g

This was made up to 200 mL with distilled H_2O and then autoclaved.

Ferric citrate, 1 M   Ferric citrate   2.5 g

This was made up to 10 mL with distilled H_2O, after slight heating to dissolve, and filter sterilised.

Antibiotics:

Table 2.1.3.1: The Antibiotics used for *E. coli* growth

Antibiotics	[Stock]	[Working]
Kanamycin	50 mg mL^{-1} in dH$_2$O	30 µg mL^{-1}
Tetracycline	5 mg mL^{-1} in ethanol	10 µg mL^{-1}
Ampicillin	100 mg mL^{-1} in dH$_2$O	100 µg mL^{-1}
Chloramphenicol	34 mg mL^{-1} in ethanol	34 µg mL^{-1}
Hygromycin B	50 mg mL^{-1}	25 µg mL^{-1}

2.1.4 Media and solutions for DNA work

TE Buffer   Tris-HCl, pH 8.0   10 mM

                     EDTA, pH 8.0   1 mM

6x DNA Loading Buffer: Bromophenol blue   0.25 %

                              Glycerol   50 %

                              TE Buffer   50 %

Molecular size marker Hyperladder[TM] I (Bioline)

Figure 2.1.4.1: DNA Hyperladder. 5 μL of the Hyperladder I applied to 1 % (w/v) agarose gel.
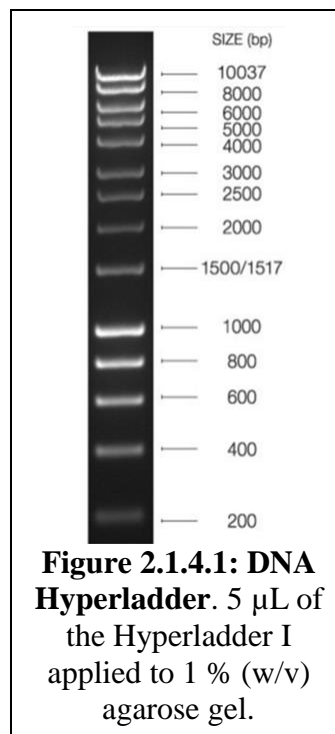
2.1.5 Media and solutions for protein work

2.1.5.1 Solutions for nickel ion affinity chromatography

Charge buffer $NiSO_4$ 100 mM

High salt binding buffer   Tris-HCl, pH 8.0   20 mM

                              NaCl   500 mM

                              Imidazole, pH 7.5   10 mM

High salt wash buffer I:	Tris-HCl, pH 8.0	20 mM
	NaCl	500 mM
	Imidazole, pH 7.5	50 mM
High salt wash buffer II:	Tris-HCl, pH 8.0	20 mM
	NaCl	500 mM
	Imidazole, pH 7.5	100 mM
High salt elution buffer:	Tris-HCl, pH 8.0	20 mM
	NaCl	500 mM
	Imidazole, pH 7.5	400 mM
Low salt binding buffer	Tris-HCl, pH 8.0	20 mM
	NaCl	100 mM
	Imidazole, pH 7.5	10 mM
Low salt wash buffer I:	Tris-HCl, pH 8.0	20 mM
	NaCl	100 mM
	Imidazole, pH 7.5	50 mM
Low salt wash buffer II:	Tris-HCl, pH 8.0	20 mM
	NaCl	100 mM
	Imidazole, pH 7.5	100 mM
Low salt elution buffer:	Tris-HCl, pH 8.0	20 mM
	NaCl	100 mM
	Imidazole, pH 7.5	400 mM
Strip buffer:	Tris-HCl, pH 8.0	20 mM
	NaCl	100 mM

EDTA, pH 8.0   100 mM

2.1.5.2 Solutions for buffer exchange

Exchange buffer: Tris-HCl, pH 8.0   20 mM

                          NaCl   100 mM

2.1.5.3 Solutions for cobalamin-agarose column chromatography

Wash buffer I: Tris-HCl, pH 8.0   20 mM

Wash buffer II: Tris-HCl, pH 8.0   20 mM

                          Cyanocobalamin   10 nM

Wash buffer III: Tris-HCl, pH 8.0   20 mM

                          Cyanocobalamin   100 nM

Wash buffer IV: Tris-HCl, pH 8.0   20 mM

                          Cyanocobalamin   1 μM

Elution buffer: Tris-HCl, pH 8.0   20 mM

                          Urea   8 M

2.1.6 Media and solutions for cobalamin intermediate purification work

2.1.6.1 Solutions for DEAE-Sephacel® column chromatography

Wash buffer I: Tris-HCl, pH 8.0   20 mM

Wash buffer II: Tris-HCl, pH 8.0   20 mM

                          NaCl   100 mM

Wash buffer III: Tris-HCl, pH 8.0   20 mM

                          NaCl   200 mM

Elution buffer: Tris-HCl, pH 8.0   20 mM

                          NaCl   600 mM

2.1.6.2 Solutions for RP-18 column chromatography

Wash buffer I:	Trifluoroacetic acid	0.1 % (v/v)
Wash buffer II:	Methanol	5 %
Wash buffer III:	Methanol	15 %
Wash buffer IV:	Methanol	25 %
Wash buffer V:	Methanol	50 %
Wash buffer VI:	Methanol	100 %

2.1.7 Solutions for protein acrylamide gel and MALDI

2.1.7.1 Sodium dodecyl sulphate (SDS) gel

2x sample buffer	500mM Tris-HCl, pH 6.8	2.5 mL
	Glycerol	2 mL
	10 % (w/v) SDS	4 mL
	β-mecaptoethanol	1 mL
	Bromophenol blue	Two tips
10x Running buffer	Tris-HCl	30 g L^{-1}
	Glycine	144 g L^{-1}
Coomassie blue stain	Trichloroacetic acid (100 %)	250 mL
	Coomassie blue R250	0.6 g
	SDS	0.1 g
	Tris-HCl	0.25 g
	Glycine	0.15 g

This was made up to 500 mL with double distilled H$_2$O.

Molecular mass marker, NEB broad range:
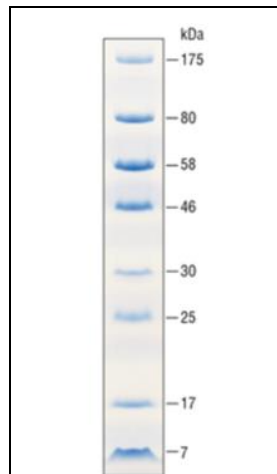
Figure 2.1.7.1.1: Broad range protein marker. 7 μL of the molecular mass marker was applied to the SDS gel.

Molecular mass marker, Spectra™ Multicolor Low Range Protein Ladder:
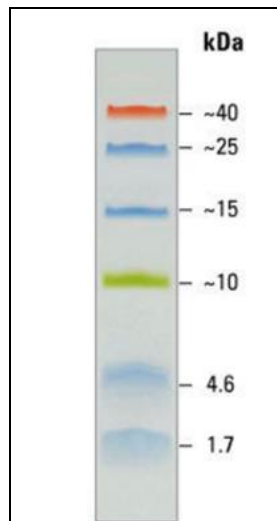
Figure 2.1.7.1.2: Low range protein marker. 7 μL of the molecular mass marker was applied to the SDS gel.

2.1.7.2 SDS gel compositions

Table 2.1.7.2.1: SDS gel compositions

SDS gels						
Running gels	8 %	10 %	12.5 %	15 %	Stacking gel	5 %
dH$_2$O (mL)	5.7	4.7	3.4	2.2	dH$_2$O (mL)	3.4
30 % Acrylamide (mL)	4	5	6.3	7.5	30 % Acrylamide (mL)	1.5
1.5 M Tris-HCl, pH 8.8 (mL)	3.8	3.8	3.8	3.8	0.5M Tris-HCl, pH 6.8 (mL)	1.9
10 % SDS (mL)	1.5	1.5	1.5	1.5	10 % SDS (mL)	0.75
10 % APS (mL)	0.15	0.15	0.15	0.15	10 % APS (mL)	0.075
TEMED (mL)	0.01	0.01	0.01	0.01	TEMED (mL)	0.01

2.1.7.3 MALDI solutions

Gel stains

5 % coomassie blue G-250 stock

        Coomassie blue G-250   0.5 g

This was made up to 10 mL with double distilled H$_2$O.

Colloidal coomassie blue G-250 dye stock solution

        Ammonium sulphate   50 g

        Phosphoric acid 85 % (w/v)   6 mL

        5 % coomassie blue G-250 stock   10 mL

This was made up to 500 mL with double distilled H$_2$O.

Colloidal coomassie blue G-250 dye working solution (prepared immediately before staining)

        Coomassie blue G-250 dye stock solution   40 mL

        Methanol   10 mL

Gel fix   Acetic acid   10 %

        Ethanol   40 %

All made up in distilled H$_2$O.

Wash solution I:	NH_4HCO_3: acetonitrile (1:1)	50 mM
Wash solution II:	NH_4HCO_3	50 mM
Swell solution I:	DTT	10 mM
	NH_4HCO_3	50 mM
Swell solution II:	NH_4HCO_3	50 mM
	Iodoacetimide	50 mM
Digest solution:	NH_4HCO_3	10 mM
	Acetonitrile	10 %

This digest solution was diluted 1 in 10 for the working solution.

Trypsin solution	Trypsin	20 µg

This was all rehydrated in 200 µL of resuspension buffer.

Extraction solutions	Acetonitrile	50 %
	Formic acid	5 %
TA solution	0.1 % Trifluoroacetic acid	
	100 % Acetonitrile	

These two were combined in a 2:1 ratio.

Matrix solution	2,5-dihydroxybenzoic acid	1 mg

This was diluted in 1 mL of TA solution.

2.1.8 Solutions for gel filtration column (size exclusion)

Buffer A, Sample buffer:	Tris-HCl, pH 8.0	20 mM
	NaCl	100 mM

2.1.9 Solutions for HPLC-MS

Elutant A: Trifluoroacetic acid   0.1 %

This was diluted in double distilled H_2O and filtered.

Elutant B: Methanol   50 %

This was diluted in double distilled H_2O and filtered.

2.1.10 Solutions for NMR

Solution for the whole intermediate: D_2O 100 %

Solution for amide recognition: 10 mM Phosphate buffer pH 7

                                        D_2O 10 %

2.1.11 Media for plant growth

   Murashige and Skoog salts (Sigma)   2.2 g L^{-1}

   Phyto-agar   0.8 % (w/v)

The pH was subsequently adjusted to 5.7 with 1 M KOH.

2.1.12 Media for *Caenorhabditis elegans* growth

   NaCl   3 g

   Agar   17 g

   Peptone   2.5 g

   H_2O 975 mL

This was autoclaved, and then the following sterile solutions were added:

   1 M potassium phosphate pH 6   25 mL

1 M MgSO$_4$	1 mL
1 M CaCl$_2$	1 mL
Cholesterol (5 mg mL^{-1} in ethanol)	1 mL

2.2 Microbacterial methods

2.2.1 Sterilisation

Every solution was sterilised for 15 minutes at 121 °C and 1 bar of pressure in the autoclave, apart from those that were temperature sensitive which were filter sterilised through a 0.21 μm filter.

2.2.2 Storage of bacteria

Glycerol stocks were made for long-term storage of bacteria. Glycerol was added to an overnight bacterial culture to a final concentration of 15 % (v/v), except for the *Mycobacterium tuberculosis* stocks which were 20 % glycerol. The stocks were then stored at -80 °C.

2.2.3 Using the bacterial stocks

The bacteria were streaked directly from the frozen stock on to LB agar plates with antibiotics where appropriate. The plates were then incubated overnight at 37 °C or 28 °C depending on the strain. *M. tuberculosis* stocks were added straight into liquid media to start the culture growing.

2.2.4 Liquid cultures

Liquid cultures were inoculated with a single colony from an agar plate culture. The medium was supplemented with antibiotics where required. The cultures were rotated at 180 rpm in baffled flasks at 37 °C, 28 °C or 25 °C, depending on the culture.

2.2.5 Preparation of competent cells

Competent *E. coli* cells were prepared based on the method described by Sambrook *et al*, 1989 (Sambrook and Green 2001). To start with, 100 mL of LB medium was inoculated

with a single colony and left to grow overnight at 37 °C at 180 rpm. The starter culture was grown to an optical density of between 0.4 and 0.6 at OD_{600}. The cells were cooled on ice for 10 minutes and pelleted at a rotation of 3,000 rpm at 4°C. The pellets were then resuspended in 20 mL of ice cold 0.1M $CaCl_2$ and incubated on ice for 40 minutes. This was spun down at 3000 rpm for 10 minutes at 4 °C. The pellet was resuspended in 2 mL of ice cold 0.1 M $CaCl_2$ with 15 % of glycerol (v/v). This cell solution was pipetted into sterile tubes in 25 µL aliquots, and frozen in the -80 °C freezer.

2.2.6 Transformation of competent cells

The competent cells were defrosted on ice for 10 minutes. To this, 1 µL of plasmid DNA was added and gently flicked to mix it through. The mixture was incubated on ice for 15 minutes, and then heat shocked by incubation in a 42 °C water bath for 50 seconds before being transferred on to ice for at least 2 minutes. To these cells, 250 µL of SOC medium was added. The cells were then incubated at 37 °C for 20 minutes to 1 hour. After this the culture was spread over an LB agar plate containing the appropriate antibiotics and incubated at 37 °C overnight.

2.2.7 Production of recombinant proteins

To produce recombinant protein the *E. coli* strains BL21(DE3) pLysS or BL21(DE3) were transformed with a vector containing the gene of interest and plated on the appropriate LB agar. The recombinant strain was grown in LB or 2YT (see Section 2.1.3) at 37 °C, 28 °C or 25 °C with 180 rpm rotation. In the 37 °C cultures protein expression was induced with 0.4 mM IPTG overnight at 19 °C. The 28 °C cultures were not induced, but grown overnight at 28 °C except for the CbiG cultures which were incubated overnight at 25 °C.

The next day cells were centrifuged down at 4000 rpm for 15 minutes at 4 °C. The pellet was resuspended in 10 mL of the appropriate binding buffer (see Section 2.1.5.1) and then frozen at -20 °C until used for protein purification (see Section 2.5.1).

The precorrin-7 producing strain was overproduced in the fermenter using the medium described in Section 2.1.3, with a 50 mL starter culture. It was grown at 28 °C overnight at 200 rpm, maintained at pH 7 using 2 M NH_4OH and 2 M phosphoric acid, with 1 % (v/v) antifoam dripped in to prevent it from bubbling over.

2.2.8 Sonication of bacteria

The harvested cells were defrosted and lysed by sonication using a Sonics Vibracell Ultrasonic processor. An output wattage between 20 and 30 was applied to the cell pellet for 30 seconds followed by a 30 second rest which was repeated for a total of 6 minutes. The sonicated cells were centrifuged at 18,000 rpm for 20 minutes at 4 °C.

2.3 Molecular biology methods

2.3.1 PCR reactions

The PCR reactions were performed in an Eppendorf AG 22331 PCR machine, as set out below:

Table 2.3.1.1: PCR reaction set up

	Without DMSO	With DMSO	With DMSO
ddH$_2$O	34.5 μL	33.5 μL	32.5 μL
10 x PCR buffer (Roche, containing 18 mM MgCl$_2$)	5 μL	5 μL	5 μL
DMSO	0 μL	1 μL	2 μL
2 mM dNTPs	5 μL	5 μL	5 μL
10 μM 5' primer	2 μL	2 μL	2 μL
10 μM 3' primer	2 μL	2 μL	2 μL
DNA template	1 μL	1 μL	1 μL
Taq polymerase	0.5 μL	0.5 μL	0.5 μL

Table 2.3.1.2: PCR reaction methods

Step	Temperature	Time Roche *Taq*	Cycles	Function
1	96 °C	2 minutes	1	Initial denaturation of the chromosomal DNA
2	96 °C	30 seconds	35	Denaturation of the amplified DNA
3	55 °C	30 seconds		Annealing of the primers
4	72 °C	1 minute per 1,000 base pairs		Elongation
5	72 °C	5 minute	1	Final elongation
Stop	4 °C	Paused		

Upon the completion of the PCR the reaction was run on an agarose gel by electrophoresis.

2.3.2 Agarose gel electrophoresis

All agarose gel electrophoreses used 1 % (w/v) agarose gel concentration in TAE buffer with ethidium bromide (0.5 µg mL^{-1} final concentration). DNA loading buffer was added to the samples at 20 % of the total sample volume, and then loaded into the gel wells. The electrophoresis was run between 70-80 V for about 1 hour.

Once this was complete the gel was exposed to UV radiation (312 nm). The ethidium bromide, which absorbs this, is a fluorescent dye intercalated between the DNA base pairs. Therefore the emission at 590 nm was used to visualise the DNA, and a photograph was taken through a red filter.

2.3.3 Extraction and purification of DNA

The band of interest was cut from the agarose gel using a scalpel blade. The extraction and purification of the DNA from the gel was performed using the microcentrifuge protocol from the Qiagen QIAquick® gel extraction kit. Purification of DNA from PCR samples was done using the same kit, but using the QIAquick® PCR purification method.

2.3.4 Construction of plasmids

The vector and the insert were both digested with sticky end compatible enzymes, usually *Nde*I and *Spe*I. These fragments were incubated for 2 hours at room temperature with T4 DNA ligase (Promega) in the following manner:

Insert	5 µL
Vector	3 µL
10x Rapid ligation buffer (Promega)	1 µL
T4 DNA ligase (Promega)	1 µL (1 U per 10 µL)

2.3.5 Amplification and checking of plasmids

The ligation mixture was used to transform *E. coli* JM109 (see Section 2.2.6) which was plated and incubated overnight on LB agar plates containing the appropriate antibiotics. A single colony was picked to inoculate 5 mL of LB, which was incubated overnight at 37 °C. This starter culture was spun at 3000 rpm and the DNA was purified from the pellet using Qiagen QIAprep® miniprep kit, miniprep purification protocol using a microcentrifuge.

The isolated plasmid DNA was digested again to confirm that the insert had ligated into the vector. The optimal buffer for the digestion was chosen using the information provided by Promega or New England Biolabs. The relevant enzymes were added to the plasmid DNA and incubated at the optimum temperature for the chosen enzymes, usually 37 °C, for 2 hours. The digested plasmid was run by agarose gel electrophoresis and the results were compared to the theoretical migration and number of the digested fragments.

Table 2.3.5.1: A typical digest protocol

	Single digest	Double digest
ddH$_2$O	3.5 μL	3 μL
Restriction enzyme 1	0.5 μL	0.5 μL
Restriction enzyme 2	-	0.5 μL
10 x digest buffer	1 μL	1 μL
Plasmid DNA	5 μL	5 μL

2.3.6 Sequencing of the plasmids

Once the insert was confirmed to be present in the vector, the plasmid was sequenced (at Source Bioscience PLC) to make sure that no mutations were introduced in the gene amplification and ligation processes.

2.4 Protein purification and analysis

2.4.1 Immobilised metal ion chromatography

In all the IMAC purifications nickel was used as the immobilised ion for isolating the hexa-histidine tagged proteins from the supernatant of the sonicated bacterial pellet (see Section 2.2.8). The Ni^{2+} was loaded on to Chelating Sepharose^{TM} Fast Flow resin whereupon the supernatant of the bacterial pellet after sonication was applied to the column. Proteins with exposed hexa-histidine tag will stick to the column whilst everything else flows through. The His-tagged protein is isolated by competitive elution with increasing concentrations of imidazole. Imidazole is a histidine mimic which displaces the hexa-histidine tagged protein from the nickel when in high enough concentrations, i.e. in elution buffer (see Section 2.1.5.1). All the proteins were purified at room temperature, but CbiH was purified anaerobically (at <

2 ppm oxygen). The purification method in an empty PD-10 column is set out below in Table 2.4.1.1.

Table 2.4.1.1: IMAC purification method

Step		Volume (mL)
1	Chelating Sepharose^{TM} resin	3-5
	Distilled H_2O	50
2	0.1 M $NiSO_4$	10
	Binding buffer	50
3	Supernatant	~40
4	Binding buffer	50
5	Wash buffer I	25
6	Wash buffer II	25
7	Elution buffer	~25
8	Strip buffer	15
9	dH_2O	50

The elution buffer was applied to the column to wash off the hexa-histidine tagged protein so the flow through was collected in 2 mL fractions. In order to identify the fractions with the most concentrated protein Bradford reagent (Bio-Rad) was used, as set out below:

dH_2O	10 μL
Bradford reagent	5 μL
Protein fraction	5 μL

This gives an arbitrary measure of protein concentration as the drop turns from red to varying degrees of blue depending on protein concentration. In this way the most concentrated elutions could be identified for further use.

The strip buffer contains EDTA which chelates the nickel ions and washes them off the resin, regenerating it for the next purification. All of the buffers were allowed to flow through according to gravity and were not pushed.

2.4.2 Buffer exchange

The most concentrated elutions were combined and desalted on a PD-10 desalting column (GE healthcare) using the protocol provided, summarised in Table 2.4.2.1.

Table 2.4.2.1: Buffer exchange protocol

Step		Volume (mL)
1	Exchange buffer	25
2	Protein elution	2.5
3	Exchange buffer	3.5
4	dH$_2$O	50

The 3.5 mL of the exchange buffer in step 3 was collected as it contained the protein.

2.4.3 Protein concentration calculations

Instead of performing a full Bradford protein assay and using the standard curve produced to find the concentration, the extinction coefficients of the proteins were used. The absorbance at 280 nm and the extinction coefficient was combined in the equation below to calculate the concentration of the protein.

$$A_{280}=\varepsilon cl$$

Where A is the absorbance at 280 nm, ε is the extinction coefficient of the protein at 280 nm, c is the concentration and l is the path length. The cuvettes used were 1 cm deep, so the path length was invariably one.

The concentration could then be converted into mg mL^{-1} using the following equation:

Protein concentration (mg mL^{-1}) = Molecular mass (Da) x Concentration (M)

2.4.4 Gel filtration

An Amersham Biosciences P-920 FPLC chromatography system was used for gel filtration purification. A superdex^{TM} G200 column (GE Healthcare) was pre-equilibrated with 1.2 column volumes of buffer. Concentrated purified protein (1-2 mL) was loaded onto the column. The protein was eluted at a flow rate of 0.5 mL min^{-1} into 1 mL fractions with real-time detection of protein concentration at 280 nm to determine when the protein eluted. Fractions were further analysed by SDS-PAGE.

2.4.5 Polyacrylamide gel electrophoresis

Single concentration polyacrylamide gels were made using the protocol described in Section 2.1.7.2. , but the 4-20 % gradient gels were from Bio-Rad (USA) and Thermo Fisher

Scientific (UK). The polyacrylamide gels were run according to the protocol set out in Laemmli, 1970 . The samples, taken from every flow through of each buffer during protein purification, the supernatant, pellet and the post-sonication total, were combined with SDS sample buffer in a 1:1 ratio and boiled in a water bath for 5 minutes to denature the proteins. Depending on protein concentration, 5-20 µL of the samples were loaded into each well, and 7 µL of the protein marker was added to the first well to provide a guide to the relative molecular mass of the proteins. Electrophoresis was performed at a constant voltage of 200 V using an ATTO 1D dual mini slab gel tank, ATTO Corporation and a Bio-Rad Power Pac 300 electrophoresis power supply, for around 90 minutes. Once this had run the gel was stained with coomassie blue stain for about 1 hour before sequential washing with dH_2O. The de-stained gel was then photographed.

2.4.6 MALDI-TOF SDS-PAGE analysis

The MALDI (Matrix-assisted laser desorption/ionisation) samples were run on SDS-PAGE and fixed for one hour before being transferred into colloidal coomassie stain. The bands of interest were excised using a clean scalpel, finely chopped into small cubes and transferred into a microcentrifuge tube. These gel pieces were then subjected to alkylation and reduction, before an overnight tryptic digest at 37 °C. The peptide fragments were extracted using an ultrasonic water bath and the supernatant was collected. The samples were mixed 1:1 with a matrix solution (10 mg mL^{-1} 2,5-dihydroxybenzoic acid in 2:1 0.1 % TFA/acetonitrile) and 1 µL of each sample was pipetted on to a MTP AnchorChip 384 plate. A peptide calibration standard was used to calibrate the machine every four samples. The samples were analysed by a Bruker Ultrafile Xtreme MALDI TOF-TOF (time-of-flight) mass spectrometer. Data was collected using flexControl and sent, using BioTools to the in house MASCOT analysis suite. The peptide mass fingerprint was run against the Uniprot database. MS-MS of every sample was also performed and run against the same database.

2.4.7 Crystallisation of CobH (T85A)

The hanging drop method of vapour diffusion was implemented for the crystallisation trials (McPherson 1976). Recombinantly produced CobH (T85A) was purified on a nickel affinity column using the hexa-histidine tag. The most concentrated elutions were buffer exchanged (Section 2.4.2) and concentrated to 1 mL using Amicon® 10 kDa centrifuge protein concentrators, which gave a concentration of 57.3 mg mL^{-1}. This was purified on the

FPLC (Section 2.4.4). The FPLC elutions were diluted from 31.5 mg mL^{-1} to 15.75 mg mL^{-1} (diluted further to 7 mg L^{-1} for the optimised tray) in buffer comprised of 20 mM Tris pH 8 containing 100 mM NaCl for the apo-protein, or with 1.3 mM of allyl-HBA (dissolved in 20 mM Tris pH 8 containing 100 mM NaCl) for the holo-protein. The protein was diluted 1:1 with the well condition on a cover slip to give a total volume of 4 μL. The cover slip was inverted, placed over the well condition and sealed with vacuum grease. The 24 well plates used in this hanging drop vapour diffusion method were incubated at 20 ºC and checked regularly for crystal formation. The initial screen used was the Molecular Dimension Structure Screen-01 (Jancarik and Kim 2000; Wooh et al. 2003). Any promising conditions were screened around with slight changes in well composition (see Appendix A.6 for an example). The crystals were picked up using a loop and dipped into the appropriate cryoprotectant solutions, as described by McFerrin and Snell (McFerrin and Snell 2002). The crystal data was collected at the Diamond Light Source synchrotron (Oxfordshire) by Prof. David Brown. The structure elucidation and refinement statistics are described in Chapter 3, 3.4.3).

2.5 Biochemical methods

2.5.1 Anaerobic techniques

All anaerobic work was performed in a glove box (Belle Technologies) nitrogen environment containing less than 2 ppm oxygen. Liquids were degassed under argon prior to transferring into the glove box, and any chemicals were taken in to the glove box as powders and dissolved within, using degassed H$_2$O, providing they did not need pH adjustment. Any resins, such as PD-10 columns, were taken into the glove box at least 24 hours prior to use, and equilibrated with 50 mL of anaerobic buffer.

2.5.2 Preparation of cobalt-factor III

The vector pET-coco-2-*cobA-hemB-hemC-hemD-sirC-cbiX^s-cbiL* was provided by Dr Simon Moore. It was modified from pET-coco-ABCDC, originally constructed by Dr Evelyne Deery and subsequently added to by Dr Stefanie Frank. These proteins were overproduced in *E. coli* and purified on a nickel affinity column. After transfer into the glove box and buffer exchange into 20 mM Tris pH 8 and 100 mM NaCl, these proteins were

incubated with 4 mM ALA, 7.5 mM SAM, 750 nM NAD^+ and 1 mM cobalt at 37 °C. To prevent precipitation of the proteins the cobalt was gradually titrated into the incubation every 10-15 minutes to allow cobalt chelation to occur. Cobalt factor III was produced up to a concentration of 1 mM (5-10 mL) using this method. After completion, 1 mM DTT could be added to maintain the cobalt (II) oxidation state, and prevent oxidation to cobalt (III).

2.5.3 Preparation of cobalt-precorrin 6b

The five glycerol stocks of *Bascillus megaterium* transformed individually with P_{xylA}-CbiH^{His}, P_{xylA}-cbiF^{His}, P_{xylA}-cbiD^{His}, P_{xylA}-cbiG^{His}, and P_{xylA}-cbiJ^{His} were provided by Dr Simon Moore. All five cultures were grown in LB in the presence of tetracycline at 28°C, 150 rpm for 3-4 hours before induction with 5 % (w/v) xylose, 1 mM of ferric citrate was also added to the CbiH culture. These litres were then left overnight at 28°C except for CbiG which was grown at 25°C. All of the proteins were purified using nickel affinity chromatography, but only CbiH was purified anaerobically in the glove box. The other proteins were purified aerobically and buffer exchanged in the glove box. The proteins were combined and incubated with 1 mM cobalt-factor III, 2.8 mM SAM, 1.7 mM NADH, 56 nM Na Dithionite, 5.56 mM DTT, NaOH, and 20 mM Tris pH 8, and incubated overnight in the dark at 37 °C. The reaction was monitored using a UV-vis spectrophotometer.

2.5.4 Preparation of C5-allyl-cobalt-precorrin 7

The CbiE, T and C proteins were overexpressed in *E. coli* and purified by nickel affinity chromatography, then buffer exchanged in the glove box. The CbiE (436 μM) was incubated with allyl-SAM (400 μM), DTT (10 mM) and cobalt-precorrin 6 overnight in the dark at 37 °C. The next day the CbiC (40 μM) and CbiT (57 μM) were added to the incubation with SAM (1 mM) and again incubated overnight in the dark at 37 °C. The product was analysed using HPLC-MS (see Section 2.5.13).

2.5.5 Preparation of precorrin-7

The vector pET3a-*cobA-cobI-cob G-cob J-cob F-cob M-cob K-cob L^C-cob E*^{His} was provided by Dr Evelyne Deery. A 4 L BioSTAT® A plus fermenter (Sartorius AG) was used to grow *E. coli* transformed with this plasmid which overproduced precorrin-7 for 24 hours at 28 °C. The pH and temperature were kept constant at 7 and 28 °C respectively, using the Sartorius BioPAT® MFCS/DA software. The cells were harvested and lysed using a B series cabinet model cell disrupter (Constant systems). This was spun again at 18,000 rpm. The

supernatants were acidified with 0.3 % (v/v) TFA and heated to 70 °C for 30 minutes to remove as many proteins from the solution as possible. This was spun at 4,000 rpm until the supernatant was clear. It was then diluted 1:1 with acetonitrile and spun again at 4,000 rpm before being applied to an anion exchange column (see Section 2.5.11) followed by further purification on a C18 column (see Section 2.5.12). The product was run on the HPLLC-MS to confirm its mass (see Section 2.5.13).

2.5.6 Preparation of C5-allyl HBA

CobL and CobH were overproduced in *E. coli* and purified on a nickel affinity column using the hexa-histidine tag. The allyl-SAM was synthesised chemically as described by Wang et al. 2011, and run on a DOWEX-50W X4-200® resin to separate the *R* and the desired *S* isomers using HCl as the elutant. This was then added to the *in vitro* protein assay composed of precorrin-7 (60 µM), cobL (6 µM), cobH (30 µM) and allyl-SAM (170 µM) performed in exchange buffer. This was incubated at 28 °C, 160 rpm, overnight in the dark. To purify, the assay was heated to 65 °C, then spun at 4,000 rpm in a benchtop centrifuge, and the resulting supernatant was applied to an anion exchange column (see Section 2.5.11).

2.5.7 Preparation of C5-allyl HBAH

This assay contained C5-allylHBA (30 µM), CobB (2 µM), CobQ (5 µM), adenosine triphosphate pH 7 (8 mM), L-Glutamine (8 mM), and $MgCl_2$ (20 mM) performed in exchange buffer. This was incubated at 28 °C, with 160 rpm rotation, overnight in the dark. To purify, the assay was heated to 80 °C, then spun at 4,000 rpm, acidified with 0.3 % trifluoroacetic acid (TFA) and then spun again. The supernatant was applied to an RP18 column (see Section 2.5.12).

2.5.8 Preparation of C5-Fluorophore Cobyric acid

The C5-allyl group was modified through a thiol-ene coupling reaction with cysteamine. A solution of C5-allyl-HBAH (30 µM) was incubated at 55 °C in acetate buffer (pH 4.0, 0.25 M) containing cysteamine (100 mM) and VA-044 (20 mM) in the dark for 15 mins. The product was purified over RP18 and analysed by HPLC-MS (see Section 2.5.10).

Cobalt insertion was achieved through incubation of the desired metal-free corrinoid (20 µM) with $CoCl_2$ (10 mM) in ammonium hydroxide (0.2 M) at 80 °C under an anaerobic N_2 atmosphere. The reaction was followed UV-visible spectroscopy until deemed complete (~

60 mins) after which it was quenched with KCN and the product purified over RP18 (see Section 2.5.12).

The fluorophore succinimidyl ester (BODIPY® TR-X, Oregon green® 514, or fluorescein) (0.5 mg) was added to C5-thioamide- cyano cobyric acid (0.1 mg) and N, N-DIPEA (10 µL) in DMSO (0.5 mL). The reaction was shaken at room temperature for 2 hours in the dark. The solvent was removed *in vacuo* and the compound was purified by semi-preparative HPLC (see Section 2.5.13).

2.5.9 Preparation of ribose Fluorophore cobalamin

The ribose-5'-hydroxyl group of commercial cyano-cobalamin (Sigma) was activated with CDI and coupled with 1, 2-diaminoethane as described previously (McEwan et al. 1999). The alkyl derivative was added to either the succinimidyl ester of Oregon Green® 514 or BODIPY® TR-X with 2 equivalents of N, N-DIPEA in DMSO. The reaction was stirred for 16 hours in the dark at room temperature. The product was precipitated with acetone and collected by filtration. Fluorescent derivatives were dissolved in 50 % acetonitrile and purified by semi-preparative HPLC (see Section 2.5.13).

2.5.10 Ultraviolet-visible (UV-vis) spectrophotometry

UV-vis spectra were recorded on a Varian Cary 50 Bio UV-vis spectrophotometer over a range of 200-800 nm, usually 280-650 nm. For UV-vis data obtained in tandem with mass spectra please see Section 2.5.13.

2.5.11 Anion-exchange chromatography

An empty PD-10 column had 5 mL of diethylaminoethyl sepharose (DEAE-Sephacel®) (Sigma) poured into it, and washed with 25 mL of wash buffer I. The solution containing the intermediate was then loaded onto the column. The column was washed with buffers containing increasing concentrations of NaCl (100-600 mM) whereupon the coloured elution fractions were collected (Section 2.1.6).

2.5.12 RP-18 chromatography

LiChroprep® RP-18 (Merck) was routinely used for bench-top preparative reverse phase chromatography. The LiChroprep® was mixed with hexane before 2 mL of the slurry was applied to the PD-10 column (in the fume cupboard). The hexane was removed by

continuous washing with methanol. Once the column was in methanol, 50 mL of 0.1 % (v/v) TFA was used to wash the column. TFA was then added to the tetrapyrrole solution at 0.1 % (v/v) in order to acidify it. This was applied to the column and washed with 0.1 % (v/v) TFA. The column was then washed with increasing concentrations of methanol until the tetrapyrrole eluted (usually ~50 %). The elutions were dried down on a vacuum centrifuge and resuspended in dH$_2$O for HPLC-MS analysis.

For intermediates with six amide groups this was done at neutral pH in order to encourage the more hydrophyllic intermediates to be retained during mobile phase, but otherwise the process was the same.

2.5.13 Reversed phase HPLC-MS analysis

Every intermediate produced was run on the HPLC-MS, and the SAM-analogues and fluorophore analogues were purified by preparative HPLC prior to use. Depending on the concentration of the sample between 10-100 μL was loaded onto the Ace 5 AQ column (2.1 x 150 mm, 5 μm, 100 Å, from Advanced Chromatography Technologies). The column was connected to an Agilent 1100 series HPLC coupled to a micrOTOF-Q (Bruker) mass spectrometer, equipped with an online UV-vis diode array and fluorescence detectors. Every protocol run was at a 0.2 mL min^{-1} flow rate using 0.1 % (v/v) TFA (solvent A) and 100 % acetonitrile (solvent B), with the column kept at 25 °C. The three methods used are shown in Figure 2.5.13.1.

Figure 2.5.13.1: **The LC timetable of solvent composition.** Three LC methods were used during HPLC analysis, all using the same solvents. Solvent A is 0.1 % (v/v) TFA and solvent B is 100 % acetonitrile. All the methods were designed by Dr Andrew Lawrence.
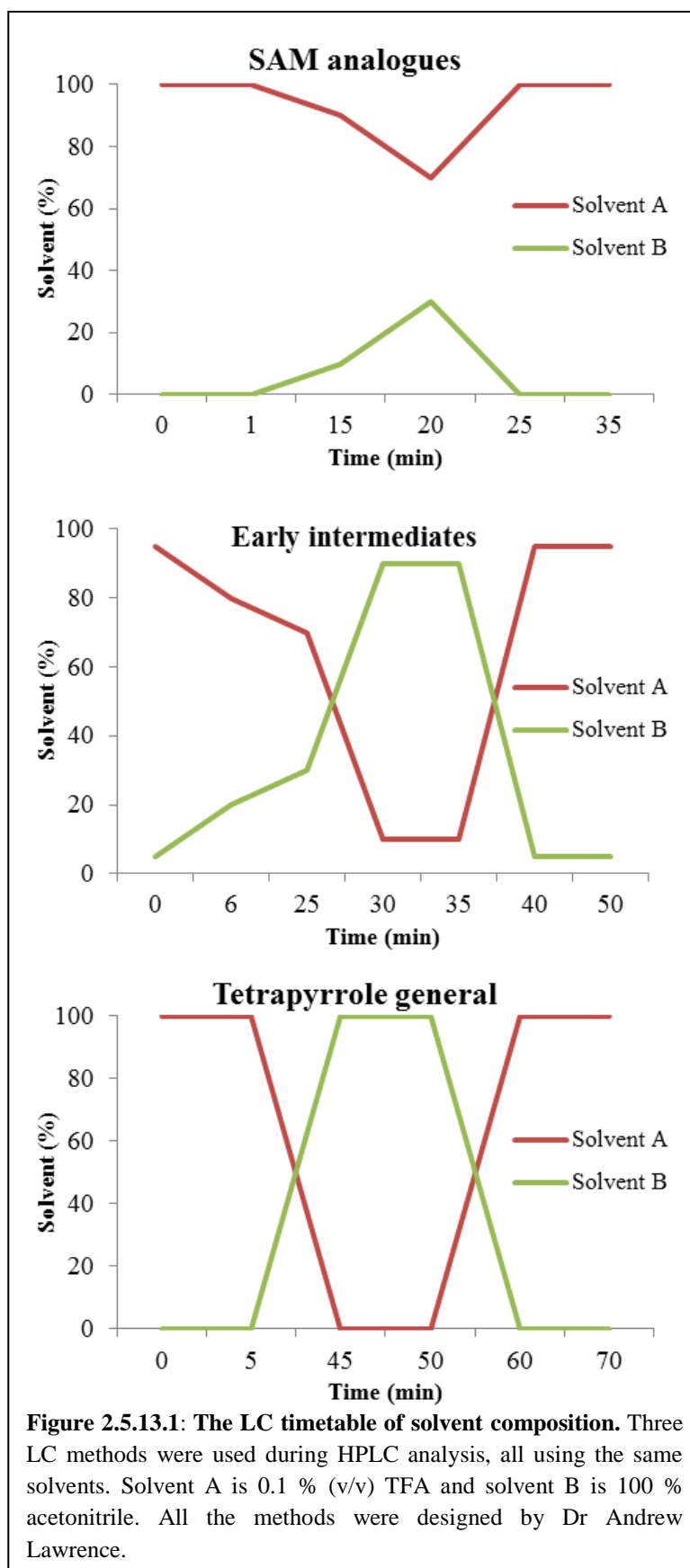
The electrospray ionisation mass spectrometry (ESI-MS) data was obtained on a Bruker micrOTOF II-MS using positive mode electrospray ionisation. The results were correlated to the UV-vis spectra acquired after the LC, using diode array and UV detection, over a 200-800 nm range.

Semi-preparative HPLC was performed on an Agilent 1000 series HPLC using a CS SIL C18 column (250 mm x 10 mm, 5 µm; Charlton Scientific) at a flow rate of 5 mL min^{-1} running the gradients described above. Elutants were identified by the UV spectra.

2.5.14 NMR analysis

NMR experiments were performed using a 600 MHz (^{1}H) Bruker Avance III spectrometer with a 5 mm QCI-F cryoprobe. ^{1}H Chemical shift referencing was based on the position of the water resonance, ^{13}C referencing used ^{1}H/^{13}C gyromagentic ratios to define indirect carrier position and all data were obtained at 25 °C (Wishart and Case 2001).

2.5.15 Allyl-SAM synthesis and DOWEX-50W X4-200 ® purification

First a 1:1 ratio of formic: acetic acid was cooled in an ice bath, 600 µL of this was used in the reaction mixture outlined below:

SAH	5 mg
Allyl Bromide	44 µL (40 equivalents)
AgClO$_4$	6 mg (2 equivalents)

This was stirred for 2.5 hours at room temperature and then the following were added:

Allyl Bromide	44 µL (40 equivalents)
AgClO$_4$	6 mg (2 equivalents)

This was stirred for a further 2.5 hours at room temperature. The reaction mixture was then centrifuged to remove the AgClO$_4$ and other precipitants. The supernatant was diluted with 3 mL of ddH$_2$O. To extract the residual allyl bromide 5 mL of diethyl ether was added to the mixture and the top layer was removed. This was repeated 3 times. The remaining solution was dried in a vacuum centrifuge and resuspended in 500 µL of 1:1 formic: acetic acid.

The DOWEX-50W X4-200® column was prepared with ddH$_2$O and then the allyl-SAM mixture was applied. This was washed with two column volumes of increasing concentrations of HCl (0.1 M, 1 M, 2.5 M, and 5 M). The *R* isomer washed off the column at 2.5 M HCl whilst the desired *S* isomer purified at 5 M HCl. The aliquots were dried on the vacuum centrifuge and resuspended in ddH$_2$O before running on the HPLC-MS (see Section 2.5.13).

2.6 Methods of experiments conducted *in vivo*

2.6.1 Bioassay plates

Quantitative *S. enterica* AR3612 bioassay plates were performed as explained in Chapter 3 (pg96) and as described previously (Raux et al. 1996).

2.6.2 Imaging in *Escherichia coli*

BL21 (DE3) *E. coli* was transformed with pLysS-*btuB* whilst OP50 *E. coli* was transformed with pET-BAD-*btuBF*. Both were grown in a 4 mL LB culture after inoculation with 16 µL of starter culture. This was grown with 1 µM of C5-fluorophore-cobyric acid overnight with induction. One mL of culture was spun down and washed three times with fresh LB to remove external fluorophore. Next, 5 µL of this culture was dried on to a 1 % (w/v) LB-agarose pad and imaged on an Olympus IX81 widefield microscope with PlanApo 150 x OTIRFM-SP 1.49 numerical aperture lens mounted on ASI stage (Applied Scientific), and illuminated using LED light sources (Cairn Research Ltd) with appropriate filters (YFP and mCherry, Chroma). The samples were visualised using a Princeton ProEM 1024 back-thinned EMCCD camera (Princeton Instruments) on Metamorph software (Molecular Devices). Each 3D-maximum projection of volume data was calculated from 13 *z*-plane images, each 0.2 µm apart.

2.6.3 Extraction of cobalamin from *Lepidium sativum* using P-Per

Lepidium sativum seeds were sterilised by washing with 70 % ethanol three times then rinsed five times with sterile water. The seeds were placed on the agar (see Section 2.1.11) containing different concentrations of cobalamin, and grown for one week in sunlight at room temperature. The cotyledons of the cress were collected and washed 5 times with water. The residual water was removed by pipette, after centrifugation. The P-Per was prepared as instructed in the enclosed manual (Thermo Fisher): 283 µL reagent A, 2.9 µL of reagent B,

and 214 µL of reagent C. A little sand was added to the cotyledons along with 100 µL of the P-Per. The cotyledons were ground using a hand held pellet pestle (Sigma Aldrich) for 2 minutes. The suspension was vortexed and then centrifuged for 3 minutes at 15,000 rpm rotation. The lower aqueous phase was collected and applied to the bioassay plate (see Section 2.6.1).

2.6.4 Imaging in Plants

Both *A. thaliana* and *L. sativum* seeds were sterilised by washing with 70 % ethanol three times then rinsed five times with sterile water. The seeds were then placed on the agar (see Section 2.1.11) with 0.5 µM Oregon green linked to either the ribose of cobalamin or to the C5 of cobyric acid, and grown at room temperature in the dark for five days. Whole cotyledons or sectioned cress were placed directly on the glass slide with water. The samples were imaged with a Leica TCS SP2 laser scanning confocal microscope (Leica Microsystems, Germany) with AOBS (Acoust-Optical Beam Splitter) detected using PMTs (photomultiplier tubes). Both the 40 x and 63 x HCX PL APO oil lenses, numerical aperture 1.25 and 1.4 respectively, were used. Samples were excited at 514 nm from an Argon–Krypton-mixed gas laser and images were acquired in the green/yellow region of the light spectrum (525–590 nm). The software used to image was LCS (Leica Confocal Software) and the images were processed using Leica Lite confocal software and FIJI (Schindelin et al. 2012).

2.6.5 Imaging in *Caenorhabditis elegans*

The OP50 pET-BAD-*btuBF* were grown in the same way as Section 2.6.2 with BODIPY® TR-X as the fluorophore in the analogues. The following day the cells were pelleted and resuspended in 1 mL of fresh LB, then pelleted again and so on until the cells had been resuspended three times in fresh LB. The final pellet was resuspended in 1 mL of fresh LB and 200 µL of the OP50 culture was pipetted on to the centre of NGM agar plates. These plates were left to dry in a sterile culture hood for 4 hours and then stored at 4 °C until used (see Section 2.1.12). Three forth larval stage (L4) *C. elegans* were transferred to the seeded plates and left to grow at 20 °C for four days before imaging. Almost all of the nematodes reported in these investigations are at L4, the stage just before adult. This is because adults are hard to image as the middle of the worm is obstructed by eggs. In contrast, the somatic structures of the nematode have formed by mid-L4 stage but eggs within the uterus have not formed, which make this the perfect stage for imaging.

For each slide, 3-5 L4 stage *C. elegans* were mounted in M9 + 0.2 % levamisole on a 2 % agarose pad and imaged within 30 min at the University of Bristol on a Leica SP8X AOBS confocal laser scanning microscope attached to a Leica DMi8 inverted epifluorescence microscope with 'Adaptive Focus Control'. The sample was excited with a white light laser at 594 nm and detected between 599-712 nm using a hybrid gated detector for the fluorophore, and excited with a 405 nm diode, with images acquired between 410-505 nm for the gut granule autofluorescence. All images were taken using a 20 x numerical aperture 0.75 dry lens. Again the images were processed using Leica Lite confocal software and FIJI .

The persistence of the fluorescence in the *C. elegans* was investigated by transferring L3 nematodes, previously grown on BODIPY® TR-X analogue containing OP50, on to plates with unfortified OP50 and imaging after 22 hours. The imaging was done on a Leica DM R fluorescence microscope using 515-560 nm (N2-1) excitation and 590 nm emission filters for the BODIPY® TR-X fluorescence and 456-490 nm (I3) excitation and 515 nm emission filters for autofluorescence. The images were obtained using a Photometrics Cool Snap HQ camera with a 20 x HC PL Fluotar lens, numerical aperture 0.5. The software used was Micro-manager (Edelstein et al. 2014).

2.6.6 Imaging in *Mycobacterium tuberculosis*

The *ΔmetE* strain of *M. tuberculosis* was grown in 150 μL in a 96 well plate with 1 μM of the BODIPY® TR-X fluorophore analogues added to the media. Two samples of each analogue were grown, one for 24 hours and the other for 48, until the cells were at a predicted optical density of 0.6 at 650 nm. The conditions were repeated for both 7H9 and Sauton's media. To image, each sample was spun down (15,000 rpm rotation) and washed three times in Sauton's medium. Drops of 50 μL were pipetted directly onto the slide and imaged live on a Leica SP5 using 594 nm excitation and imaging between 600-700 nm, in a CAT-III (BSL-3) laboratory at the NIH (Bethesda, MD, USA). The images were processed using Leica Lite confocal software and FIJI (Schindelin et al. 2012).

2.7 Other *Mycobacterium tuberculosis* methods

2.7.1 Cobalamin-agarose column

Every strain used (H37Rv (WT), H37RvJO (H37Rv), and *ΔbacA*H37RvJO (*ΔbacA*)) was grown in 3 x 450 mL for five days in 7H9 medium until an optical density of 0.6. The

cultures were then spun down at 3500 rpm for 10 minutes and the pellets were resuspended in 7H9 before being transferred into microcentrifuge tubes containing small beads. These were then cooled on dry ice and bead beaten at 7000 for 45 seconds. This was repeated 5 times with cooling between each beating. The samples were centrifuged for a minute at 15,000 rpm and the supernatant was filtered through 0.22 μm filters into clean microcentrifuge tubes.

The supernatant was removed from CAT-III (BSL-3) and applied to the ~2 mL cobalamin-agarose column, synthesised as described in Bioconjugate technologies (Hermanson 2013), which had been pre-equilibrated with 20 mM Tris pH 8. This was washed with buffers containing increasing concentrations of cobalamin (10 nM to 1 μM) and then washed with 8 M urea to strip any remaining proteins off the column, before washing again with 50 mL 20 mM Tris pH 8. Every flow through was collected and, after buffer exchanging the urea elution into 20 mM Tris pH 8, all the samples were concentrated using 10 kDa Pierce™ Protein Concentrator centrifuge concentrator (Thermo Fisher Scientific) to 500 μL from 50 mL. The samples were then run on 4-20 % SDS-PAGE (see Section 2.4.5) and analysed by MALDI-TOF (see Section 2.4.6).

2.7.2 Corrinoid supplemented growth

Three strains were used: H37Rv (WT), H37RvJO (H37Rv), and *ΔbacA*H37RvJO (*ΔbacA*). Each strain was grown with five different concentrations of corrinoid: 0 μM, 100 nM, 500 nM, 1 μM, and 10 μM. The cultures were grown for five days in 7H9 medium until an optical density of 0.6 at 650 nm. The cultures were then spun down at 3,500 rpm for 10 minutes and the pellets were resuspended in 1 mL of 7H9 before being transferred into a microcentrifuge tube. These were spun for 30 seconds at 15000 rpm and washed three times with 7H9. The pellet was resuspended in 1 mL of H$_2$O and transferred into a microcentrifuge containing small beads. These were then cooled on dry ice and bead beaten at 7,000 for 45 seconds, repeated 5 times. The samples were centrifuged for a minute at 15,000 rpm and the supernatant was filtered through 0.22 μm filters into clean microcentrifuge tubes. These samples were then plated on AR3612 *S. enterica* bioassay plates and subsequently analysed.

2.7.3 RNA sequencing

Five strains were used: H37Rv (WT), H37RvJO (H37Rv), *ΔbacA*H37RvJO (*ΔbacA*), *ΔmetE*H37RvJO (*ΔmetE*), and *ΔmetH*H37RvJO (*ΔmetH*). All RNA sequencing cultures were set up with and without 1 μM of corrinoid added to the 7H9 media except for *ΔmetE* which

was set up in three cultures: + corrinoid + 2 mM L-methionine, + corrinoid - 2 mM L-methionine, and - corrinoid + 2 mM L-methionine. The samples were subcultured five times, maintaining an optical density under 0.2 at 650 nm, and grown at 37 °C at 200 rpm. After 5 subcultures the cells were harvested and TRIzol® (Thermo Fisher Scientific) was rapidly added to the pellet. Once resuspended, all the samples were transferred to 0.2 μm bead containing microcentrifuge tubes (about one fifth of a 1 mL tube). The samples were cooled on dry ice and then bead beaten at 7,000 for 45 seconds at a time for at least three minutes. They were then spun down at 15,000 rpm for 40 seconds and the supernatant was transferred to a clean microcentrifuge tube. The RNA was then extracted using Quick-RNA™ MiniPrep Plus (Zymo research) RNA purification protocol. The purified RNA samples were sent to the microarray services at the NIH (Bethesda, MD, USA) to be analysed. Samples containing the corrinoid were compared to those without, and with *ΔmetE* the sample with corrinoid was compared to the sample with corrinoid and methionine, as well as the sample with just methionine.

2.7.4 *ΔmetE* rescue

In a 24 well plate, six different cobalamin intermediates were added to a 1 μM concentration in 500 μL of 7H9 media. Three strains were tested: H37RvJO (H37Rv), *ΔmetE* H37RvJO (*ΔmetE*), *ΔmetH* H37RvJO (*ΔmetH*). These were diluted in the well so a predicted OD_{650} of 0.6 would be reached on the fifth day. Each well was mixed every day by pipetting, and a photograph was taken. The plate, set up as shown in Table 2.7.4.1, was grown at 37 °C with 100 rpm rotation for one week.

Table 2.7.4.1: Set up of the *ΔmetE* rescue experiment

Strain	PC7	HBAD	HBAH	Cobyric Acid	Blank	Cobalamin
metE::Hyg 25 μgmL^{-1}[Final]	6 μL Culture 2.8 μL PC7 491.2 μL 7H9	6 μL Culture 3.2 μL HBAD 490.8 μL 7H9	6 μL Culture 3.2 μL HBAH 490.8 μL 7H9	6 μL Culture 11.6 μL CyA 482.8 μL 7H9	6 μL Culture 494 μL 7H9	6 μL Culture 0.5 μL B$_{12}$ 493.5 μL 7H9
H37Rv	1.3 μL Culture 2.8 μL PC7 495.9 μL 7H9	1.3 μL Culture 3.2 μL HBAD 495.5 μL 7H9	1.3 μL Culture 3.2 μL HBAH 495.5 μL 7H9	1.3 μL Culture 11.6 μL CyA 487.5 μL 7H9	1.3 μL Culture 498.7 μL 7H9	1.3 μL Culture 0.5 μL B$_{12}$ 498.2 μL 7H9
metH	1.2 μL Culture 2.8 μL PC7 496 μL 7H9	1.2 μL Culture 3.2 μL HBAD 495.6 μL 7H9	1.2 μL Culture 3.2 μL HBAH 495.6 μL 7H9	1.2 μL Culture 11.6 μL CyA 487.6 μL 7H9	1.2 μL Culture 498.8 μL 7H9	1.2 μL Culture 0.5 μL B$_{12}$ 498.3 μL 7H9
✘	✘	✘	✘	✘	✘	✘

The blank denotes that no corrinoid was added to the media and wells containing X were filled with 1 mL of ddH$_2$O to prevent the plate from drying out.

Chapter 3

Synthesising C5-cobalamin analogues using a combination of chemical and native biological methods

3.0 Introduction

The aim of the research discussed in this chapter was to determine whether it is possible to attach a reporter group directly onto the corrin macrocycle of cobalamin, and to assess whether such an analogue would remain biologically active. There are three main ways in which to approach such a construction: chemically, biologically (enzymatically), and a combination of the two. The total chemical synthesis of cobalamin is difficult as it involves around 70 steps with very low yield (Eschenmoser 1988). The complete *de novo* biological synthesis of cobalamin involves about 30 enzymatic steps, many different cofactors and roughly 25 molecules of ATP (Deery et al. 2012; Galperin and Grishin 2000; Heldt et al. 2005). The hybrid method of synthesis is relatively new. It uses recombinantly expressed cobalamin biosynthesis enzymes to produce an intermediate, which can then be chemically converted into cobalamin. The primary advantage of the hybrid technique in analogue synthesis is that the intermediate can be chemically modified and then fed back into the biological synthesis system by incubating *in vitro* with subsequent native pathway enzymes. If these analogues are recognised by the native enzymes it gives a good indication that they will also be accepted by the cobalamin uptake system *in vivo*.

The analogues synthesis detailed in this chapter utilises the hybrid synthesis route. This can be broken down into three sections: overproduction and extraction of the required intermediate from cell culture, modification of the intermediate and progression of biosynthesis by native biosynthesis enzymes *in vitro*, and chemical extension and further modification of the introduced appendage (Figure 3.0.1). It is ideal for the position of modification to be on the corrin ring because it is less likely to be cleaved *in vivo* and specific modification is possible. Previous modification to cobalamins have involved alkylation of the central cobalt ion, esterification of the ribose on the lower nucleotide loop, or modification of one of the amidated side chains (Clardy et al. 2011). Any cobalt coordinated β ligand is at risk of being exchanged to a methyl or adenosyl group *in vivo*. It is also undesirable to modify the side chains off the macrocycle as they are implicated in uptake and transport (Clardy et al. 2011). Additionally, the side chains are very similar and therefore, hard to specifically modify, often resulting in preparative HPLC purification to attain the desired product, which is very time consuming. The nucleotide loop itself can be modified, but this could affect base on/ off configuration which is important in some binding conformations, and some organisms exchange the lower loop altogether (Fedosov et al. 2007; Helliwell et al. 2016; Yi et al. 2012). With all this considered, a direct modification of the corrin ring is preferable. The chosen site

of modification is the C5 position in between rings A and B, the recipient of the final SAM mediated methylation. Therefore, the alternative group on the C5 will not interfere with any modifications on the corrin ring as they have already taken place. As the intermediates to be used for C5 modification have completed corrin macrocycles this tends to make them more stable. Additionally, the C5 position methyl group is in the plane of the ring unlike the *b* and *c* sidechains which flank it on either side, therefore a small new chemical group will also be in the plane of the ring which will decrease the likelihood of it hindering sidechain recognition.

Figure 3.0.1: The abridged late insertion pathway from precorrin-7 until the BODIPY® TR-X C5 analogue. Uroporphorinogen (Uro'gen) III is the modified tetrapyrrole precursor, *Av* denotes that CobQ is from *Allochromatium vinosum* unlike all the other proteins which are from *R. capsulatus*. The enzymes and chemicals are in blue whilst the cofactors are in black. The modifications which have taken place during the reaction are highlighted in red.

There are two pre-C5 methylation intermediates which can be used because there are broadly two different pathways for cobalamin synthesis: The Early Insertion (or anaerobic) pathway; and The Late Insertion (or aerobic) pathway, so called in reference to the timing of cobalt insertion (see Figure 3.0.2). Both pathways were investigated for their potential in constructing cobalamin analogues. The main advantage of the Early insertion pathway is that the intermediates already contain cobalt, whereas the Late insertion pathway intermediates do not, and are therefore, stable in aerobic environments. The required intermediate from the Early insertion pathway is cobalt-precorrin 7, whilst the Late insertion pathway equivalent is precorrin-7. These are the intermediates which were incubated with their respective C5 methyltransferases and a SAM analogue, allyl-SAM, in an attempt to modify the C5 position.

An allyl group was chosen as the transfer group because it is small and yet easily modifiable at a later stage. This increases the chances of the methyltransferase enzymes attaching it to the C5 position whilst limiting the detrimental effects the presence of this group may have on binding and recognition.

Precorrin-7 is not a natural intermediate of the Late insertion pathway. Normally, precorrin-6 is methylated at C5, C15 and decarboxylated at C12 by CobL, to form precorrin-8. However, a recent paper discovered that CobL is a fused protein, and that the C-terminus of CobL (CobL[C]) can function independently to methylate C15 and decarboxylate C12 (Deery et al. 2013). CobL[C] can therefore, be incubated with precorrin-6 to produce a new intermediate, precorrin-7, which is unmethylated at C5. Hence, precorrin-7 can be used to synthesise the C5 analogues.

The lack of the cobalt ion in the Late insertion pathway intermediates means that it will have to be inserted into the corrin macrocycle after the C5 modification. Previous unpublished data from the Warren group showed that CobNST, the cobalt chelation complex, is inefficient in *in vitro* assays, so chemical insertion of cobalt is preferable, even though much of the substrate is lost in the reaction due to the harsh conditions. In the Late insertion pathway *in vivo* cobalt chelation occurs after the first amidase reaction. However, as the cobalt will be inserted chemically there is no reason not to do this later, after the final amidations. This may serve to stabilise the corrin macrocycle and increase the yield of the chemical cobalt insertion reaction. CobQ, the final amidation enzyme, usually requires the presence of cobalt and the adenosylation of the upper ligand. However, *Allochromatium vinosum*, a halotolerant purple sulphur reducing bacteria, has been proposed to make metal free cobalamin, so called hydrogenobalamin (Toohey 1965). It follows that the CobQ from *A. vinosum* should not require a cobaltous intermediate, so this was used instead of the *R. capsulatus* homologue.

Once the sidechains have been modified and the cobalt inserted, the allyl group can be extended to a thioamine. This will leave a terminal primary amine which will react with activated N-hydroxysuccinimidyl ester groups commonly found on fluorophores, to produce a fluorescent C5-cobyric acid analogue.

The aims set out for this chapter were:

1. To investigate the Early and Late insertion cobalamin biosynthesis pathways to determine which was better suited for analogue synthesis.

2. To synthesise a C5 modified analogue.

3. To attach a fluorophore to the C5 position.

4. To test the recognition of the C5 fluorophore analogue in *S. enterica*.
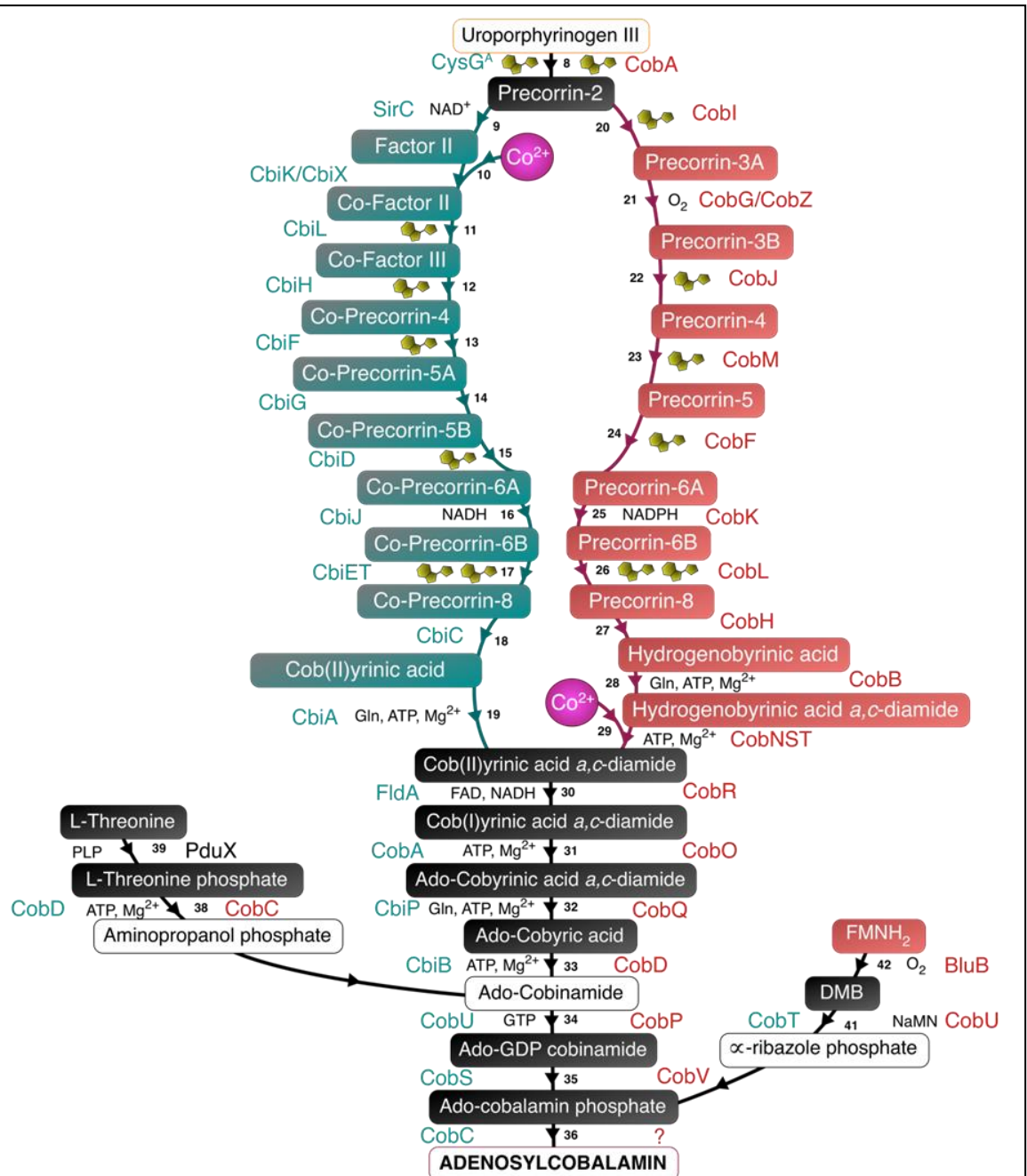
Figure 3.0.2: **The two cobalamin biosynthesis pathways.** The late insertion is shown on the right in red and early on the left in green. The intermediates of both pathways where they converge are shown in black. This Figure is taken from Dr Deery's research poster.

3.1 Results

3.2 Synthesis of allyl-SAM

An allyl group was chosen as the transfer group because it is small enough to limit any interference with the native enzymes, but easily extendable with a thioamine, to produce a primary amine which can subsequently react with many compounds, like fluorophores. The synthesis of allyl-SAM is shown in Figure 3.2.1, A (Chapter 2.5.15). The product is an enantiomeric mix of *R* and *S* isomers which were separated during purification on DOWEX-50W X4-200® cation exchange column. This separation is possibly due to shielding of cations in the *R*-isomer which prevent it from binding as tightly as the *S*-isomer. Only the *S*-isomer is biologically active. The mass of the allyl-SAM was confirmed by HPLC MS (see Figure 3.2.1, B and C).

Figure 3.2.1: The synthesis of allyl-SAM and HPLC-MS confirmation. A: The analogue was sythesised using allyl bromide as the allyl donor and silver perchlorate as the catalyst. The reaction was conducted in an acetic acid/ formic acid mixture at room temperature (see Chapter 2, 2.5.15) and then purified on a DOWEX-50W X4-200® column followed by confirmation on the HPLC-MS. B: The HPLC trace shows two peaks corresponding to the two entioisomers, R and S respectively, prior to DOWEX-50W X4-200® purification. C: The later HPLC peak at 6.3 minutes is the S isomer which shows an absorbance at 257 nm with the expected 425.1639 m/z.

3.3 Early insertion (anaerobic) pathway

The Early insertion or anaerobic pathway, as the names suggests, inserts cobalt into the macrocycle at a very early stage in the synthesis of cobalamin. It is sometimes referred to as the oxygen independent pathway, as the ring contraction mechanism does not require molecular oxygen unlike the late insertion pathway (Moore, Biedendieck, et al. 2013;

Schroeder et al. 2009). Partly as a result of early cobalt insertion, intermediates isolated from this pathway are very oxygen sensitive. Consequently work on the early insertion pathway has to be conducted in anaerobic glove boxes at <2 ppm oxygen (Moore, Biedendieck, et al. 2013).

3.3.1 Producing the substrate for modification: Cobalt precorrin-6b

The C5 position was chosen as the point of modification on the cobalamin macrocycle. The methylation of the C5 position is one of the two last methylation reactions on the macrocycle and, in the early insertion pathway. Cobalt precorrin-7 is the intermediate used as the substrate in the C5 methylation reaction. However, cobalt precorrin-7 is very unstable, therefore, the more stable intermediate one step before cobalt precorrin-7, cobalt precorrin-6b, was isolated. This was produced in two stages: synthesis of cobalt factor-III, and conversion of this into cobalt precorrin-6b. Cobalt precorrin-6b can subsequently be converted into cobalt precorrin-7 *in vitro* (Figure 3.3.1.1).

Figure 3.3.1.1: The strategy for synthesising C5-allyl analogues using Early insertion pathway intermediates. Cobalt-Factor-III was isolated from cultures recombinantly exressing the first seven enzymes. it was subsequnetly incubated *in vitro* with CbiH, F, G, D, and J to produce Cobalt-precorrin-6B. CbiE allylates the C5 position using the synthesised allyl-SAM as the donor. CbiT methylates the C15 position using SAM as the methyl donor and decarboxylates C12, after which CbiC migrates the methyl group from C11 to C12.

The plasmid pETcoco-2-*cobA, hemB, hemC, hemD, sirC, cbiX[S], cbiL* (see Chapter 2, 2.1.2.2) was over expressed in *E. coli* to produce the enzymes for the conversion of aminolevulinic acid (ALA) to cobalt factor-III. The enzymes were purified aerobically on the bench using a nickel affinity column and then buffer exchanged in the glove box into 20 mM Tris pH 8, 100 mM NaCl. The proteins were incubated with 4 mM ALA, 7.5 mM SAM, 750 nM NAD$^+$. After 30 minutes at 37 °C, 1 mL of 10 mM CoCl$_2 \cdot 6H_2O$ was added dropwise, with constant stirring. The reaction mixture was left overnight at 37 °C. During this time the three Hem proteins convert ALA into uroporphorinogen-III (uro'gen III), the modified tetrapyrrole progenitor (Lobo et al. 2009). CobA subsequently methylates uro'gen III at positions C2 and C7 to make precorrin-2. SirC oxidises the corrin ring using NAD$^+$ as a cofactor, to generate sirohydrochlorin (factor-II), the substrate for the cobalt chelatase CbiX.

CbiL methylates the C20 position to produce the last uncontracted ring intermediate, cobalt factor-III (Kadish et al. 2003). The cobalt factor-III was purified on an anion exchange column and eluted in 20 mM Tris pH 8 containing 600 mM NaCl. Cobalt factor-III is a green pigment and these distinctive, coloured fractions were collected.

The cobalt factor-III was converted into cobalt precorrin-6b using the enzymes of the anaerobic pathway from *Bacillus megaterium*, as the pathway had recently been elucidated using this suite of enzymes (Moore and Warren 2012). From cobalt factor-III there are 5 enzyme reactions to cobalt precorrin-6b. These reactions were performed in one *in vitro* incubation with CbiH, F, G, D, and J, together with 2.8 mM SAM, 1.7 mM NADH. CbiH was purified in the glove box, but the other enzymes were purified aerobically on the bench using a nickel affinity column. All of the enzymes were buffer exchanged in the glove box into 20 mM Tris pH 8 containing 400 mM NaCl. The incubation was set up at 37 °C overnight in the presence of the reducing agents sodium dithionite and DTT. The colour change resulting from the extrusion of the C20 carbon catalysed by CbiH can be followed by UV-vis spectrophotometry (Appendix A.3).

There were a couple of issues with this assay: The 4Fe-4S centre that is found in CbiH makes it difficult to overproduce and purify in great quantities as it is unstable, and the protein is prone to proteolytic cleavage (see Appendix A.2). This meant that it had to be purified just before it was added to the assay. The bigger issue was CbiF, which gave a low yield when purified from frozen pellets. However, freshly grown CbiF purified in good yields (see Appendix A.3).

3.3.2 Specifically modifying the C5 position

The native enzyme that modifies the C5 position in *B. megatarium* is CbiET, a fusion of CbiE and CbiT. CbiET is responsible for methylation of C5, C15 and decarboxylation of C12. However, this is not useful for specific modification of C5 only. Therefore, the unfused CbiE and CbiT from *Methanothermobacter thermautotrophicus* were used. CbiE is the protein responsible for the methylation of C5 and CbiT for the methylation of C15 and decarboxylation of C12 (Moore, Lawrence, et al. 2013). The proposed mechanism of C5 modification involves the allyl-SAM analogue together with the CbiE methyltransferase to allow the transfer of an allyl group instead of a methyl one (Figure 3.3.1.1). Methyltransferases have previously been shown to work with SAM analogues (Wang et al. 2011).

CbiE, T, and C were all overproduced in *E. coli* and purified by nickel affinity column in the glovebox. Cobalt precorrin-6b was incubated with 400 µM allyl-SAM and CbiE overnight. The following morning 1 mM SAM was added, to flood the reaction with SAM, along with CbiT and CbiC to try to produce C5-allyl cobyrinic acid.

The incubation of allyl-SAM and CbiE was attempted many times, potentially producing C5-allyl cobalt precorrin-7. However, due to the instability of C5-allyl cobalt precorrin-7 the production of intermediate was not confirmed by HPLC-MS. Subsequent incubations with CbiT and CbiC resulted in the production of extremely low yields of a different corrinoid (Figures 3.3.2.1). The compound detected is a corrinoid as it expresses a corrin-like UV absorbance (Figure 3.3.2.1) with a characteristic peak around 350 nm and two around 490/ 510 nm. However, it is not C5-allyl cobyrinic acid, which has a mass of 963.3 m/z, as it has a mass of 923.3 m/z. This corresponds to the mass of cobalt precorrin-7 where the C12 position is decarboxylated and the C15 position is methylated.

The failure to produce C5-allyl cobyrinic acid, combined with the temperamental nature of all the assays, led to the conclusion that the Early insertion pathway, despite the benefits of cobalt containing intermediates, is not the best pathway for analogue synthesis. Therefore, the late insertion pathway was investigated for synthesis potential.

Figure 3.3.2.1: HPLC-MS data for the synthesis of C5-Allyl-Cobyric acid using the Early insertion pathway. A: The HPLC trace for the C5-allyl cobyrinic acid reaction at 450 nm. There are many peaks and the sample is not clean so an extracted ion chromatogram (EIC) was taken, shown below, using the mass of the predicted C5-allyl cobyrinic acid. However, there are no mass peaks corresponding to the theoretical mass of the C5-allyl cobyrinic acid. B shows the UV spectrum at 21.2 minutes and the corresponding mass spectra. Although the mass of C5-allyl cobyrinic acid, 963.3 m/z, is not clear in the mass spectra, the characteristic cobalamin UV peaks at 352 nm is present and the 494 nm peak is close to the 510/530 nm peaks. The shift to 488 nm may be caused by   DTT so this compound is likely to be a corrinoid. However, the m/z of 923.3 corresponds to cobalt precorrin-7 where C5 is unmodified (inset).

3.4 Late insertion (aerobic) pathway

The late insertion pathway was elucidated in the early 1990s in *Pseudomonas denitrificans* (Blanche et al. 1995) and the naming of the enzymes is therefore specific for this organism. This pathway was characterised before the Early insertion pathway was elucidated. The Late insertion pathway incorporates the cobalt ion after the first amidase reaction, into HBAD. Therefore, the cobalt has to be inserted after the C5 modification has taken place. Two significant benefits of the late insertion pathway are that the proteins are a lot easier to work with recombinantly in *E. coli*, and the later intermediates are oxygen stable, so they can be manipulated on the bench instead of in an anaerobic chamber.

3.4.1 Making the substrate: Precorrin-7

The key intermediate for modification at the C5 position in the late insertion pathway is precorrin-7. This intermediate can be synthesised recombinantly *in vivo* in *E. coli* through the expression of CobA, I, G, J, M, F, K, L[C], (Figure 3.4.1.1, A, inset) and harvested from the bacterial pellet. The proteins were coproduced with CobE which binds precorrin-7 in the *E. coil*. After growth, the precorrin-7 was purified from the lysed cells by binding it to a diethylaminoethyl (DEAE-Sephacel®) column followed by a C18 column, as detailed in Chapter 2, 2.5.5 (Figure 3.4.1.1). The plasmid containing all of the necessary genes was provided by Dr Evelyne Deery.

Figure 3.4.1.1: The HPLC and MS data of purified precorrin-7 and the plasmid used to make it. A: The UV trace of precorrin-7 at 400 nm. Both peaks represent precorrin-7, but only the first (13.2 minutes) mass spectra data is shown (B). The mass, 867.4134 m/z is exactly what was expected. The large plasmid containing all the genes to recombinantly produce precorrin-7 is shown inset in A. CobL[c] indicates it is just the C-terminus that is present, whilst CobE[His] indicates that the protein has a hexahistidine tag, which is used to purify the precorrin-7 bound to CobE on a nickel affinity column. AmpR encodes ampicillin resistance and ori shows the origin of replication for this plasmid.

3.4.2 The synthesis of C5-allyl-HBA

The strategy for ananlogue synthesis using the Late insertion pathway intermediates and enzymes is shown in Figure 3.4.2.1. The synthesis of C5-allyl-HBA firstly requires the transformation of precorrin-7 into C5-allyl-precorrin-8 by incubating precorrin-7 with allyl-SAM and CobL. C5-allyl-precorrin-8 is then incubated with CobH to generate C5-allyl-HBA. C5-allyl-precorrin-8 less is stable than C5-allyl-HBA, so CobL and H were incubated together. This reaction has a clear colour change (Figure 3.4.2.2) caused by the shift in the macrocycle double bond conjugation instigated by CobH. The T85A mutant of CobH was found to give higher yields of C5-allyl-HBA than the wild type enzyme. This increase in activity likely reflects an improved ability of the substrate to be accommodated within the active site. To demonstrate this clearer, the T85A mutant was crystallised in the presence and absence of its product.

Figure 3.4.2.1: The strategy for synthesising C5-allyl analogues using Late insertion pathway intermediates. The purified precorrin-7 can be incubated with CobL and allyl-SAM to specifically modify the C5 position, alongside incubation with CobH to migrate the C11 methyl group to C12. The *a*, *b*, *c*, *d*, *e*, and *g* sidechains are amidated by CobB and CobQ (from *A. vinosum*) in another *in vitro* incubation to produce C5-allyl-HBA- *a*, *b*, *c*, *d*, *e*, *g*-hexamide (C5-allyl-HBAH).

Figure 3.4.2.2: C5-allyl- HBA HPLC-MS data and the colour change from precorrin-7. A: The colour change which occurs during the incubation of CobL and CobH from precorrin-7 (left) and C5-allyl- HBA (right). B: The UV trave aat 328 nm showing two peaks for C5-allyl-HBA. Both peaks have the expected mass of C5-allyl-HBA, 907.4 m/z. The UV spectra have changed from 391 nm to have maxima at 328 nm and either side of 500 nm, characteristic of cobalt-less corrins.

3.4.3 The crystal structure of CobH (T85A) with C5-allyl-HBA

A crystal structure of the T85A CobH mutant will not only give insight into the reasons this mutation is more efficient for work with analogues, but the co-crystal will confirm the structure of the C5-allyl-HBA analogue too. In order to form crystals of the *R. capsulatus* CobH (T85A) mutant it was overproduced in *E. coli* BL21 (DE3) with a hexa-histidine tag and purified on a nickel affinity column as described in Chapter 2, 2.4.1. The most concentrated elutions were pooled and concentrated further using a 15 mL protein concentrator (Millipore) to 1 mL of 57.3 mg mL^{-1}. This millilitre was further purified on the FPLC (fast protein liquid chromatography) machine using a size exclusion column to remove any aggregates. The most concentrated elution from the FPLC was 31.5 mg mL^{-1}, or 1.3 mM of protein. In an earlier crystallisation trial 14 mg mL^{-1} of protein was used, so this elution was diluted to 15.75 mL^{-1} either in buffer (20 mM Tris pH 8 containing 100 mM NaCl) for the apo-protein or with 1.3 mM of allyl-HBA for the holo-protein. This gives a 1:1 ratio of CobH (T85A): allyl-HBA concentration. Both of these were used in the hanging drop vapour diffusion experiments on siliconised cover slips (Hampton research) in 24-well XRL plates (Molecular Dimensions Ltd.), as described in Chapter 2.5.6, with two drops per condition: one apo-protein, one holo-protein. Structure screen 1^{TM} (Molecular Dimensions Ltd.) was used to identify conditions that either produced crystals or crystalline precipitant. These trays were incubated aerobically at 20 °C and checked regularly for crystal growth.

The condition from structure screen 1 which produced the best crystals consisted of 0.1 M sodium cacodylate pH 6.5, 18 % w/v of polyethylene glycol (PEG) 8000 with 0.2 M calcium acetate hydrate as the salt. This condition was optimised by changing either the concentration of calcium acetate hydrate and PEG 8000, or adding volatile compounds to the well (Appendix A.4). The concentration of the protein was halved to 7 mg mL^{-1}. These optimisation trays were also incubated aerobically at 20 °C. The condition in the custom screen which yielded the best crystals was 20 % PEG 8000, 0.2 M calcium acetate hydrate, 0.1 M sodium cacodylate pH 6.5 with 20 % methanol in the well. An arm was isolated from the apo-crystal (Figure 3.4.3.1, A) and the whole co-crystal was taken for diffraction (Figure 3.4.3.1, B). This was achieved by scooping the crystal up on a loop, dipping it into cryo-protectant and flash freezing in liquid nitrogen. The initial hit screen was checked again and a diffraction quality crystal was found in condition 3: 0.2 M ammonium sulphate, 0.1 M sodium acetate, 25 % PEG 4000. This was also picked and frozen in its own cryo-protectant. The cryo-protectant was the same as the well condition, minus the methanol in the case of the

optimised condition, with 20 % glycerol added (McFerrin and Snell 2002). This serves to prevent crystalline ice formation which will interfere with subsequent x-ray diffraction.

Figure 3.4.3.1: CobH (T85A) crystals. A: The apo-protein, the arm of which was broken off (circled) for diffraction. B: A few of the co-crystals which formed, some of these were frozen to be taken for diffraction.
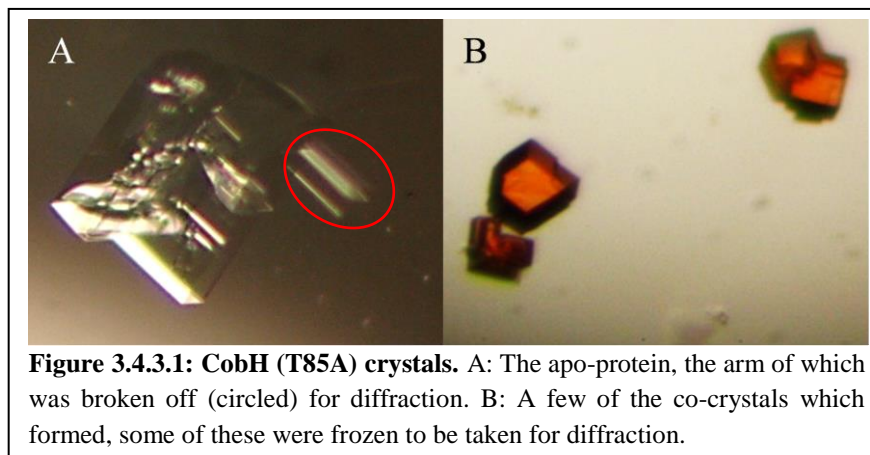
All of the data were collected at Diamond Light Source (Oxfordshire, UK) in collaboration with Prof. David Brown (University of Kent/ Argenta, Charles River) on the IO4-1 beamline using an Pilatus 2M (289x254 mm^2) detector. The data was auto-integrated and scaled using the Xia2 package using XDS (X-ray data software) and XSCALE (3dii) at the beamline (Winter and Waterman 2012). The apo-crystal diffracted to 1.2 Å and the co-crystal diffracted to 1.6 Å. They both belonged to space group C 1 2 1 which is the same as the published structure of *R. capsulatus* CobH co-crystallised with HBA (PDB 4AU1) (Deery et al. 2012). This crystal was solved in the Warren Lab (Deery et al. 2012). The parameters of the collected data were similar to the published data (Table 3.4.3.2, apo-crystal data in Appendix A.5). As there is only one amino acid substitution between this solved wild-type CobH and the T85A mutant, instead of molecular replacement, the initial model was generated by refining the dataset in Refmac5 against 4FDV (Murshudov et al. 1997). This model was further refined using Refmac5 in the CCP4i suite and manual model improvement was performed in Coot to produce the final model (Murshudov et al. 1997; Winn et al. 2011; Emsley et al. 2010).

Table 3.4.3.2: Comparison of the mutant-allyl-HBA co-crystal data collected to the previously published wild type-HBA

Single wavelength anomalous dispersion/diffraction (SAD)		
	Co-crystal CobH (T85A) allyl-HBA	**Co-crystal CobH HBA (4FDV)**
Wavelength (Å)	0.92819	0.97630
High resolution limit	1.57 (7.02 - 1.57)	1.68
Low resolution limit	48.38 (48.38 - 1.61)	34.7
Completeness	98.8 (91.6 - 99.7)	97.4
Multiplicity	3.3 (3.4 - 3.2)	4.1 (redundancy)
I/sigma	6.3 (12.6 - 1.3)	16.9 (6.6)
R_{merge}	0.237 (0.249 - 0.899)	0.059
Anomalous completeness	85.8 (87.3 - 81.9)	
Anomalous multiplicity	1.5 (1.9 - 1.7)	
Unit cell dimensions: a (Å)	71.230	70.250
b (Å)	66.630	66.030
c (Å)	48.920	48.480
α (°)	90.000	90.000
β (°)	99.030	98.980
γ (°)	90.000	90.000
Spacegroup	C 1 2 1	C 1 2 1
Sfcheck twinning score	2.13 Data does not appear twinned	

The data used in Table 3.4.3.2 for the wild type crystal complex are a combination of the data published in the paper and online in the protein database (PDB) (Deery et al. 2012). The final refinement statistics are shown in Table 3.4.3.3.

Table 3.4.3.3: The final refinement statistics of the mutant-allyl-HBA co-crystal data

	Final
R factor	0.1723
R free	0.2384
Rms Bond Length (Å)	0.0304
Rms Bond Angle (°)	2.7310

The Ramachandran plot for this data is shown in Figure 3.4.3.4 and Table 3.4.3.5.
Ramachandran plots show the distribution of the residues in the solved crystal structure
against the theoretically favoured regions. Only one residue, glycine 161, lies outside the
theoretically favoured region.

Figure 3.4.3.4: The Ramachandran plot of the final co-crystal model.
Of all the residues 98 % are in preferred regions, 1.5 % are in allowed
regions and 0.5 %, one residue GLY 161 is an outlier shown in the red
triangle on the plot. This high level of residues in the preferred regions
validates the model of the CobH (T85A) protein which validates the
structure if the ligand bound, thereby confirming the structure of allyl-
HBA. Plot was generated in Coot (Emsley et al. 2010).

Table 3.4.3.5: The Ramachandran Plot data

Residues	Number	Percentage (%)
In preferred regions	195	97.99
In allowed regions	3	1.51
Outliers	1	0.50

The high number of residues in preferred regions validates the final model of CobH (T85A) and confirms the structure of allyl-HBA. T85A CobH is a dimer, as is wild type CobH, and two molecules of allyl-HBA are bound within the dimer (Figure 3.4.3.6). The active site pocket is located around the dimer interface and is composed of residues from both of the CobH monomers in the complex (Figure 3.4.3.6, A). The resulting substrate binding pocket is very tight and there is very little space in the vicinity of the C5 position (Figure 3.4.3.6, B). The reason the T85A mutant was more efficient than the wild type CobH when synthesising C5-allyl-HBA is likely to be due to there being more space in the active site around the C5 position. The T85A CobH structure shows that allyl group is bent acutely round in order to fit into the binding pocket, even though the substitution of alanine for threonine has already increased the space available (Figure 3.4.3.7). In all the Figures with CobH and allyl-HBA the allyl group is shown in black to distinguish it from the rest of the tetrapyrrole.

Figure 3.4.3.6: Allyl-HBA in the binding cleft of CobH(T85A). The close fitting dimer of CobH (T85A) can been seen in A, inset, with one dimer predominantly in blue and the other in green. B is a zoomed in image of the black box in A showing the cramped situation of the allyl-HBA (turquoise) with the allyl group in black.

Figure 3.4.3.7: The binding pocket of wildtype CobH (PDB:4FDV) and CobH (T85A) complexed with the natural product, HBA, or analogue, C5-allyl-HBA. A and B show the wild type CobH, whilst C and D are CobH (T85A), as labelled on the figure. The mesh represents the protein surface at the THR85 or ALA85 residue. As seen in B, the THR85 mesh clashes with the allyl group (black), modelled in using PyMol (Delano 2002), however it does not in the T85A mutant, D. The wild type CobH is coloured with carbon in orange, nitrogen in blue and oxygen in red. CobH (T85A) has carbon in yellow, nitrogen in blue and oxygen in red. HBA shows carbons in grey, nitrogen in blue and oxygen in red, with the C5 carbon and methyl group in cyan. In C5-allyl-HBA carbon is cyan, nitrogen is blue and oxygen is red, with the allyl group in black.

The threonine residue in wild type CobH is shown in the structure in panels A and B (Figure 3.4.3.7), PDB reference 4FDV. The mesh which indicates the predicted surface of this residue shows a clash with the allyl group whereas the alanine 85 mutation (Figure 3.4.3.7, C and D) allows the allyl group to fit. The contortion of the allyl group suggests it would fit better if given more space and further mutations will be necessary if transferring a larger group to the C5 position.

The presence of the allyl group at the C5 position confirms the structure of the C5-allyl-HBA (Figure 3.4.3.8), the first of the novel cobalamin analogues.

Figure 3.4.3.8: The structure of C5-allyl-HBA. This was confirmed by X-ray diffraction in the CobH (T85A) protein crystal. Carbon is shown in cyan, nitrogen in blue and oxygen in red.
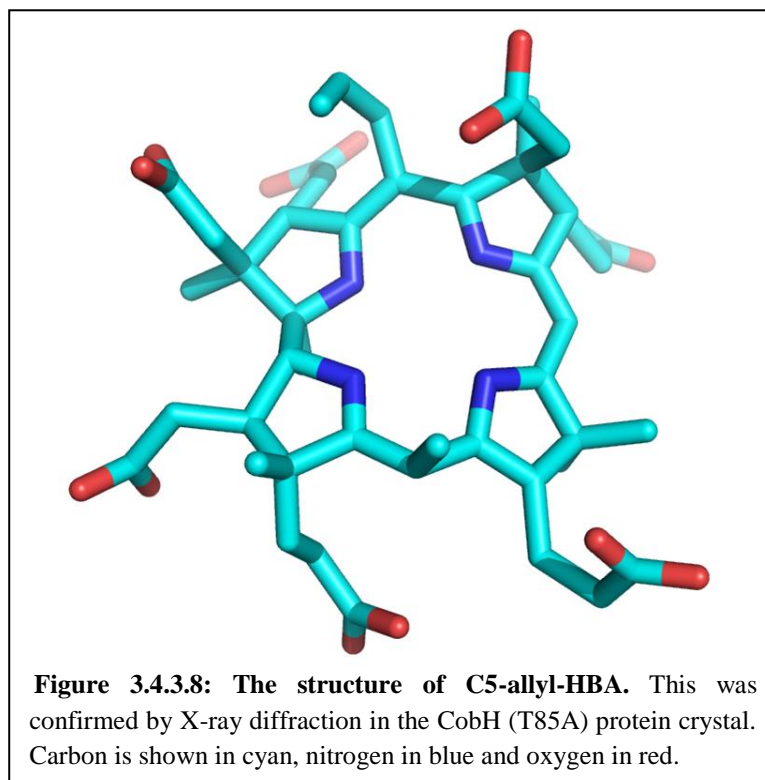
3.4.4 Amidation of the sidechains

There are two amidases which amidate all of the propionate and acetate sidechains of the corrin, except for sidechain *f*. The first reaction, catalysed by CobB, uses glutamine as the amide donor to amidate sidechains *a* and *c* (Chapter 2, 2.5.7). These are the first two of the six amidations that occur around the macrocycle. This reaction also requires ATP, 8 mM, and magnesium chloride, 20 mM, and was incubated overnight at 37 °C.

The reaction progressed efficiently to form C5-allyl-HBA-*a,c*-diamide (C5-allyl-HBAD) (HPLC-MS data Figure 3.4.4.1). The other four amidation reactions are catalysed by CobQ which, in *R. capsulatus*, requires cobalt insertion, reduction of the metal ion and adenosylation before it will amidate the sidechains. To avoid cobalt insertion at this stage the CobQ from *A. vinosum* was used as it has no requirement for cobalt or the adenosyl upper ligand, even though there is hardly any sequence discrepancy between the two CobQs (Appendix A.9).

The reaction assay was set up much like the CobB reaction, and the *A. vinosum* CobQ amidated the *b*, *d*, *e* and *g* sidechains, producing a completely new intermediate (even excluding the allyl group) C5-allyl-HBA-*a,b,c,d,e,g*-hexamide (C5-allyl-HBAH) (Figure 3.4.4.1). The structure of this intermediate was confirmed by NMR (Figure 3.4.4.1 with HPLC-MS data) (Additional data in Appendix A.11). Subsequent modifications to C5-allyl-

HBAH were performed using chemistry and not *in vitro* assays with native proteins (Figure 3.4.4.2).

Figure 3.4.4.1: HPLC-MS data for C5-allyl-HBAD and C5-allyl-HBAH as well as C5-allyl HBAH NMR data. A: Top top panel shows C5-allyl HBAD trace with the corresponding UV absorbance and MS. The mass is the expected value of 907.4. C5-allyl HBAH is on the lower panel and elutes earlier than C5-allyl HBAD. Again, the mass was the expected 901.5. This sample of C5-allyl HBAH was also run on the NMR (B), confirming the structure of six amides, the C5 allyl group and the cobalt-less corrin ring.

Figure 3.4.4.2: Strategy for chemical modifications to C5-allyl-HBAH for fluorophore attachment. The allyl can be extended using cysteamine to give a primary amine at the C5 moiety. The cobalt can then be inserted into the macrocycle and a fluorophore, in this instance BODIPY-TR-X succinimidyl ester can be attached.

3.4.5 Extending the linker and cobalt insertion

To extend the allyl group a thiol-ene reaction was implemented (Figure 3.4.5.1) with cysteamine as the donor molecule and VA044 (chemical name 2,2'-Azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride) as the radical initiator. This left the C5 group with a terminal amine which reacts with succinimidyl ester groups on commercial fluorophores (Wang et al. 2011) (Chapter 2, 2.5.8).

Figure 3.4.5.1: **The thiol-ene reaction mechanism used to extend the allyl group at C5.** The attachment of cysteamine to the C5 allyl results in a terminal primary amine which can be exploited for further modifications. VA044 is the radical initiator used to form the thiol radical.
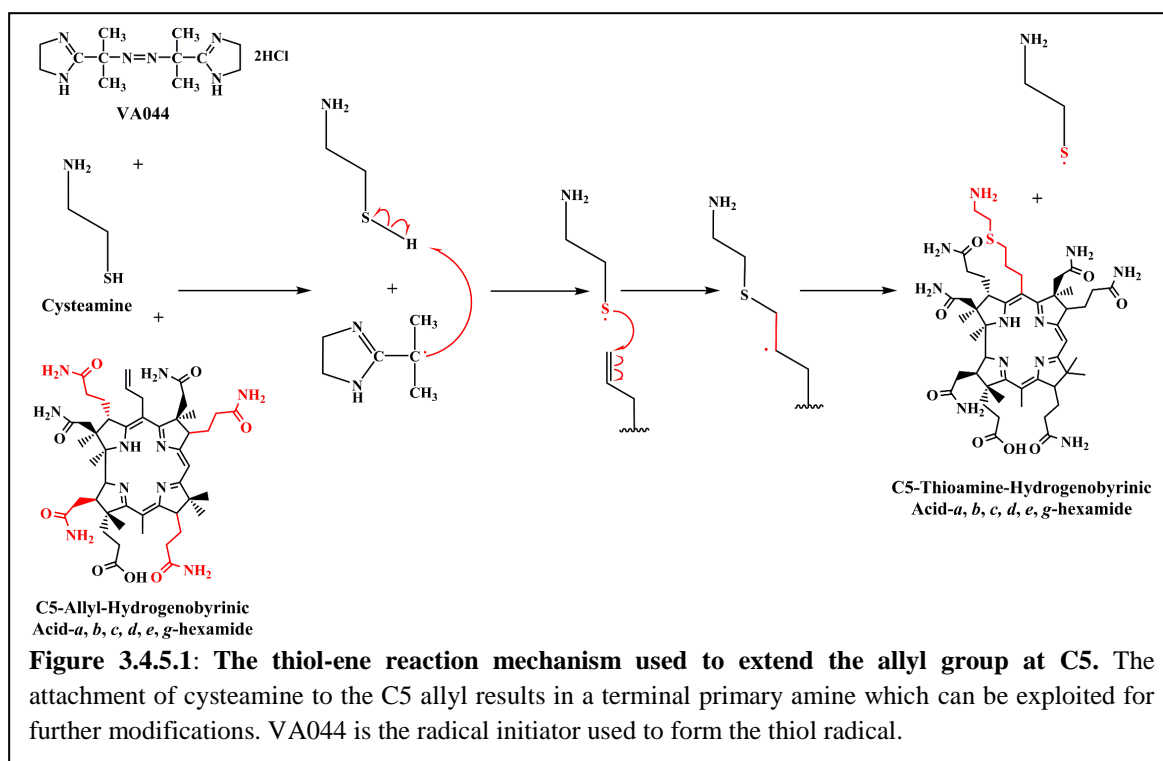
The extension of the allyl group gave rise to C5-thioamine-HBAH (HPLC-MS data in Figure 3.4.5.2). This intermediate analogue was then used in the cobalt insertion reaction. The chemical insertion method was used. The intermediate was incubated with cobalt chloride

hexahydrate and ammonium hydroxide, and then stirred for 1.5 hours at 80 °C in the dark in an anaerobic glove box at < 2 ppm oxygen (Moore, Biedendieck, et al. 2013; Kräutler 2006) (HPLC-MS and NMR data in Figure 3.4.5.2. Additional NMR data is in Appendix A.8).

Figure 3.4.5.2: HPLC-MS data for C5-thioamine-HBAH and C5-thioamine-cobyric acid as well as C5-thioamine-cobyric acid NMR data. A: Top panel shows C5-thioamine-HBAH trace with corresponding UV and mass spectra. The m/z is one more than the expected 977.5. The bottom panel of A is C5-thioamine-cobyric acid with the UV and mass spectra. It is important to note the shift of maxima from 328 nm to 352 nm. This is caused by the incorporation of cobalt into the macrocycle. The NMR data in B confirms the structure of this new analogue C5-thioamine-cobyric acid, in particular the extension of the allyl group at C5 and the incorporation of cobalt.

3.4.6 Attaching the fluorophore

Once the allyl group had been extended to a primary amine and the cobalt inserted, the C5-thioamine-cobyric acid was ready for fluorophore conjugation. It is also possible to attach the lower loop at this stage but this was not done as many cobalt containing intermediates of cobalamin are recognised by the uptake proteins of different organisms regardless of the presence of the lower loop (Nielsen et al. 2012; Fedosov et al. 2007; Lildballe et al. 2012). The fluorophore conjugate cobyric acid analogues can be traced *in vivo* upon uptake, and the localisation within the organisms can be discovered. Initially the fluorophore used was

fluorescein (Figure 3.4.6.1 A, HPLC-MS data in Appendix A.15), but this photobleached quickly (see Chapter 4, 4.3 *E. coli*) so BODIPY® TR-X and Oregon green® 514 nm cobyric acid analogues were also synthesised (Figure 3.4.6.1 B) as they are known to be more photo-stable (Hinkeldey et al. 2008).

Figure 3.4.6.1: Strategy for the attachment of the fluorophore. A shows the formation of the peptide bond linking the fluorescein to the cobyric acid analogue to synthesise a fluorescent analogue. B contains the two other cobyric acid analogues: Oregon-green on the left and BODIPY on the right. They were both synthesised in the same way as shown in panel A.
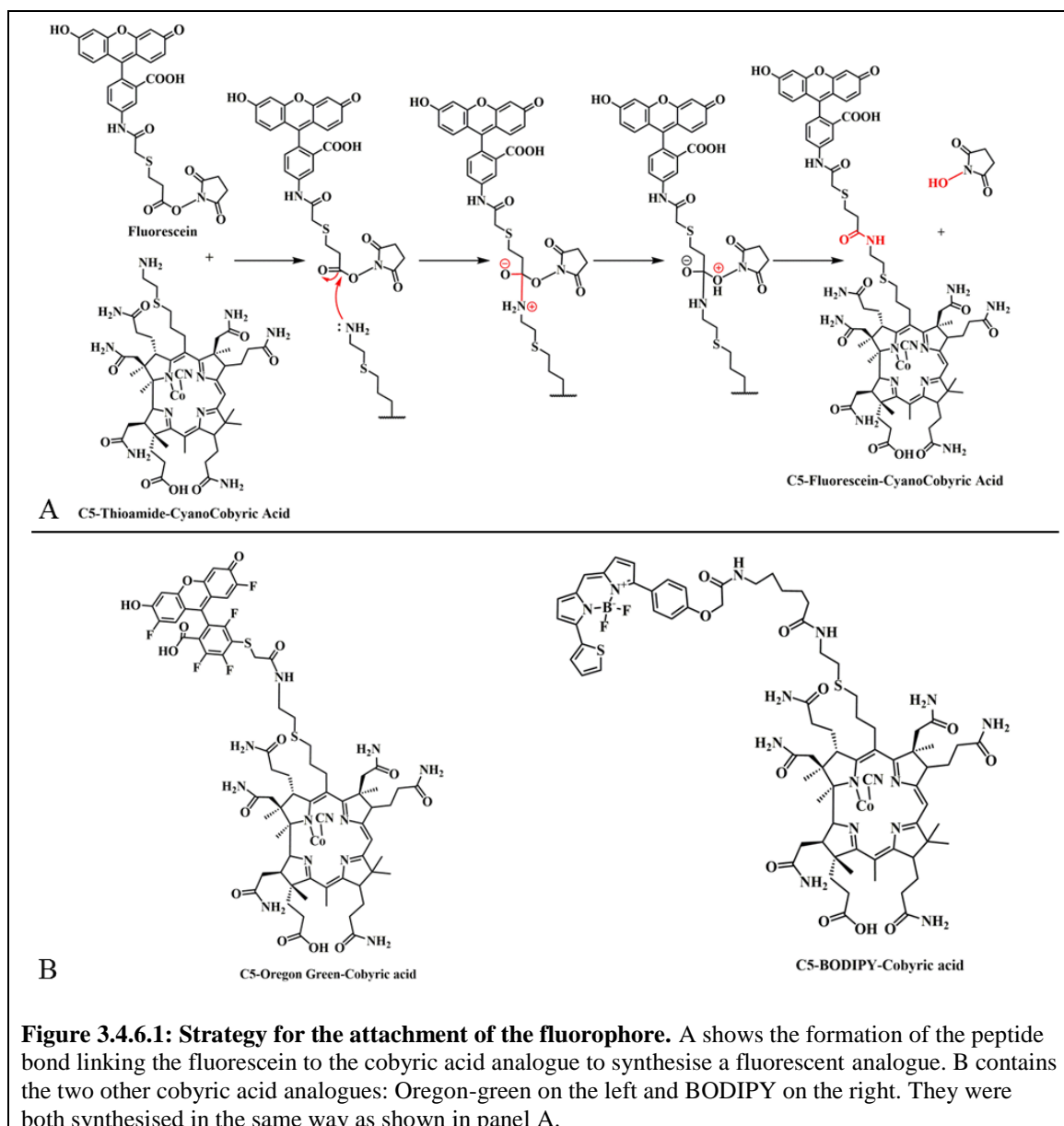
The C5-fluorescein cobyric acid was tested for recognition *in vivo* on a bioassay plate. This was done by dotting the analogue on a *S. enterica* AR3612 bioassay plate (Figure 3.4.6.2) (Raux et al. 1996). This *S. enterica* has both *cysG* and *metE* genes deleted. The *metE* gene encodes the cobalamin independent methionine synthase protein, so the deletion of *metE* means *S. enterica* relies on the cobalamin dependent *metH* for methionine synthesis making the growth dependent on cobalamin (or exogenous methionine). The *cysG* gene encodes the enzyme responsible for the conversion of uro'gen III to precorrin-2, the first committed step

in cobalamin biosynthesis (Figure 1.2.1.1), deletion of this means *S. enterica* cannot produce its own cobalamin. The first cobalamin intermediate that the *S. enterica* cobalamin transport proteins can take up is cobyric acid, which it can convert to cobalamin. Therefore, *S. enterica* AR3612 is dependent on cobalamin intermediates from cobyric acid onwards for growth (Raux et al. 1996). If the *S. enterica* in the agar grows around the dot of the analogue then it can not only take the analogue up, but can convert it into an active cofactor for MetH, the cobalamin dependent methionine synthase. The reason *S. enterica* is in the agar is because it can only synthesise cobalamin anaerobically (Jeter et al. 1984).

Figure 3.4.6.2: *S. enterica* **bioassay plate grown with analogue intermediates** *S. enterica* bioassay plate showing the growth induced by the presence of C5-fluorescein cobyric acid (top two), allyl-cobyric acid and cobalamin.
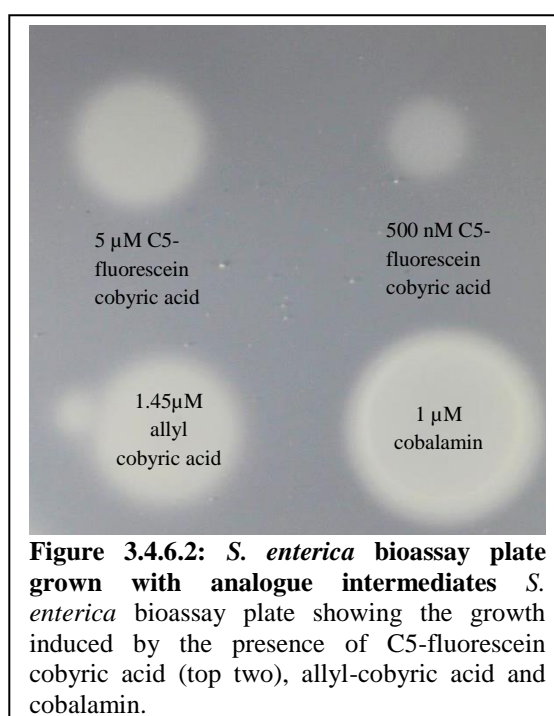
The samples were dotted on to the plate in 10 µL drops, and incubated overnight at 37 °C. Figure 3.4.6.2 shows that the fluorescein analogue is taken up by *S. enterica* and facilitates its growth. The lower left spot is the C5-allyl cobyric acid which supports growth more effectively than the C5 fluorescein equivalent. The 1 µM cobalamin standard is the lower right spot. There is a clear decrease in either recognition of the C5 analogues or utilisation *in vivo* compared to the cobalamin standard, which is less concentrated than any of the analogues and yet induces a larger area of growth.

3.5 Discussion

The specific modification of the C5 position using a combination of an allyl-SAM analogue and the C5 methyltransferase enzyme works well, resulting in the synthesis of C5 analogues of cobyric acid. However, the Early insertion pathway was not viable. The

substrate intermediate, cobalt precorrin-6b, was difficult to synthesise and there is no proof that C5-allyl analogues were made. The HPLC-MS data shows that a un-allylated cobalt precorrin-7 was synthesised. This intermediate has the C12 position decarboxylated, a reaction catalysed by CbiT which also methylates the C15 position. Therefore, this intermediate is likely to have the C5 position undecorated. This indicates that the CbiE protein may not be active in the incubation, or unable to use allyl-SAM as a donor molecule. This was not the only issue with using the Early insertion pathway to synthesise analogues. The main issue was the instability of the intermediates used and produced. This meant that analysis and confirmation of the products on the aerobically run HPLC-MS was impossible. Due to these problems the Late insertion pathway was investigated.

In the Late insertion pathway the recent discovery that the C15 methylated, C12 decarboxylated intermediate precorrin-7 could be overproduced revolutionised this research (Deery et al. 2012). The C5 position could now be specifically targeted for allylation by CobL and allyl-SAM. The following methyl migration reaction, catalysed by CobH, was improved by the T85A mutation. This mutation was shown in the crystal structure to increase the space within the binding pocket around the C5 position, thereby allowing the allyl group to fit easier. The allyl group was highly bent to fit into the enlarged active site, which means that this reaction can be improved further by expanding the active site. The crystal structure also confirmed the presence of the allyl group on the C5-allyl-HBA. The amidation reactions catalysed by CobB and CobQ^{Av} worked efficiently to produce C5-allyl-HBAH. The structure of this was confirmed by NMR which showed clear peaks corresponding to the six amides and the allyl group. The allyl was extended using cysteamine to give a terminal amine which can react with many chemicals. Cobalt was chemically inserted into the macrocycle of this thioamine analogue. The one drawback of using this pathway was this need for subsequent cobalt insertion as there is a significant loss of material, but this also meant that the (cobalt-less) intermediates produced were much more stable than those of the Early insertion pathway.

The fluorophores were attached to the thioamine linker to complete the cobyric acid analogue synthesis. This reaction and the cobalt insertion reaction resulted in multiple products and the desired analogue had to be purified by preparative HPLC. The biological activity of these analogues was confirmed on bioassay plates containing a cobyric acid dependent strain of *S. enterica*. The C5-fluorescein cobyric acid analogue supported the growth of *S. enterica*. This proved that the analogue was recognised by the *S. enterica*

cobalamin uptake proteins, and converted into functional methyl-cobalamin cofactors *in vivo*, which were subsequently used by MetH to produce methionine.

These fluorescent analogues can be used to image the uptake and localisation of cobyric acid in bacteria and in higher organisms as well. Cobalt β ligand fluorophore conjugates of cobalamin have been imaged in human cells, but C5 analogues have never been made before (Shell and Lawrence 2015). These C5 analogues allow investigations into the earlier intermediates of the pathway as well as cobalamin. In the next chapter these analogues are imaged in various organisms, something that has never been done before.

Chapter 4

Characterising molecular recognition of cobalamin analogues in a variety of organisms

4.0 Introduction

Cobalamin is an essential vitamin for many organisms, such as humans, but higher plants and fungi do not require it (Croft et al. 2005; Moore and Warren 2012; Helliwell et al. 2011). However there is evidence that plants, for example lettuce and soya beans, can absorb exogenous cobalamin (Watanabe et al. 2013). There is thus a need to confirm cobalamin uptake as opposed to bacterial contamination and hence *L. sativum* (garden cress) and *A. thaliana* were included in the investigations. Although *L. sativum* is not a standard model organism for plants, it is small and easily cultivatable in the lab alongside *A. thaliana*.

Parasitic worms such as *D. latum* can cause cobalamin deficiency in the host organism due to the parasite stealing cobalamin from the host (Allen 2008). A recent paper has shown that *C. elegans*, a nematode worm, displays a cobalamin deficiency phenotype when grown in the absence of cobalamin which causes loss of fertility, extended life cycle and a reduced lifespan (Bito et al. 2013). However, little is known about cobalamin uptake and distribution, and a *C. elegans* cobalamin transport protein has not yet been identified. Other model organisms, such as zebrafish, have cobalamin binding proteins which display similar binding affinities to the human transport proteins TCII, IF and HC for cobalamin recognition (Greibe, Fedosov, and Nexo 2012). It is likely that *C. elegans* will behave in a comparable way as bioinformatic analyses suggest that 60-80 % of human genes have orthologues in *C. elegans* (Kaletta and Hengartner 2006). *C. elegans* have two cobalamin-dependent enzymes: MetH and MCM, and are easy to nurture in the lab on agar plates with *E. coli* as their food source, but they can also eat other microorganisms (Bito et al. 2013; Corsi et al. 2016).

The lifecycle of *C. elegans* at 22 °C is around 3 days long (Figure 4.0.1), with the first eggs laid around 60 hours after the nematodes hatch. These eggs can obstruct the internal detail around the vulva when imaging the nematodes. Therefore, L4 stage nematodes were used for the imaging investigations, as the somatic structures of the nematode have formed but eggs within the uterus have not.

The occurrence of a cobalamin deficiency phenotype, the homology with humans and the ease of culturing and imaging all make *C. elegans* an optimal laboratory model for animals, specifically humans. In addition, cobalamin metabolism related human homologue gene mutants in *C. elegans* are available which can provide insight into cobalamin related diseases arising in humans (Froese and Gravel 2010).
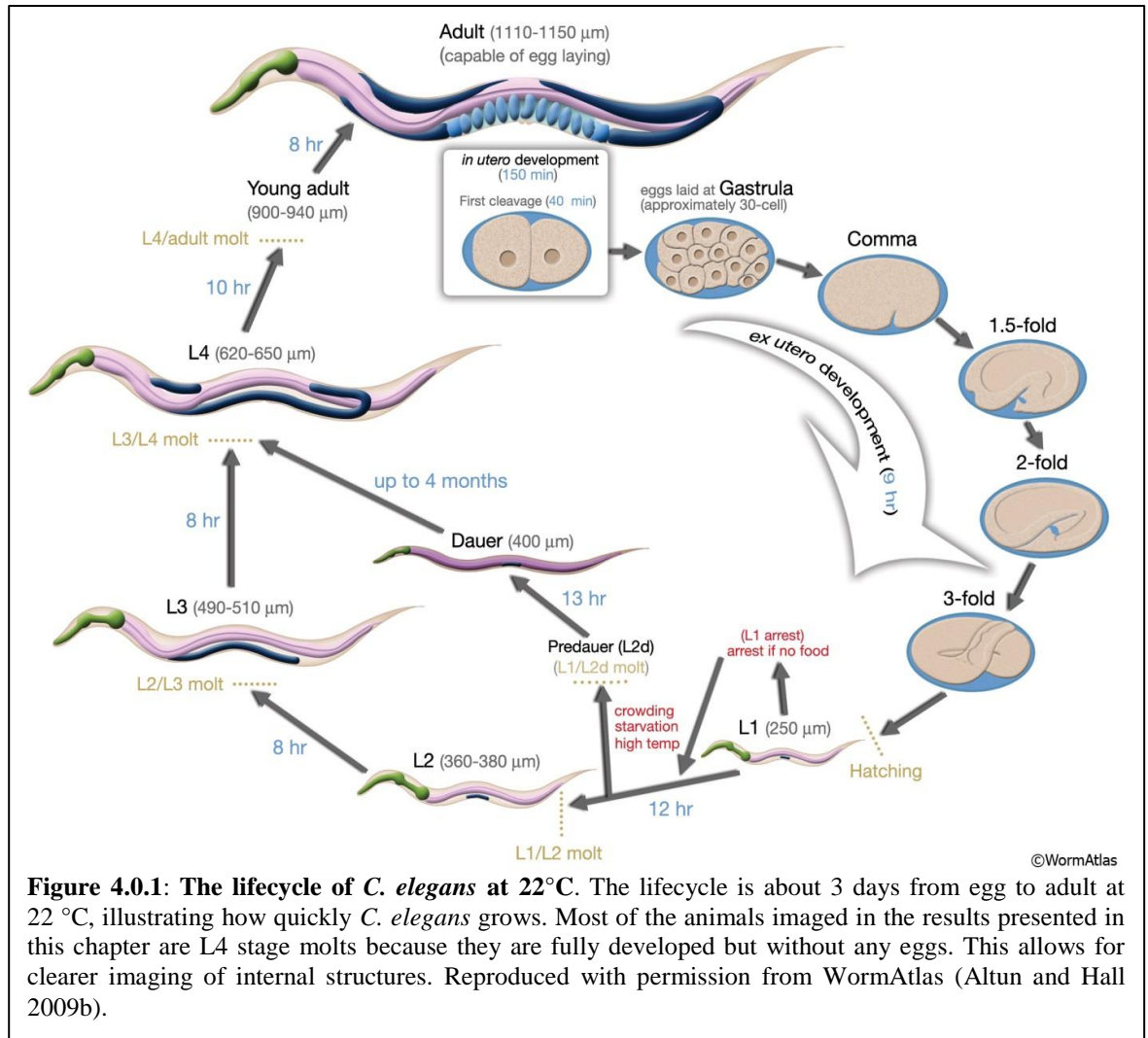
Figure 4.0.1: **The lifecycle of** *C. elegans* **at 22°C**. The lifecycle is about 3 days from egg to adult at 22 °C, illustrating how quickly *C. elegans* grows. Most of the animals imaged in the results presented in this chapter are L4 stage molts because they are fully developed but without any eggs. This allows for clearer imaging of internal structures. Reproduced with permission from WormAtlas (Altun and Hall 2009b).

Of all the uptake systems associated with cobalamin absorption, the *E. coli* cobalamin uptake mechanism has been the best studied in terms of molecular detail (Figure 4.0.2). This system allows for the uptake of both complete cobalamin molecules as well as the salvage of incomplete corrinoids such as cobinamide (K. Kadish et al. 2003). The outer membrane protein, BtuB, has a cobalamin riboswitch upstream of the gene which stops gene expression in the presence of cobalamin, limiting the cellular concentrations of corrinoids (Gallo et al. 2008). *E. coli* can convert cobinamide into cobalamin, but it cannot convert cobyric acid into cobinamide (Figure 4.0.2) even though it can scavenge it (Raux et al. 1996). *S. enterica* has similar cobalamin uptake proteins to *E. coli*, and can synthesise cobalamin *de novo* so it can complete the cofactor synthesis from any scavenged intermediates (Kanehisa and Goto 2000; Kanehisa et al. 2016).

Figure 4.0.2: The cobalamin uptake system and biosynthesis pathway enzymes in *E. coli*. A: BtuB recognises corrinoids on the extracellular surface and transports them into to periplasm via a TonB mechanism. BtuF then binds the periplasmic corrinoids and forms a complex with BtuCD on the inner membrane. This initiates the release of the corrinoid from BtuF and allows BtuCD to transport it to the cytoplasm. Reproduced with permission from (Kräutler and Puffer 2012). B: The boxes in green highlight the genes present in *E. coli*. It has all the enzymes necessary to convert cobinamide into various cobalamins, starting with BtuR, the enzyme responsible for adenosylating cobinamide, but has none of the earlier biosynthesis enzymes.

The aims set out for this chapter were:

1. To compare the molecular recognition/ interference of the fluorophore analogues on the uptake mechanism in *S. enterica*.

2. To compare the uptake of C5 cobyric acid analogues to the ribose linked cobalamin analogues in *E. coli* and *C. elegans*.

3. To ascertain where the corrinoid analogues localise in *C. elegans*.

4. To examine C5 cobyric acid analogue and ribose linked cobalamin analogue uptake in *L. sativum* and *A. thaliana*.

4.1 Results

4.2 Chemical synthesis of ribose linked cobalamin analogues

Having made a number of cobalamin analogues, the natural next step was to investigate how, or if, these compounds are taken up in different biological systems. The organisms listed above may not take up C5-cobyric acid intermediate analogue; therefore cobalamin ribose analogues were made in the same way as previously reported (Lee and Grissom 2009; Pathare et al. 1996). These ribose linked cobalamin analogues have been shown to bind to cobalamin transport proteins, and allow a comparison between cobyric acid and cobalamin analogue uptake (Clardy et al. 2011).

4.2.1 Chemical synthesis of BODIPY® TR-X ribose linked cobalamin

The chemical synthesis of ribose linked cobalamin analogues was performed as previously reported (Lee and Grissom 2009; Pathare et al. 1996). Firstly CDI (1,1-carbonyl diimidazole) coupling of 1,2-Diaminoethane on commercial vitamin B_{12} (Sigma) formed a primary amine on the 5' hydroxyl of the ribose. The attachment of the fluorophore to the primary amine is the same reaction as for the C5 analogues (Figure 4.2.1.1). The resulting product was dried in a vacuum centrifuge and resuspended in water before HPLC purification.

Figure 4.2.1.1: The chemical synthesis of BODIPY® TR-X ribose linked cobalamin. The experiment was repeated with Oregon green® 514 as well.
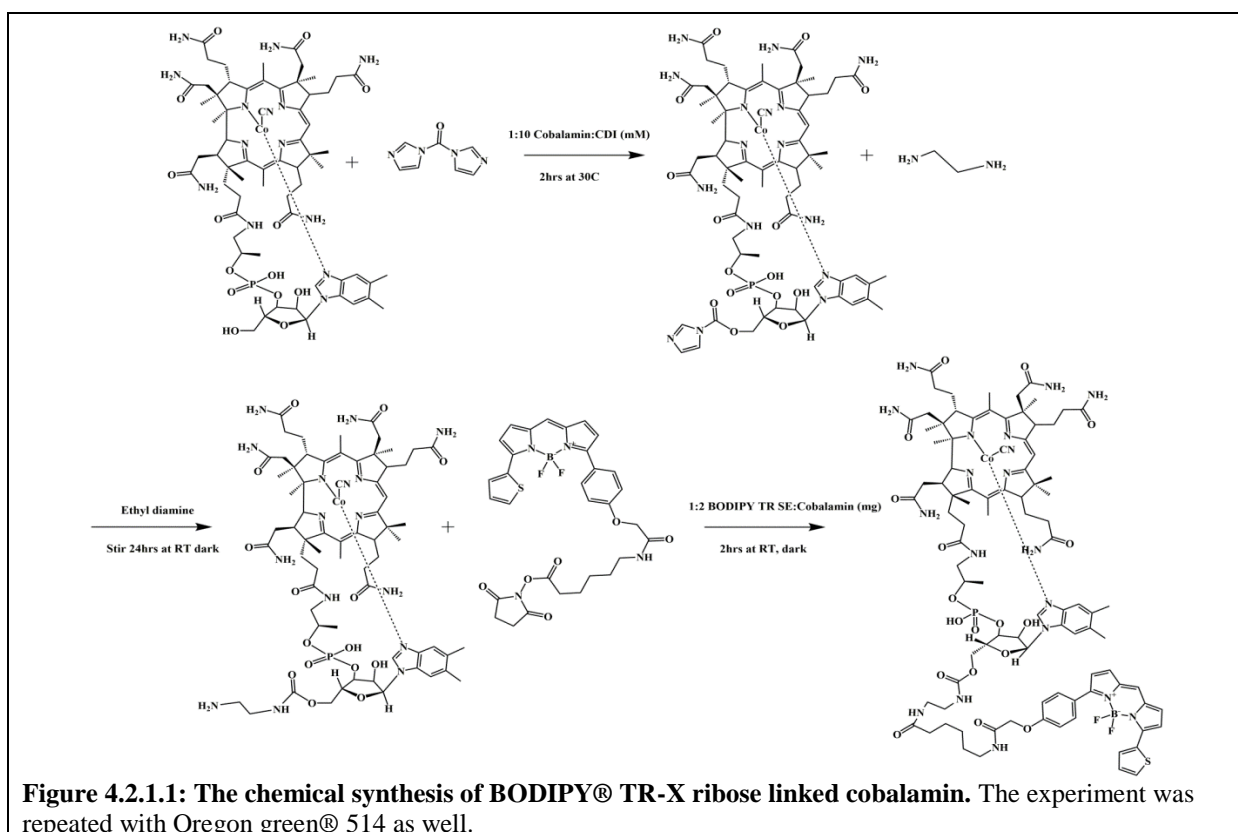
The mass was confirmed by HPLC-MS (Figure 4.2.1.2). The recognition of ribose linked cobalamin analogues by IF, HC and TCII, the human uptake proteins, has been confirmed in previous studies (McEwan et al. 1999; Clardy et al. 2011).

Figure 4.2.1.2: HPLC-MS of BODIPY TR-X® ribose linked cobalamin (BOB$_{12}$). This sample was run before purification using preperative HPLC. Panel A is the absorbance exhibited by the analogue at 355 nm. The m/z and the UV absorbance correlate exactly with the predicted mass of the BODIPY cobalamin analogue, and the absorbance peaks of cobalamin (~355 nm) and BODIPY (~588 nm) are present.

4.2.2 Chemical synthesis of Oregon green® 514 ribose linked cobalamin

The Oregon green® 514 ribose linked cobalamin (Figure 4.2.2.1, A) was synthesised in the same way as the BODIPY® TR-X analogue, but using Oregon green® 514 instead of BODIPY® TR-X as the fluorophore conjugate (Figure 4.2.1.1). The resulting product was

dried in a vacuum centrifuge and resuspended in water before HPLC purification, and the mass was confirmed by HPLC-MS (Figure 4.2.2.1).

Figure 4.2.2.1: The structure of ribose linked Oregon green® 514nm cobalamin (OGB$_{12}$) and the HPLC-MS data. This sample was run after purification using preparative HPLC. Panel A shows the absorbance exhibited by the analogue at 355 nm, with the structure of OGB$_{12}$ inset. The m/z and the UV absorbance also correlate exactly with the predicted mass of OGB$_{12}$ and the absorbance peaks of cobalamin (~355 nm) and Oregon green 514® (~514 nm) are present.

4.3 Recognition and uptake of C5-cobyric acid and ribose linked cobalamin analogues in *Salmonella enterica*

The BODIPY® TR-X and Oregon green® 514 cobyric acid analogues and the equivalent ribose linked analogues were tested for uptake into *S. enterica* via the cobalamin uptake system. This was done by applying the analogues on to *S. enterica* containing bioassay plate. This *S. enterica* (AR3612) has both *cysG* and *metE* deleted. The *metE* gene encodes the cobalamin independent methionine synthase, so the deletion of *metE* means *S. enterica* relies on the cobalamin-dependent MetH protein for methionine synthesis, making the growth of the *S. enterica* dependent on cobalamin. The *cysG* gene is responsible for the conversion of uro'gen III to precorrin-2, the first committed step in cobalamin biosynthesis (Chapter 1, Figure 1.2.1.1), so *S. enterica* cannot produce its own cobalamin. The first cobalamin intermediate that the *S. enterica* can take up is cobyric acid, which it can convert to cobalamin, therefore the *S. enterica* AR3612 is dependent on cobalamin intermediates from cobyric acid onwards for growth (Raux et al. 1996).

Each analogue was dotted in 10 µL drops of 5 µM on to the bioassay plate along with a 50 nM reference dot of cobalamin (Figure 4.3.1, A). Dots of 10 nM, 50 nM, 100 nM and 1 µM of cobalamin were placed on another plate. This is the standards plate (Figure 4.3.1, B) which is incubated alongside the analogues plate, and is then used to calculate a standard curve of the diameter of growth against the concentration of corrinoid provided (Graph 4.3.2). The equation of the line is used to calculate the concentration of analogue taken up compared to cobalamin. In this equation, *x* is substituted for the diameter of growth measured and thus *y*, the concentration of analogue, can be solved. The concentration is normalised to the 1 µM reference point on the plate. Figure 4.3.1 shows that the analogues are recognised by *S. enterica*, but are not taken up as efficiently as cobalamin.
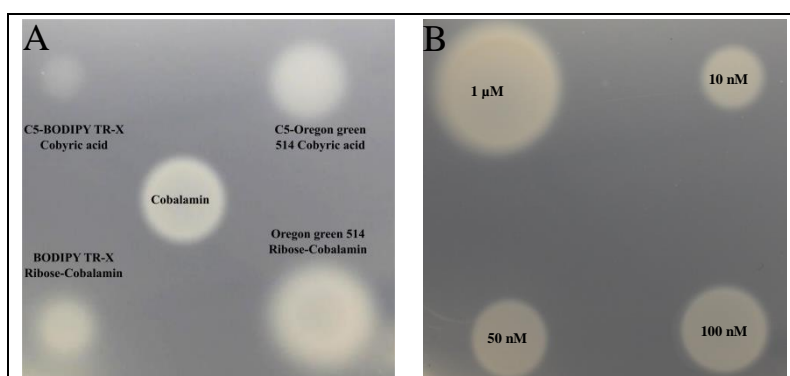
Figure 4.3.1: The *S. enterica* bioassay plate of all the Oregon green® 514 and BODIPY® TR-X fluorophore intermediates and the cobalamin standards. A: The analogues were dropped at a 5 µM concentration, with a reference dot of 50 nM cobalamin. B: The cobalamin reference plate which is used to calculate the concentration of cobalamin equivalent to the growth facilitated by the analogues by measuring the diameter of growth and plotting this against concentration to produce an equation. All drop sizes were 10 µL.

Graph 4.3.2: Relation between cobalamin concentration and diameter of growth on the *S. enterica* bioassay plate. The standard curve generated from measuring the diameter of growth of each standard in relation to the diameter of the plate, then plotted against the concentration of cobalamin dotted on to the plate. This was used to calculate the concentration of cobalamin equivalent to the growth facilitated by the analogues.

The diameter of the whole plate and the diameter of growth with the analogues were measured. The diameter of growth was divided by the diameter of the plate to give the size of growth in relation to the plate. This was regarded as *x* and substituted into the equation of the line shown in Graph 4.3.2. The equation was solved to give *y*, the concentration in nM. The 50 nM cobalamin reference drop was divided by the calculated concentration of the 50 nM standard, and the calculated concentrations of the analogues was normalised to this.

All of the analogues were 5 µM, 100 times more concentrated than the reference drop, but yielded growth spots which represented significantly less concentrated solutions. However, it is evident that the analogues are taken up and can be used to synthesise functional cobalamin cofactors to support the growth of cobalamin-dependent *S. enterica*.

Both of the BODIPY® TR-X analogues facilitate less growth than their Oregon green® 514 equivalents and the ribose-linked analogues support more growth than their cobyric acid counterparts. This is evident in the diameter of *S. enterica* rings of growth each analogue can induce when the same volume and concentrations are dropped on the bioassay plate. Table 4.3.4 shows that the Oregon green® 514 ribose cobalamin supports the *S. enterica* growth equivalent to 180 nM of cobalamin whereas C5- Oregon green® 514 cobyric acid is only equivalent to around 37 nM of cobalamin. BODIPY® TR-X ribose cobalamin is comparable to around 13 nM cobalamin which is about a third of the C5- Oregon green® 514

cobyric acid, but the C5-BODIPY ® TR-X cobyric acid equates to around 5 nM of cobalamin. The discrepancy between the cobyric acid analogues and the cobalamin analogues is likely to be because the cobyric acid intermediate must be converted into a functional cofactor before growth can occur. The difference between the two fluorophores may be caused by the fluorophore itself interfering with either the cofactor synthesis or binding to the cobalamin-dependent enzymes.

Table 4.3.4: The concentrations of the C5 analogues calculated using Graph 4.3.2

	Pixels	Pix/plate	Concentration (nM)	Normalised (nM)
C5- Oregon green® 514 cobyric acid	346	0.14	40.49	37.33
C5-BODIPY ® TR-X cobyric acid	189	0.08	5.17	4.77
Oregon green® 514 ribose cobalamin	466	0.19	195.15	179.93
BODIPY® TR-X ribose cobalamin	268	0.11	14.57	13.43
50 nM Cobalamin ref	365	0.15	51.94	47.89
Whole plate	2496			

4.4 Construction of *Escherichia coli* overproducing the outer membrane protein of the cobalamin transport system

Without the *btuB* plasmid (Figure 4.4.1) in the *E. coli* cells the analogues are not taken up in sufficient concentrations to produce detectable fluorescence (data not shown). This is due to a nonsense mutation in the *btuB* gene in BL21(DE3) which causes early termination of translation (Figure 4.4.1). Therefore, to ensure maximum uptake of the corrinoids and overcome the riboswitch regulation, *E. coli* were transformed with pLysS-*btuB*. These BL21 *E. coli* cells containing functional copies of *btuB* were used to image the corrinoid analogues *in vivo*.

Figure 4.4.1: The pLysS plasmid containing the *E. coli btuB* gene and the amino acid alignment of BtuB in BL21 (DE3) and JM109. A: On the pLysS plasmid containing the *btuB* gene (orange) the T7 promoter and terminator are in white whilst the chloramphenicol resistance gene, CmR is in green. There is a His tag present at the start of the gene which is not on the plasmid map. The T7 lysozyme hydrolyses the T7 RNA polymerase produced by the activity of the weak T7 RNA polymerase promoter Φ3.8 shown in white above the lysozyme, but it is still leaky enough to transcribe and translate the T7 RNA polymerase. B: There is a missense mutation at position 162 in the BtuB from BL21 (DE3), but the most detrimental mutation is the nonsense mutation at position 58, resulting in an early termination instead of glutamine. Therefore, BL21 (DE3) does not have a native functional BtuB.

4.5 Imaging of uptake of C5-cobyric acid and ribose linked cobalamin analogues in *Escherichia coli*

To test whether the analogues were recognised by native cobalamin transport systems they were incubated with BL21 (DE3) *E. coli* transformed with pLysS-*btuB*. This was grown in 4 mLs of LB (Luria-Broth, Chapter 2, 2.2.2) with 34 µg mL^{-1} of chloramphenicol and either 6 µM of the C5-fluorescein cobyric acid or 1 µM of one of the Oregon green® 514 analogues. The cultures were induced with 2 mM of IPTG and left overnight at 37 °C in a shaking incubator before imaging.

The next day the cells were pelleted, and resuspended in LB to wash off any of the analogue attached to the surface of the cells. This process was repeated three times. The cells were mounted on LB-agarose pads and imaged on an Olympus IX81 widefield microscope mounted on ASI stage (Applied Scientific) (see Chapter 2, 2.6.2) using PlanApo 100 x OTIRFM-SP 1.49 numerical aperture lens giving a total magnification of 160 x with the camera. The GFP filter (488±10 nm excitation/ 530±15 nm emission) was used for the C5-fluorescein cobyric acid containing cells and the YFP filter (500±10 nm excitation/ 535±15 nm emission) was used for the Oregon green® 514 analogues. Although this is not the maxima for Oregon green® 514, which is excited at 514 nm and emits at 530 nm, it was the closest filter available but it means the fluorescence is diminished compared to what it would have been if excited at 514 nm.

The C5-fluorescein cobyric acid can be seen localising to the cells but the fluorophore photobleached quickly (Figure 4.5.1, C). This is why the concentration of the C5-fluorescein cobyric acid added to the *E. coli* culture is six times more than that of Oregon green® 514: to allow the camera to image the fluorescence before it bleached completely. No DIC image was taken of the cells grown with C5-fluorescein cobyric acid, so there is no overlay, but the fluorescence is in discreet oblongs uniformly across the image which are about the size of *E. coli*.
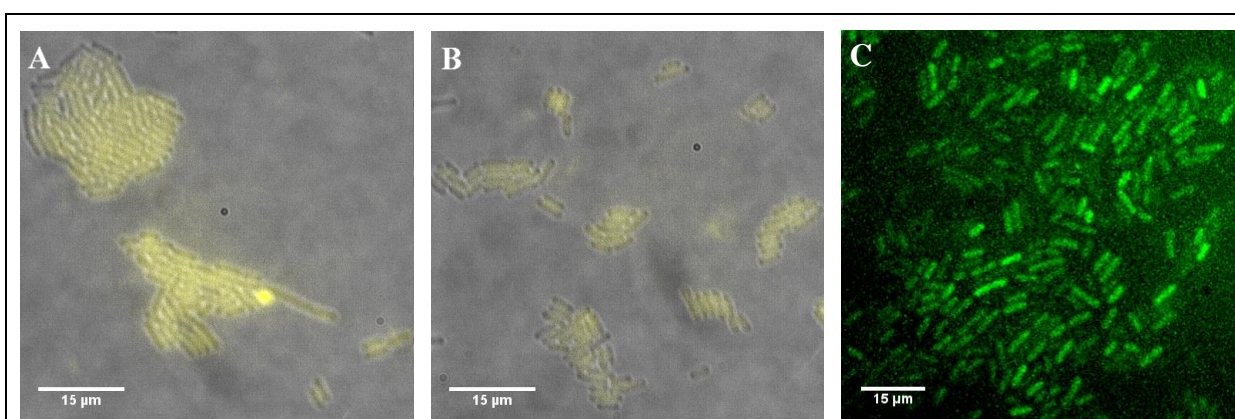
Figure 4.5.1: Cobalamin ribose linked Oregon green® 514 and C5- Oregon green® 514 cobyric acid and C5-fluorescein cobyric acid, internalised by BL21 (DE3) expressing BtuB. All three of the analogues are internalised by BL21 (DE3) expressing BtuB which proves that they are all recognised by BtuB and the subsequent cobalamin transport proteins, BtuFCD. Some of the cells are very long indicating stress possibly caused by overexpression of BtuB in the membrane.

BODIPY® TR-X and Oregon green® 514 fluorophores have been shown to be more photo-stable and this was observed in the cells grown in the presence of Oregon green® 514 (Hinkeldey et al. 2008). The Oregon green® 514 analogues were both taken up by *E. coli* (Figure 4.5.1, A and B) but it appears that the C5-cobyric acid analogue is not taken up as well in comparison to the cobalamin ribose analogue.

4.6 Imaging of uptake of C5-cobyric acid and ribose linked cobalamin BODIPY® TR-X analogues in *Escherichia coli* OP50

OP50 is the *E. coli* strain used to feed *C.elegans*. Due to the broad auto-fluorescence around blue/violet spectrum in the nematodes, the BODIPY® TR-X analogues were used as the excitation and emission wavelengths avoid any overlap with the autofluorescence. OP50, are a uracil requiring mutant of the Berkley strain and have been the food source for *C. elegans* since Sydney Brenner proposed *C. elegans* as a model organism in the 1970s (Brenner 1974; May et al. 2009). As these *E. coli* are the route by which the fluorescent analogues were introduced to *C. elegans* the cells needed to contain as much fluorophore-

analogue as possible. To do this, both BtuB and BtuF, which is the *E. coli* periplasmic cobalamin binding protein, were overexpressed in these cells using a pET-BAD vector (Figure 4.6.1, A).

The OP50 was transformed with the pET-BAD *btuB btuF* plasmid. It was grown for 4 hours and then induced with 2 % (w/v) *L*-arabinose. The BODIPY® TR-X analogues and controls were added to a final concentration of 1 µM at the same time as the arabinose. Eight cultures of OP50 were grown, all with pET-BAD *btuB btuF* but four were uninduced and four were induced. Four different supplements were provided to the two different induction backgrounds: the first had BODIPY® TR-X ribose linked cobalamin, the second just BODIPY® TR-X, the third had C5- BODIPY® TR-X cobyric acid, and the forth got nothing. These were left overnight in the dark at 28 °C in a shaking incubator. The next day the cells were pelleted (Figure 4.6.1, B) and resuspended 3 times with LB to wash off any exogenous BODIPY® TR-X analogues stuck to the outer cell membrane. These cells were resuspended in LB and imaged in the same way as the C5-fluorescein cobyric acid and the two Oregon green® 514 analogues.

Figure 4.6.1: **The pET-BAD plasmid containing the *E. coli btuBF* genes and the colour of the OP50 pellets after three washes with LB.** A: The *btuBF* genes (orange and yellow respectively) in this plasmid are under a BAD promoter inducible by *L*-arabinose (light blue). The BAD promoter was used instead of the T7 promoter because it is weaker, resulting in less protein production so the cell membrane should be less stressed. AmpR encodes ampicillin resistance. B: The pellet of cells incubated with BODIPY TR-X ribose cobalamin and induced have an intense dark blue colour from the BODIPY TR-X, but the equivalent uninduced cell pellet is cream, as if there was no supplementation in the media. Therefore, induction is necessary to obtain a high level of analogue in the cells for ease of imaging. Cells grown in the presence of BODIPY TR-X alone showed no difference between the induced and uninduced backgrounds, and the slight blue seen in the pellets is likely due to the hydrophobicity of BODIPY TR-X allowing it to interact with the cell membrane.

Figure 4.6.1, B shows the colour of the different pellets after washing the external fluorophore analogues off. The blank cultures with nothing added are a normal cream colour, the two supplemented with BODIPY® TR-X alone both show the same blue tinge regardless of whether the transporters are present or not, this is probably due to the hydrophobic BODIPY® TR-X sticking in the bacterial membrane. In contrast, the BODIPY® TR-X ribose cobalamin pellet is dark blue in the OP50 which were induced, but a similar cream colour to the blanks without induction. The native copies of BtuB and BtuF are present in the OP50 proteasome so there should be some BODIPY® TR-X ribose cobalamin inside these cells. The C5- BODIPY® TR-X cobyric acid cultures looked like the BODIPY® TR-X ribose cobalamin (data not shown).

The fluorescence microscopy images show that the C5-BODIPY® TR-X cobyric acid is taken up into the OP50 when the transport proteins are overproduced (Figure 4.6.2, A). The ribose linked BODIPY® TR-X cobalamin is also taken up in the same conditions (Figure 4.6.2, B). As there is no difference between the cobalamin analogue and the C5-cobyric acid analogue this suggests that the position of the fluorophores does not impede uptake or recognition. It also proves that *E. coli* cobalamin uptake proteins can scavenge cobyric acid even though they cannot use it (Raux et al. 1996). However, in the un-induced cultures (Figure 4.6.2, D and E), neither of the corrinoid analogues were taken up well, with only a small percentage of cells fluorescing enough to be visualised. The cultures which were supplied with only the BODIPY® TR-X fluorophore showed little difference between induced and un-induced backgrounds (Figure 4.6.2, C and F). The slight fluorescence seen in the BODIPY® TR-X samples is likely to be due to it sticking in the membrane, or reacting with outer membrane proteins, and not washing off.

The three induced backgrounds shown in Figure 4.6.2: C5- BODIPY® TR-X cobyric acid; BODIPY® TR-X ribose and BODIPY® TR-X fluorophore only were the three used as *C. elegans* food in the cobalamin uptake experiments in *C. elegans* reported in the next section.

Figure 4.6.2: Cobalamin ribose linked BODIPY® TR-X and C5- BODIPY® TR-X cobyric acid internalised by OP50 *E. coli* expressing *btuB* and *btuF*. A: OP50 pET-BAD-*btuB btuF* incubated with C5- BODIPY® TR-X cobyric acid. B is OP50 pET-BAD-*btuB btuF* incubated with BODIPY® TR-X ribose cobalamin. C is OP50 pET-BAD-*btuB btuF* incubated with BODIPY® TR-X. The lower panels are the uninduced cultures. D is C5- BODIPY® TR-X cobyric acid, E is BODIPY® TR-X ribose cobalamin and **F** is BODIPY® TR-X.

4.7 Recognition and localisation of C5- BODIPY® TR-X cobyric acid and BODIPY® TR-X ribose cobalamin in *Caenorhabditis elegans*

OP50 *E. coli* saturated with the BODIPY® TR-X analogues were used to seed nematode growth medium (NGM) plates (Chapter 2, 2.2.2). The OP50 was transformed with pET-BAD-*btuBF* and grown in 5 mL cultures with 1 µM of BODIPY® TR-X analogue added to the media. The plasmid was induced with *L*-arabinose and incubated overnight at 28 °C in the dark. The following day the cells were pelleted and resuspended in 1 mL of fresh LB, then pelleted again and so on until the cells had been resuspended three times in fresh LB. The final pellet was resuspended in 1 mL of fresh LB and 200 µL of the OP50 culture was pipetted on to the centre of NGM agar plates. These plates were left to dry in a sterile culture hood for 4 hours and then stored at 4 °C until used. Three L4 nematodes of the N2 Bristol strain of *C. elegans* were transferred on to plates seeded with OP50 grown with either C5-BODIPY® TR-X cobyric acid, BODIPY® TR-X ribose cobyric acid, or BODIPY® TR-X only. These plates were left for four days at 20 °C and the L4 *C. elegans* were imaged on a confocal microscope. The microscope used in this experiment was the Leica SP8 (Chapter 2, 2.6.5) at Bristol University. The gut autofluorescence of *C. elegans*, excitation 405 nm, emission 410-505 nm, was used to orientate the nematode (Coburn et al. 2013). All of the images were taken using a 20 x lens at 1 x zoom for the whole nematode, and 2.5 x zoom for the more detailed images. They were 1024 x 1024 pixels and then further zoomed to 500 x 500 pixels in the 2.5 x zoomed images shown below.

The BODIPY® TR-X ribose cobalamin (Figure 4.7.1) fluorescence is localised to discrete spots in the head, vulva and tail regions. These rings of fluorescence are not in the intestine which suggests that the analogues have been taken up and specifically transported into these fluorescent regions. The C5- BODIPY® TR-X cobyric acid did not show much fluorescence, only a scant glow at the end of the intestine (Figure 4.7.2). The same is true of the BODIPY® TR-X on its own, although this can be seen in the intestinal granules near the head (Figure 4.7.3). This proves that the *C. elegans* cobalamin uptake protein from the intestine to the pseudocoelem (discussion Figure 4.9.1) cannot scavenge a cobyric acid intermediate. In Figure 4.7.4 these fluorescent areas were expanded in all three conditions to show a direct comparison. These six cells to which the BODIPY® TR-X ribose cobalamin associates are thought to be the coelomocytes. There is a pair in the head, tail and vulva regions. The coelomocytes 3 and 4 in panel A clearly show a ring like structure consisting of many smaller spherical compartments. What these compartments correspond to is unknown.

Figure 4.7.1: *C. elegans* **given OP50 transformed with pET-BAD-*btuB btuF* provided with BODIPY® TR-X ribose cobalamin at 1 x zoom.** The autofluorescence is shown in cyan and the BODIPY® TR-X ribose cobalamin analogue is shown in red. The red fluorescence localises to discrete cells in three areas behind the head. tail and near the vulva.

Figure 4.7.2: *C. elegans* **given OP50 transformed with pET-BAD-*btuB btuF* provided with C5- BODIPY® TR-X cobyric acid at 1 x zoom.** The autofluorescence is shown in cyan and the BODIPY® TR-X ribose cobalamin analogue is shown in red. The red fluorescence is only present at the top of the gut where there is a high concentration of *E.coli*, but nowhere in the *C. elegans*. Therefore is appears that the *C. elegans* cobalamin uptake machinery does not recognise cobyric acid.

Figure 4.7.3: *C. elegans* **given OP50 transformed with pET-BAD-*btuB btuF* provided with BODIPY® TR-X only at 1 x zoom.** The autofluorescence is shown in cyan and the BODIPY® TR-X ribose cobalamin analogue is shown in red. There is no fluorescence for BODIPY® TR-X so the uptake seen in figure 4.7.1 is due to the presence of cobalamin.
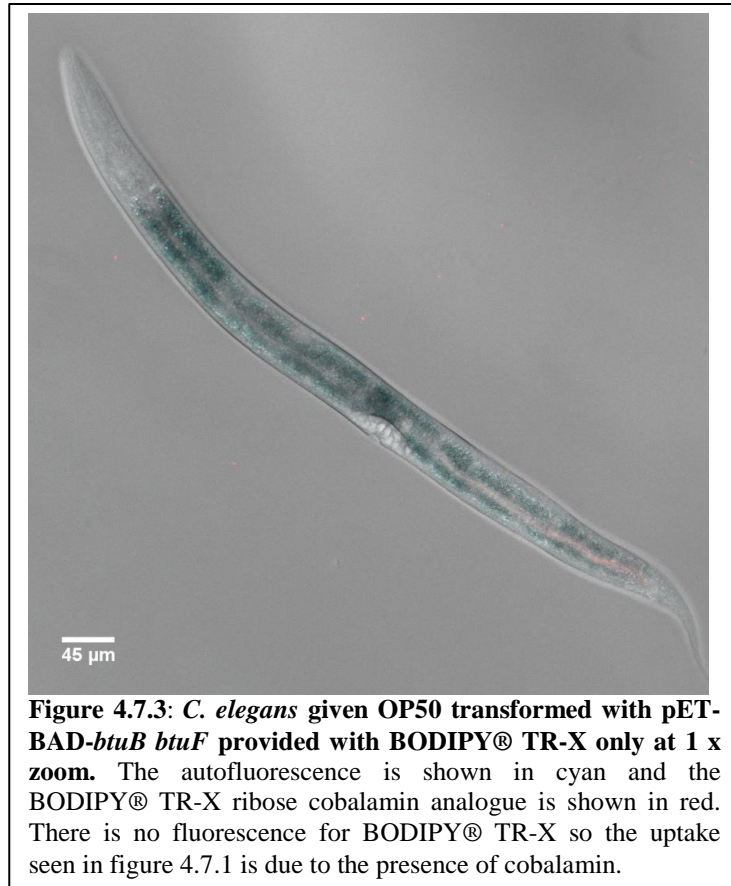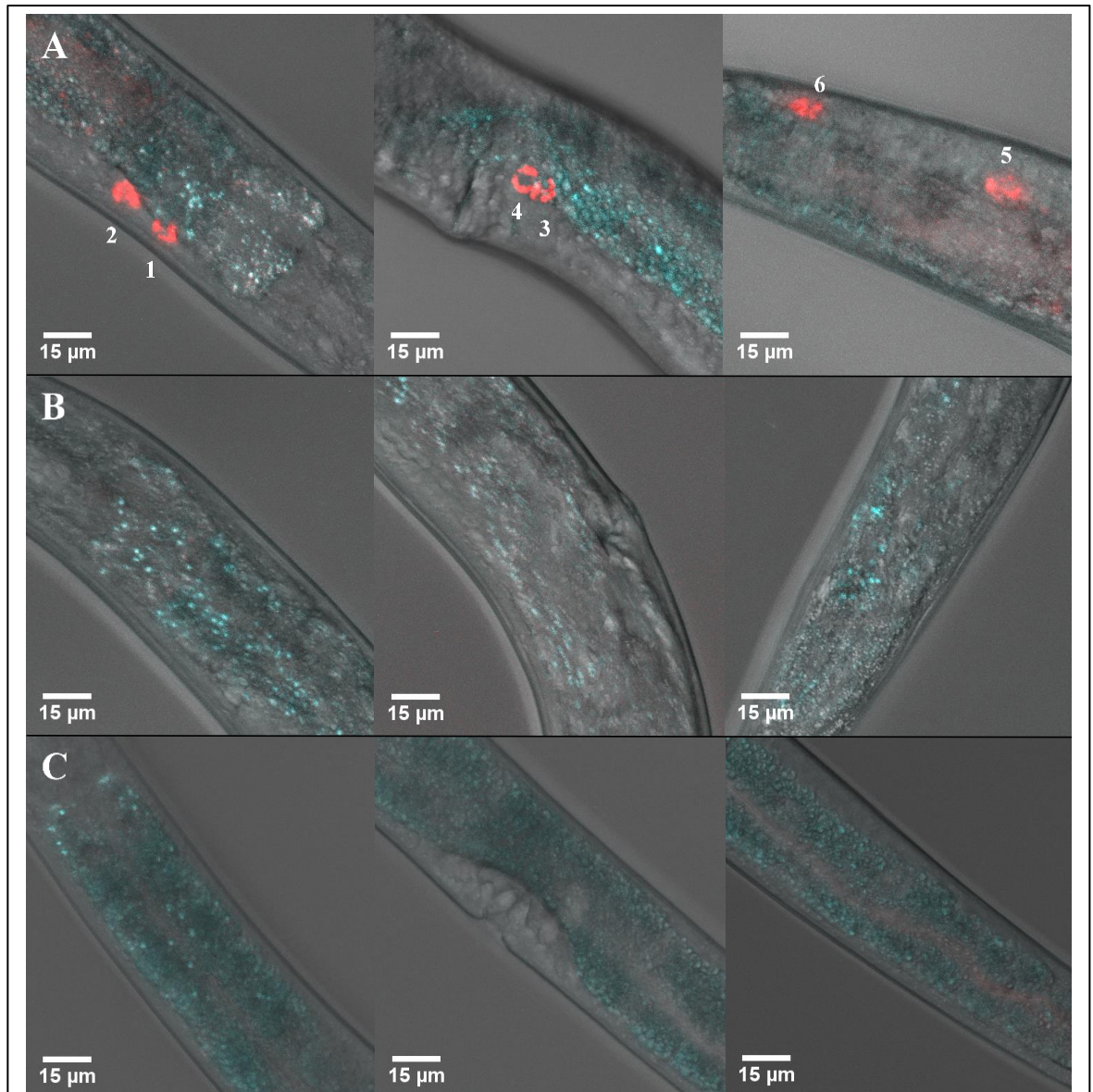
Figure 4.7.4: *C. elegans* **fed with OP50 containing either BODIPY® TR-X ribose linked cobalamin, C5- BODIPY® TR-X cobyric acid, or BODIPY® TR-X zoomed in on the coelomocytes.** Panel **A** shows the BODIPY® TR-X ribose linked cobalamin taken up into the coelomocytes (labelled 1 to 6) from the gut into discrete structures within the coelomocytes, what these compartments are is unknown. Panel **B** is C5- BODIPY® TR-X cobyric acid and **C** is BODIPY® TR-X on its own. Neither of these conditions shows any internal BODIPY® TR-X fluorescence proving that the *C. elegans* cobalamin uptake machinery does not recognise cobyric acid. The three panels from left to right show the head, vulva and tail regions respectively. These are all zoomed in regions of the nematodes shown in Figures 4.7.1-3, but some of the images have been rotated. The autofluorescence of the gut granules is shown in cyan and the BODIPY® TR-X fluorescence in red.

4.7.1 Investigating the persistence of the BODIPY® TR-X cobalamin analogue in

Caenorhabditis elegans coelomocytes

In order to test how persistent the fluorescence is in the coelomocytes, nematodes grown on BODIPY® TR-X ribose cobalamin were transferred to unsupplemented OP50 plates and imaged at intervals. There is very little time for the L4 nematodes to grow on the

control plates before they mature into adults and become harder to image, therefore L3 worms were also transferred. However, after 22 hours even they were adults, which is evident in Figure 4.7.1.1. These images were taken on a Leica DM R fluorescence microscope using 515-560 nm (N2-1) excitation and 590 nm emission filters for the BODIPY® TR-X fluorescence, and 456-490 nm (I3) excitation and 515 nm emission filters for autofluorescence.

The BODIPY® TR-X fluorescence is still present in all six coelomocytes. These are fluorescence microscope images so there is no *z* dimension to take an image stack and condense them together to get a clearer image like the confocal. This also means that the intensity of the signal is reduced to that of one image and less detail is apparent when compared to Figure 4.7.4. However it is evident that the fluorescence in the coelomocytes is still present. In Figure 4.7.1.1 eggs can be seen both inside and outside the nematode where they were laid prior to being picked for imaging. No fluorescence is observed in these eggs.
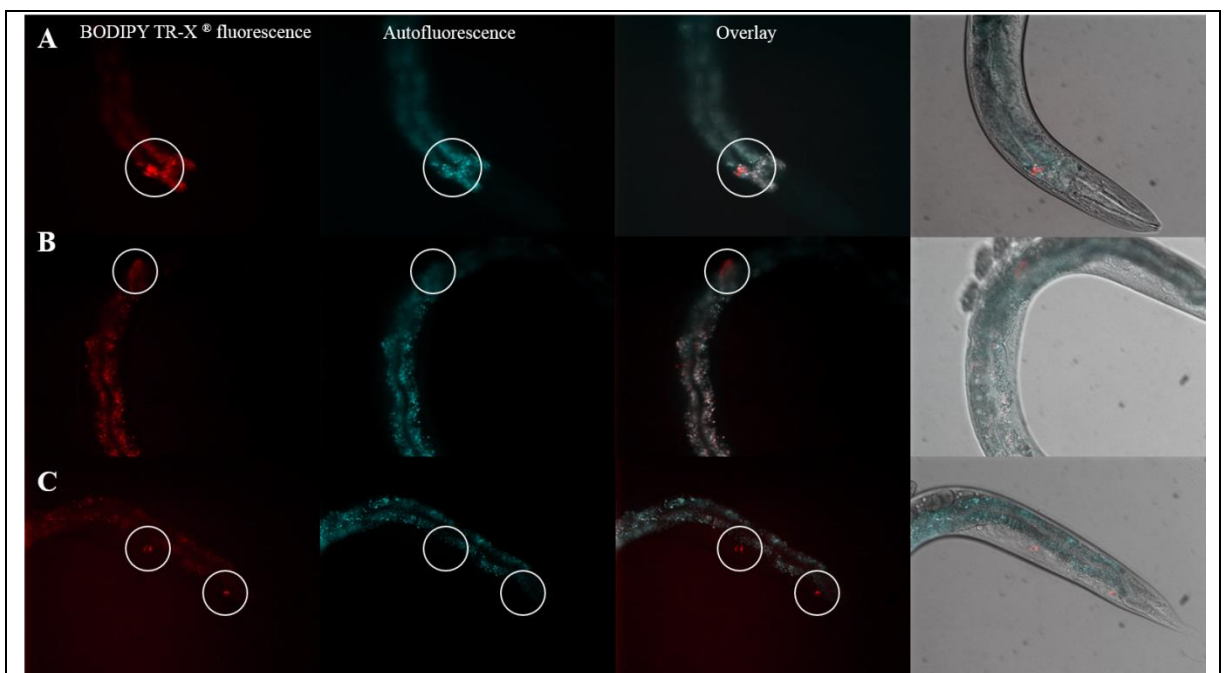
Figure 4.7.1.1: Persistence of fluorescence in the coelomocytes. A: The head of worm 1 after being fed un-supplemented OP50 *E. coli* for 22 hours, shows the coelomocytes, circled in white, at the side of the head are still fluorescing, so the analogues have not been hydrolysed. The vulva coelomocytes are harder to see due to the presence of the ova, but there is some fluorescence seen in the overlay which corresponds to the coelomocytes. The pair of coelomocytes in the tail are visible too, but could be slightly diminished in intensity. It is hard to draw any conclusions about the depletion in fluorescence as the images are a single image and not a projection through *z* as the previous images have been. The red of the coelomocytes are not associated with the intestine autofluorescence and do not co-localise. The two fluorescences overlays are shown in white, but the coelomocytes only appear red where the BODIPY TR-X® fluorophore is present and are therefore not caused by autofluorescence of the gut. These images were taken at 20 x zoom with 300 ms exposure.

4.8 Investigating C5-Oregon green® 514 cobyric acid and ribose linked cobalamin analogue uptake in *Lepidium sativum* and *Arabidopsis thaliana*

In light of suggestions that plants can be cobalamin enriched by fertilising with cobalamin containing manure, the uptake of cobalamin into *L. sativum*, garden cress, was tested as part of an outreach programme at a local school arranged by the Warren Lab (Watanabe et al. 2013). This work was collated by Dr. Evelyne Deery.

4.8.1 Cobalamin enrichment in *Lepidium sativum*

The garden cress seeds were sterilised in 10 % (v/v) bleach for 20 minutes, then spun and the bleach was pipetted off. The seeds were further washed in 70 % (v/v) ethanol-water, and spun before removing the liquid. This ethanol wash was repeated three times. The seeds were placed on Murashige and Skoog media containing increasing concentrations of cobalamin and grown for one week. Four conditions were used: no cobalamin, 0.1 mg L^{-1}, 1 mg L^{-1}, and 10 mg L^{-1}. After one week's growth, the cotyledons of two plants were taken from each condition and washed 5 times with water. Once the water was removed, P-Per (Thermo Fisher) was added to the leaves along with sand, and the mixture was ground for 2 minutes. This was centrifuged for 3 minutes at 15000 rpm and the lower aqueous phase was removed as the sample. The samples were applied to the bioassay plate in 10 µL drops (Figure 4.8.1.1) and the plate was left overnight at 37 °C.

Figure 4.8.1.1: Bioassay plate of extracts from the cress grown with different concentrations of cobalamin supplement. The cress appears to take up the cobalamin in a concentration dependent manner as the concentration of cobalamin supplement increases.

Table 4.8.1.2: Cobalamin content of *Lepidium sativum* cotyledons

Cobalamin concentration supplied (mg L^{-1})	Weight after the wash and spin (mg)	Diameter of bioassay growth (cm)	Cobalamin in leaves in (pg mg^{-1})
0	23	0	0
0.1	36	0.8	3.3
1	27	1.125	30.9
10	23	1.75	712.5

The diameter of growth on the bioassay plate was used to determine the concentration of cobalamin in each extract. This was calculated using a cobalamin standard curve in the same way as the *S. entrica* plates in Section 4.3. The concentration was converted into picogrammes and the uptake of cobalamin was calculated per milligramme of cress. Cobalamin uptake into the cotyledons increased in a concentration dependent manner with increased cobalamin in the media. A 10 fold increase from 0.1 mg mL^{-1} to 1 mg mL^{-1} resulted in a near-equivalent increase of 3.3 pg mg^{-1} to 30. 9 pg mg^{-1} in the cotyledons. However, the 10 mg mL^{-1} sample contained around 20 times the cobalamin in the cotyledons compared to the 1 mg mL^{-1} sample. The results were plotted on the Graph 4.8.1.3 on log-log scales which gives a straight line of best-fit. This reiterates that the concentration of cobalamin in the cotyledons is proportional to the cobalamin added in the medium.
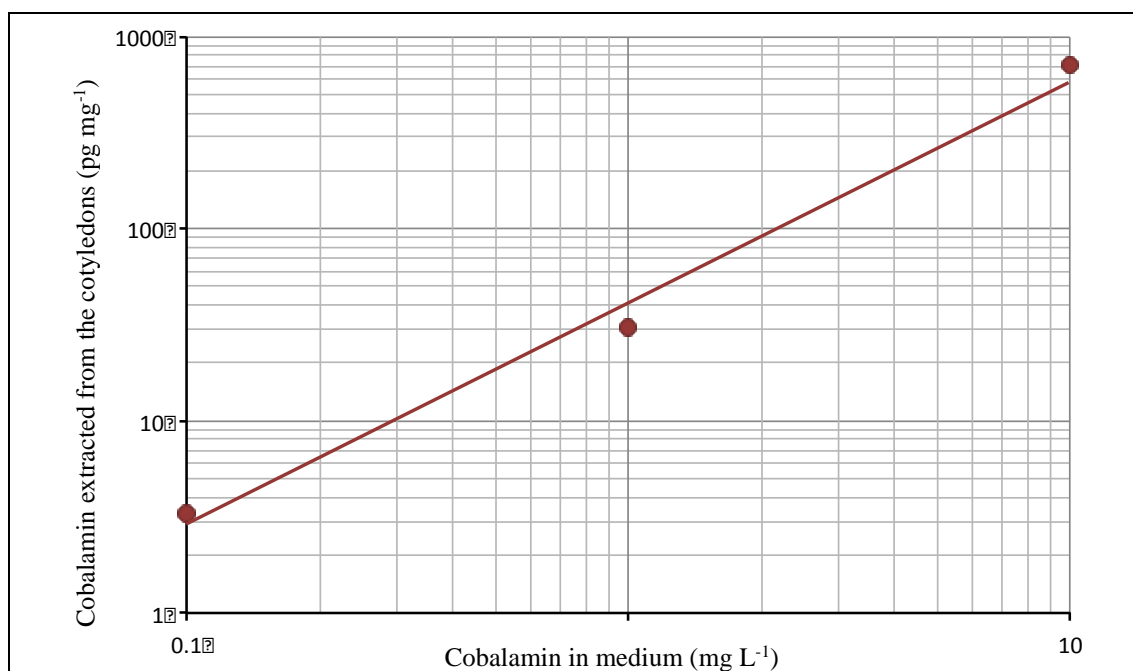
Graph 4.8.1.3: **Graph plotting the cobalamin content of the *L. sativum* cotyledons against the concentration of cobalamin provided in the media.** The graph shows that the concentration of cobalamin in the cotyledons is proportional to the cobalamin added in the medium.

4.8.2 Uptake of Oregon green® 514 analogues in *Lepium sativum* and *Arabidopsis thaliana*

The previous experiment shows that cobalamin is taken into the cotyledons. To confirm this, and to see if it is accumulated anywhere else in the plant, *L. sativum* was grown with Oregon green® 514 ribose cobalamin and then imaged. The greatest issue with imaging plants is the chlorophyll autofluorescence (Figure 4.8.2.1). Oregon green® 514 is excited at 514 nm and emits at 530 nm (see Chapter 2, 2.2.2). This is in the 'dark region' of chlorophyll fluorescence which allows the Oregon green® 514 to be imaged without too much bleed through from the chlorophyll.

The seeds of the plants were sterilised in 10 % (v/v) domestic bleach for 20 minutes. The bleach was removed with three 70 % ethanol washes followed by the same in distilled water. These were placed on Murashige and Skoog media (Chapter 2, 2.2.2) supplemented with 0.5 µM of either the Oregon green® 514 ribose linked cobalamin or cobalamin and dried around a flame. The glass pots used for the *L. sativum* were placed in the dark at room temperature for 5 days. *A. thaliana* requires vernalisation so after the seeds were washed and placed on the agar, the *A. thaliana* petri dishes were wrapped in aluminium foil and left in the fridge for at least 4 days. They were then exposed to a bright light for 30 minutes to synchronise growth. Only after this were they grown.

Figure 4.8.2.1: **The absorption and emission spectra of chlorophyll.** This graph was reproduced with permission from Leiper 2010.
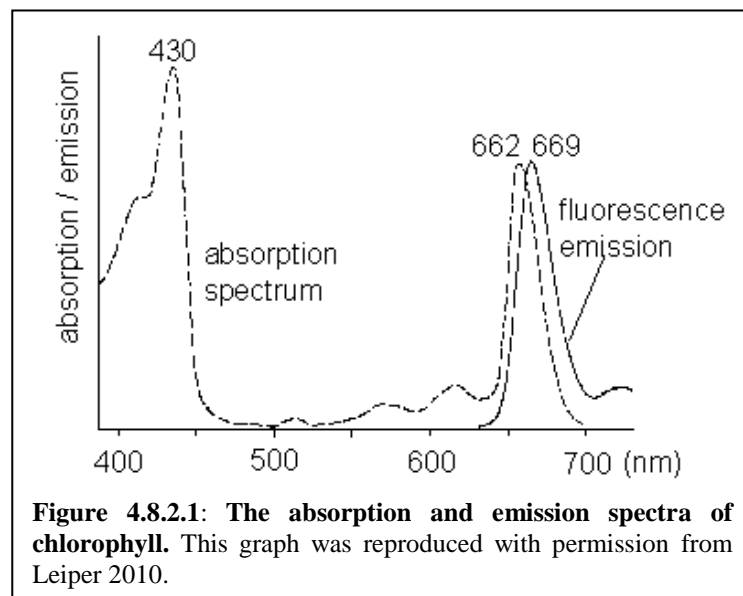
On the fifth day of growth the plants were harvested and imaged using Leica TCS SP2 confocal microscope (see Chapter 2, 2.6.4). The excitation was at 514 nm and emission was detected between 525–590 nm. The *A. thaliana* were imaged as whole plants but the *L.*

sativum were sectioned at the root, hypocotyl and cotyledon to increase light penetration. No part of the *A. thaliana* plants took up Oregon green® 514 ribose linked cobalamin (data not shown). However, *L. sativum* did (Figure 4.4.2.2, A and C). The Oregon green® 514 ribose linked cobalamin is localising to the vacuoles of the cotyledons (Figure 4.4.2.2, C) which cannot be seen in the *L. sativum* when provided with cobalamin only (Figure 4.4.2.2, D). The roots did not show any difference in fluorescence between the Oregon green® 514 ribose linked cobalamin and the cobalamin sample (Figure 4.8.2.2, A and C). This shows that the cobalamin analogue is deposited in the cotyledons and not in any other part of the seedlings.

The C5-Oregon green® 514 cobyric acid analogue also appears to be taken up into the cotyledons (Figure 4.8.2.2, E). However, these images were taken on a different microscope (Chapter 2, 2.6.4) with slightly different parameters and should be repeated to confirm that the fluorescence is present in the vacuoles and is not due to bacterial contamination.
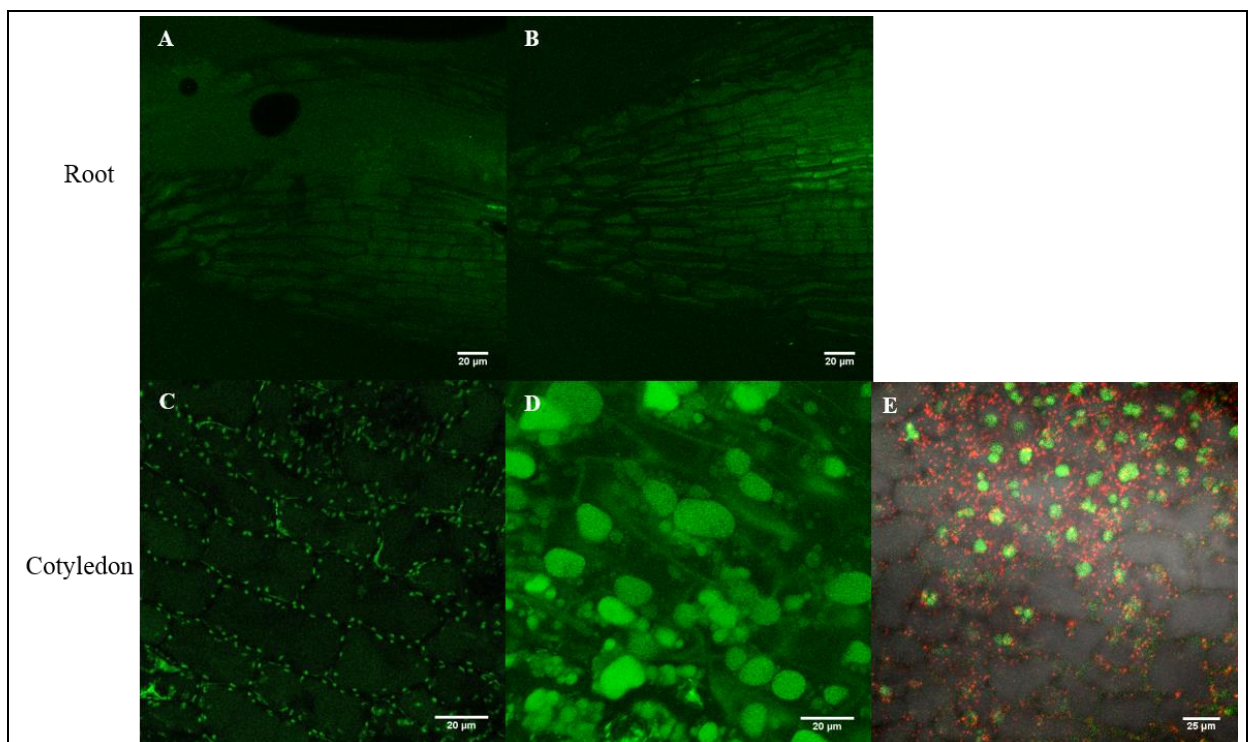
Figure 4.8.2.2: Confocal microscope images of Oregon green® 514 linked analogue uptake in *L. sativum* cotyledon and root. All of the images shown are in *L. sativum*. From left to right: The first column has cobalamin added to the media, the second has Oregon green 514® ribose linked cobalamin and the last column has C5-Oregon green 514® cobyric acid. There is no great difference between the Oregon green® 514 B_{12} and the B_{12} in the root tip but there is a noticeable difference in the cotyledons. The Oregon green® 514 cobalamin fluorescence is clearly localising in the vacuoles of the cotyledons (D) and is not present in the cobalamin (vitamin B_{12}) sample (C). The cotyledons of the seedlings given C5-Oregon green 514® cobyric acid (E) show some fluorescence (green) which is within the ring of chloroplast fluorescence (red) which suggests it is also localising to the vacuoles of the cotyledons. However, this experiment was conducted on a different microscope and needs to be repeated.

4.9 Discussion

The research reported in this chapter shows that fluorescent analogues of cobalamin can be taken up into a range of organisms. Analysis of the cobalamin-dependent *S. enterica* AR3612 bioassay plates demonstrate that the strain can be rescued by the C5 and ribose analogues, therefore coenzyme synthesis can be completed so as to facilitate growth on a bioassay plate (Figure 4.3.1). However, this growth is poor compared to the cobalamin standard which indicates that the presence of the fluorophore is causing a disruption to the natural cobalamin metabolism *in vivo*. There is a possibility that the linker joining the corrinoid to the fluorophore can be hydrolysed *in vivo* but the difference in efficacy between the two fluorophore analogues suggest that it is the fluorophore which has caused the varied reduction in growth. This infers that the fluorophore is still present on the analogue and has not been cleaved.

E. coli also shows clear uptake and localisation to the cells (Figure 4.5.1 and 4.6.2) of all the analogues, regardless of which fluorophore is used or where it is conjugated. This confirms that *E. coli* can recognise cobyric acid and that the position of the fluorophore on the corrin ring does not halt uptake.

In *C. elegans*, which requires cobalamin, the ribose linked cobalamin analogues localised to six distinct cells thought to be the coelomocytes. The function of the coelomocytes has not been fully defined but they are thought to be involved in endocytosis and coupled with lysosomal degradation (Treusch et al. 2004; Fares and Grant 2002). The coelomocytes are not essential as the *C. elegans* still reproduce after toxin ablation of them all (Fares and Greenwald 2001). There is no known cobalamin related reason for the BODIPY® TR-X ribose cobalamin to localise to here, but little is known about cobalamin storage and transport in *C. elegans*, and it could be that this is where cobalamin is sequestered until required. Alternatively, after the initial uptake of the analogue, the next transport protein recognition could be blocked by the presence of the BODIPY® TR-X on the ribose so the analogue remains in the pseudocoelom until the coelomocytes endocytose it on the basis it is an unnatural substance to the *C. elegans*. Although the fluorescence in the coelomocytes is not affected within 22 hours, it gives no insight into the state of the cobalamin analogue: the BODIPY® TR-X may have been cleaved off. Also it is unclear whether the vesicles containing the BODIPY® TR-X ribose linked cobalamin within the coelomocyte are lysosomes or storage vesicles. In either case, the initial intestinal lumen to pseudocoelom

(Figure 4.9.1) transporter recognises the BODIPY® TR-X ribose cobalamin analogue. This experiment also proves that *C. elegans* cannot recognise cobyric acid. *C. elegans* have no cobalamin biosynthesis genes and cannot complete cofactor synthesis from earlier intermediates. It is therefore unlikely that the uptake proteins will recognise any early intermediates of cobalamin synthesis.
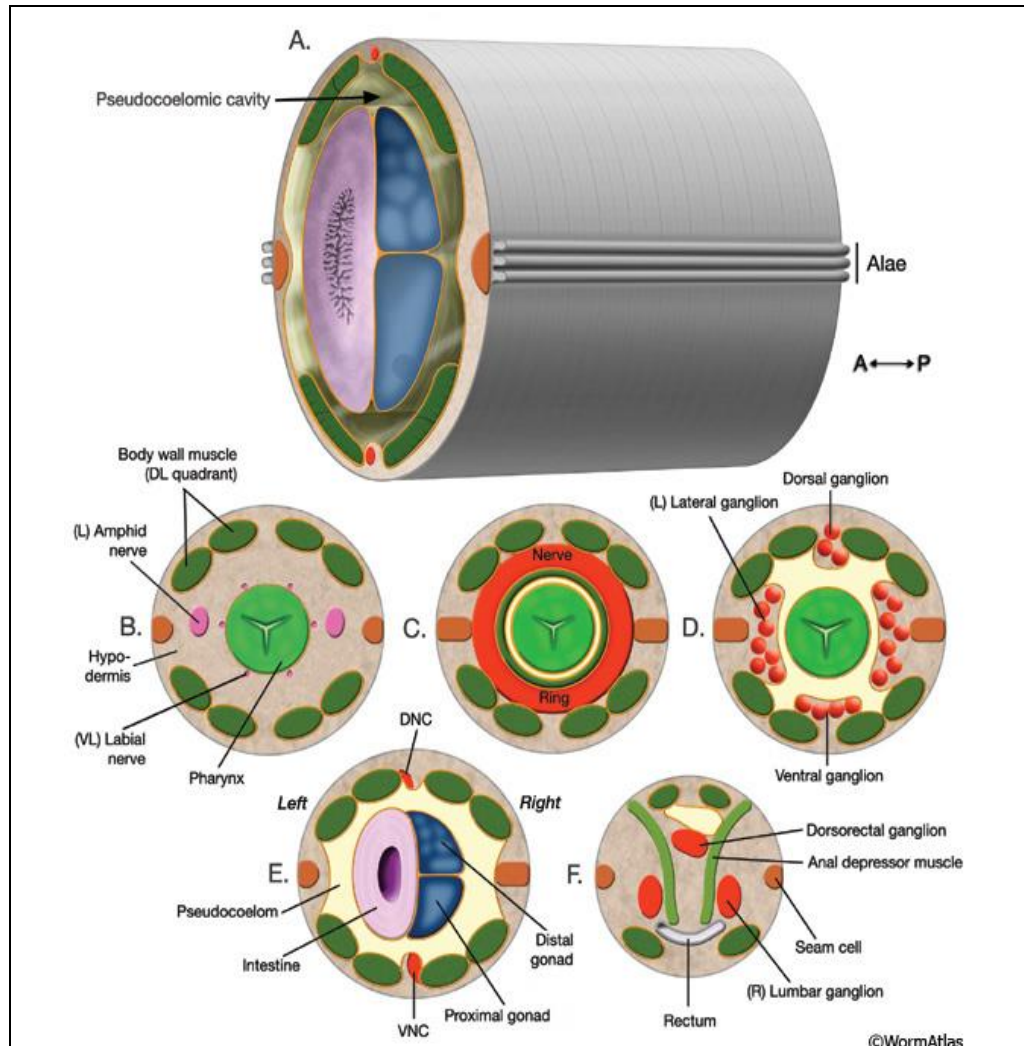
Figure 4.9.1: Sections through *C. elegans*.
A. Posterior body region. Body wall (outer tube) is separated from the inner tube (alimentary system, gonad) by a pseudocoelom.
B. Section through anterior head. The narrow space between the pharynx and the surrounding tissues anterior to the NR (Nerve Ring) can be considered an accessory pseudocoelom because the main pseudocoelom is sealed off at the NR level.
C. Section through the middle of head.
D. Section through posterior head.
E. Section through posterior body. Dorsal nerve cord (DNC); ventral nerve cord (VNC).
F. Section through tail, rectum area.

The BODIPY TR-X® ribose linked cobalamin analogue must be taken up from the intestine into the pseudocoelemic cavity through a cobalamin-dependent transport protein which is currently unidentified.
This figure was reproduced with permission from WormAtlas ©(Altun and Hall 2009b)

Even though plants do not require cobalamin, *L. sativum* showed uptake of the cobalamin in the cotyledons in a concentration dependent manner. Oregon green 514® cobalamin analogues were also taken up into the cotyledons where it was stored. This transport and storage of cobalamin in plants could be exploited to produce plants with a higher nutritional value.

The lack of uptake in the *A. thaliana* compared to the *L. sativum* suggests that, although plants do not require cobalamin and are therefore unlikely to have specific uptake mechanisms, there appears to be different uptake processes for cobalamin. These are not likely to be cobalamin specific but may regulate macromolecule traffic in some plants.

Chapter 5

Cobalamin and *Mycobacterium tuberculosis*

5.0 Introduction

Little is known about the relationship between *M. tuberculosis* and cobalamin. It was only relatively recently that the cobalamin membrane transporter, BacA, was identified (Gopinath, Venclovas, et al. 2013). Bioinformatics analyses has shown that *M. tuberculosis* has cobalamin biosynthetic enzyme candidates for almost the whole pathway (Rodionov et al. 2003). However, it is still not known whether *M. tuberculosis* is a facultative scavenger of cobalamin, if it can synthesise cobalamin *de novo*, or scavenge earlier intermediates and complete cofactor synthesis. The genome of *M. tuberculosis* encodes for three cobalamin-dependent enzymes: MetH, RNR-II, and MCM, none of which appear to be essential (Table 5.0.1) (see Chapter 1, Section 1.7) (Gopinath, Moosa, et al. 2013).

Table 5.0.1: Cobalamin dependent enzymes of *M. tuberculosis* and their independent alternatives

Protein	Gene	Cobalamin independent alternative
Methionine synthase	*metH*	*metE*
Ribonucleotide reductase class II	*nrdZ*	Ribonucleotide reductase class I
Methyl malonyl coenzymeA mutase	*mutAB*	Methylcitrate pathway

It has been reported that the expression of the adenosyl-cobalamin requiring RNR-II, encoded by *nrdZ,* is upregulated during dormancy (Gopinath, Moosa, et al. 2013; Boshoff and Barry 2005). However, it is not known whether this is for chromosomal integrity upkeep or DNA repair upon exit from non-replicating persistence 2 (NRP2) dormancy. NRP2 is the second stage of dormancy where oxygen levels are at about 0.06 % saturation. Exit of this stage is termed "disease reactivation". When *M. tuberculosis* is replicating in a low oxygen environment it is reliant on RNR-II enzymes for the production of dNTPs. This upregulation of the adenosyl-cobalamin dependent RNR-II is potentially exploitable as a means of shepherding cobalamin conjugated drugs into *M. tuberculosis* either at the reactivation point or even during dormancy.

In order to understand better the relationship between *M. tuberculosis* and cobalamin there is a need to study the roles of these enzymes and their effect on *M. tuberculosis*.

The aims set out for this chapter were:

1. To determine if cobalamin analogues, cobalamin itself or intermediates of cobalamin biosynthesis are exogenously taken up by *M. tuberculosis*.

2. To ascertain which form of cobalamin *M. tuberculosis* makes, and if it can synthesise it from an earlier intermediate.

3. To investigate if corrinoids regulate genes via the cobalamin riboswitch.

4. To identify any cobalamin binding proteins by purification on a cobalamin column.

Figure 5.0.2: Immunological lifecycle of tuberculosis infection in humans. The period of latency during the immunological equilibrium is where the low levels of oxygen can occur. Exit of this phase into the reactivation stage is thought to be the point at which *nrdZ* will be most active, providing a route of access for cobalamin conjugated drugs. This figure was reproduced with permission from The immunological life cycle of tuberculosis by Ernst 2012.

5.1 Results

5.2 Cobalamin and other Corrinoid uptake in *Mycobacterium tuberculosis*

5.2.1 Cobalamin and Cobinamide uptake

The aims of this experiment are two-fold: The first data required is the confirmation that cobalamin can be exogenously taken up in detectable quantities, and to ascertain whether

the BacA cobalamin transporter is also responsible for transport of earlier intermediates as well. For this three different strains of *M. tuberculosis* were chosen to test for uptake (referred to by the underlined lettering): Wild type (WT) ATCC 2729Y *M. tuberculosis* (clinically isolated in 1934), H37Rv ATCC *M. tuberculosis* (clinically isolated in 1998), and *ΔbacA*, a *bacA* deletion of H37Rv. These strains were incubated with 0 nM, 100 nM, 500 nM, 1 µM, and 10 µM of either cobalamin or cobinamide for five days at 37 ℃ in 50 mL cultures. After this the cultures were spun, washed with fresh 7H9 growth media, lysed, spun again and 10 µL of the supernatant was plated on *S. enterica* AR3612 containing minimal media plates (Chapter 2, 2.6.1 and Chapter 3, 3.4.6) (Raux et al. 1996). The three *M. tuberculosis* strain samples, with the same concentration of exogenous corrinoid, were plated on the same bioassay plate with an extra 50 nM reference drop of commercial cobalamin or cobinamide where appropriate. Dots of 10 nM, 50 nM, 100 nM and 1 µM of the relevant commercial corrinoid were dropped on another plate. These plates were then incubated overnight at 37 ℃ to allow the *S. enterica* to grow.

The standards plate was used to calculate a standard curve of the diameter of growth against the concentration of corrinoid provided (as peformed in Chapter 4, Section 4.3). The concentration of the corrinoids detected for each culture in the three different strains were normalised to the 50 nM reference point on each plate, and compared once the repeats were averaged (Graphs 5.2.1.1 and 5.2.1.2).

The *ΔbacA* strain shows severe impairment of cobalamin transport (Graph 5.2.1.1.). The uptake is between 3 and 5 times worse in *ΔbacA* apart from the culture provided with 10 µM of cobalamin were it is closer to 10 times poorer. This could indicate that there is another mode of uptake, perhaps promiscuous transport or diffusion, but its efficiency is limited to just over 20 µM of cobalamin over 5 days.

In contrast, the uptake of cobinamide (Graph 5.2.1.2) is lower across all three strains when fortified with 10 µM of cobinamide but equivalent at lower supplement concentrations compared to the cobalamin uptake. The adverse effects of the *bacA* deletion are less pronounced in all the cultures which imply that there is another uptake mechanism for cobinamide, independent of BacA. However, the loss of BacA does hamper cobinamide uptake more in the 10 µM supplemented culture. This increase in impairment is also true in the cobalamin sample and could indicate that the alternative route of uptake, postulated previously, is negatively regulated by the concentration of corrinoid in the cell. Overall, the concentration of cobalamin taken up in the 10 µM culture of the *ΔbacA* strain is less than that

of cobinamide which suggests that either there is a preference for cobinamide over cobalamin or that the negative feedback is more sensitive to cobalamin levels (please note that Graphs 5.2.1. 1 and 5.2.1.2 are on different scales). All three of the tested *M. tuberculosis* strains can take up cobalamin and cobinamide in a concentration dependent manner.

There are no previous reports concerning cobinamide uptake into *M. tuberculosis*, but this experiment clearly demonstrates that both cobalamin and cobinamide can be exogenously taken up into *M. tuberculosis* in a concentration dependent manner. It also shows that BacA may not be the sole route of cobalamin transport, as both cobalamin and cobinamide are taken up in the *ΔbacA* strain. This could indicate the presence of an unknown corrinoid transport module in *M. tuberculosis*.

Graph 5.2.1.1: The concentration of corrinoid detected in three different strains of *M. tuberculosis* when provided with increasing concentrations of cobalamin in culture. The uptake increases in a concentration dependent manner across all three strains even though the cobalamin uptake in *ΔbacA* is 10 times worse when supplemented with 10 µM of cobalamin.

Graph 5.2.1.2: The concentration of corrinoid detected in three different strains of *M. tuberculosis* when provided with increasing concentrations of cobinamide in culture. The uptake increases in a concentration dependent manner across all three strains the same as with cobalamin, but the loss of BacA has less impact, perhaps due to the cobinamide uptake being uniformly worse across all strains. Interestingly, in both cases *ΔbacA* corrinoid uptake is most hindered when supplied with 10 µM of cobalamin or cobinamide which could be linked to a potential feedback mechanism in an alternative uptake route.

5.2.2 Pseudo-cobalamin

Pseudo-cobalamin is a naturally occurring analogue of cobalamin. Instead of a DMB as the lower axial ligand it has adenine (Figure 5.2.2.1) (Fieber et al. 2002). This difference results in a 500 times lower binding affinity for human IF compared to cobalamin. IF is one of the human cobalamin transport proteins responsible for ferrying cobalamin from the duodenum to the blood and it is this reduction in binding affinity that is thought to contribute to pseudo-cobalamin lack of function in humans (Nielsen et al. 2012; Taga et al. 2008). TCII and HC, two other human cobalamin transport proteins, do not have such a drastically low binding affinity for pseudo (Greibe, Fedosov, and Nexo 2012). In fact some bacteria, e.g. *Synechococcus* species and *Lactobacillus reuteri*, synthesise pseudo-cobalamin instead of cobalamin, whilst, more commonly, others can recognise pseudo-cobalamin and then swap the lower loop to the DMB version so cobalamin can be used as a cofactor e.g. *Pavlova lutherii*, *Chlamydomonas reinhardtii*, *S. enterica* and *Dehalococcoides mccartyi* (Helliwell et al. 2016; Santos et al. 2007; Yi et al. 2012).

This experiment establishes whether or not pseudo-cobalamin can be taken up in *M. tuberculosis*. The investigation was set up in the same way as the cobalamin and cobinamide, above, except that it was done in 20 mL rather than 50 mL and was only performed once.

Figure 5.2.2.1: **The chemical structure of pseudo-cobalamin.** The X denotes CN; Adenosyl or other upper axial ligands. The adenine, which differentiates it from cobalamin, is in the red circle.
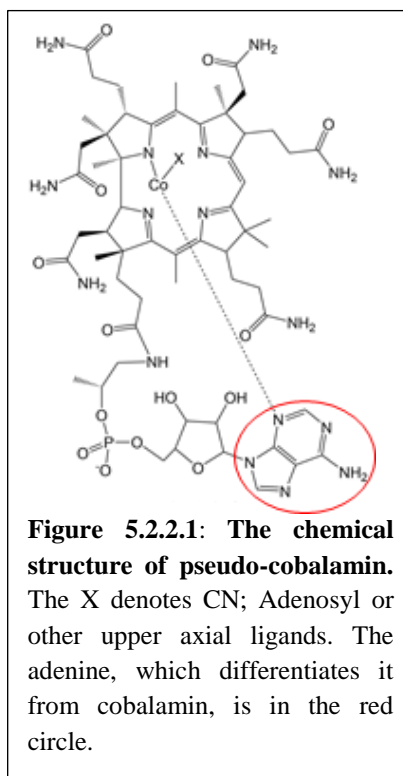
Graph 5.2.2.2 shows that pseudo-cobalamin can be taken up by all three strains of *M. tuberculosis*, and in a concentration dependent fashion in the WT and *ΔbacA* strains. This can be seen by the increase in the concentration of corrinoid detected when 1 µM and 10 µM of pseudo-cobalamin is in the media. In the WT strain there is a 3 to 4 fold increase each time the concentration of exogenous cobalamin doubles, whereas in *ΔbacA* it only increases by 50 %. It is impossible to know if the H37Rv strain increases uptake of pseudo-cobalamin in a concentration dependent manner as only the 10 µM culture extract enabled the *S. enterica* AR3612 to grow, and this may be due to contamination rather than uptake.

The decrease in concentration of corrinoid detected in the 10 µM culture of *ΔbacA* is about 12 times lower than WT or H37Rv strains, equivalent to the decrease seen in the cobalamin experiment. This suggests that BacA recognises the lower ligands of corrinoids, especially as its absence had less effect on the transport of cobinamide which has no lower loop. However, the concentration of pseudo-cobalamin detected in the *ΔbacA* strain with 10 µM exogenously added is about the same, around 50 nM, as the concentration detected in the equivalent cobinamide sample. This suggests that the possible BacA independent cobalamin

uptake route recognises pseudo-cobalamin as well as cobinamide, meaning that pseudo-cobalamin is taken up via both the BacA and the non-BacA mechanisms. If this alternative route is negatively regulated, it is o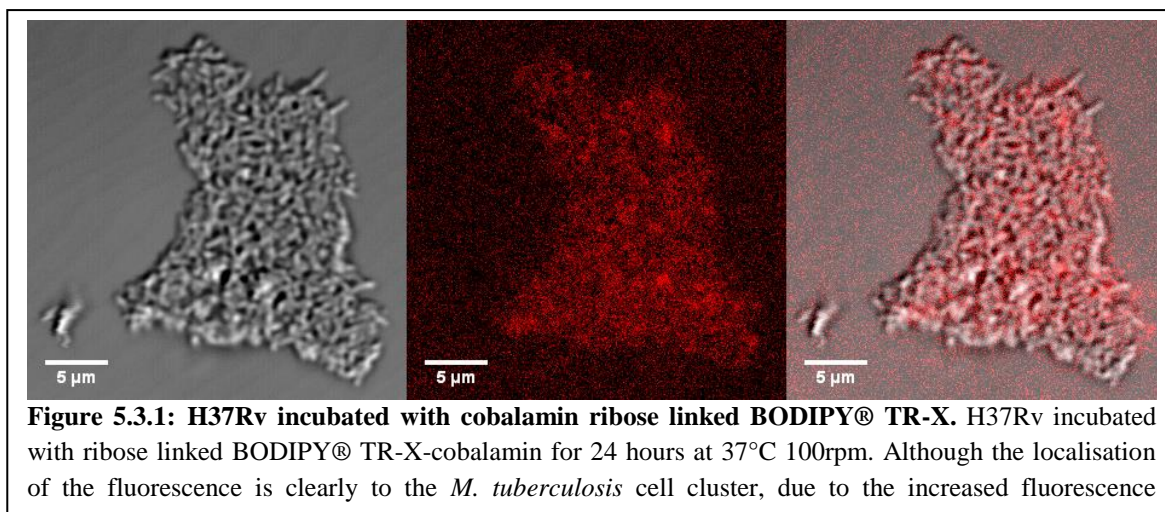nly done so by cobalamin and not by either pseudo-cobalamin or cobinamide. A graph comparing all three corrinoid uptake in all strains is in Appendix C.2.

Pseudo-cobalamin uptake in three strains when given different concnetrations of exogenous pseudo-cobalamin

■ WT Pseudo Cobalamin   ■ H37Rv Pseudo Cobalamin   □ BacA Pseudo Cobalamin

Graph 5.2.2.2: **The concentration of corrinoid detected in three different strains of *M. tuberculosis* when provided with the increasing concentrations of pseudo-cobalamin in culture.** The uptake increases in a concentration dependent manner across all three strains the same as with cobalamin, but the loss of BacA has less impact, perhaps due to the cobinamide uptake being uniformly worse across all strains. Interestingly, in both cases *ΔbacA* corrinoid uptake is most hindered when supplied with 10 μM of cobalamin or cobinamide which could be linked to a potential feedback mechanism in an alternative uptake route.

5.3 Fluorescent Corrinoid uptake and detection in *Mycobacterium tuberculosis*

In Chapter 3 two broadly different type of corrinoid analogue were synthesised: the ribose linked cobalamin analogues and the C5 linked cobyric acid analogues. A number of different fluorophores can be attached at these sites but for the purposes of this experiment the BODIPY® TR-X fluorophore analogues were used as they contrast with *M. tuberculosis* autofluorescence (detectable in the cyan range: excitation 405 nm; emission, 475 nm) better than the other fluorophore analogues made (Patiño et al. 2008). Both the C5 and ribose linked BODIPY® TR-X analogues were incubated with *M. tuberculosis* to ensure that they were still taken with the fluorophore attached. In this way the uptake can be visualised.
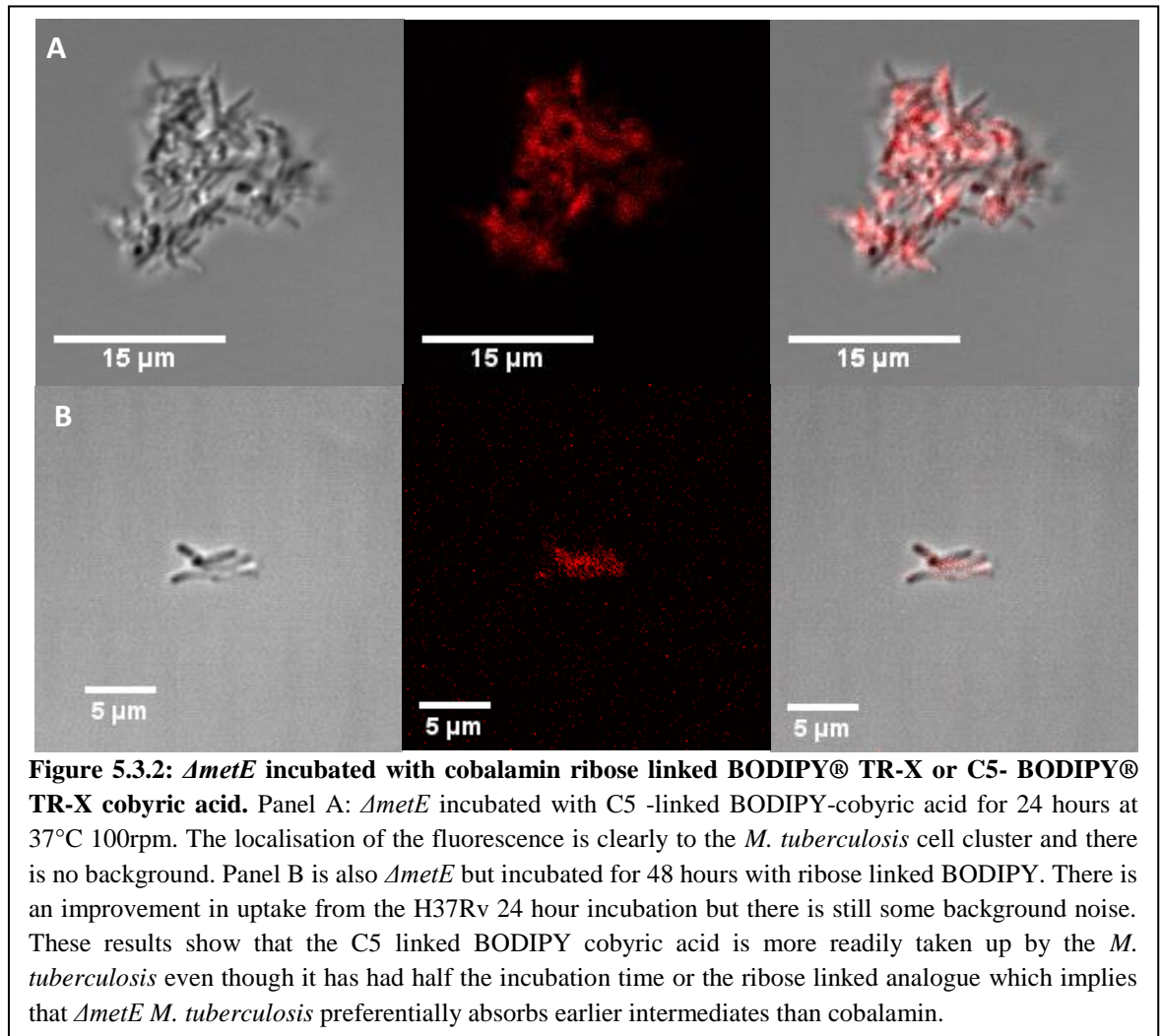
The H37Rv strain and the *ΔmetE* strain (a deletion in H37Rv) were incubated with 1 µM of both BODIPY® TR-X corrinoid analogues for 24 hours at 37 °C and then imaged. In the H37Rv sample the fluorescence shown in the middle of Figure 5.3.1 has a clear outline which marries up to the DIC. Although the correlation of the fluorescence with the DIC was good, the brightness of the fluorescence had to be artificially increased so much that unspecific background fluorescence artefacts became visible (Figure 5.3.1).

Figure 5.3.1: H37Rv incubated with cobalamin ribose linked BODIPY® TR-X. H37Rv incubated with ribose linked BODIPY® TR-X-cobalamin for 24 hours at 37°C 100rpm. Although the localisation of the fluorescence is clearly to the *M. tuberculosis* cell cluster, due to the increased fluorescence

The cobalamin-dependent *ΔmetE* strain was grown in Sauton's defined medium (Chapter 2, 2.1.3) which is a minimal media used in this instance to limit potential contamination with endogenous cobalamin. In addition to the change in media, a 48 hour incubation was included as well as a 24 hour one. Apart from these alterations the *ΔmetE* cultures were set up in the same way as the H37Rv experiment. The 24 and 48 hour incubations were both imaged at the same time (Figure 5.3.2).

These changes succeeded in improving the uptake of the analogues into *M. tuberculosis*. This was particularly true of the C5-BODIPY® TR-X cobyric acid. Panel A (Figure 5.3.2) shows the C5-BODIPY® TR-X cobyric acid after 24 hours of incubation. The fluorescence clearly associates with the cells seen in the DIC. Panel B shows the ribose linked BODIPY® TR-X cobalamin after 48 hours of incubation. The fluorescence again matches the DIC well, although the fluorescence again had to be increased artificially, resulting in some background interference. After 24 hours the cobalamin analogue sample showed no detectable fluorescence (data not shown) whereas the C5 linked BODIPY® TR-X cobyric acid samples showed good correlation in both the 24 hour (Panel A, Figure 5.3.2) and 48 hour incubations (data not shown). This adds to the previous observation that earlier intermediates in the

cobalamin biosynthesis pathway are more readily taken up than cobalamin. In both cases the analogue does localise to the *M. tuberculosis*, and the cultures of *ΔmetE* continue to grow on analogue supplementation alone. Therefore, the *M. tuberculosis* can finish the lower loop formation, attach the upper ligand and then use the 'completed' analogue as a cofactor for the methionine synthase MetH.

Figure 5.3.2: *ΔmetE* incubated with cobalamin ribose linked BODIPY® TR-X or C5- BODIPY® TR-X cobyric acid. Panel A: *ΔmetE* incubated with C5 -linked BODIPY-cobyric acid for 24 hours at 37°C 100rpm. The localisation of the fluorescence is clearly to the *M. tuberculosis* cell cluster and there is no background. Panel B is also *ΔmetE* but incubated for 48 hours with ribose linked BODIPY. There is an improvement in uptake from the H37Rv 24 hour incubation but there is still some background noise. These results show that the C5 linked BODIPY cobyric acid is more readily taken up by the *M. tuberculosis* even though it has had half the incubation time or the ribose linked analogue which implies that *ΔmetE M. tuberculosis* preferentially absorbs earlier intermediates than cobalamin.

5.4 What is the earliest intermediate to rescue *ΔmetE* and which cobalamin does *Mycobacterium tuberculosis* make?

5.4.1 *ΔmetE* rescue

The two previous sections have shown that cobinamide can be taken up by *M. tuberculosis* and that a cobyric acid analogue can be used to compensate the corrinoid dependent *ΔmetE* strain for a lack of cobalamin. Is it possible for earlier intermediates to do the same? To investigate this, a number of intermediates were made and studied for rescue.

The intermediates used were precorrin-7, HBAD, HBAH, and cobyric acid. A final concentration of 1 µM of each intermediate was added to the 500 µL 7H9 media into which *M. tuberculosis* was added so an optical density of 0.6 would be reached on the fifth day (*M. tuberculosis* has a 17 hour doubling time at 37 ℃). Three strains of *M. tuberculosis* used were: *ΔmetE*, *ΔmetH*, and H37Rv.

The experiment was conducted in a 24-well plate at 37 °C, revolving at 100 rpm. This was photographed every day. A blank, in which no intermediate was added, served as a negative control, and cobalamin was used as the positive control. Figure 5.4.1.1 is a photograph of the plate on the 7th day of incubation. The H37Rv and *ΔmetH* wells are all growing regardless of the presence or absence of any intermediate. However, there is no growth in precorrin-7, HBAD, HBAH or the blank wells with *ΔmetE*. The slight growth observed in these wells is due to the original inoculum. The cobalamin and cobyric acid supplemented wells both support *ΔmetE* growth.
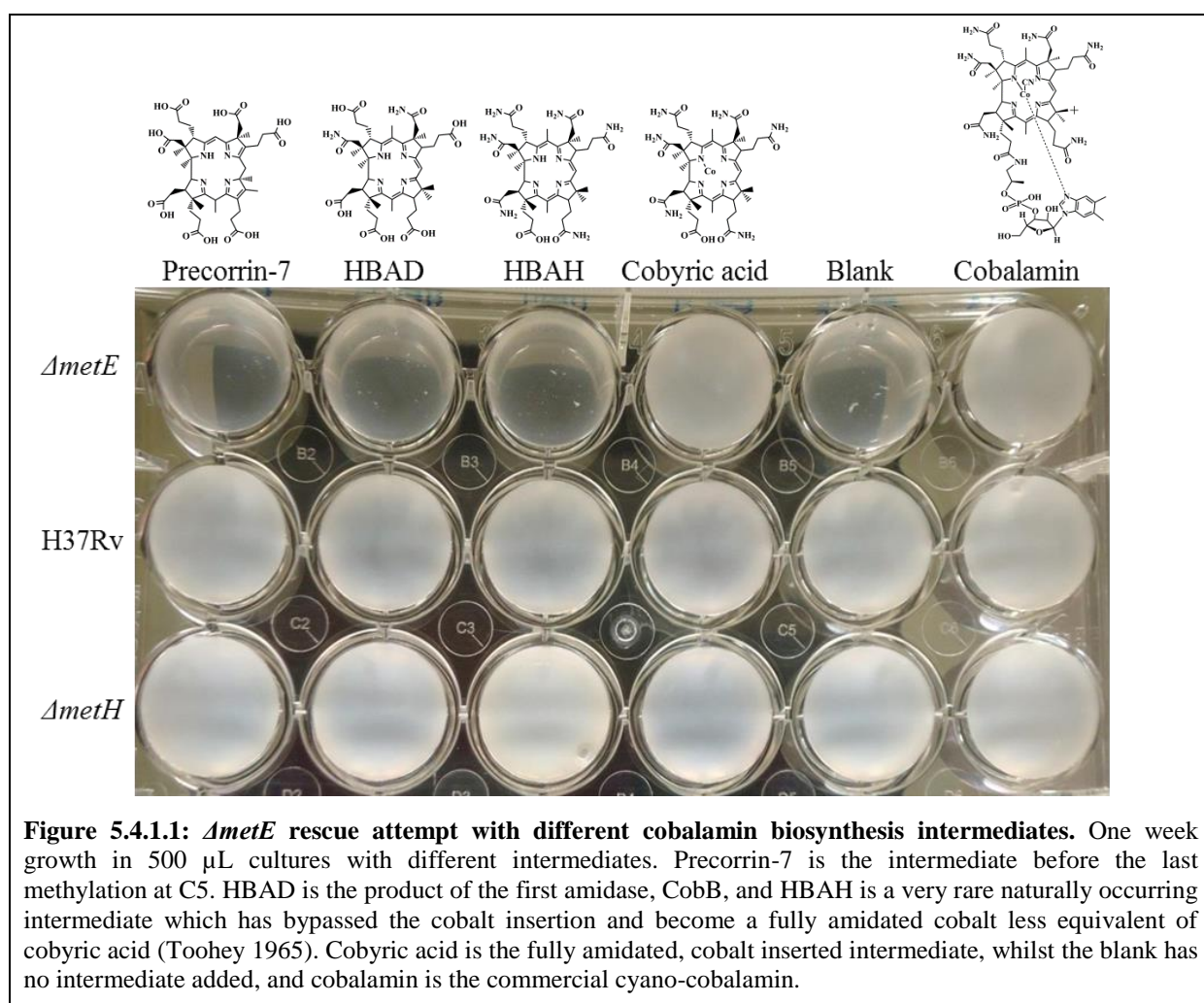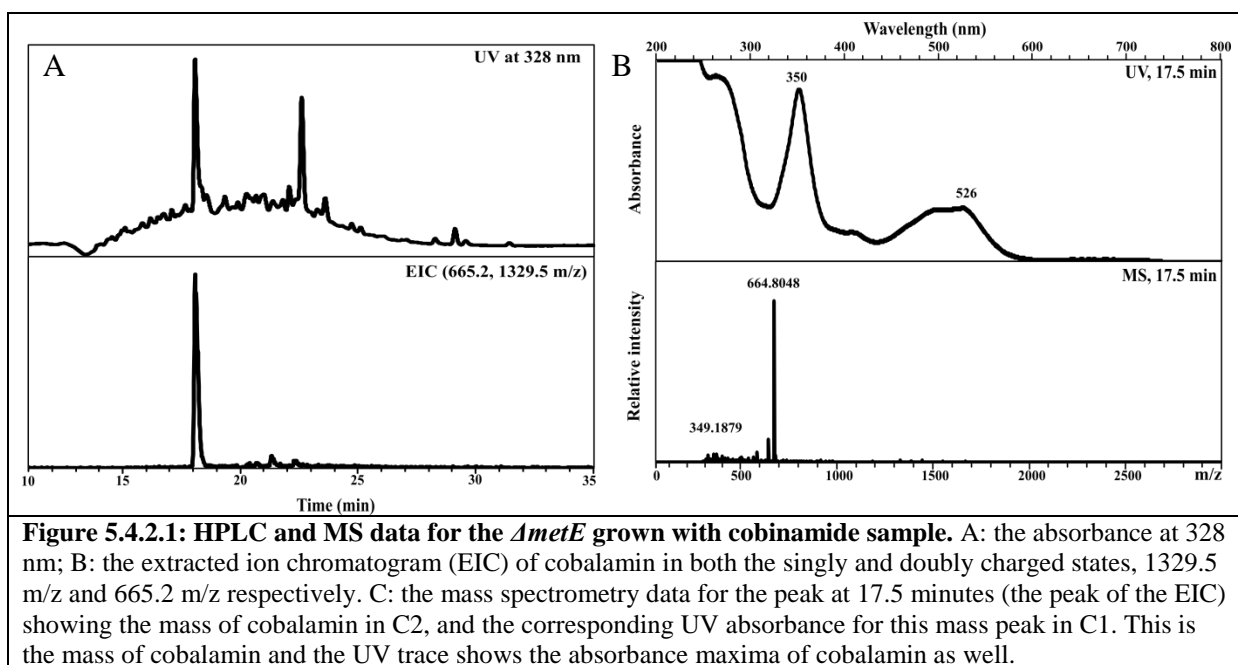

Figure 5.4.1.1: *ΔmetE* **rescue attempt with different cobalamin biosynthesis intermediates.** One week growth in 500 µL cultures with different intermediates. Precorrin-7 is the intermediate before the last methylation at C5. HBAD is the product of the first amidase, CobB, and HBAH is a very rare naturally occurring intermediate which has bypassed the cobalt insertion and become a fully amidated cobalt less equivalent of cobyric acid (Toohey 1965). Cobyric acid is the fully amidated, cobalt inserted intermediate, whilst the blank has no intermediate added, and cobalamin is the commercial cyano-cobalamin.

5.4.2 Which cobalamin does *Mycobacterium tuberculosis* make?

The previous sections of this chapter have shown that *M. tuberculosis* can recognise cobyric acid, cobalamin and pseudo-cobalamin. Both cobalamin and pseudo-cobalamin are potential final products for corrinoid biosynthesis, but it is unknown which *M. tuberculosis* makes. To find out, 1.35 L of *ΔmetE* was grown for 5 days with 1 µM of dicyano-cobinamide in the 7H9 media. The culture was spun, lysed and the supernatant extracted. This supernatant was applied to a nickel affinity column with the His-tagged *E. coli* periplasmic binding protein BtuF preloaded. After a sequence of washes the BtuF bound compounds were eluted in 8 M urea which was then passed down a reverse phase C18 column to remove the urea, and eluted in 50 % methanol. This was concentrated on a vacuum centrifuge, left in the light to remove the upper ligand, and run on the HPLC-MS.

Figure 5.4.2.1 shows that the mass detected was 664.8 m/z and has absorption maxima at 350 nm and 526 nm. This is the expected mass of cobalamin with no upper ligand and the UV trace corresponds to that of cobalamin as well. Therefore *M. tuberculosis* makes cobalamin and not pseudo-cobalamin. The sample was also checked for residual cobinamide (Appendix C.3) but none was found.

Figure 5.4.2.1: HPLC and MS data for the *ΔmetE* grown with cobinamide sample. A: the absorbance at 328 nm; B: the extracted ion chromatogram (EIC) of cobalamin in both the singly and doubly charged states, 1329.5 m/z and 665.2 m/z respectively. C: the mass spectrometry data for the peak at 17.5 minutes (the peak of the EIC) showing the mass of cobalamin in C2, and the corresponding UV absorbance for this mass peak in C1. This is the mass of cobalamin and the UV trace shows the absorbance maxima of cobalamin as well.

5.5 Corrinoid RNA Regulation

There are two predicted cobalamin riboswitches in *M. tuberculosis*: one before *metE* and one prior to *ppe2* (Vitreschak et al. 2003).The *metE* gene encodes the cobalamin independent methionine synthase mentioned previously, whilst the *ppe2* gene encodes a predicted seven transmembrane domain cobalt transporter (Gopinath, Moosa, et al. 2013). This, combined with the evidence that suggests *nrdZ*, the cobalamin-dependent ribonucleotide reductase, is upregulated in hypoxic conditions induced during granuloma development, all indicate some cobalamin mediated/linked RNA regulation in *M. tuberculosis* (Boshoff and Barry 2005).

5.5.1 Cobalamin

In order to ascertain the effect of cobalamin on the RNA landscape of *M. tuberculosis* 1 μM of cobalamin was incubated with different strains and subcultured four times, never exceeding an optical density of 0.2 at 650 nm. RNA sequencing data was taken and compared to the non-supplemented cultures which were grown alongside the fortified strains. Five strains were used in these experiments: WT, H37Rv, *ΔbacA*, *ΔmetE*, and *ΔmetH*. As established earlier in the chapter, *ΔmetE* does not grow without supplementation with cobalamin which means that the RNA sequencing cannot be performed by comparing a culture with added exogenous cobalamin to one without, as the one without will not grow. To circumvent this issue three cultures of *ΔmetE* were set up: with 1 μM cobalamin, with 2 mM methionine, and with both 1 μM cobalamin and 2 mM methionine. By providing methionine the need for a functional methionine synthase is eradicated, but in order to compare the RNA profiles impartially the third culture which has been given both supplements must be compared to the other two cultures separately.

During the course of the experiment the *ΔmetH* culture with cobalamin struggled to grow more and more the further it was subcultured. The *ΔmetH* strain relies upon MetE for methionine synthesis, one of the genes proposed to have a cobalamin riboswitch preceding it. If the cobalamin riboswitch downregulates *metE* in the presence of cobalamin, as is predicted, then as the levels of *metE* RNA from the starter culture fall with subsequent subculturing and *ΔmetH* strain continues to proliferate there will be fewer *metE* transcripts to translate into protein, leaving the *ΔmetH* cells with no methionine synthase. This decrease in active

methionine synthase will result in reduced methionine production which will limit protein production, leading to a slower growth rate and potentially cell death (Ron 1975). Fortunately the *ΔmetH* cultures with cobalamin survived long enough to extract sufficient RNA for quantification by sequencing technology.

5.5.2 Cobinamide

The experiment was also conducted using cobinamide instead of cobalamin. There were two reasons for this: to confirm that cobinamide could support the growth of *ΔmetE*, and to investigate whether cobinamide could bind to the cobalamin riboswitch. Figure 5.5.2.1 shows that cobinamide does facilitate *ΔmetE* growth. Cobinamide can bind to the cobalamin riboswitch as the *ΔmetH* cultures showed that cultures provided with cobinamide diminished in the same way they did with cobalamin indicating that cobinamide can also stop transcription of *metE*.

Figure 5.5.2.1: *ΔmetE* cultures grown with cobinamide and/ or methionine. *ΔmetE* cultures grown at 37°C at 100 rpm supplemented with (from left to right) 1 µM cobinamide; 1 µM cobinamide and 2 mM methionine; and 2 mM methionine.

The RNA sequencing data presented in Table 5.5.2.3 combines both the cobalamin and cobinamide results. Differences in RNA levels of over ±1 were considered to be significant particularly if it is so across all of the strains. Increases in RNA levels compared to the control cultures of +1 and over are highlighted in green and decreases of -1 and under are in red, non-significant results are paler hues thereof. There were no high impact effects of either cobalamin or cobinamide on the RNA levels overall. Significant results relating to known cobalamin related genes or genes implicated in other experiments are summarised in Table 5.5.2.3.There were some that had markedly different regulation in the presence of cobalamin or cobinamide, for example *mutB* (Rv1493), the large subunit of the MCM was upregulated in all the strains in the presence of cobinamide compared to the control, but not with cobalamin. Oddly neither corrinoid had any significant effect on *mutA* (Rv1492) even though the two genes are next to each other in the *M. tuberculosis* genome. The *metE* gene

levels are increased in the WT, H37Rv and *ΔmetH* strains with cobalamin, and cobinamide has the same effect in both the WT and H37Rv, compared to their respective control cultures. In the *ΔbacA* strain on the other hand, cobinamide significantly decreases the level of *metE*. As this is the *ΔbacA* strain the internal concentration of the corrinoid should be lower than those with the transporter and yet this is the only strain in which *metE* transcripts decrease compared to the control. The *metH* gene transcription is not affected by the corrinoids in any strain. In contrast, the gene encoding ketol acid reductoisomerase, *ilvC*, is downregulated compared to the control in all strains in the presence of either corrinoid, but it is involved in amino acid synthesis and has no obvious link to cobalamin metabolism in any way. The *cobQ1* gene is the only cobalamin biosynthesis gene upregulated in the cobinamide cultures compared to the control cultures. None of the other biosynthesis genes are significantly up or downregulated in either corrinoid cultures (Appendix C.5). *M. tuberculosis* has two *cobQ* genes, the one in question is Rv0255c which is the longer of the two, and the other Rv3713. There is no great similarity between the two in amino acid sequence (Figure 5.5.2.2) or DNA (Appendix C.4), and there is no known reason as to why there are two.

The other cobalamin riboswitch regulated gene, *ppe2*, is upregulated compared to the
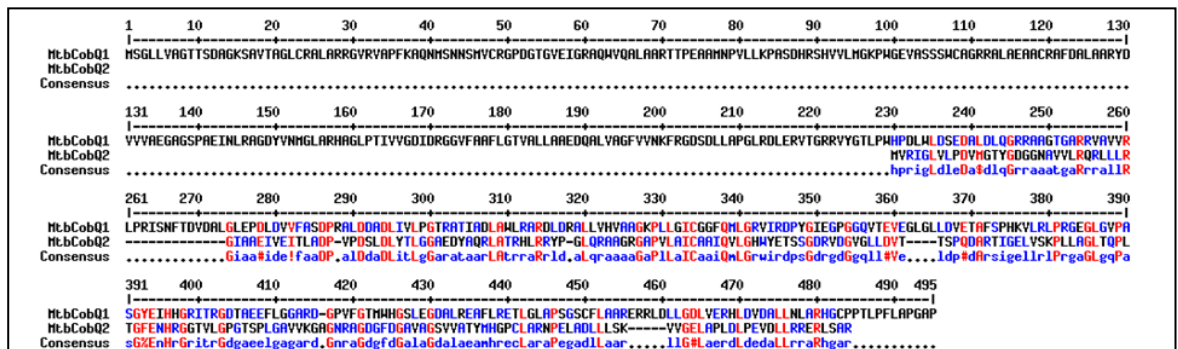

Figure 5.5.2.2: The amino acid alignment of the two CobQs from *M. tuberculosis*. There is no known reason why there are two and they are not similar in amino acid sequence or in DNA (Appendix C.4).

control in both experiments, but more uniformly in the cobinamide cultures. Both the *ΔbacA* cultures show a diluted response compared to the other strains.
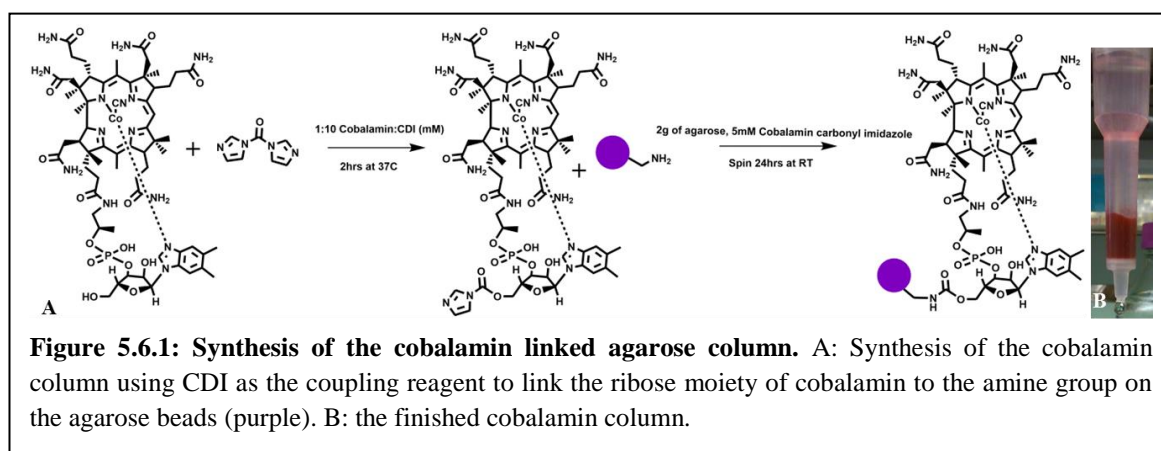
Table 5.5.2.3: The change in RNA levels caused by the presence of cobalamin or cobinamide

| Cobalamin | | | | | Cobinamide | | | | | | | |
ΔmetE	WT	H37Rv	*ΔmetH*	*ΔBacA*	*ΔmetE*	H37Rv	*ΔmetH*	*ΔBacA*	WT	Gene name	Gene identity	Protein function
0.161	0.3839	0.3041	1.1885	0.5626	0.9867	0.9103	1.2792	0.9871	0.9202	*mutB*	Rv1493	Large subunit of MCM
-0.749	-0.1119	-0.2598	0.721	0.012	-0.4299	-0.2293	-0.1027	-0.2316	-0.112	*mutA*	Rv1492	Small subunit of MCM
0.428	1.3186	1.4111	3.8119	-0.4685	-0.1469	2.8173	-0.2046	-1.1669	3.4224	*metE*	Rv1133c	Cobalamin independent methionine synthase
0.774	0.3076	0.5418	-0.254	0.4614	0.8137	0.5784	0.407	0.6081	0.6003	*metH*	Rv2124c	Cobalamin dependent methionine synthase
-0.6136	-0.534	-0.7537	-1.4403	-0.6129	-0.9874	-1.1826	-1.1932	-1.4606	-1.3111	*ilvC*	Rv3001c	Ketol acid reductoisomerase
0.3257	0.2448	0.0277	0.7251	0.2846	1.4511	0.6745	1.6493	1.1899	0.8317	*cobQ1*	Rv0255c	Cobyric acid synthase
1.4875	0.5375	0.7729	1.4572	0.3699	2.669	1.5796	0.8967	0.5667	1.5498	*ppe2*	Rv0256c	Cobalt transporter

The gene identity is the number assigned to the gene within the *M. tuberculosis* genome whereas the gene name is how it is referred to commonly; it is usually based on the function of the protein. To provide a comparison with *metE*, *metH* is also shown in the table. The values shown in the table are the Q values comparing RNA levels with supplementation to cultures without supplementation. The more red the colour of the box, the lower the RNA level compared to the unsupplemented culture, and the more green the box is, the higher the RNA level is compared to the unsupplemented cultures.
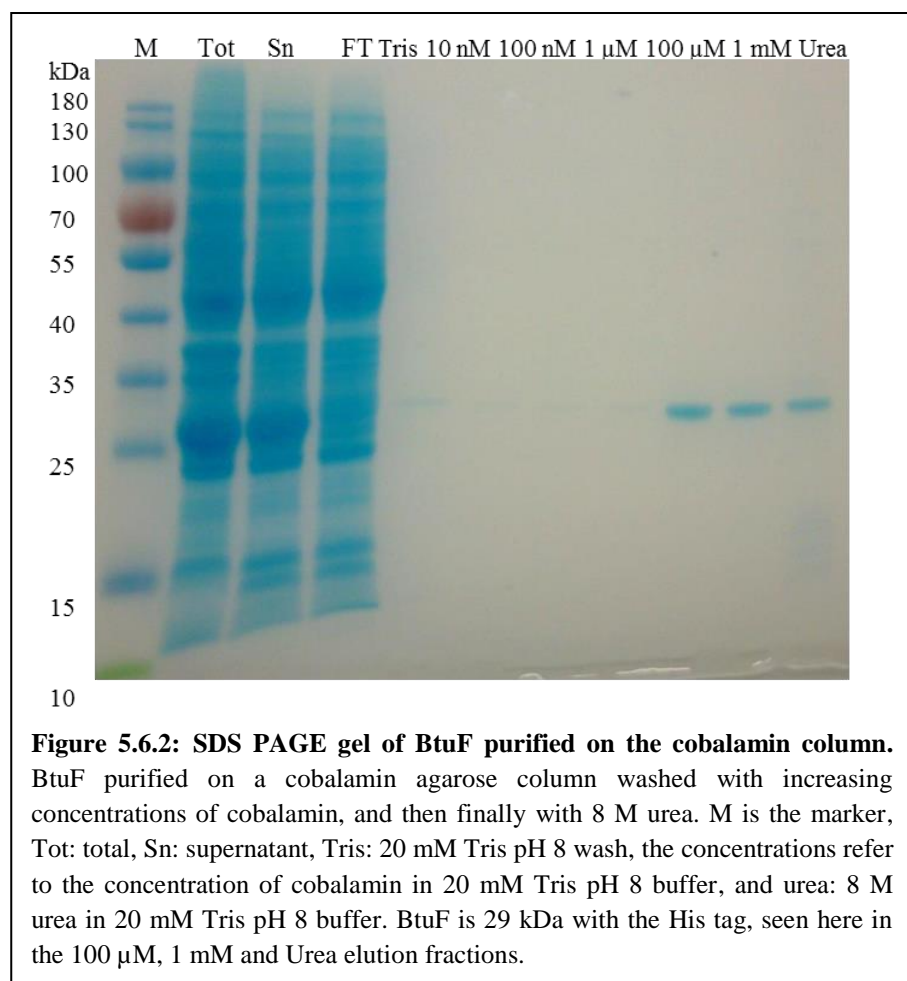
5.6 Cobalamin binding proteins

There are three known cobalamin-dependent enzymes and one known cobalamin transporter but there may be more cobalamin binding proteins (Gopinath, Moosa, et al. 2013; Gopinath, Venclovas, et al. 2013). In order to investigate this at the protein level a cobalamin resin was synthesised by linking cobalamin to agarose beads via the ribose (Figure 5.6.1). This was then packed into a column.

Figure 5.6.1: Synthesis of the cobalamin linked agarose column. A: Synthesis of the cobalamin column using CDI as the coupling reagent to link the ribose moiety of cobalamin to the amine group on the agarose beads (purple). B: the finished cobalamin column.

The viability of the column was checked by purifying BtuF from *E. coli*. BtuF was overexpressed in *E. coli* and the supernatant of the cells, after lysis and spinning, was purified on the cobalamin column (Figure 5.6.2). The binding constant of BtuF is known to be between 10 and 30 nM, but Figure 5.6.2 shows 100 µM of cobalamin was needed to elute it from the column (Lewinson et al. 2010; Cadieux et al. 2002). To ensure that the column was washed sufficiently 50 mL of each wash buffer was passed through the 2 mL column followed by an 8 M urea wash to remove anything left on the column.

The strains of *M. tuberculosis* used in this investigation were H37Rv, WT, *ΔmetE* and *ΔbacA*. A total of 1.35 L of the strain was grown, then spun and lysed. The supernatant was

either applied directly to the column, or incubated overnight at 4°C with gentle rocking to maximise binding. The column was then washed with increasing concentrations of cobalamin until the protein eluted. Every elution was concentrated to ensure a clear band, and run on SDS-PAGE followed by MALDI-TOF analysis to identify the proteins. However, even after concentrating the elutions on a 5 kDa concentrator from 50 mL to 500 µL, bands were only observed in some samples on the SDS-PAGE. Most elutions had faint bands or none at all. A 5 kDa filter was selected for the concentrator as the first purification showed a faint band at just over the 10 kDa marker band (Appendix C.6). In order to accommodate the 5 kDa to 180 kDa sized proteins all of the fractions were run on 4-20 % SDS gels.

Figure 5.6.2: SDS PAGE gel of BtuF purified on the cobalamin column. BtuF purified on a cobalamin agarose column washed with increasing concentrations of cobalamin, and then finally with 8 M urea. M is the marker, Tot: total, Sn: supernatant, Tris: 20 mM Tris pH 8 wash, the concentrations refer to the concentration of cobalamin in 20 mM Tris pH 8 buffer, and urea: 8 M urea in 20 mM Tris pH 8 buffer. BtuF is 29 kDa with the His tag, seen here in the 100 µM, 1 mM and Urea elution fractions.

All of the strains were subjected to the same purification. The clearest elutions from all the purifications were chosen for MALDI-TOF analysis (Figure 5.6.3). MALDI-TOF was performed by excising the SDS-PAGE bands of interest (highlighted in Figure 5.6.3 in red boxes). These are washed thoroughly with ammonium carbonate, acetonitrile and iodoacetomide, and then left digesting overnight in a trypsin solution. This results in the production of a mixture peptides unique to each protein. The molecular mass of these peptides

were measured in the mass spectrometer which generates a peptide mass fingerprint (PMF) dataset (Walker 2009). This dataset was compared with a database of theoretical PMFs produced by *in silico* trypsin digestion. The closest match is flagged, but a scoring method is used to show how similar the match is (the coloured column in Table 5.4.4, the higher the score the better the match) as there are often more than one possible hit. MS/MS uses the initial PMF dataset and selects a range of mass to charge (m/z) values to fragment further. These are detected generating a second dataset for the same sample. By combining these two datasets from the original PMF and the MS/MS a more accurate database match can be performed.
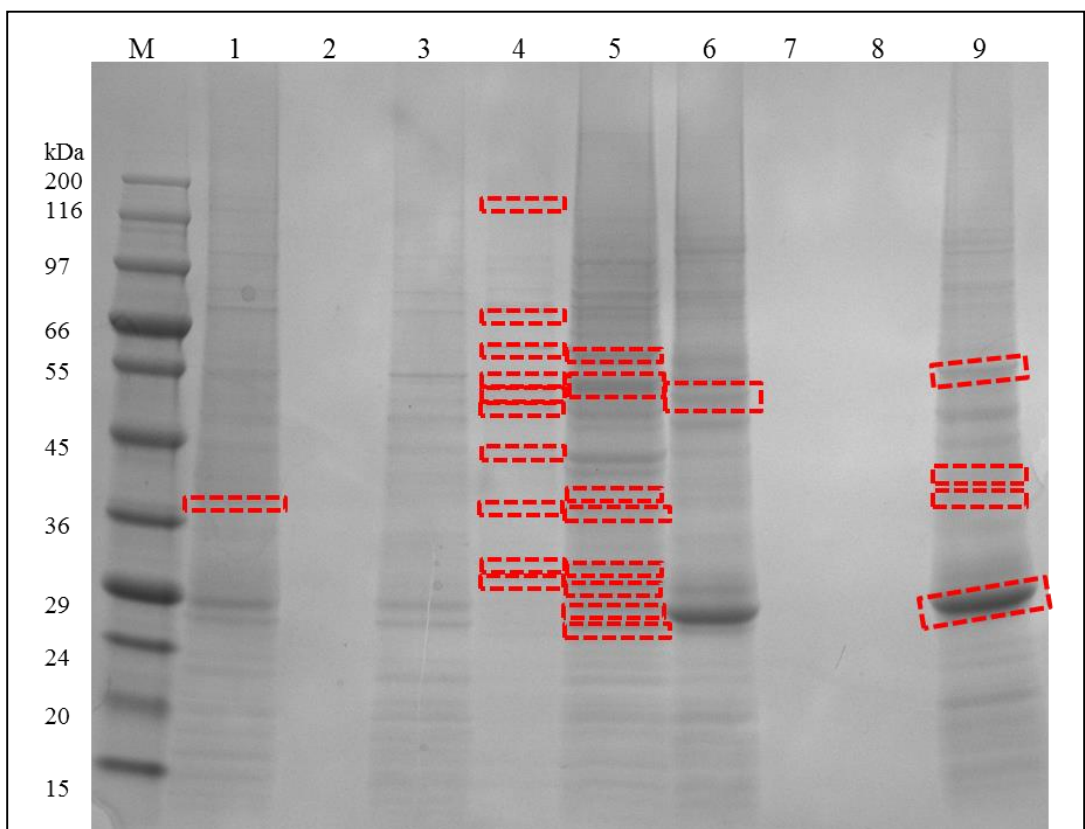

Figure 5.6.3: MALDI reference gel. 1: *ΔmetE* overnight with the resin urea wash. 2: H37Rv overnight with the resin 100 nM wash. 3: H37Rv overnight with the resin urea wash. 4: *ΔbacA* overnight with the resin 1 µM wash. 5: *ΔbacA* overnight with the resin urea wash. 6: WT urea wash. 7: H37Rv 10 nM wash. 8: H37Rv 1 µM wash. 9: H37Rv urea wash. The only wash that shows up on the gel apart from the urea washes is the 1 µM *ΔbacA* wash which was 100 times less concentrated wash than the one overexpressed BtuF elutes in, but is still promising. The red dotted boxes indicate bands which were excised for MALDI-TOF. These were numbered from top to bottom and prefixed with the name of the elution. No bands were cut from elution 3 because they looked identical to elution 1. The results are in Table 5.6.4.

The MS/MS results can be seen in Table 5.6.4, whilst the initial mass spectrometry results can be found in Appendix C.7. The bands were numbered from the top down and

prefixed with the name of the elution e.g. *ΔbacA* O/N 1 µM B$_{12}$-1 refers to the top band of lane 4. *ΔbacA* is the strain, O/N means it was incubated overnight at 4 °C with the cobalamin resin and 1 µM B$_{12}$ is the concentration of cobalamin in the wash buffer which eluted this sample.

Many of these results are background proteins which often appear in MALDI, such as the various 30s ribosomal proteins and DNA K, a chaperone protein. Others such as bovine serum albumin (BSA) are figments of the experiment as this is present in the media. The presence of *E. coli* BtuF in the WT elution must be cross contamination as this was purified before the column was tested with *E. coli* BtuF. The more distinct results are glutamine synthetase I, citrate synthase I, and ketol-acid reductoisomerase. Citrate synthase I (gene identification Rv0896, gene name *gltA2*) is one of the enzymes in the glyoxylate cycle which is part of the cobalamin independent propionate metabolism route in *M. tuberculosis* (see Chapter 1, Figure 1.7.1.1). Glutamine synthetase I (gene identification Rv2220, gene name *glnA1*) converts glutamate into glutamine, but glutamate can also be converted into glyoxylate, which feeds into the glyoxylate cycle. The link between ketol-acid reductoisomerase and cobalamin is unclear, although both the gene and the protein have been identified in the RNA sequencing screen (Section 5.5) and the cobalamin binding investigation. MetE was also identified, even though it is MetH which binds cobalamin. The amino acid sequences of these two proteins are not similar at all (Appendix C.8) even after trypsin digestion. The lowest band in lane 6 (Figure 5.6.3) was not cut out because it was too close to the excised bands in lane 5 to cut out cleanly. This was sacrificed in favour of the bands in lane 5 as the lane 6 band looks to be the same size as the band in lane 9, which is not present in lane 5.

Table 5.6.4: MALDI-TOF results (MS/MS)

Lane	Band		MS-MS	Organism	Score value
1	*ΔmetE* O/N urea	1	30S ribosomal protein S2	*M. tuberculosis* (strain ATCC 25618 / H37Rv)	141
4	*ΔbacA* O/N 1 µM B12	1	Serum albumin	Bos taurus	111
4	*ΔbacA* O/N 1 µM B12	2	**5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase (metE)**	*M. tuberculosis* (strain ATCC 25618 / H37Rv)	476
4	*ΔbacA* O/N 1 µM B12	3	Chaperone protein DnaK	*M. tuberculosis* (strain ATCC 25618 / H37Rv)	221
4	*ΔbacA* O/N 1 µM B12	4	30S ribosomal protein S1	*M. tuberculosis* (strain ATCC 25618 / H37Rv)	244
4	*ΔbacA* O/N 1 µM B12	5	Did not run	Did not run	Did not run
4	*ΔbacA* O/N 1 µM B12	6	**Glutamine synthetase I**	*M. tuberculosis* (strain ATCC 25618 / H37Rv)	178
4	*ΔbacA* O/N 1 µM B12	7	**Citrate synthase I**	*M. tuberculosis* (strain ATCC 25618 / H37Rv)	202
4	*ΔbacA* O/N 1 µM B12	8	Alcohol dehydrogenase B	*M. tuberculosis* (strain ATCC 25618 / H37Rv)	80
4	*ΔbacA* O/N 1 µM B12	9	Electron transfer flavoprotein subunit alpha	*M. tuberculosis* (strain ATCC 25618 / H37Rv)	464
4	*ΔbacA* O/N 1 µM B12	10	Electron transfer flavoprotein subunit beta	*M. tuberculosis* (strain ATCC 25618 / H37Rv)	372
5	*ΔbacA* O/N urea	1	Chaperone protein DnaK	*M. tuberculosis* (strain ATCC 25618 / H37Rv)	610
5	*ΔbacA* O/N urea	2	Glutamine--fructose-6-phosphate aminotransferase [isomerising]	*M. tuberculosis* (strain ATCC 25618 / H37Rv)	36
5	*ΔbacA* O/N urea	3	DNA-directed RNA polymerase subunit alpha	*M. tuberculosis* (strain ATCC 25618 / H37Rv)	216
5	*ΔbacA* O/N urea	4	**Ketol-acid reductoisomerase**	*M. tuberculosis* (strain ATCC 25618 / H37Rv)	197
5	*ΔbacA* O/N urea	5	Electron transfer flavoprotein subunit alpha	*M. tuberculosis* (strain ATCC 25618 / H37Rv)	332
5	*ΔbacA* O/N urea	6	Electron transfer flavoprotein subunit beta	*M. tuberculosis* (strain ATCC 25618 / H37Rv)	440
5	*ΔbacA* O/N urea	7	Probable enoyl-CoA hydratase echA8	*M. tuberculosis* (strain ATCC 25618 / H37Rv)	56
5	*ΔbacA* O/N urea	8	30S ribosomal protein S4	*M. tuberculosis* (strain ATCC 25618 / H37Rv)	224
6	WT urea	1	Vitamin B12-binding protein	Escherichia coli (strain K12)	220
9	H37Rv urea	1	30S ribosomal protein S1	*M. tuberculosis* (strain ATCC 25618 / H37Rv)	249
9	H37Rv urea	2	Vitamin B12-binding protein	Escherichia coli (strain K12)	98
9	H37Rv urea	3	30S ribosomal protein S3	*M. tuberculosis* (strain ATCC 25618 / H37Rv)	107
9	H37Rv urea	4	Vitamin B12-binding protein	Escherichia coli (strain K12)	253

This table shows a summary of the MALDI-TOF MS/MS results. The column entitles 'Lane' refers to the lane number this sample was run in on the MALDI reference gel (Figure 5.6.3). The column 'Band' details the background of the *M. tuberculosis* strain, e.g. *ΔbacA*, the wash it eluted from the cobalamin- column in, e.g. 1 µM B_{12}, and the number band it was within this sample, e.g. 1 would indicate it was the top band in this SDS sample.

5.7 Discussions

Cobalamin and cobinamide are both taken up into *M. tuberculosis* in a concentration dependent manner with increasing concentrations of exogenous corrinoid. Pseudo-cobalamin appears to be absorbed in a similar fashion but because the *S. enterica* bioassay plate is less sensitive to pseudo-cobalamin, many of the samples taken from cells grown in lower exogenous concentrations of pseudo-cobalamin do not show any growth on the bioassay plate. In the *ΔbacA* strain when provided with 10 µM exogenous corrinoid all three are taken up into the *M. tuberculosis* cells. This indicates that another corrinoid uptake system may be present in *M. tuberculosis*. Cobinamide and pseudo-cobalamin are taken up to about 50 nM in the cells whereas cobalamin is half of this, which either means that the alternative uptake route prefers pseudo-cobalamin and cobinamide, or it may be that this transport route is negatively regulated by cobalamin, but not by pseudo-cobalamin and cobinamide. This negative regulation could limit the internal cell concentration of cobalamin, but not restrict pseudo-cobalamin and cobinamide.

The internal concentration of pseudo-cobalamin is much higher than either cobalamin or cobinamide when the same concentration of corrinoid is provided. This could be a figment of this experiment, for example *M. tuberculosis* may selectively import pseudo-cobalamin in preference to cobalamin and cobinamide. This would be strange if so because *M. tuberculosis* synthesises cobalamin, not pseudo-cobalamin when *ΔmetE* is grown solely on cobinamide (Section 5.3.2).

The two C5 and ribose linked BODIPY® TR-X fluorophore analogues were both taken up into *M. tuberculosis*. However, the ribose linked analogue was not taken up as well as the C5 analogue. This could be because the C5 analogue is smaller and impedes the uptake proteins less. It could also be because the BacA transport route is not as efficient at transporting the analogues as the hypothetical alternative route, which is less effective at transporting cobalamin compared to earlier intermediates. The superior uptake of the C5 analogue into *M. tuberculosis* indicates a potential advantage for synthesising earlier

intermediate analogues rather than cobalamin ones. In humans TCII, the transport protein which mediates cell uptake, only binds cobalamin (Greibe, Fedosov, and Nexo 2012). Therefore, in a *M. tuberculosis* infected individual if an earlier intermediate is in the blood this should only be taken up by *M. tuberculosis* cells and not human cells. This presents a possible way of targetting the analogues to *M. tuberculosis* cells in a clinical environment.

M. tuberculosis recognises cobalamin intermediates containing cobalt as both cobyric acid and cobinamide can rescue the cobalamin-dependent strain *ΔmetE*. The chemical equivalent of cobyric acid without the cobalt, HBAH, does not rescue *ΔmetE*. This suggests that either the HBAH is not recognised and taken up into the cell due to the lack of the cobalt ion, or that it is taken up but the cobalt chelation complex of *M. tuberculosis* and subsequent cobalamin biosynthesis enzymes are not sufficient to complete cofactor synthesis in concentrations needed to support growth. This could be due to limited cobalt availability in the media. The inability of the Δ*metE M. tuberculosis* strain to grow without exogenous cobalamin supplementation infers that *M. tuberculosis* does not produce enough cobalamin *de novo* to support itself under standard *in vitro* conditions, although it does not conclusively prove that *M. tuberculosis* cannot synthesise cobalamin *de novo* (Gopinath, Moosa, et al. 2013).

The only cobalamin biosynthesis gene flagged in the RNA sequencing experiment which was affected by the presence of a corrinoid was *cobQ1*, which had a higher concentration when cobinamide was added to the media (Appendix C.5). Cobalamin did not produce the same effect. The large subunit of MCM, *mutB*, has the same reaction to cobinamide as *cobQ1*. However, *mutA*, the smaller subunit of MCM does not. There is no known reason why only one subunit should be upregulated, especially as MCM is a cobalamin-dependent enzyme and yet cobalamin as no effect on RNA production. The two genes with known cobalamin riboswitches, *metE* and *ppe2* are both in Table 5.5.2.3, along with the cobalamin-dependent methionine synthase encoding gene, *metH*. Of these genes, only *ppe2* shows uniform increases in RNA when in the presence of cobinamide. This effect is less in the cobalamin samples. The reason cobinamide appears to have more effect on the genes than cobalamin may be due to the cobinamide entering the cells faster than the cobalamin (as with the fluorescent analogues) and, therefore, eliciting an effect on the genes earlier and more consistently than cobalamin.

Proteins or genes identified in the cobalamin column experiment were mostly associated with cobalamin-dependent processes, if not already known to be linked to cobalamin. Glutamine synthetase I converts glutamate into glutamine Glutamate can also be converted into glyoxylate, which feeds into the glyoxylate cycle. Citrate synthase I is one of the enzymes in the glyoxylate cycle which is part of the cobalamin independent propionate metabolism route in *M. tuberculosis*, circumventing the methylmalonyl pathway (see Chapter 1, Figure 1.7.1.1). Ketol-acid reductoisomerase, which was also highlighted in the RNA sequencing experiment, is connected to the metabolism of valine, leucine and isoleucine. The link between ketol-acid reductoisomerase and cobalamin is unclear, but the gene encoding it, *ilvC* (Rv3001c), is downregulated in the presence of both cobalamin and cobinamide, whilst the protein eluted off the cobalamin binding column in a 1 µM cobalamin wash. This suggests that ketol-acid reductoisomerase binds cobalamin but is also downregulated by it.

Chapter 6

Following the journey of cobalamin

6.0 General discussion

The objectives of the research outlined in this thesis were to synthesise a corrin-ring modified analogue of cobalamin, to attach a fluorophore or drug to this position, to analyse the efficacy of the resultant analogues in different organisms, and to identify corrinoid specificity and localisation in *M. tuberculosis* and *C. elegans*. Corrin-ring modified analogues have potential to be used to target cells infected with *M. tuberculosis*, certain cancer cells, or parasitic worms, all of which have an increased cobalamin requirement (Mclean et al. 1997; Waibel et al. 2008; Allen 2008). Previous cobalamin analogues have been made by modifying the peripheries of the molecule but none have directly altered the macrocycle ring component (Clardy et al. 2011). Such modifications can be done either chemically or enzymatically using the cobalamin biosynthetic pathway. Chemically, it is hard to precisely modify one residue when the same functional groups occur at many different positions on the ring. Enzymatically, it is difficult because many of the biosynthesis enzymes which modify positions on the corrin macrocycle use the same cofactors. This means if a cofactor analogue is used to alter a position, other biosynthesis enzymes could use the same cofactor analogue and target a different position to the intended one. The results presented in the previous chapters show that by using a combination of the native cobalamin biosynthesis enzymes and chemical techniques, individual macrocycle modifications are possible, and result in biologically active analogues. Corrin fluorophore conjugates prove the biological viability of these analogues and show where they localise *in vivo*.

6.1 Cobalamin analogue synthesis

The method chosen for corrin modification was a combination of *in vitro* incubation with cobalamin biosynthesis enzymes, and chemical modifications. There are two routes of cobalamin biosynthesis, termed the Early and the Late cobalt insertion pathways, but the Early insertion pathway proved to be unusable for analogue synthesis. Although the intermediates of this pathway had the benefit of the cobalt ion already in the macrocycle, this actually made them very oxygen sensitive and unstable. The instability of the intermediates resulted in degradation of the desired product after synthesis.

The Late insertion pathway was successful in synthesising a C5 corrin analogue of cobyric acid by using allyl-SAM incubated with CobL to transfer the allyl group on to the C5 position of precorrin-7 in place of a methyl group. The cobalamin biosynthesis enzymes CobH, B, and Q, along with their cofactors, converted the C5-allyl precorrin-8 into C5-allyl
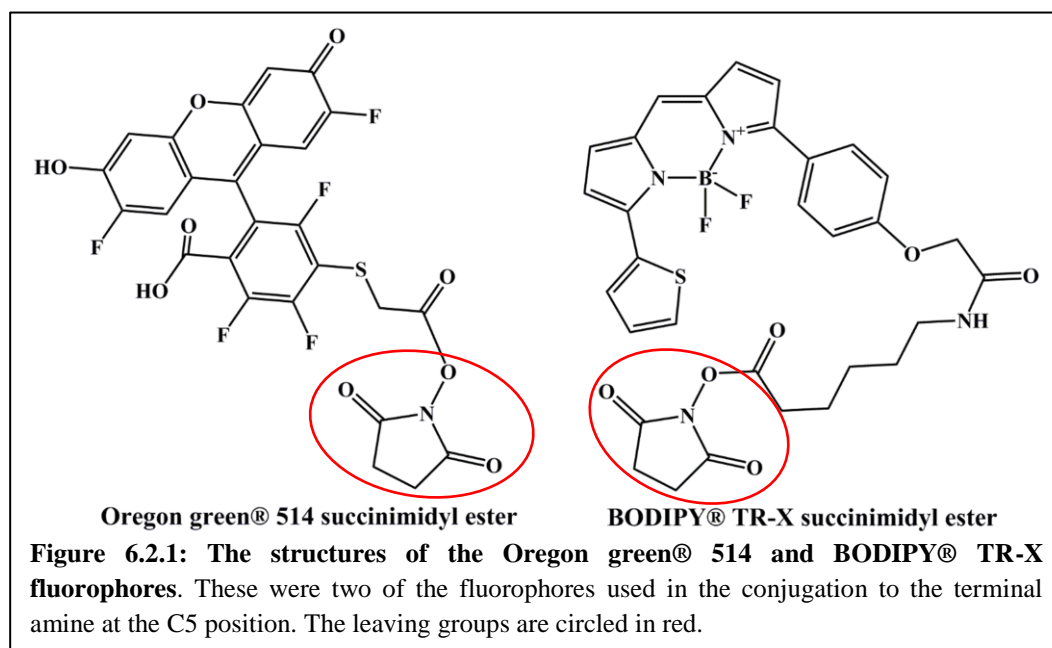
HBAH. The mass and absorbance of each intermediate was confirmed by HPLC MS. The CobH enzyme used in this synthesis is a T85A mutant as this was found to convert more C5-allyl precorrin-8 into C5-allyl HBA than the wild type protein. The reason for this could be seen in the crystal structure of CobH T85A when co-crystallised with C5-allyl HBA which revealed that the alanine residue, being much smaller than the threonine, creates more space in the active site around the C5 position, allowing the allyl group to fit. The co-crystal also confirmed the structure of C5-allyl HBA, conclusively showing that the allyl group has been transferred to the C5 position. Following this, the amidation reactions catalysed by CobB and CobQ both worked well, producing C5-allyl HBAH. The structure of this intermediate was verified by NMR, which showed the presence of the C5 allyl and the six amides on the *a*, *b*, *c*, *d*, *e*, and *g* sidechains. To produce the metal-free HBAH analogues involves using CobQ out of sequence. The CobQ used is from *A. vinosum* as *R. capsulatus* CobQ only amidates after metal insertion and adenosylation of the cobalt. The NMR confirmation of this C5-allyl HBAH not only validates the structure of this novel intermediate, but also demonstrates that a cobalt-less hexa-amidated intermediate, HBAH, is a stable intermediate.

The allyl group was extended and the cobalt ion inserted using previously published chemical methods (Wang et al. 2011; Kräutler 2006). The fluorophore was attached to the primary amine of the extended C5 group using a peptide reaction. Fluorescein was initially used, but it was found to photobleach very quickly so BODIPY® TR-X and Oregon green® 514 equivalents were synthesised. Alongside the synthesis of the C5 analogues, ribose linked analogues were also manufactured. This was necessary as the C5 analogues were not taken further than cobyric acid and many organisms which require cobalamin but cannot make it, do not recognise any corrinoid bar cobalamin (Nielsen et al. 2012). Thus the ribose linked analogues were synthesised to allow a direct comparison to the C5 cobyric acid compounds.

6.2 Cobalamin recognition in *Salmonella enterica* and *Escherichia coli*

The three C5 cobyric acid derivatives and the two ribose linked conjugates were tested for functionality on both *S. enterica* bioassay plates and for uptake in *E. coli*. To work as a cofactor for *S. enterica* MetH and induce growth, the analogues must have a completed lower nucleotide loop and a methyl group as the β ligand for the cobalt ion. All of the synthesised analogues supported growth on the bioassay plates, proving that they were all taken up by the bacteria, and converted into cofactor precursors which could be functionalised *in vivo*. On the *S. enterica* bioassay plates the BODIPY® TR-X analogues generated less growth than their

Oregon green® 514 counterparts. As they are identical in all but the fluorophore it must be this which differentiates them, indicating that the fluorophores are still attached to the analogues at this stage. It could be that Oregon green® 514 is easier to cleave off the cobyric acid than BODIPY® TR-X, and this is what causes the growth difference on the bioassay plate. On the other hand, this may be caused by the BODIPY® TR-X having a longer linker compared to the Oregon green® 514 (Figure 6.2.1), resulting in the BODIPY® TR-X conjugates requiring more space to fit into enzyme binding pockets than the Oregon green® 514 ones. Alternatively this difference may be due to the differing chemical properties of the two fluorophores or the flexibility of the molecules themselves. A more flexible molecule could bend and pack into the binding site; much like the allyl group does in the CobH (T85A) crystal, whereas a more rigid molecule would bind less favourably or not at all. In terms of size, the BODIPY® TR-X fluorophore has a marginally larger molecular weight of 634.5 gmol^{-1} compared to the 609.4 gmol^{-1} of Oregon green® 514. This molecular weight discrepancy is due to difference in the length of the linker to the succinimidyl ester so, unless this is the issue, the molecular weight difference is of little relevance.

Figure 6.2.1: The structures of the Oregon green® 514 and BODIPY® TR-X fluorophores. These were two of the fluorophores used in the conjugation to the terminal amine at the C5 position. The leaving groups are circled in red.

The ribose linked cobalamin derivatives produced bigger rings of growth on the *S. enterica* bioassay plates, signifying that they were converted into the active cofactor form better than the C5-cobyric acid analogues. The reason the cobalamin analogues are taken up better than the cobyric acid analogues is almost certainly because the ribose linked analogues are 'complete', only needing to swap the upper axial ligand in order to form the active cofactor. Cobyric acid, on the other hand, requires the addition of the entire nucleotide loop,

catalysed by some 6 enzymes in *S. enterica*, plus cofactors, as well as DMB synthesis, to generate the same cofactor molecule (Maggio-Hall and Escalante-Semerena 1999; Warren et al. 2002). Even though *S. enterica* can complete cofactor synthesis from cobyric acid it is prudent in a metabolic sense to absorb preferentially the cobalamin form (Maggio-Hall and Escalante-Semerena 1999).

In *E. coli* both the C5 ring modified and ribose linked corrinoid analogues were taken up and exhibited excellent co-localisation of the fluorophore fluorescence with DIC (differential interference contrast) imaging of the bacterial cells. *E. coli* cannot convert cobyric acid into the cofactor form as it lacks the *cobD* gene responsible for the synthesis of aminopropanol-*O*-2-phosphate which is subsequently attached to the *f* sidechain (Warren et al. 2002). However, cobyric acid was observed to be absorbed. This means that the *E. coli* BtuB-F-CD transport system has no selection against cobyric acid and it is essentially taking up a dead-end metabolite.

6.3 Cobalamin analogue recognition in *Caenorhabditis elegans*

C. elegans requires cobalamin for a normal life cycle, but the cobalamin transporter expressed in the intestine is unknown as *C. elegans* has no orthologues of human HC, TCII, or IF (Bito et al. 2013). Remarkably a megalin orthologue has been identified (Christensen and Willnow 1999). Megalin is the endocytic receptor of TCII-cobalamin complex present on cell surfaces which is necessary for renal retention of cobalamin in humans (Moestrup et al. 1996). Megalin is not exclusively used for cobalamin endocytosis as it is also involved in hormone (e.g. insulin (Orlando et al. 1998)), protein (e.g. PAI-1(Stefansson et al. 1996)), and other vitamins (e.g. vitamin D and retinol (Nykjaer et al. 1999; Christensen et al. 1999)) uptake in humans (Christensen and Willnow 1999). As there is no TCII orthologue in *C. elegans*, megalin is unlikely to have a role in cobalamin endocytosis.

There are a number of intracellular cobalamin processing proteins in humans most of which have been shown to have orthologues in *C. elegans* (Figure 6.3.1) (Froese and Gravel 2010). These are involved in functionalising cobalamin into the cofactor and coenzyme forms. Significantly, there has been no orthologue identified for CblF, the protein responsible for trafficking cobalamin from the lysosome into the cytosol. This indicates that the lysosome may not be involved in cobalamin internalisation. In humans TCII is degraded in the lysosome in order to release cobalamin, but, as mentioned above, *C. elegans* does not have a TCII orthologue. Together this information implies that *C. elegans* imports cobalamin in a novel way without utilising the lysosome, and reiterates that megalin is probably not involved in cobalamin uptake.
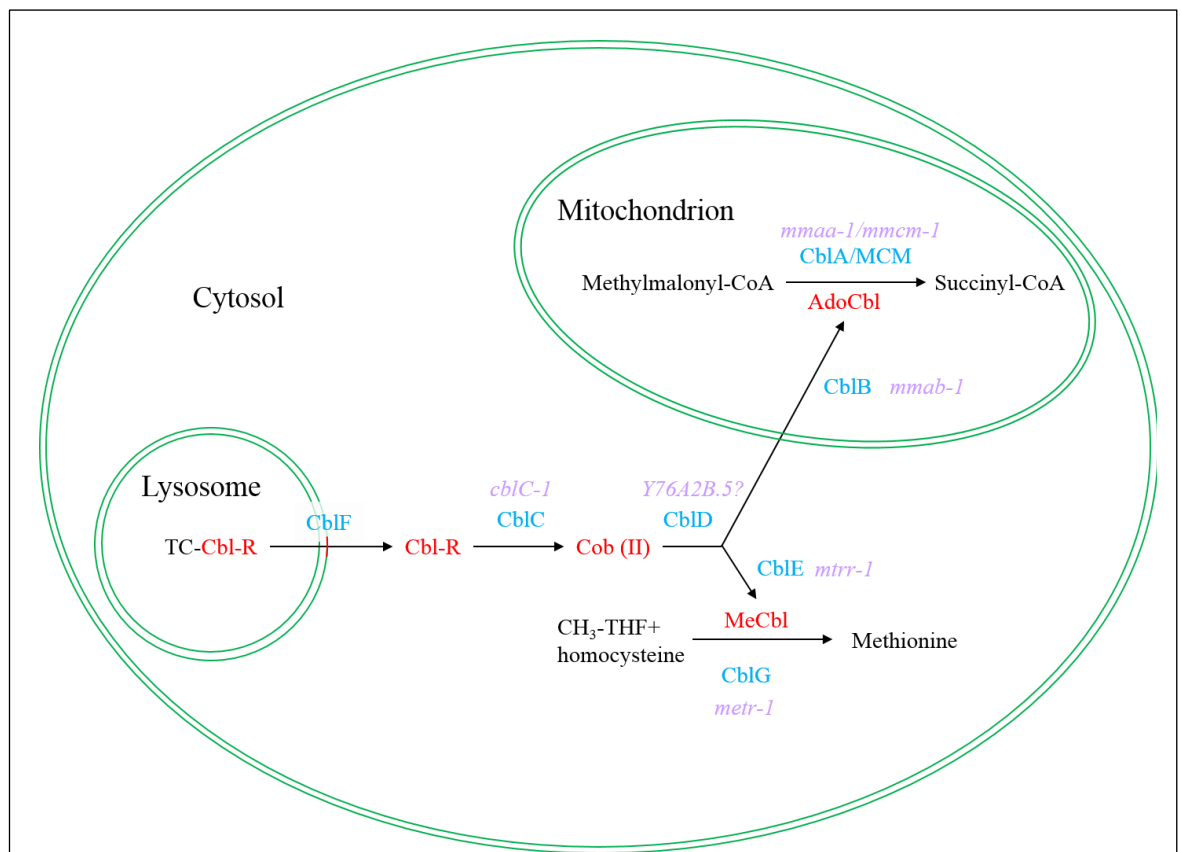
Figure 6.3.1: A schematic representation of the human intracellular processing of cobalamin and the *C. elegans* gene orthologues. Cobalamin intermediates are in red. Cbl-R denotes a cobalamin with an undefined group at the β position on the cobalt ion, whilst Me and Ado refer to methyl and adenosyl respectively. The human proteins are in light blue with the *C. elegans* gene orthologues in light purple. *Y76A2B*.5 has only been implicated in a bioinformatics search and has not been proven to be the CblD orthologue which is why there is a question mark after the gene (Froese et al. 2015). All of the other genes have been shown to catalyse their reactions in assays and to exhibit either methylmalonyl acidemia or homocysteinemia or both, as appropriate, in knock out mutants (J. Park and Kim 2015; Kuwabara and O'Neil 2001; Chandler et al. 2006; Chandler and Venditti 2005; Froese et al. 2015).

Results presented in Chapter 4 show that the C5 modified cobyric acid analogue is not recognised by the *C. elegans* transport protein(s). As previously stated, this was expected as these nematodes have no cobalamin biosynthesis genes and cannot use any corrinoid earlier than cobalamin. In some animals, e.g. humans (*H. sapiens*), rainbow trout (*O. mykiss*), and zebrafish (*D. rerio*), earlier intermediates can bind to the transport proteins, but cannot be used as they cannot be converted into cofactors in these organisms (Banerjee et al. 2009; Greibe, Fedosov, Sorensen, et al. 2012; Greibe, Fedosov, and Nexo 2012). The *C. elegans* gut transporter is not so promiscuous. Although it is possible that it will take up later intermediates than cobyric acid this is improbable as most transport proteins which do not recognise the earlier intermediates bind the nucleotide loop region, and it is only cobalamin which has a complete DMB loop (Mathews et al. 2007; Furger et al. 2013; Wuerges et al. 2006). Human HC recognises cobalamin and earlier intermediates but changes conformation depending on the bound ligand, binding cobalamin tighter than cobinamide (Furger et al. 2013). In fact, all the incomplete corrinoid binding transport proteins bind cobalamin more efficiently than the earlier intermediates (Furger et al. 2013; Greibe, Fedosov, and Nexo 2012; Greibe, Fedosov, Sorensen, et al. 2012). If the *C. elegans* transporter did bind the C5-cobyric acid analogue there would have been a high enough concentration of C5- BODIPY® TR-X cobyric acid in the nematode to image, but as it is not observed the *C. elegans* transporter is almost certainly cobalamin specific.

Upon uptake, the BODIPY® TR-X-ribose linked cobalamin localised exclusively in the coelomocytes when fed to *C. elegans*. As this is the first experiment of this kind, and only one of a handful of investigations into the relationship between cobalamin and *C. elegans*, there is little information available from which to draw conclusions. Coelomocytes are theorised to be rudimentary immune system, endocytosing foreign substances from the pseudocoelom (Altun and Hall 2009a). There are six in every nematode, in 3 pairs: behind the pharynx, near the vulva and by the anus. They contain many membrane bound vesicles and are known to endocytose injected foreign substances such as GFP, India ink etc. which are deposited in the lysosome for degradation or stored if that does not work (Fares and Greenwald 2001; Treusch et al. 2004; Fares and Grant 2002). Surprisingly, ablation of a few coelomocytes does not affect *C. elegans* fertility or survival, and even after a total loss the nematodes will develop and procreate (Fares and Greenwald 2001; Altun and Hall 2009a). It has been shown that yolk particles, which are not normally endocytosed by coelomocytes, can be induced to do so if tagged with GFP (Paupard et al. 2001). This is most probably due to the

presence of the GFP which adds credence to the argument that the BODIPY® TR-X-ribose linked cobalamin analogue is taken up into the coelomocytes because of the BODIPY® TR-X and not the cobalamin molecule.

C. elegans are cobalamin-dependent and display a loss of fertility, a shortened life span and an elongated life cycle when deficient (Bito et al. 2013). There are two cobalamin-dependent proteins identified in *C. elegans*, MCM and MS (Bito et al. 2013). The localisation of these proteins is ubiquitous, so why there is a link to the coelomocytes is currently a mystery. A possible explanation for this is that the cobalamin analogue may have been recognised by the gut transport protein as cobalamin but not by the next binding protein. This could either have resulted in the analogue being released in the pseudocoelom until the BODIPY® TR-X moiety was recognised as foreign and the molecule was targeted/ engulfed by the coelomocytes. Alternatively, the pseudocoelom transport protein also recognised the cobalamin analogue and transported it to the coelomocytes where the cobalamin is stored in the nematode, whereupon it was targeted into vesicles. The nature of these vesicles is open to speculation; they might be storage vesicles or they could be lysosomal/ degradation vesicles.

The BODIPY® TR-X-ribose linked analogue persists in the coelomocyte through many life stages. It is, however, not present in the ova at any stage, before or after expulsion from the adult (Chapter 4, Figure 4.7.1.1). Even though *C. elegans* is a hermaphrodite it still produces unfertilised oocytes which are subsequently fertilised in the spermatheca (Altun and Hall 2009b). Interestingly in *O. mykiss* (rainbow trout) the cobalamin binding protein, HIT, is present in immature oocytes but not in mature ones, although it was not investigated as to whether or not this coincided with a high cobalamin concentration (Greibe, Fedosov, Sorensen, et al. 2012). In *O. mykiss* the immature ova may require high levels of cobalamin where the mature ova do not, if the presence of the transporter correlates to that of cobalamin. In *C. elegans* the lack of fluorescence in the egg area suggests that the parent does not provide the ova or zygote with cobalamin, although the unidentified transporter may be present.

Overall, it is not clear whether the BODIPY® TR-X analogue is in the coelomocytes due to their accumulation of foreign material or because cobalamin is stored in the coelomocytes. *C. elegans* has no endoderm derived organs, like the liver, where cobalamin is stored in more complex organisms, e.g. humans, so cobalamin is likely to be stored somewhere novel (McGhee et al. 2014; Nielsen et al. 2012). To see if the analogues are localising to the same place as foreign compounds, nematodes fed with *E. coli* containing

BODIPY® TR-X-ribose linked cobalamin could be injected with India ink, and the fluorescence of the BODIPY® TR-X and the India ink can be compared *in vivo* (Fares and Greenwald 2001). To determine if the BODIPY® TR-X-ribose linked cobalamin is functioning as a cofactor *C. elegans* could be grown with the ribose analogue as the only source of cobalamin for many generations to ascertain whether a cobalamin deficiency phenotype is induced. Considering that *C. elegans* can survive without coelomocytes, feeding *E. coli* containing BODIPY® TR-X-ribose linked cobalamin to the nematode after coelomocyte ablation would allow the visualisation of the localisation of the analogue without the presence of the potential storage cell (Fares and Greenwald 2001; Altun and Hall 2009a).

6.4 Cobalamin analogue recognition in *Arabidopsis thaliana* and *Lepidium sativum*

Land plants do not synthesise or require cobalamin, owing to their lack of the cobalamin-dependent enzymes (Helliwell et al. 2011). The investigation into the uptake of cobalamin in plants was initiated due to data showing some plants could take up cobalamin and a growing concern that people who restricted their diets to omit meat (and eggs) could inadvertently self-impose cobalamin deficiency (Watanabe et al. 2013). It was thought that if plants could be 'fortified' with cobalamin, in the same way most breakfast cereals are, these people could obtain their necessary cobalamin quota from plants. *L. sativum* was shown in Chapter 4 to take up commercial cobalamin (cyanocobalamin) in a concentration dependent manner. Therefore it can be fortified with cobalamin. In order to know where the cobalamin localised in the plant, Oregon green® 514 analogues were added to the media. Remarkably, the fluorescence was only visible in the *L. sativum* vacuoles of the cotyledons and not in the roots or hypocotyl (data not shown). *A. thaliana* did not take it up at all. It is surprising that there is a difference in uptake between the two plants. There should be no cobalamin specific selectivity variation in plants at all and yet these two plants have root absorption regulation differences, which result in cobalamin exclusion in *A. thaliana*. It is feasible that the analogues are being taken up promiscuously in *L. sativum* by a separate transport system which is either under tighter regulation in *A. thaliana* or is not present.

Ultimately these experiments prove that *L. sativum* will take up cobalamin into the cotyledons. Therefore, in principle this proves that cobalamin enriched plants is a possible way to provide this vitamin to people with restricted diets which are prone to causing cobalamin insufficiency, potentially reducing the incidence of diet-mediated cobalamin

deficiency. For commercialisation and product development, the uptake of cobalamin in major cereal crops and other legumes should be investigated.

6.5 Corrinoid uptake, regulation and characterisation in *Mycobacterium tuberculosis*

A link between cobalamin and *M. tuberculosis* has been speculated for many years but there is much still unknown about the role of cobalamin in relation to the lifecycle of this bacterium (Corcino et al. 1971; Herbert 1983). Cobalamin, cobinamide and pseudocobalamin are all absorbed by *M. tuberculosis* in a concentration dependent way. The *ΔbacA* strain, a gene knockout of the one known cobalamin transporter BacA, shows a reduced uptake but not a complete loss of absorbance. The results presented in Chapter 5 show that there is uptake of cobalamin, cobinamide and pseudocobalamin, although cobalamin uptake is halved compared to the other two corrinoids. If the absorbance of these corrinoids was serendipitous then there should not be a tangible reduction of cobalamin absorption in relation to cobinamide or pseudocobalamin. As there is, this absorption is possibly due to a second uptake mechanism. This alternative route could favour cobinamide and pseudocobalamin absorption over cobalamin uptake, but as cobalamin is the corrinoid that *M. tuberculosis* produces when given cobinamide in the growth media, it is most likely to select cobalamin over the other two. Therefore, this alternative uptake passage could be negatively regulated by cobalamin but not by cobinamide or pseudocobalamin. This would mean that the alternative route can recognise all three corrinoids, but once internal cell cobalamin concentrations reach around 20 nM it is down regulated. Meanwhile, cobinamide and pseudocobalamin, unrecognised by the regulation system, continue to be absorbed.

As with *E. coli* and *S. enterica*, the C5 and the ribose analogues are both taken up in *M. tuberculosis*, although the uptake mechanisms are different. Curiously in *M. tuberculosis* the C5-BODIPY® TR-X cobyric acid analogue was taken up more readily than the ribose linked BODIPY® TR-X cobalamin analogue. The superior uptake of the C5-cobyric acid analogues may simply be due to their smaller size; *M. tuberculosis* has a single membrane but also thick waxy layer surrounding this which has been postulated to affect transport into and out of the cell (T. Smith et al. 2013). Owing to the hydrophobic nature of BODIPY® TR-X these analogues may be held up in this waxy coat more than cobyric acid or cobalamin would. This would mean that the smaller of the two slips through easier resulting in faster uptake.

It may be that the C5 analogue is taken up via the possible alternative uptake route, which limits cobalamin absorption (or does not absorb cobalamin as well as cobyric acid), as

well. If this is the case then the C5 analogue enters the *M. tuberculosis* cell via two routes, making the accumulation faster. The position of the fluorophore may also have an effect on the uptake: the BODIPY® TR-X ribose could interfere more with the uptake proteins than the BODIPY® TR-X at the C5 position on the macrocycle.

If there is a separate uptake pathway for earlier intermediates of cobalamin synthesis in *M. tuberculosis*, this is a potentially exploitable route for drug trafficking into *M. tuberculosis* cells. Therapeutics linked to earlier cobalamin intermediates introduced intravenously to humans will not be recognised and taken up into healthy human cells, as TCII only binds cobalamin (Nielsen et al. 2012). This means the drug-corrinoid conjugate will specifically enter infecting cells. Some cancers have also been implicated in taking up earlier cobalamin intermediates, so treatment of these cancers is another possible application for early intermediate conjugated therapeutics (Waibel et al. 2008; Sah et al. 2014). The viability of such compounds can be tested in cell cultures of *M. tuberculosis* infected macrophages or cancer cell lines, as well as in model organisms, e.g. mice.

6.6 Corrinoid functionality in *Mycobacterium tuberculosis*

M. tuberculosis has never been shown to synthesise cobalamin *de novo* although it has all of the cobalamin biosynthesis genes apart from *cobF*. It may be possible for *de novo* synthesis of cobalamin to occur if another, undefined, protein compensates for CobF. However, a cobalamin-dependent mutant of *M. tuberculosis*, *ΔmetE*, cannot grow without exogenous supplementation of corrinoids (Chapter 5, Section 5.2.1). Cobyric acid and cobinamide can both rescue *ΔmetE* which proves that these intermediates can be transported into the bacterium and converted into an active cofactor form, capable of supporting growth. This early intermediate rescue is dependent on the presence of the cobalt ion as HBAH, the synthetic cobalt less equivalent of cobyric acid, does not elicit growth of the bacteria. It is not explicit as to whether the cobalt-less intermediates are taken up into the *M. tuberculosis*. The lack of cobalt in the media may have prevented the insertion of cobalt into these early intermediates, and, therefore, their conversion into cofactor form. To discover if *M. tuberculosis* is capable of inserting cobalt into early intermediates, such as HBAH, cobalt can be added to the media of a culture containing *ΔmetE*. If the bacteria reproduce then the HBAH has been converted into an active cofactor form.

6.7 Cobalamin binding proteins and corrinoid mediated regulation in *Mycobacterium tuberculosis*

In the RNA sequencing experiment (Chapter 5, 5.5), the expectation was that the riboswitch regulated genes *metE* and *ppe2* would be highlighted, but only the *ppe2* gene showed a significant increase in transcripts in the presence of cobinamide, and only *ΔmetH* and *ΔmetE* strains showed increases with cobalamin. The effect on the *metE* gene was not as distinct. The RNA transcripts decreased in the *ΔbacA* strain, but increased in the WT and H37Rv strains with both cobinamide and cobalamin. This difference could be a result of lower internal concentration of corrinoid in *ΔbacA* not being sufficient to cause the same response as in the WT and H37Rv strains.

Curiously, *ilvC*, the gene encoding a ketol acid reductoisomerase, was downregulated in all of the tested *M. tuberculosis* strains supplemented with both cobalamin and cobinamide. This was the only gene which was significantly affected by the presence of the corrinoids, yet without any known link to cobalamin homeostasis in *M. tuberculosis*. In itself this is not particularly interesting, but the protein was also identified as binding to a cobalamin column. This would mean that the gene is suppressed by the corrinoids and yet the protein binds cobalamin. It could be that this gene and protein have some unknown connection to cobalamin. At present all that is known is that ketol acid reductoisomerase is involved in amino acid synthesis.

The two other proteins highlighted in the cobalamin binding column experiment were Glutamine synthetase I and Citrate synthase I, both of which are involved in MCM independent propionate metabolism pathways in *M. tuberculosis* (Figure 1.7.1) (Savvi et al. 2008). Strangely, the cobalamin independent methionine synthase MetE also purified off the cobalamin binding column. These three proteins are all involved in cobalamin independent processes, but were implicated in cobalamin binding. There is no obvious reason as to why they should bind cobalamin.

This cobalamin column binding experiment excludes membrane bound cobalamin binding proteins of which there is at least one known, BacA. Although some proteins were identified as potential cobalamin binding, there are none that could unequivocally be defined as cobalamin binding proteins.

6.8 Developing analogue synthesis

6.8.1 Clickable SAM analogues

There is a potential problem with the C5 modified corrin-ring analogues synthesised. The linker used to conjugate the fluorophore to the corrin-ring can theoretically be cleaved by native proteases. Therefore, there is a need to change this linker to a non-hydrolysable equivalent. A possible alternative is to transfer a terminal alkyne moiety to the C5 position instead of the allyl group. This alkyne can be used in copper catalysed click chemistry with terminal azides which would form a non-labile bond between the corrin macrocycle and the conjugate (Kolb and Sharpless 2003). Fluorophores with terminal azides are available for purchase so fluorescent analogue synthesis is still possible. However, as these alkyne groups are rapidly hydrolysed, the hydrocarbon chain will probably have to be longer to increase stability. Consequently, the activity of the biosynthesis proteins may be compromised because of the increase in size of the transferred group.

6.8.2 C5-analogue synthesis

The modifications at the C5 position of the corrin-ring can have some adverse effects on the reaction rate of some of the biosynthesis enzymes. This issue was touched upon relating to CobH in Chapter 3. The T85A substitution of CobH increased the volume of the active site but, even though it catalysed the reaction, the allyl group was strained. A T85G substitution would increase the space even more as glycine is smaller than alanine, but this may have reduced activity compared to the T85A mutant. The T85 residue is part of an α helix, and is, in fact, close to the terminus. It may be possible to remodel this helix so it terminates earlier, generating a larger active site volume which would allow the C5 analogues to fit better, especially if the longer alkyne moieties are conjugated. The early termination of this α helix may result in a lack of enzyme functionality and any new mutants must be tested for activity.

6.8.3 Early intermediate analogue biomedical applications

It has been shown in mice that analogues of cobalamin with radioactive conjugates attached to the sidechains no longer bind TCII or IF but retain recognition by HC. These analogues have increased specificity of tumour labelling in mice compared to analogues which can be bound by all three uptake proteins (Waibel et al. 2008). In this previous experiment the analogue is a full cobalamin with a secondary macrocycle attached to the *b*

sidechain coordinating ^{99m}Tc with a butyl linker from the amide. The authors of this paper postulated that tumours have an increased expression of HC to supply the cobalamin needed for a high proliferation rate (Waibel et al. 2008). HC is the cobalamin transporter which does bind incomplete corrinoids, such as cobinamide, and it was shown that competition between cobinamide and the ^{99m}Tc analogue results in the complete blocking of tumour binding of the analogue as effectively as cobalamin competition. Immunohistochemical staining of HC has been shown to localise to the surface membrane and cytoplasm in various tumours of the digestive tract, urinary tract, lung, salivary gland, and breast (Kim et al. 1993). This means that C5-cobyric acid analogues can be used to specifically target tumours that express a high level of HC as a way of increasing their cobalamin uptake. A paper published in 1971 postulated the link between cobalamin deficiency and leukaemia retardation after observing a patient diagnosed with both chronic myeloid leukaemia and pernicious anaemia before and after withholding therapeutic cobalamin (Corcino et al. 1971). In 1983 the same group showed that an aniline conjugated to a sidechain of cobalamin could treat acute myelogenous leukaemia, again in a single patient (Herbert 1983). These investigations have not been furthered since.

The same overexpression of HC may occur in *M. tuberculosis* infected cells, and, as *M. tuberculosis* absorbs the C5-cobyric acid analogues more efficiently than cobalamin analogues, an early intermediate conjugated drug could be specifically targeted to *M. tuberculosis* infected cells. Current knowledge of the relationship between cobalamin and *M. tuberculosis* is that the biosynthesis genes *cobK* and *cobL* are expressed during *M. tuberculosis* infection of immune-competent mice, and *cobI* and *Rv2067c* (a putative precorrin-3b methyltransferase) have been detected in human pulmonary patients (Gopinath, Moosa, et al. 2013). This is odd as the results presented in Chapter 5 shows that *M. tuberculosis* does not synthesise cobalamin *de novo* and these genes are not required in completing cobalamin biosynthesis from scavenged corrinoids. This combined data begs the question, has *M. tuberculosis* somehow compensated for the lack of *cobF* and is able to synthesise cobalamin during host infection?

Tumour specificity linked to HC expression could be immensely beneficial in targeted drug-conjugated corrinoid therapeutics. HC is the only human uptake protein which binds incomplete corrinoids, so indiscriminate absorbance could be removed by using an earlier intermediate conjugate instead of cobalamin. The incomplete corrinoid-drug conjugate should localise to the tumour(s) in much the same as the radioactive sidechain *b* conjugate did (Waibel et al. 2008). In terms of therapeutics, this means that a broader range of drugs can be used, not just the cyto-accumulation toxic ones. Using incomplete corrinoid conjugates is also beneficial in drug conjugate biosynthesis as there are fewer steps required to make them. These corrinoid analogues would have to be tested in different disease backgrounds to tailor the best intermediate for drug conjugation in each case.

6.9 Concluding remarks

The investigations reported in this thesis have shown a novel mechanism of cobalamin modification which can be further adapted to carry a fluorophore. The synthesised analogues were used to elucidate cobalamin uptake and storage across different living systems including bacteria, plants and animals. The fluorophore conjugated analogues can be used to further understanding the role of cobalamin in health and disease.

The synthesised analogues can also be recognised and used by *M. tuberculosis* which ultimately could lead to a new therapeutic role for cobalamin in targetting drugs to cells with high proliferation rates. Although the analogues currently have fluorophore conjugates, these can be replaced with drugs, such as Rifampicin. Rifampicin is a similar size to BODIPY ® TR-X (Figure 6.9.1); therefore, the activities of the fluorophore conjugates are indicative of how the drug equivalents could work.

BODIPY® TR-X succinimidyl ester   **Rifampicin**

Figure 6.9.1: The structures of BODIPY® TR-X and Rifampicin The BODIPY® TR-X fluorophore used (left) compared to the rifampicin, a standard first-line treatment drug for *M. tuberculosis* infection.
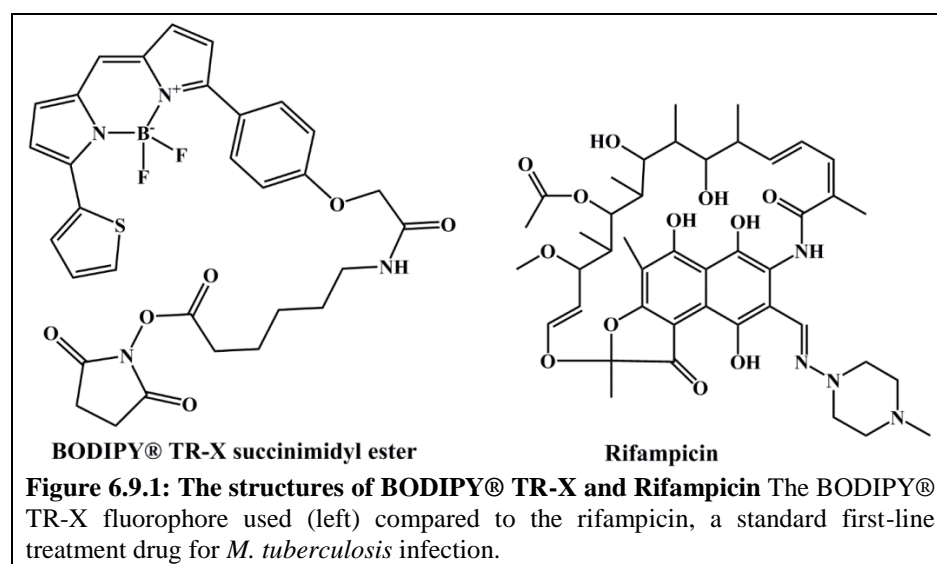
Bibliography

Bibliography

Allen, LH. 2008. Causes of Vitamin B 12 and Folate Deficiency. *Food Nutr. Bull.* 29 (2): 20–34.

Altun, ZF, and DH Hall. 2009a. Coelomocyte System. In WormAtlas.

2009b. Introduction In WormAtlas.

Ames, BN. 2001. DNA Damage from Micronutrient Deficiencies Is Likely to Be a Major Cause of Cancer. *Mutat. Res. Mol. Mech. Mutagen.* 475 (1–2): 7–20.

Andrès, E, NH Loukili, E Noel, G Kaltenbach, M Ben Abdelgheni, AE Perrin, M Noblet-Dick, F Maloisel, J-L Schlienger, and J-F Blicklé. 2004. Vitamin B12 (Cobalamin) Deficiency in Elderly Patients. *Can. Med. Assoc. J.* 171 (3): 251–59.

Anton, DL, HPC Hogenkamp, TE Walker, and NA Matwiyoff. 1980. Carbon- 13 Nuclear Magnetic Resonance Studies of the Monocarboxylic Acids of Cyanocobalamin . Assignments of the B-, D -, and E-Monocarboxylic Acids. *J. Am. Chem. Soc.* 102 (7): 2215–2219.

Bakhshi, SS, J Hawker, and S Ali. 2010. The Epidemiology of Tuberculosis by Ethnic Group in Birmingham and Its Implications for Future Trends in Tuberculosis in the UK. *Ethn. Health* 2 (3): 147–53.

Banerjee, R, C Gherasim, and D Padovani. 2009. The Tinker, Tailor, Soldier in Intracellular B12 Trafficking. *Curr. Opin. Chem. Biol.* 13 (4): 484–91.

Banerjee, R, and SW Ragsdale. 2003. The Many Faces of Vitamin B12: Catalysis by Cobalamin-Dependent Enzymes. *Annu. Rev. Biochem.* 72 (January): 209–47.

Bansal-mutalik, R, and H Nikaido. 2014. Mycobacterial Outer Membrane Is a Lipid Bilayer and the Inner Membrane Is Unusually Rich in Diacyl Phosphatidylinositol Dimannosides. *Proc. Natl. Acad. Sci. U. S. A.* 111 (13): 4958–63.

Barron, M. 1933. Pernicious Anemia and Tuberculosis: Is There an Antagonism? *J. Am. Med. Assoc.*, 1590–92.

Battaglia-Hsu, S, N Akchiche, N Noel, J-M Alberto, E Jeannesson, CE Orozco-Barrios, D Martinez-Fong, J-L Daval, and J-L Guéant. 2009. Vitamin B12 Deficiency Reduces Proliferation and Promotes Differentiation of Neuroblastoma Cells and up-Regulates PP2A, proNGF, and TACE. *Proc. Natl. Acad. Sci. U. S. A.* 106 (51): 21930–35.

Bito, T, Y Matsunaga, Y Yabuta, T Kawano, and F Watanabe. 2013. Vitamin B12 Deficiency in Caenorhabditis Elegans Results in Loss of Fertility, Extended Life Cycle, and Reduced Lifespan. *Fed. Eur. Biochem. Soc. Open Bio* 3 (January). The Authors: 112–17.

Bito, T, Y Yabuta, T Ichiyanagi, T Kawano, and F Watanabe. 2014. A Dodecylamine Derivative of Cyanocobalamin Potently Inhibits the Activities of Cobalamin-Dependent Methylmalonyl-CoA Mutase and Methionine Synthase of Caenorhabditis Elegans. *Fed. Eur. Biochem. Soc. Open Bio* 4: 722–29.

Blanche, F, B Cameron, J Crouzet, L Debussche, D Thibaut, M Vuilhorgne, FJ Leeper,

and AR Battersby. 1995. Vitamin B12: How the Problem of Its Biosynthesis Was Solved. *Angew. Chemie Int. Ed.* 34 (1): 383–411.

Boshoff, HIM, and CE Barry. 2005. Tuberculosis - Metabolism and Respiration in the Absence of Growth. *Nat. Rev. Microbiol.* 3 (January): 70–80.

Boyes, DC, AM Zayed, R Ascenzi, AJ McCaskill, NE Hoffman, KR Davis, and J Görlach. 2001. Growth Stage-Based Phenotypic Analysis of Arabidopsis: A Model for High Throughput Functional Genomics in Plants. *Plant Cell* 13 (7): 1499–1510.

Brenner, S. 1974. The Genetics of Caenorhabditis Elegans. *Genetics* 77 (1): 71–94.

Brown, KL. 2005. Chemistry and Enzymology of Vitamin B 12. *Chem. Rev.* 105: 2075–2149.

Cadieux, N, C Bradbeer, E Reeger-Schneider, W Köster, AK Mohanty, MC Wiener, and RJ Kadner. 2002. Identification of the Periplasmic Cobalamin-Binding Protein BtuF of Escherichia Coli. *J. Bacteriol.* 184 (3): 706–17.

Chanarin, I, and E Stephenson. 1988. Vegetarian Diet and Cobalamin Deficiency: Their Association with Tuberculosis. *J. Clin. Pathol.* 41 (7): 759–62.

Chandler, RJ, V Aswani, MS Tsai, M Falk, N Wehrli, S Stabler, R Allen, M Sedensky, HH Kazazian, and CP Venditti. 2006. Propionyl-CoA and Adenosylcobalamin Metabolism in Caenorhabditis Elegans: Evidence for a Role of Methylmalonyl-CoA Epimerase in Intermediary Metabolism. *Mol. Genet. Metab.* 89 (1–2): 64–73.

Chandler, RJ, and CP Venditti. 2005. Genetic and Genomic Systems to Study Methylmalonic Acidemia. *Mol. Genet. Metab.* 86 (1–2): 34–43.

Christensen, EI, JO Moskaug, H Vorum, C Jacobsen, TE Gundersen, a Nykjaer, R Blomhoff, TE Willnow, and SK Moestrup. 1999. Evidence for an Essential Role of Megalin in Transepithelial Transport of Retinol. *J. Am. Soc. Nephrol.* 10 (4): 685–95.

Christensen, EI, and TE Willnow. 1999. Essential Role of Megalin in Renal Proximal Tubule for Vitamin Homeostasis. *J. Am. Soc. Nephrol.* 10 (Figure 1): 2224–36.

Clardy, SM, DG Allis, TJ Fairchild, and RP Doyle. 2011. Vitamin B 12 in Drug Delivery : Breaking through the Barriers to a B 12 Bioconjugate Pharmaceutical. *Expert Opin. Drug Delivery*, 1–14.

Coburn, C, E Allman, P Mahanti, A Benedetto, F Cabreiro, Z Pincus, F Matthijssens, et al. 2013. Anthranilate Fluorescence Marks a Calcium-Propagated Necrotic Wave That Promotes Organismal Death in C. Elegans. *PLoS Biol.* 11 (7): 1–17.

Cohen, GN. 2014. Biosynthesis of Cobalamins Including Vitamin B12. In *Microb. Biochem. Third Ed.*, 555–65. Springer.

Cole, ST, R Brosch, J Parkhill, T Garnier, C Churcher, D Harris, S V. Gordon, et al. 1998. Deciphering the Biology of Mycobacterium Tuberculosis from the Complete Genome Sequence. *Nature* 393 (June): 537–44.

Cole, ST, K Eiglmeier, J Parkhill, KD James, NR Thomson, PR Wheeler, N Honoré, et al. 2001. Massive Gene Decay in the Leprosy Bacillus. *Nature* 409 (6823): 1007–11.

Corcino, JJ, R Zalusky, M Greenberg, and V Herbert. 1971. Coexistence of Pernicious

Anaemia and Chronic Myeloid Leukaemia: An Experiment of Nature Involving Vitamin B12 Metabolism. *Br. J. Haematol.* 20 (5). Wiley Online Library: 511–20.

Corsi, A, B Wightman, and M Chalfie. 2016. A Transparent Window into Biology: A Primer on Caenorhabditis Elegans. *Wormbook*. Accessed August 1.

Croft, MT, AD Lawrence, E Raux-Deery, MJ Warren, and AG Smith. 2005. Algae Acquire Vitamin B12 through a Symbiotic Relationship with Bacteria. *Nat. Lett.* 438 (7064): 90–93.

Dawes, SS, DF Warner, L Tsenova, J Timm, JD Mckinney, G Kaplan, H Rubin, and V Mizrahi. 2003. Ribonucleotide Reduction in Mycobacterium Tuberculosis : Function and Expression of Genes Encoding Class Ib and Class II Ribonucleotide Reductases. *Infect. Immun.* 71 (11): 6124–31.

Deery, E, S Schroeder, AD Lawrence, SL Taylor, A Seyedarabi, J Waterman, KS Wilson, et al. 2012. Supplementary-An Enzyme-Trap Approach Allows Isolation of Intermediates in Cobalamin Biosynthesis. *Nat. Chem. Biol.* 8 (11): 933–40.

Deery, E, S Schroeder, AD Lawrence, SL Taylor, A Seyedarabi, J Waterman, KS Wilson, D Brown, and MA Geeves. 2013. An Enzyme-Trap Approach Allows Isolation of Intermediates in Cobalamin Biosynthesis. *Nat. Chem. Biol.* 8 (11): 933–40.

Delano, WL. 2002. The PyMOL Molecular Graphics System.

Dereven'kov, IA, DS Salnikov, R Silaghi-Dumitrescu, S V. Makarov, and OI Koifman. 2016. Redox Chemistry of Cobalamin and Its Derivatives. *Coord. Chem. Rev.* 309: 68–83.

Dobson, CM, T Wai, D Leclerc, H Kadir, M Narang, TJ Hudson, DS Rosenblatt, and RA Gravel. 2002. Identification of the Gene Responsible for the cblB Complementation Group of Vitamin B12 -Dependent Methylmalonic Aciduria. *Hum. Mol. Genet.* 11 (26): 3361–69.

Edelstein, AD, MA Tsuchida, N Amodaj, H Pinkard, RD Vale, and N Stuurman. 2014. Advanced Methods of Microscope Control Using µManager Software. *J. Biol. Methods* 1 (2): 1–10.

Emsley, P, B Lohkamp, WG Scott, and K Cowtan. 2010. Features and Development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 66 (4): 486–501.

Ernst, JD. 2012. The Immunological Life Cycle of Tuberculosis. *Nat. Rev. Immunol.* 12 (August). Nature Publishing Group: 581–91.

Eschenmoser, A. 1988. Vitamin B12 : Experiments Concerning the Origin of Its Molecular Structure. *Angew. Chem. Int. Ed. Engl.* 27 (1): 5–39.

Fares, H, and B Grant. 2002. Deciphering Endocytosis in Caenorhabditis Elegans. *Traffic* 3 (1): 11–19.

Fares, H, and I Greenwald. 2001. Genetic Analysis of Endocytosis in Caenorhabditis Elegans: Coelomocyte Uptake Deficient Mutants. *Mol. Cell. Biol.*

Fedosov, SN, NU Fedosova, B Kraeutler, E Nexo, and TE Petersen. 2007. Mechanisms of Discrimination between Cobalamins and Their Natural Analogues during Their Binding

to the Specific B12-Transporting Proteins. *Biochemistry* 46 (21): 6446–58.

Fenech, M. 2001. The Role of Folic Acid and Vitamin B12 in Genomic Stability of Human Cells. *Mutat. Res. Mol. Mech. Mutagen.* 475 (1): 57–67.

Fieber, W, B Hoffmann, W Schmidt, E Stupperich, R Konrat, and B Kraeutler. 2002. Pseudocoenzyme B 12 and Adenosyl-Factor A : Electrochemical Synthesis and Spectroscopic Analysis of Two Natural B 12 Coenzymes with Predominantly − Base-off × Constitution. *Helv. Chim. Acta* 85.

Flodh, H, and S Ullberg. 1968. Accumulation of Labelled Vitamin B12 in Some Transplanted Tumours. *Int. J. Cancer* 3: 694–99.

Frank, S, AA Brindley, E Deery, P Heathcote, AD Lawrence, HK Leech, RW Pickersgill, and MJ Warren. 2005. Anaerobic Synthesis of Vitamin B12: Characterization of the Early Steps in the Pathway. *Biochem. Soc. Trans.* 33 (Pt 4): 811–14.

Froese, DS, and RA Gravel. 2010. Genetic Disorders of Vitamin B12 Metabolism: Eight Complementation Groups--Eight Genes. *Expert Rev. Mol. Med.* 12 (November): e37.

Froese, DS, J Kopec, F Fitzpatrick, M Schuller, TJ McCorvie, R Chalk, T Plessl, et al. 2015. Structural Insights into the MMACHC-MMADHC Protein Complex Involved in Vitamin B12 Trafficking. *J. Biol. Chem.* 290 (49): 29167–77.

Furger, E, DC Frei, R Schibli, E Fischer, and AE Prota. 2013. Structural Basis for Universal Corrinoid Recognition by the Cobalamin Transport Protein Haptocorrin * 288 (35): 25466–76.

Fyfe, JA, and HC Friedmann. 1969. Pseudo B12 Biosynthesis. *J. Biol. Chem.* 131 (7): 1667–72.

Gallo, S, M Oberhuber, RKO Sigel, and B Kräutler. 2008. The Corrin Moiety of Coenzyme B12 Is the Determinant for Switching the btuB Riboswitch of E. Coli. *ChemBioChem* 9 (9): 1408–14.

Galperin, MY, and N V. Grishin. 2000. The Synthetase Domains of Cobalamin Biosynthesis Amidotransferases CobB and CobQ Belong to a New Family of ATP-Dependent Amidoligases, Related to Dethiobiotin Synthetase. *Proteins Struct. Funct. Genet.* 41 (2): 238–47.

Gonzalez, JC, R V Banerjee, S Huang, JS Sumner, and RG Matthews. 1992. Comparison of Cobalamin-Independent and Cobalamin-Dependent Methionine Synthases from Escherichia Coli: Two Solutions to the Same Chemical Problem. *Biochemistry*, 6045–56.

Gopinath, K, A Moosa, V Mizrahi, and DF Warner. 2013. Vitamin B 12 Metabolism in Mycobacterium Tuberculosis, 1405–18.

Gopinath, K, C Venclovas, TR Ioerger, JC Sacchettini, JD McKinney, V Mizrahi, and DF Warner. 2013. A Vitamin B12 Transporter in Mycobacterium Tuberculosis. *Open Biol.* 3 (2): 120175.

Greibe, E, S Fedosov, and E Nexo. 2012. The Cobalamin-Binding Protein in Zebrafish Is an Intermediate between the Three Cobalamin-Binding Proteins in Human. *PLoS One* 7 (4).

Greibe, E, S Fedosov, BS Sorensen, SS Poulsen, and E Nexo. 2012. A Single Rainbow

Trout Cobalamin-Binding Protein Stands in for Three Human Binders. *J. Biol. Chem.* 287 (40): 33917–25.

Grossman, A. 2016. Nutrient Acquisition: The Generation of Bioactive Vitamin B12 by Microalgae. *Curr. Biol.* 26 (8): R319–21.

Gruber, K, B Puffer, and B Kraeutler. 2011. Vitamin B12-Derivatives—enzyme Cofactors and Ligands of Proteins and Nucleic Acids. *Chem Soc Rev* 40.

Hazra, AB, JLA Tran, TS Crofts, and ME Taga. 2013. Article Analysis of Substrate Specificity in CobT Homologs Reveals Widespread Preference for DMB , the Lower Axial Ligand of Vitamin B 12. *Chem. Biol.* 20 (10). Elsevier Ltd: 1275–85.

Heldt, D, AD Lawrence, M Lindenmeyer, E Deery, P Heathcote, SE Rigby, and MJ Warren. 2005. Aerobic Synthesis of Vitamin B 12 : Ring Contraction and Cobalt Chelation. *Biochem. Soc. Trans.*, 815–19.

Helliwell, KE, AD Lawrence, A Holzer, DJ Scanlan, MJ Warren, AG Smith, KE Helliwell, et al. 2016. Cyanobacteria and Eukaryotic Algae Use Different Chemical Variants of Vitamin B 12. *Curr. Biol.* 26: 1–10.

Helliwell, KE, GL Wheeler, KC Leptos, RE Goldstein, and AG Smith. 2011. Insights into the Evolution of Vitamin B 12 Auxotrophy from Sequenced Algal Genomes. *Mol. Biol. Evol.* 28 (10): 2921–33.

Herbert, V. 1983. The Inhibition of Some Cancers and the Promotion of Others by Folic Acid, Vitamin B12, and Their Antagonists. In *Bristol Myers Nutr. Symp.* Vol. v. 2.

Hermanson, GT. 2013. *Bioconjugate Techniques*. Edited by J Audet. 3rd ed. Elsevier Inc. doi:10.1016/B978-0-12-382239-0.00025-X.

Hinkeldey, B, A Schmitt, and G Jung. 2008. Comparative Photostability Studies of BODIPY and Fluorescein Dyes by Using Fluorescence Correlation Spectroscopy. *ChemPhysChem* 9 (14): 2019–27.

Hodgkin, DC, J Kamper, M Mackay, J Pickworth, KN Trueblood, and JG White. 1956. Structure of Vitamin B12. *Nature* 177: 64–66.

Hogenkamp, HPC, DA Collins, CB Grissom, and FG West. 1999. Diagnostic and Therapeutic Analogues of Cobalamin. In *Chem. Biochem. B12; John Wiley Sons New York,* 385–410.

Jancarik, J, and S-H Kim. 2000. Sparse Matrix Sampling: A Screening Method for Crystallization of Proteins. *J. Appl. Crystallogr.* 24: 409–11.

Jeter, RM, BM Olivera, JR Roth, RM Jeter, BM Olivera, and JR Roth. 1984. Salmonella Typhimurium Synthesizes Cobalamin (Vitamin B12) de Novo under Anaerobic Growth Conditions . Salmonella Typhimurium Synthesizes Cobalamin (Vitamin B12) De Novo Under Anaerobic Growth Conditions 159 (1): 206–13.

Kadish, K, KM Smith, and R Guilard. 2003. *The Porphyrin Handbook: The Iron and Cobalt Pigments: Biosynthesis, Structure and Degradation, Volume 12,*. Academic press. https://books.google.co.uk/books?id=xjqtusETxCIC&printsec=frontcover&hl=ja#v=one page&q&f=false.

Kaletta, T, and MO Hengartner. 2006. Finding Function in Novel Targets: C. Elegans as a Model Organism. *Nat. Rev. Drug Discov.* 5 (5): 387–98.

Kanehisa, M, and S Goto. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 28 (1): 27–30.

Kanehisa, M, Y Sato, M Kawashima, M Furumichi, and M Tanabe. 2016. KEGG as a Reference Resource for Gene and Protein Annotation. *Nucleic Acids Res.* 44 (D1): D457–62.

Kim, Y, K Ogawa, Y Wakatsuki, Y Nakashima, K Takasu, H Yamabe, and E Immunoreactivity. 1993. Immunohistochemical Localisation of Vitamin B12 R-Binder in Uterine Cervical and Endometrial Adenocarcinomas. *Acta Hist Cyto* 26 (6): 507–14.

Kolb, HC, and KB Sharpless. 2003. The Growing Impact of Click Chemistry on Drug Discovery. *Drug Discov. Today* 8 (24): 1128–37.

Kolberg, M, KR Strand, P Graff, and KK Andersson. 2004. Structure, Function, and Mechanism of Ribonucleotide Reductases. *Biochim. Biophys. Acta* 1699: 1–34.

Kolhouse, JF, SP Stabler, and RH Allen. 1993. Identification and Perturbation of Mutant Human Fibroblasts Based on Measurements of Methylmalonic Acid and Total Homocysteine in the Culture Media. *Arch. Biochem. Biophys.* 303 (2): 355–60.

Kräutler, B. 2006. Cobalt: B 12 Enzymes & Coenzymes. *Encycl. Inorg. Chem.* 12 (Figure 3).

Kräutler, B, and B Puffer. 2012. Handbook of Porphyrin Sciences. In *21-25*, edited by KM Kadish, KM Smith, and R Guilard, 218. World Scientific Publishing Co. Pte. Ltd.

Kuwabara, P, and N O'Neil. 2001. The Use of Functional Genomics in C . Elegans for Studying Human Development and Disease. *J. Inherit. Metab. Dis.* 24: 127–38.

Kuzminski, BAM, EJ Del Giacco, RH Allen, SP Stabler, and J Lindenbaum. 2016. Effective Treatment of Cobalamin Deficiency With Oral Cobalamin 92 (4): 1191–98.

Lai, SC, Y Nakayama, JM Sequeira, and E V. Quadros. 2011. Down-Regulation of Transcobalamin Receptor TCblR/CD320 by siRNA Inhibits Cobalamin Uptake and Proliferation of Cells in Culture. *Exp. Cell Res.* 317 (11): 1603–7.

Lee, M, and CB Grissom. 2009. Design, Synthesis, and Characterization of Fluorescent Cobalamin Analogues with High Quantum Efficiencies. *Org. Lett.* 11 (12): 2499–2502.

Leiper, J. 2010. Chlorophyll a Fluorescence Measurements in Plant Biology.

Lewinson, O, AT Lee, KP Locher, and DC Rees. 2010. A Distinct Mechanism for the ABC Transporter BtuCD – BtuF Revealed by the Dynamics of Complex Formation. *Nat. Publ. Gr.* 17 (3). Nature Publishing Group: 332–38.

Lildballe, DL, E Mutti, H Birn, and E Nexo. 2012. Maximal Load of the Vitamin B12 Transport System: A Study on Mice Treated for Four Weeks with High-Dose Vitamin B12 or Cobinamide. *PLoS One* 7 (10): e46657.

Lobo, S a L, A Brindley, MJ Warren, and LM Saraiva. 2009. Functional Characterization of the Early Steps of Tetrapyrrole Biosynthesis and Modification in Desulfovibrio Vulgaris Hildenborough. *Biochem. J.* 420 (2): 317–25.

Maggio-Hall, LA, and JC Escalante-Semerena. 1999. In Vitro Synthesis of the Nucleotide Loop of Cobalamin by *Salmonella Typhimurium* Enzymes. *Proc. Natl. Acad. Sci. U. S. A.* 96 (21): 11798–803.

Marsh, ENG. 1999. Coenzyme B12 (Cobalamin) Dependent Enzymes. *Essays Biochem.* 34: 139–54.

Martens, JH, H Barg, MJ Warren, and D Jahn. 2002. Microbial Production of Vitamin B12. *Appl. Microbiol. Biotechnol.* 58 (3): 275–85.

Martin, BD, and RG Finke. 1992. Bonds and Bond Dissociation Enthalpies : A Homolysis Rate Enhancement Following One-Antibonding-Electron Reduction of Methylcobalamin, no. 4: 585–92.

Mathews, FS, MM Gordon, Z Chen, KR Rajashankar, SE Ealick, DH Alpers, and N Sukumar. 2007. Crystal Structure of Human Intrinsic Factor: Cobalamin Complex at 2.6-A Resolution. *Proc. Natl. Acad. Sci. U. S. A.* 104 (44): 17311–16.

May, R, NJ Loman, AS Haines, MJ Pallen, C Boehnisch, CW Penn, CH Lee, and J Kim. 2009. The Genome Sequence of E. Coli OP50. *Worm Breeders Gaz.*

McEwan, JF, HS Veitch, and GJ Russell-Jones. 1999. Synthesis and Biological Activity of Ribose-5'-Carbamate Derivatives of Vitamin B 12. *Bioconjug. Chem.* 10 (6). American Chemical Society: 1131–36.

McFerrin, MB, and EH Snell. 2002. The Development and Application of a Method to Quantify the Quality of Cryoprotectant Solutions Using Standard Area-Detector X-Ray Images. *J. Appl. Crystallogrphy* 35: 538–45.

McGhee, JD, T Wiesenfahrt, and AE Dineen. 2014. The Caenorhabditis Elegans Intestine. In *Transl. Res. Discov. Gastroenterol. Organog. to Dis.*, edited by DL Gumucio, LC Samuelson, and JR Spence. John Wiley & Sons.

Mclean, GR, PM Pathare, DS Wilbur, AC Morgan, CS Woodhouse, JW Schrader, and J Ziltener. 1997. Cobalamin Analogues Modulate the Growth of Leukemia Cells in Vitro Cobalamin Analogues Modulate the Growth of Leukemia Cells in Vitro1. *Cancer*, no. 16: 4015–22.

McPherson, A. 1976. Crytsllization of Proteins from Polyethylene Glycol. *J. Biol. Chem.* 251 (20): 6300–6303.

Moestrup, SK, H Birn, PB Fischer, CM Petersen, PJ Verroust, RB Sim, EI Christensen, and E Nexo. 1996. Megalin-Mediated Endocytosis of Transcobalamin-Vitamin-B12 Complexes Suggests a Role of the Receptor in Vitamin-B12 Homeostasis. *Proc. Natl. Acad. Sci. U. S. A.* 93 (16): 8612–17.

Moore, SJ, R Biedendieck, AD Lawrence, E Deery, MJ Howard, SEJ Rigby, and MJ Warren. 2013. Characterization of the Enzyme CbiH 60 Involved in Anaerobic Ring Contraction of the Cobalamin (Vitamin B 12) Biosynthetic 288 (1): 297–305.

Moore, SJ, AD Lawrence, R Biedendieck, E Deery, S Frank, MJ Howard, SEJ Rigby, and MJ Warren. 2013. Elucidation of the Anaerobic Pathway for the Corrin Component of Cobalamin (Vitamin B12). *Proc. Natl. Acad. Sci. U. S. A.* 110 (37): 14906–11.

Moore, SJ, and MJ Warren. 2012. The Anaerobic Biosynthesis of Vitamin B12. *Biochem. Soc. Trans.* 40 (3): 581–86.

Murooka, Y, Y Piao, P Kiatpapan, and M Yamashita. 2003. Production of Tetrapyrrole Compounds and Vitamin B12 Using Genetically Engineering of Propionibacterium Freudenreichii: A Review. *EDP Sci.* 85 (10): 9–22.

Murshudov, GN, AA Vagin, and EJ Dodson. 1997. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 53 (3). International Union of Crystallography: 240–55.

Nielsen, MJ, MR Rasmussen, CBF Andersen, E Nexø, and SK Moestrup. 2012. Vitamin B12 Transport from Food to the Body's Cells--a Sophisticated, Multistep Pathway. *Nat. Rev. Gastroenterol. Hepatol.* 9 (6). Nature Publishing Group: 345–54.

Nykjaer, A, D Dragun, D Walther, H Vorum, C Jacobsen, J Herz, F Melsen, EI Christensen, and TE Willnow. 1999. An Endocytic Pathway Essential for Renal Uptake and Activation of the Steroid 25-(OH) Vitamin D3. *Cell* 96 (4): 507–15.

Orlando, R a, K Rader, F Authier, H Yamazaki, BI Posner, JJ Bergeron, and MG Farquhar. 1998. Megalin Is an Endocytic Receptor for Insulin. *J. Am. Soc. Nephrol.* 9 (10): 1759–66.

Park, J, and J Kim. 2015. Characterization of a B12 Trafficking Chaperone Protein from Caenorhabditis Elegans. *Protein Pept Lett* 22 (1): 31–38.

Park, SH, and a Bendelac. 2000. CD1-Restricted T-Cell Responses and Microbial Infection. *Nature* 406 (August): 788–92.

Pathare, PM, DS Wilbur, S Heusser, E V Quadros, P McLoughlin, and a C Morgan. 1996. Synthesis of Cobalamin-Biotin Conjugates That Vary in the Position of Cobalamin Coupling. Evaluation of Cobalamin Derivative Binding to Transcobalamin II. *Bioconjug. Chem.* 7 (2): 217–32.

Patiño, S, L Alamo, M Cimino, Y Casart, F Bartoli, MJ Garcia, and L Salazar. 2008. Autofluorescence of Mycobacteria as a Tool for Detection of Mycobacterium Tuberculosis. *J. Clin. Microbiol.* 46 (10): 3296–3302.

Paupard, M, A Miller, B Grant, D Hirsh, and DH Hall. 2001. Immuno-EM Localization of GFP-Tagged Yolk Proteins in C . Elegans Using Microwave Fixation. *J. Histochem. Cytochem.* 49 (8): 949–56.

Ramagopalan, S V, R Goldacre, A Skingsley, C Conlon, and MJ Goldacre. 2013. Associations between Selected Immune-Mediated Diseases and Tuberculosis : Record-Linkage Studies. *BMC Med.* 11 (1). BMC Medicine: 1.

Raux, E, A Lanois, F Levillayer, MJ Warren, E Brody, A Rambach, and C Thermes. 1996. Salmonella Typhimurium Cobalamin (Vitamin B12) Biosynthetic Genes: Functional Studies in S. Typhimurium and Escherichia Coli. *J. Bacteriol.* 178 (3): 753–67.

Raux, E, HL Schubert, and MJ Warren*. 2000. Biosynthesis of Cobalamin (Vitamin B12): A Bacterial Conundrum. *Cell. Mol. Life Sci.* 57 (13): 1880–93.

Robinson, F. 1966. CHAPTER X – VITAMIN B12. In *Vitam. Co-Factors Enzym. Syst.*,

682–798.

Rodionov, D a., AG Vitreschak, A a. Mironov, and MS Gelfand. 2003. Comparative Genomics of the Vitamin B12 Metabolism and Regulation in Prokaryotes. *J. Biol. Chem.* 278 (42): 41148–59.

Roessner, CA, HJ Williams, and AI Scott. 2005. Genetically Engineered Production of 1-Desmethylcobyrinic Acid , 1-Desmethylcobyrinic Acid a , c -Diamide , and Cobyrinic Acid a , c -Diamide in Escherichia Coli Implies a Role for CbiD in C-1 Methylation in the Anaerobic Pathway to Cobalamin * □. *J. Biol. Chem.* 280 (17): 16748–53.

Ron, EZ. 1975. Growth Rate of Enterobacteriaceae at Elevated Temperatures: Limitation by Methionine. *J. Bacteriol.* 124 (1): 243–46.

Sah, B-R, R Schibli, R Waibel, L von Boehmer, P Bläuenstein, E Nexo, A Johayem, et al. 2014. Tumor Imaging in Patients with Advanced Tumors Using a New 99mTc-Radiolabeled Vitamin B12 Derivative. *J. Nucl. Med.* 55 (1): 43–49.

Sambrook, J, and M Green. 2001. *Molecular Cloning: A Laboratory Manual.* Cold Harbor Laboratory Press, New York.

Santos, F, JL Vera, P Lamosa, GF de Valdez, WM de Vos, H Santos, F Sesma, and J Hugenholtz. 2007. Pseudovitamin B12 Is the Corrinoid Produced by Lactobacillus Reuteri CRL1098 under Anaerobic Conditions. *FEBS Lett.* 581 (25): 4865–70.

Savvi, S, DF Warner, BD Kana, JD McKinney, V Mizrahi, and SS Dawes. 2008. Functional Characterization of a Vitamin B12-Dependent Methylmalonyl Pathway in Mycobacterium Tuberculosis: Implications for Propionate Metabolism during Growth on Fatty Acids. *J. Bacteriol.* 190 (11): 3886–95.

Schindelin, J, I Arganda-Carreras, E Frise, V Kaynig, M Longair, T Pietzsch, S Preibisch, et al. 2012. Fiji: An Open-Source Platform for Biological-Image Analysis. *Nat. Methods* 9 (7): 676–82.

Schroeder, S, AD Lawrence, R Biedendieck, RS Rose, E Deery, RM Graham, KJ McLean, AW Munro, SEJ Rigby, and MJ Warren. 2009. Demonstration That CobG, the Monooxygenase Associated with the Ring Contraction Process of the Aerobic Cobalamin (Vitamin B12) Biosynthetic Pathway, Contains an Fe-S Center and a Mononuclear Non-Heme Iron Center. *J. Biol. Chem.* 284 (8): 4796–4805.

Selhub, J. 1999. Homocysteine Metabolism. *Annu. Rev. Nutr.* 19 (January): 217–46.

Shell, TA, and DS Lawrence. 2015. Vitamin B12: A Tunable, Long Wavelength, Light-Responsive Platform for Launching Therapeutic Agents. *Acc. Chem. Res.* 48 (11): 2866–74.

Smeltzer, CC, MJ Cannon, PR Pinson, JD Munger, FG West, and CB Grissom. 2001. Synthesis and Characterization of Fluorescent Cobalamin (CobalaFluor) Derivatives for Imaging. *Org. Lett.* 3 (6): 799–801.

Smith, EL, LFJ Parker, and DE Gant. 1955. Anti-Metabolites from Vitamin B12-Proceedings of The Biochemical Society. *Biochem. J.* 104 (3): 14.

Smith, T, KA Wolff, and L Nguyen. 2013. Molecular Biology of Drug Resistance in Mycobacterium Tuberculosis. *Curr. Top. Microbiol. Immunol.* 358 (January): 3–32.

Smith, WJ, NP Oien, RM Hughes, CM Marvin, ZL Rodgers, J Lee, and DS Lawrence. 2014. Cell-Mediated Assembly of Phototherapeutics. *Angew. Chem. Int. Ed. Engl.* 53 (41).

Spence, JD. 2016. Metabolic Vitamin B12 Deficiency: A Missed Opportunity to Prevent Dementia and Stroke. *Nutr. Res.* 36 (2). Elsevier Inc.: 109–16.

Stabler, SP, EP Brass, PD Marcell, and RH Allen. 1991. Inhibition of Cobalamin-Dependent Enzymes by Cobalamin Analogues in Rats. *J. Clin. Invest.* 87 (4): 1422–30.

Stefansson, S, DA Lawrence, and WS Argraves. 1996. Plasminogen Activator Inhibitor-1 and Vitronectin Promote the Cellular Clearance of Thrombin by Low Density Lipoprotein Receptor-Related Proteins 1 and 2. *J. Biol. Chem.* 271 (14): 8215–20.

Stubbe, J. 1994. Binding Site Revealed of Nature ' S Most Beautiful Cofactor.

Sulston, JE, E Schierenberg, JG White, and JN Thomson. 1983. The Embryonic Cell Lineage of the Nematode Caenorhabditis Elegans. *Dev. Biol.* 100 (1): 64–119.

Szumowski, JD, KN Adams, PH Edelstein, and L Ramakrishnan. 2013. Antimicrobial Efflux Pumps and Mycobacterium Tuberculosis Drug Tolerance: Evolutionary Considerations. *Curr. Top. Microbiol. Immunol.* 374 (January): 81–108.

Taga, ME, GC Walker, ME Taga, and GC Walker. 2008. Pseudo-B 12 Joins the Cofactor Family. *Am. Soc. Microbiol.* 190 (December 2007): 10–13.

Toh, B-H, IR van Driel, and PA Gleeson. 1997. Pernicious Anaemia. *N. Engl. J. Med.* 341 (10): 738–46.

Toohey, JI (University of CB. 1965. A Vitamin B12 Compound Containing No Cobalt. *Biochemistry* 394 (1954): 934–42.

Treusch, S, S Knuth, S a Slaugenhaupt, E Goldin, BD Grant, and H Fares. 2004. Caenorhabditis Elegans Functional Orthologue of Human Protein H-Mucolipin-1 Is Required for Lysosome Biogenesis. *Proc. Natl. Acad. Sci. U. S. A.* 101 (13): 4483–88.

Vitreschak, AG, D a Rodionov, and A a Mironov. 2003. Regulation of the Vitamin B 12 Metabolism and Transport in Bacteria by a Conserved RNA Structural Element Regulation of the Vitamin B 12 Metabolism and Transport in Bacteria by a Conserved RNA Structural Element. *Rna* 9: 1084–97.

Waibel, R, H Treichler, NG Schaefer, DR van Staveren, S Mundwiler, S Kunze, M Küenzi, et al. 2008. New Derivatives of Vitamin B12 Show Preferential Targeting of Tumors. *Cancer Res.* 68 (8): 2904–11.

Walker, JM. 2009. *Methods in Molecular Biology. Life Sci.* Vol. 531. doi:10.1007/978-1-62703-239-1_1.

Wang, R, W Zheng, H Yu, H Deng, and M Luo. 2011. Labeling Substrates of Protein Arginine Methyltransferase with Engineered Enzymes and Matched S-Adenosyl-L-Methionine Analogues. *J. Am. Chem. Soc.* 133: 7648–51.

Warner, DF, S Savvi, V Mizrahi, and SS Dawes. 2007. A Riboswitch Regulates Expression of the Coenzyme B 12 -Independent Methionine Synthase in Mycobacterium Tuberculosis : Implications for Differential Methionine Synthase Function in Strains

H37Rv and CDC1551. *J. Bacteriol.* 189 (9): 3655–59.

Warren, MJ, E Raux, HL Schubert, and JC Escalante-Semerena. 2002. The Biosynthesis of Adenosylcobalamin (Vitamin B12). *Nat. Prod. Rep.* 19 (4): 390–412.

Watanabe, F, Y Yabuta, Y Tanioka, and T Bito. 2013. Biologically Active Vitamin B 12 Compounds in Foods for Preventing Deficiency among Vegetarians and Elderly Subjects. *J. Agric. Food Chem.* 61: 6769–75.

Werder, SF. 2010. Cobalamin Deficiency, Hyperhomocysteinemia, and Dementia. *Neuropsychiatr. Dis. Treat.* 6: 159–95.

Whipple, GH, and FS Roschelle-Robbins. 1925. Favourable Influence of Liver, Heart and Skeletal Muscle in Diet of Blood Regeneration Anaemia. *Am. J. Physiol.* 72 (3): 408–418.

Widner, FJ, AD Lawrence, E Deery, D Heldt, S Frank, K Gruber, K Wurst, MJ Warren, and B Kräutler. 2016. Total Synthesis , Structure , and Biological Activity of Adenosyl- Rhodibalamin, the Non-Natural Rhodium Homologue of Coenzyme B12. *Angew. Chem. Int. Ed. Engl.*, 1–7.

Willard, HF, IS Mellman, and LE Rosenberg. 1978. Genetic Complementation among Inherited Deficiencies of Methylmalonyl-CoA Mutase Activity: Evidence for a New Class of Human Cobalamin Mutant. *Am J Hum Genet* 30 (1): 1–13.

Winn, MD, CC Ballard, KD Cowtan, EJ Dodson, P Emsley, PR Evans, RM Keegan, et al. 2011. Overview of the CCP4 Suite and Current Developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 67 (4): 235–42.

Winter, G, and D Waterman. 2012. xia2 Manual, 1–15.

Wishart, D, and D Case. 2001. Use of Chemical Shifts in Macromolecular Structure Determination. *Methods Enzymol.* 338: 3–34.

Wooh, JW, RD Kidd, JL Martin, and B Kobe. 2003. Comparison of Three Commercial Sparse-Matrix Crystallization Screens. *Acta Crystallogr. - Sect. D Biol. Crystallogr.* 59 (4): 769–72.

World Health Organisation. 2015. Global Tuberculosis Report 2015.

Wuerges, J, G Garau, S Geremia, SN Fedosov, TE Petersen, and L Randaccio. 2006. Structural Basis for Mammalian Vitamin B12 Transport by Transcobalamin. *Proc. Natl. Acad. Sci. U. S. A.* 103 (12): 4386–91.

Yi, S, EC Seth, Y Men, SP Stabler, RH Allen, L Alvarez-cohen, and ME Taga. 2012. Versatility in Corrinoid Salvaging and Remodeling Pathways Supports Corrinoid-Dependent Metabolism in Dehalococcoides Mccartyi. *Appl. Environ. Microbiol.* 78 (21): 7745–52.

Young, DB, I Comas, and LPS de Carvalho. 2015. Phylogenetic Analysis of Vitamin B12-Related Metabolism in Mycobacterium Tuberculosis. *Front. Mol. Biosci.* 2 (March): 1–14.

Zhang, Y, NW Hodgson, MS Trivedi, and HM Abdolmaleky. 2016. Decreased Brain Levels of Vitamin B12 in Aging , Autism and Schizophrenia, 1–19.

Appendices
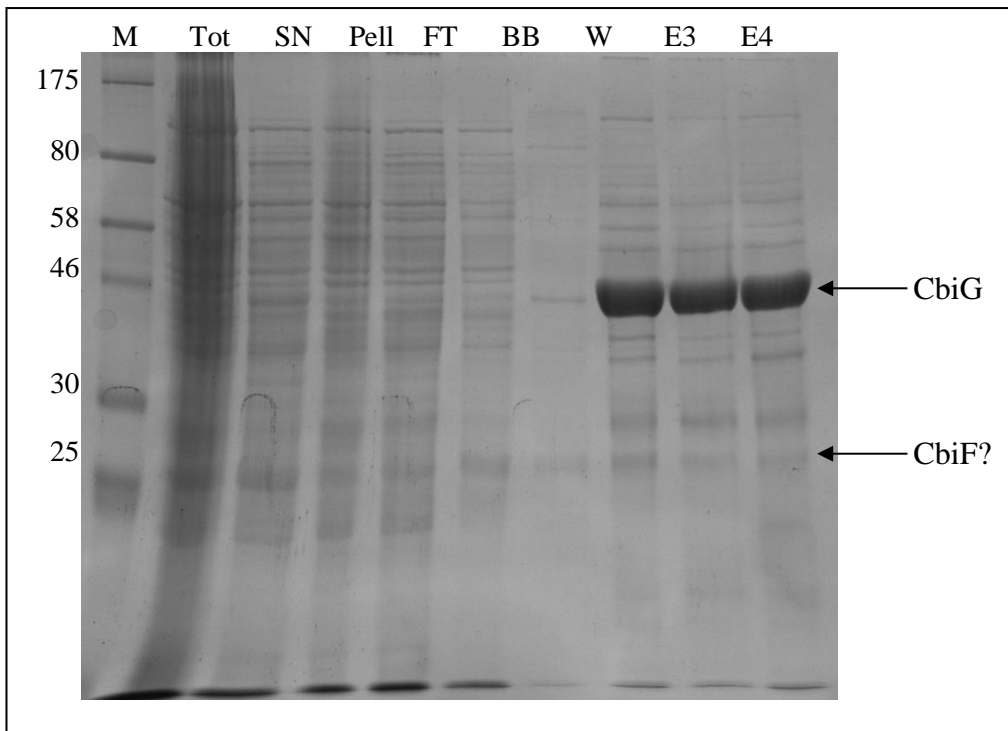
Appendix A

A.1: SDS gel of CbiH purification
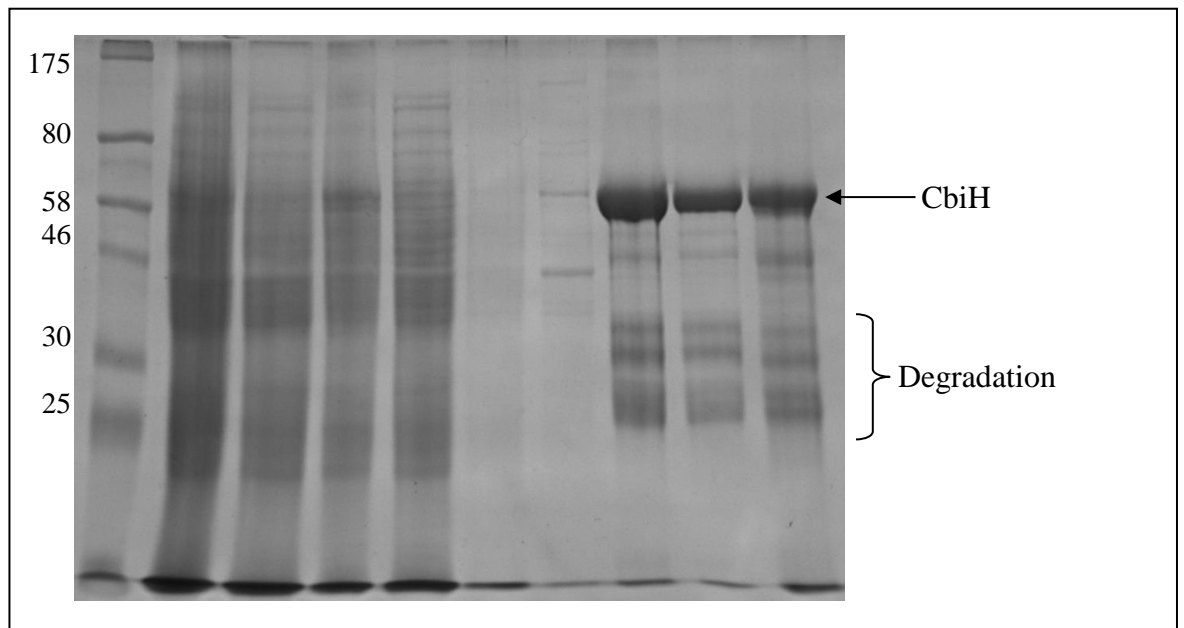
A.2: Graph showing the change in spectra during the formation of Cobalt-precorrin-6B from Cobalt-factor-III
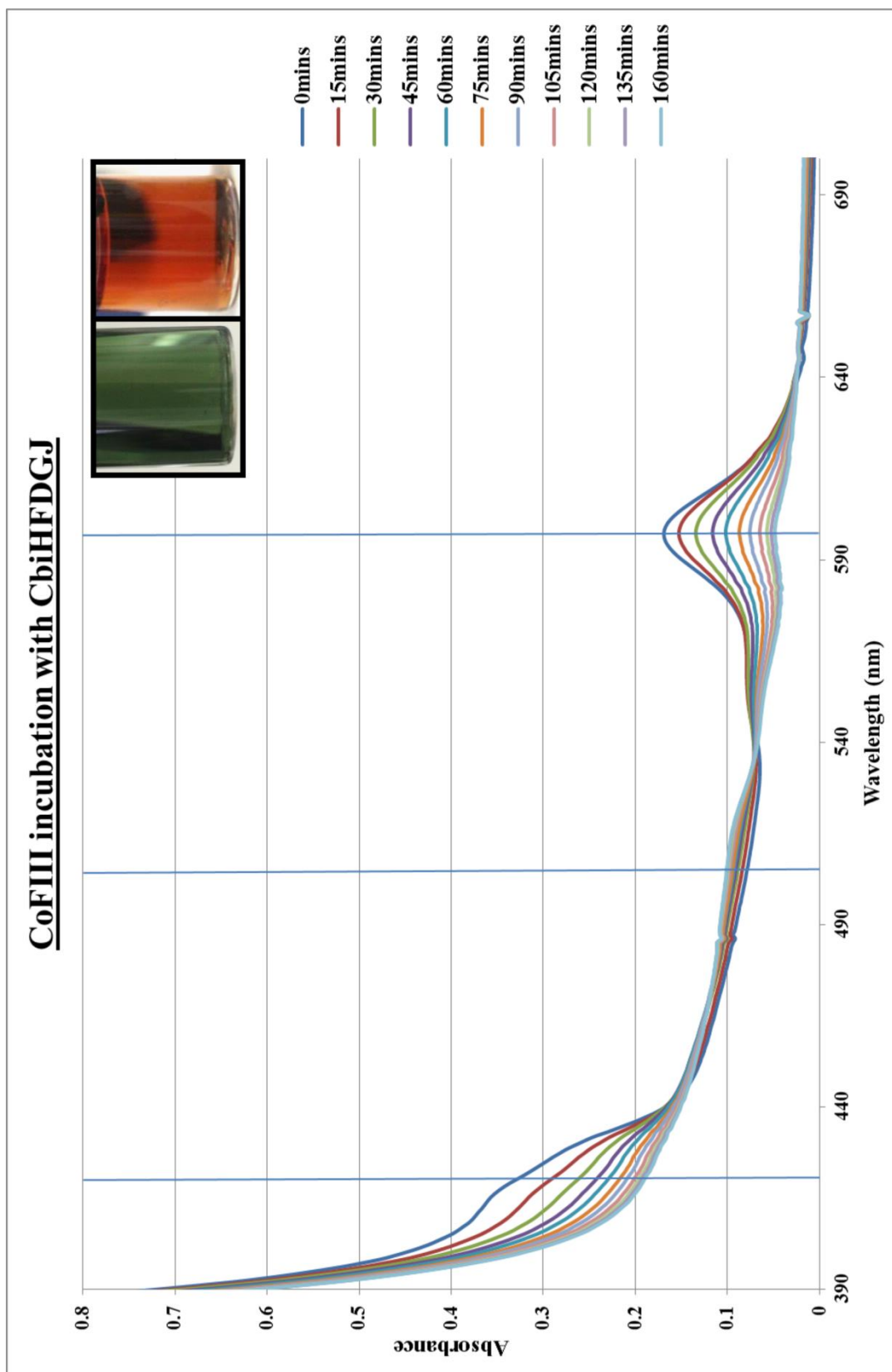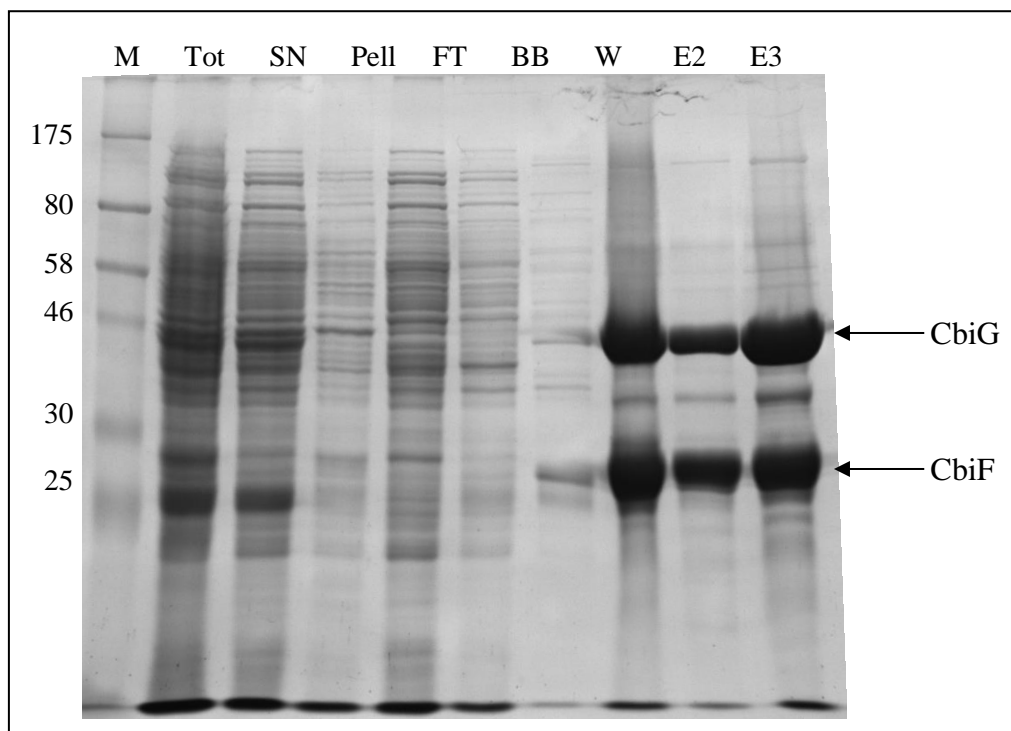
A.3: SDS gel of CbiF purification frozen and fresh pellets

From a frozen pellet

From a fresh pellet

A.4: Optimisation tray for CobH(T85A) apo and holo-protein

% PEG 8000	0.1 M Sodium Cacodylate pH6.5 (NaCac) Calcium acetate hydrate (CaAH)					
	0.2 M	0.17 M	0.14 M	0.11 M	0.08 M	0.05 M
5 %	100 µl 1 M CaAH 50 µl NaCac 62.5 µl PEG8000	85 µl 1 M CaAH 50 µl NaCac 62.5 µl PEG8000	70 µl 1 M CaAH 50 µl NaCac 62.5 µl PEG8000	55 µl 1 M CaAH 50 µl NaCac 62.5 µl PEG8000	40 µl 1 M CaAH 50 µl NaCac 62.5 µl PEG8000	25 µl 1 M CaAH 50 µl NaCac 62.5 µl PEG8000
H_2O	287.5	302.5	317.5	332.5	347.5	362.5
10 %	100 µl 1 M CaAH 50 µl NaCac 125 µl PEG8000	85 µl 1 M CaAH 50 µl NaCac 125 µl PEG8000	70 µl 1 M CaAH 50 µl NaCac 125 µl PEG8000	55 µl 1 M CaAH 50 µl NaCac 125 µl PEG8000	40 µl 1 M CaAH 50 µl NaCac 125 µl PEG8000	25 µl 1 M CaAH 50 µl NaCac 125 µl PEG8000
H_2O	225	240	255	270	285	300
15 %	100 µl 1 M CaAH 50 µl NaCac 187.5 µl PEG8000	85 µl 1 M CaAH 50 µl NaCac 187.5 µl PEG8000	70 µl 1 M CaAH 50 µl NaCac 187.5 µl PEG8000	55 µl 1 M CaAH 50 µl NaCac 187.5 µl PEG8000	40 µl 1 M CaAH 50 µl NaCac 187.5 µl PEG8000	25 µl 1 M CaAH 50 µl NaCac 187.5 µl PEG8000
H_2O	162.5	177.5	192.5	207.5	222.5	237.5
	0.1 M Sodium Cacodylate pH6.5 (NaCac) 0.2 M Calcium acetate hydrate (CaAH)					
Additive	15 % Glycerol	30 % Glycerol	Dioxane	Ethanol	Propanol	Methanol
20 % PEG 8000	100 µl 1 M CaAH 50 µl NaCac 250 µl PEG8000	100 µl 1 M CaAH 50 µl NaCac 250 µl PEG8000	100 µl 1 M CaAH 50 µl NaCac 250 µl PEG8000 100 µl Dioxane	100 µl 1 M CaAH 50 µl NaCac 250 µl PEG8000 100 µl Ethanol	100 µl 1 M CaAH 50 µl NaCac 250 µl PEG8000 100 µl Propanol	100 µl 1 M CaAH 50 µl NaCac 250 µl PEG8000 100 µl Methanol
H2O	100 µl	100 µl	0 µl	0 µl	0 µl	0 µl
Glycerol	0.6 µl	1.2 µl				

The alcohols were added to the wells and the glycerol to the drops themselves.

A.5: Collected data for the apo-crystal CobH (T85A) compared to the co-crystal and the previously published wild type CobH from *Rhodobacter capsulatus*

Single wavelength anomalous dispersion/diffraction (SAD)		
	Co-crystal CobH (T85A) allyl-HBA	**Apo-crystal CobH (T85A)**	**Co-crystal CobH HBA (4FDV)**
Wavelength (Å)	0.92819	0.92819	0.97630
High resolution limit	1.57 (7.02 - 1.57)	1.17 (4.53 - 1.17)	1.68
Low resolution limit	48.38 (48.38 - 1.61)	48.34 (48.34 - 1.21)	34.7
Completeness	98.8 (91.6 - 99.7)	98.8 (99.9 - 96.7)	97.4
Multiplicity	3.3 (3.4 - 3.2)	2.9 (3.3 - 1.9)	4.1 (redundancy)
I/sigma	6.3 (12.6 - 1.3)	12.6 (53.4 - 1.4)	16.9 (6.6)
R_{merge}	0.237 (0.249 - 0.899)	0.026 (0.017 - 0.626)	0.059
Anomalous completeness	85.8 (87.3 - 81.9)	79.4 (93.7 - 49.1)	
Anomalous multiplicity	1.5 (1.9 - 1.7)	1.3 (1.7 - 1.3)	
Unit cell dimensions: a (Å)	71.230	71.415	70.250
b (Å)	66.630	66.491	66.030
c (Å)	48.920	49.101	48.480
α (°)	90.000	90.000	90.000
β (°)	99.030	99.652	98.980
γ (°)	90.000	90.000	90.000
Spacegroup	C 1 2 1	C 1 2 1	C 1 2 1
Sfcheck twinning score	2.13 Your data do not appear twinned	1.92 Your data do not appear twinned	

A.6: The sequence alignment of *Allochromatium vinosum* and *Rhodobacter capsulatus* CobQ

Alignment produced using MultAlin

A.7: C5-allyl-HBAH NMR data to confirm the structure of the novel intermediate
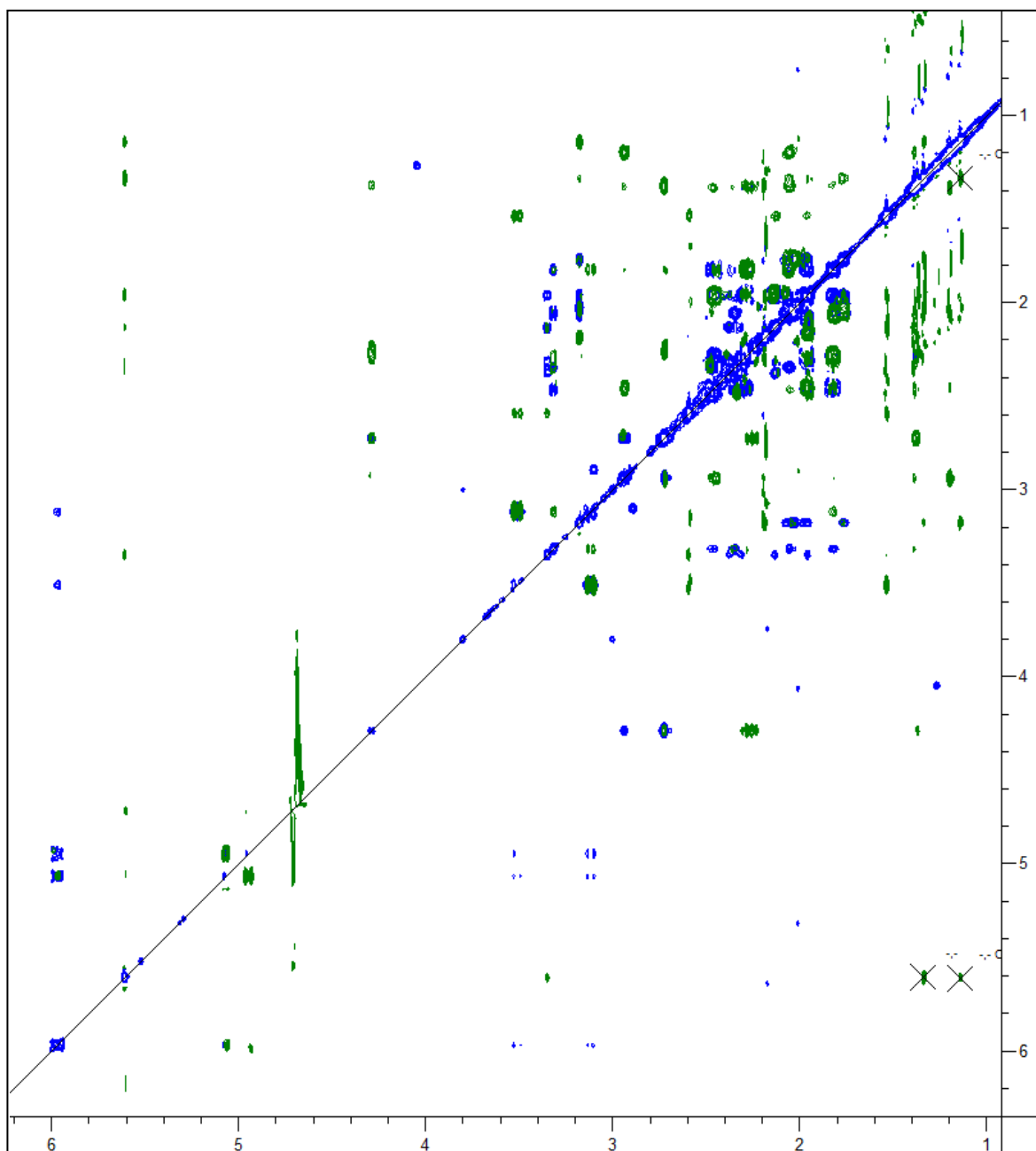
Chemical shift assignments:

Eight carbons were not assignable (mostly due to lack of signal). The terminal vinyl protons (C53) and C19 were only observable in the sensitivity enhanced HSQC (possibly due to water suppression suppressing the signals).

Assignment	δ(13C) [ppm]	δ(1H) [ppm]		
C1	83.2	-		
C1A	24.4	1.20		
C2	48.8	-		
C2A	18.4	1.39		
C21	43.6	2.27		
C22	?	-		
C3	57.8	3.31		
C31	28.0	1.82	/	2.05
C32	37.4	2.35	/	2.47
C33	?	-		
C4	?	-		
C5	?	-		
C51	33.1	3.11	/	3.51
C52	140.2	5.97		
C53	119.0	4.96	/	5.07
C6	162.5	-		
C7	49.1	-		
C7A	22.9	1.53		
C71	47.7	2.59		
C72	177.7	-		
C8	55.6	3.35		
C81	27.8	1.96	/	2.14
C82	35.0	2.31	/	2.37
C83	?	-		
C9	171.4	-		
C10	92.1	5.61		
C11	193.3	-		
C12	51.7	-		
C12A	21.6	1.33		
C12B	33.6	1.14		
C13	56.0	3.18		
C131	28.0	1.76	/	2.02
C132	34.6	1.96	/	2.07
C133	?	-		
C14	?	-		
C15	105.7	-		
C151	16.6	2.18		
C16	182.1	-		
C17	60.1	-		
C17B	22.2	1.37		
C171	34.9	1.84	/	2.27
C172	36.1	1.95	/	2.45
C173	?	-		
C18	42.2	2.94		
C181	35.7	2.72		
C182	178.9	-		
C19	69.5	4.29		

Observed ROE interactions

From	To
C1A	C31, C32, C18, C19
C2A	C21, C3, C18, C31, C32, C181
C21	C2A, C3, C19
C3	C2A, C21, C51
C31	C1A, C2A, C51
C32	C1A, C2A
C51	C3, C31, C7A, C71
C52	
C53	C71
C7A	C51, C71, C8, C81, C82
C71	C51, C53, C7A, C8
C8	C7A, C71, C10
C81	C7A, C10
C82	C7A, C10
C10	C8, C81, C82, C12A, C12B
C12A	C10, C12B, C13, C131
C12B	C10, C13, C12A
C13	C12A, C12B, C151
C131	C12A
C132	C172
C151	C13, C17B, C171
C17B	C151, C171, C181, C19
C171	C151, C17B, C18
C172	C132, C18
C18	C1A, C2A, C171, C172
C181	C2A, C171, C17B
C19	C1A, C21, C17B

A.8: C5-thioamine-cobyric acid NMR

Chemical shift assignments

	δ(1H) [ppm]	δ(13C) [ppm]
C1		85.34
C1A	1.43	24.28
C2		49.27
C2A	1.46	19.19
C21	2.21, 2.28	43.74
C22		178.09
C3	3.89	58.16
C31	1.93, 2.24	28.56
C32	2.45, 2.53	37.55
C33		183.99
C4		179.48
C5		108.28
C51	2.35	17.90
C52	1.61, 1.84	34.42
C53	3.02	41.58
C55	2.77	32.96
C56	3.06	41.48
C6		166.71
C7		52.09
C7A	1.7	22.60
C71	2.28, 2.54	47.23
C72		177.67
C8	3.28	57.35
C81	1.67, 2.24	28.96
C82	2.64, 2.68	35.45
C83		178.83
C9		174.70
C10	5.82	93.80
C11		180.65
C12		49.78
C12A	1.37	21.07
C12B	1.13	32.89
C13	3.24	55.67
C131	1.75, 2.06	28.68
C132	2.07, 2.29	34.45
C133		181.83
C14		165.44
C15		106.92
C151	2.26	17.59
C16		180.71
C17		61.92
C17A	1.24	20.00
C171	1.96, 2.32	35.96
C172	1.80, 2.43	35.87

C173		180.63
C18	2.82	41.83
C181	2.24	44.51
C182		180.60
C19	3.68	78.28

A.9: C5-fluorescein-cobyric acid HPLC MS to check conversion
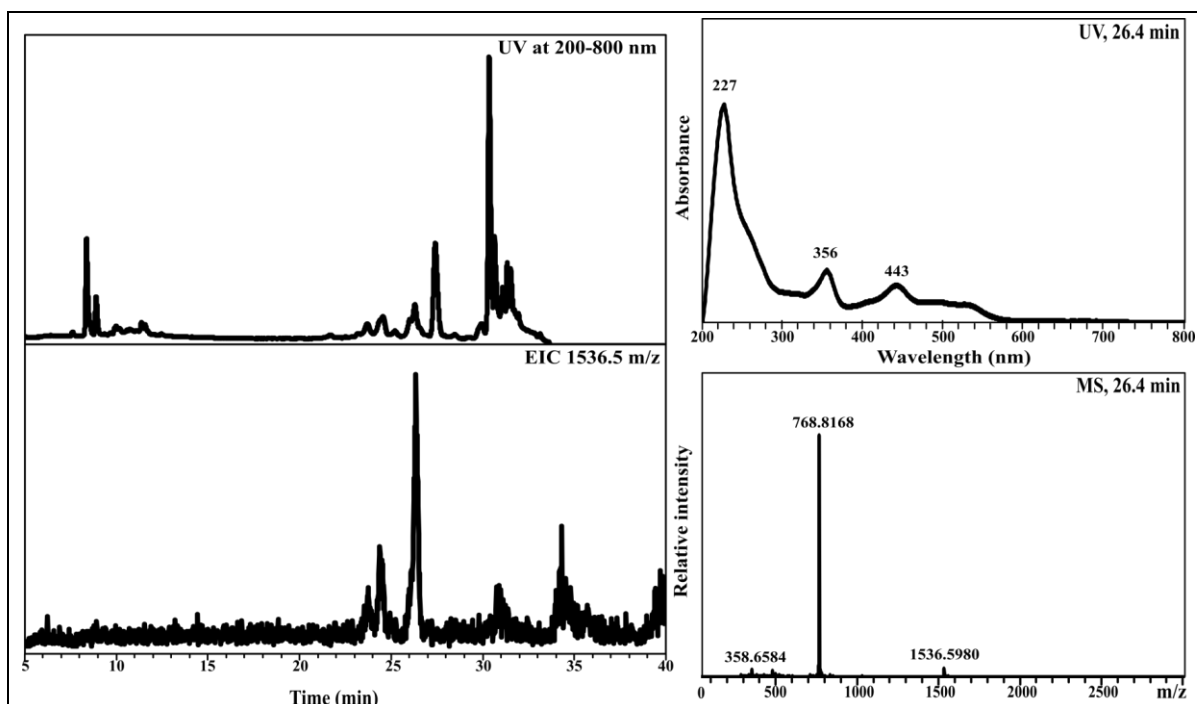
The underlined extracted underlined ion underlined chromatogram (EIC) detects a m/z of 1536.5 (±0.2) in the mass spectrometry data. The peaks correspond to the smaller peaks either side of 25 minutes. The larger peaks at 30 minutes in the UV chromatogram had no clear cobalamin absorption and a cacophony of mass spectrometry values.

Appendix B

B.1 pET14b-*E. coli* btuF without the periplasmic sequence
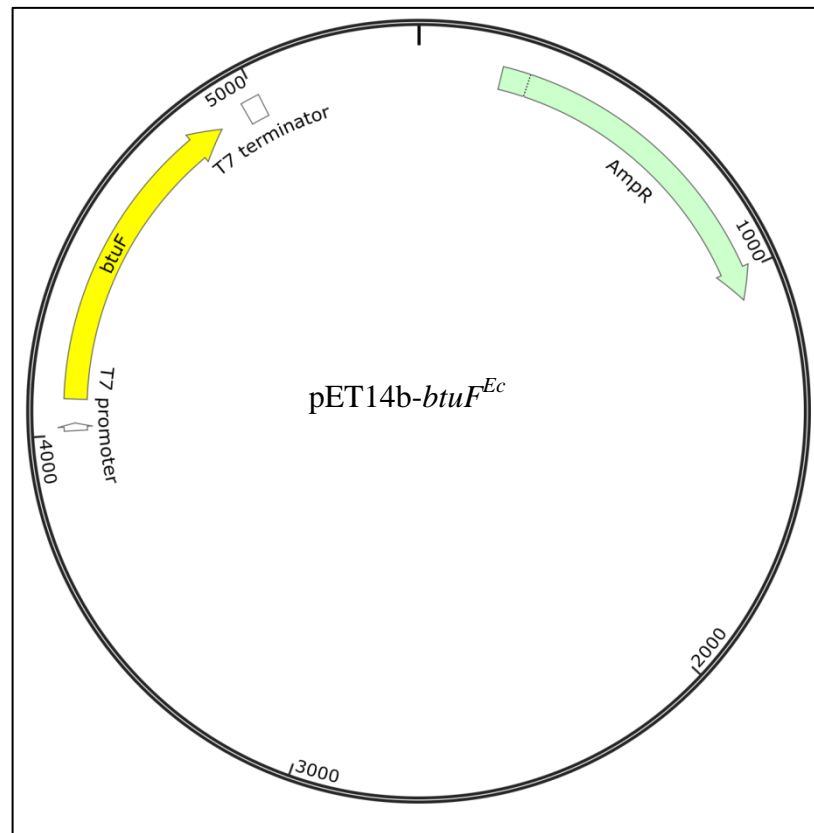
The T7 promoter and terminator are in white whilst the ampicillin resistance gene, AmpR is in green. BtuF is shown in yellow. There is a His tag present at the start of the gene which is not shown on the Figure. The *Ec* after *btuF* is present because the gene was taken from *E. coli*.

B.2 cobalamin standards grown on a *S. enterica* bioassay plate at the same time as Figure 4.3.4.1
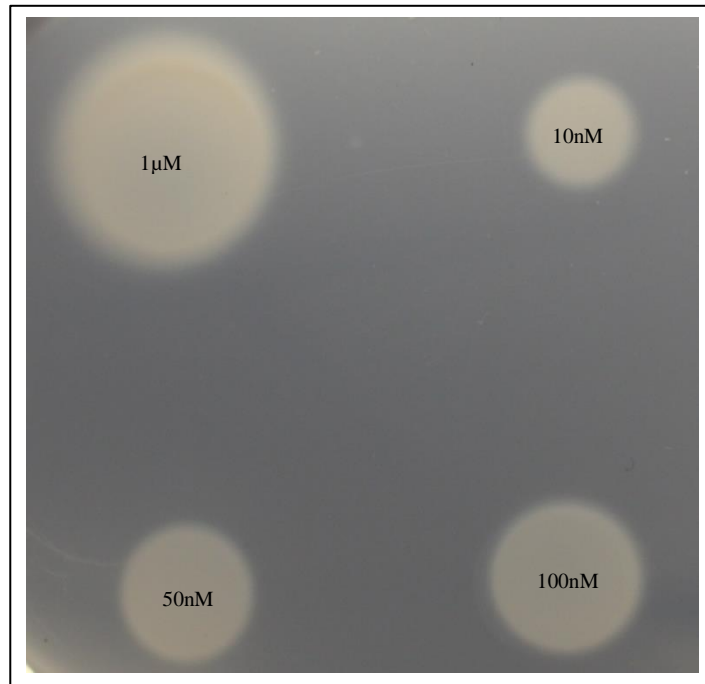
Appendix C

C.1 Bioassay standards plates
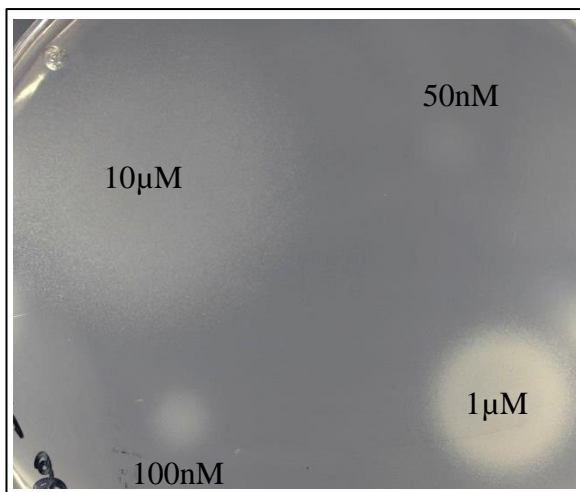
C1.1 Pseudo-cobalamin

Pseudo-cobalamin standards repeated without the 10nM sample which did not grow. The 1μM standard is roughly equivalent to the 50nM standard of cobalamin or cobinamide.
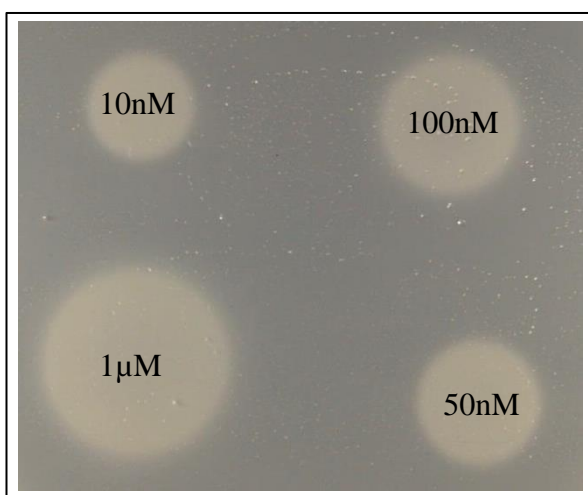
C1.2 Cobalamin

A typical bioassay plate of cobalamin standards.
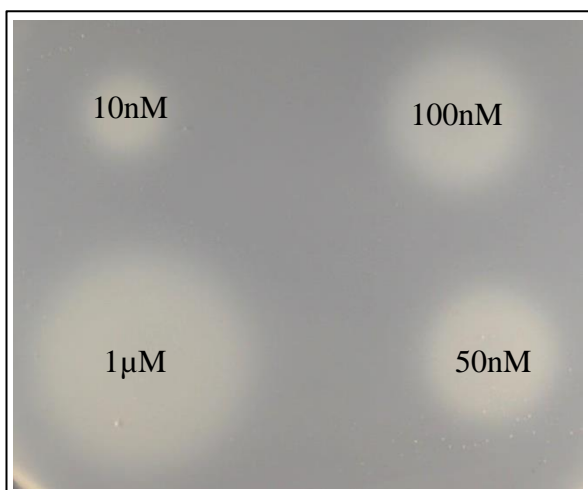
C1.3 Cobinamide

A typical bioassay plate of cobinamide standards.

C.2 Graph to show the change in different corrinoid uptake between WT, H37Rv and *ΔbacA*
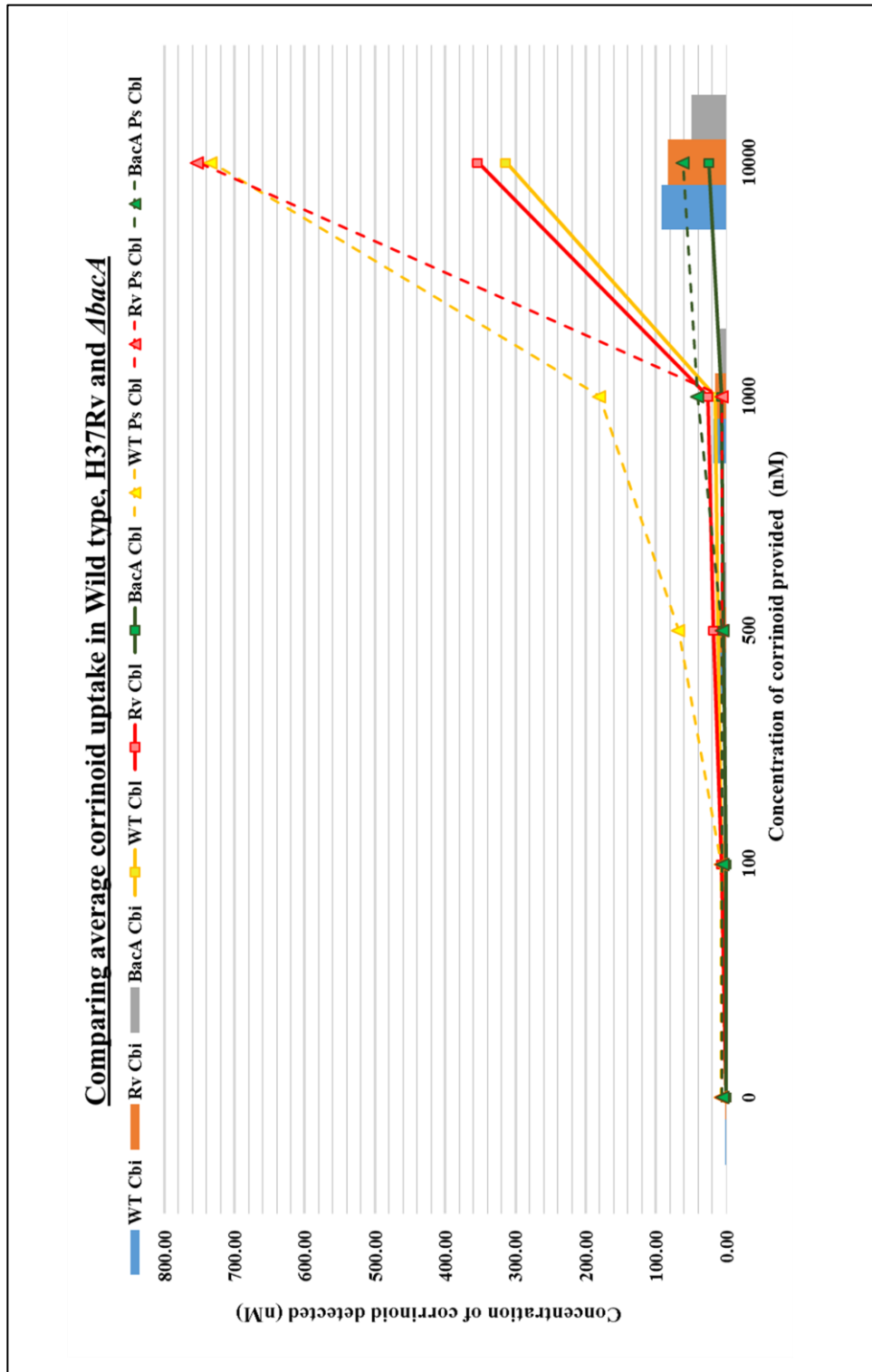
Pseudo-cobalamin without being scaled up to 50 ml, but still using the 20 ml values.

C.3 Extracted ion chromatogram for cobinamide in the sample *ΔmetE* grown on cobinamide and run on the HPLC-MS
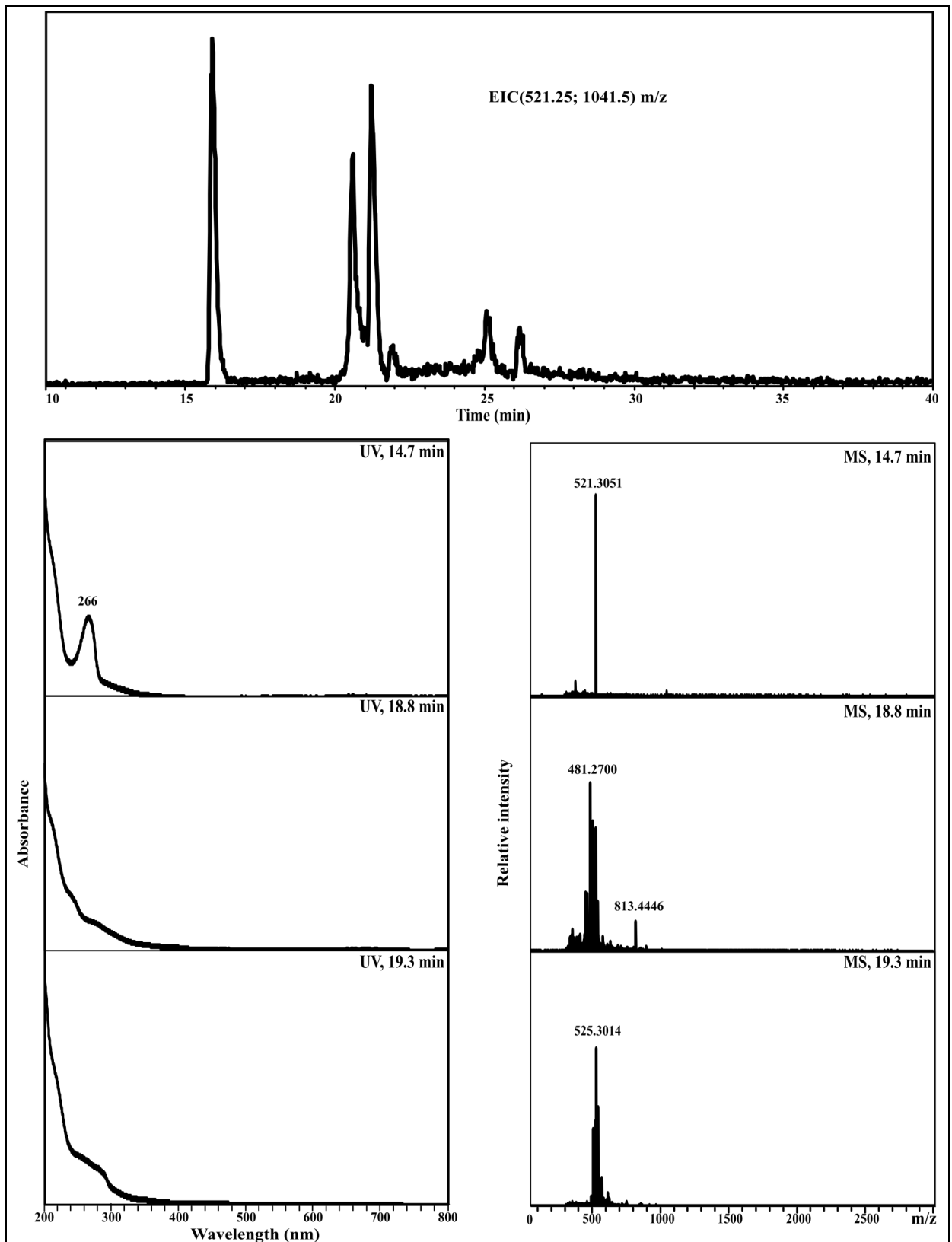
The extracted ion chromatogram for cobinamide in the singly and doubly charged state: 1041.5 and 521.25 respectively. The mass spectrometry data that follows is for the three biggest peaks. None of the show the m/z for cobinamide or the 350nm, 526nm absorbance maxima observed in corrinoids. They are likely to be contaminants from the purification process.

C.4 Alignment of the DNA from the two *M. tuberculosis cobQ* genes

Alignment produced using MultAlin

C.5 Table of the effect of cobalamin and cobinamide on the cobalamin biosynthesis gene RNA levels

Protein function	Gene identity	Gene name	Cobinamide					Cobalamin				
			ΔmetE	ΔBacA	ΔmetH	H37Rv	WT	ΔBacA	ΔmetH	H37Rv	WT	ΔmetE
Cob(I)alamin_adenosyltransferase	Rv2849c	cobA	-0.0234	-0.0566	-0.0546	-0.1879	-0.1889	-0.0276	0.195	-0.1165	-0.0535	-0.073
C20 and C17 methyltransferase	Rv2066	cobI-J	-0.197	-0.1818	-0.1442	-0.148	-0.1208	-0.1573	0.1495	-0.3261	-0.3121	-0.0682
Precorrin reductase	Rv2064	cobG	-0.4795	-0.2931	-0.075	-0.1868	-0.206	-0.0965	0.1447	0.1036	0.0337	-0.5438
Precorrin-3, C11 methyltransferase	Rv2071c	cobM	0.3399	0.2925	0.4124	0.3557	0.2215	0.3363	0.8075	0.3719	0.3152	0.0979
Precorrin6a reductase	Rv2070c	cobK	-0.2836	-0.7981	-0.3881	-0.4534	-0.441	0.1828	0.3415	0.1766	0.129	-0.0408
C5, C15 methyltransferase, C12 decarboxylase	Rv2072c	cobL	0.0218	-0.1613	0.0834	-0.2936	-0.2814	-0.1062	0.2009	-0.1046	-0.1304	-0.217
Precorrin8 isomerase	Rv2065	cobH	-0.5209	-0.5145	-0.4138	-0.779	-0.5496	-0.6822	-0.1596	-0.4557	-0.1905	-0.5565
HBA-*a*, *c*-amidase	Rv2848c	cobB	0.0604	-0.1646	0.0568	-0.1262	-0.1932	-0.1506	-0.1393	-0.24	-0.1599	-0.1023
Cobalt chelatase	Rv2062c	cobN	-0.4401	0.1533	0.1412	0.0035	0.145	-0.0283	0.2519	-0.0702	0.0835	-0.2155
Cobyric acid synthase	Rv0255c	cobQ1	1.4511	1.1899	1.6493	0.6745	0.8317	0.2846	0.7251	0.0277	0.2448	0.3257
Possible cobyric acid synthase	Rv3713	cobQ2	-0.8565	0.0699	0.1874	-0.2346	-0.1879	-0.3184	0.2403	-0.2595	-0.1591	-0.2863
Cobinamide synthase	Rv2236c	cobD	-0.1395	-0.0086	0.1226	-0.0773	-0.0711	-0.0946	-0.1481	0.0542	0.0591	0.0207
Cobinamide kinase	Rv0254c	cobU	-0.0811	-0.3847	-0.3631	-0.1139	-0.317	-0.2821	-0.0797	-0.0596	0.0593	0.2281
Cobalamin (5'-phosphate) synthase	Rv2208	cobS	-0.0904	0.0098	0.2133	-0.2418	-0.3345	-0.113	0.4259	0.0516	-0.0179	-0.3685
Nicotinate nucleotide dimethylbenzimidazole transferase	Rv2207	cobT	-0.1529	0.1389	0.3872	-0.137	-0.1145	-0.3641	0.3244	-0.0819	-0.07	-0.5835

C.6 SDS gel of *ΔbacA* purification on the cobalamin column
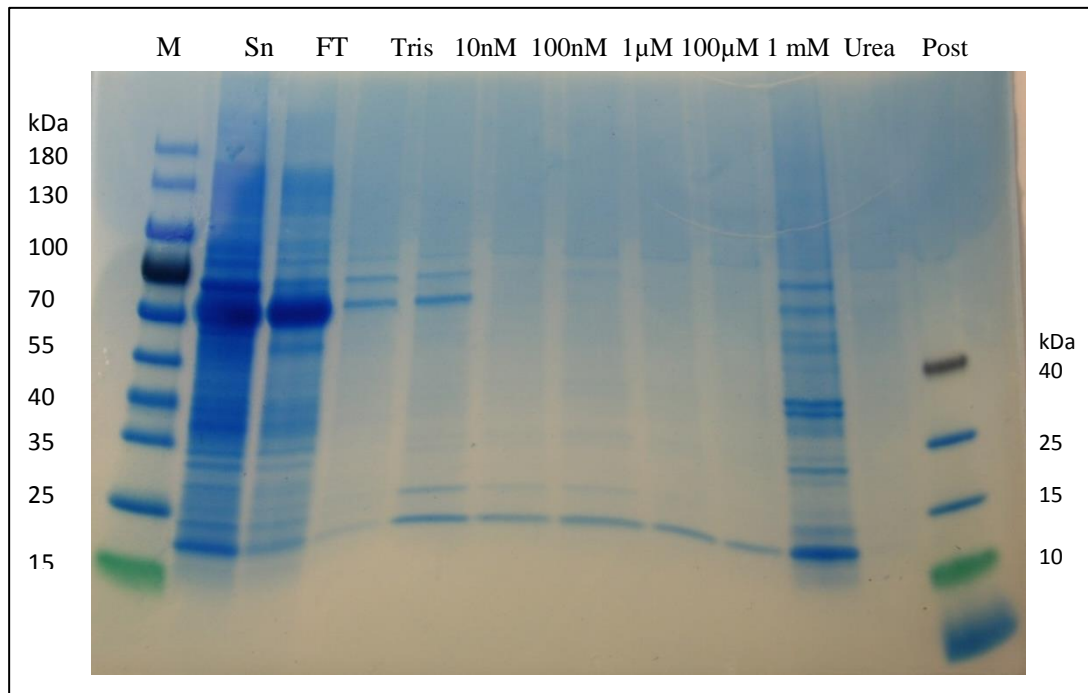
The first purification of ΔbacA run on a 4-20 % gel after the elutions were concentrated. M is the marker; Sn: supernatant; Tris: 20 mM Tris pH 8 wash; the concentrations refer to the concentration of cobalamin in 20 mM Tris pH 8 buffer; and urea: 8 M urea in 20 mM Tris pH 8 buffer; Post is a 20 mM Tris pH 8 wash done after the urea elution; LM: Low range marker.

C.7 Initial Mass Spectrometry results of MALDI from the cobalamin column

Band		MS	Organism	E value
ΔmetE O/N urea	1	30S ribosomal protein S2	M. tuberculosis (strain ATCC 25618 / H37Rv)	0.00012
ΔbacA O/N 1 µM B12	1	Serum albumin	Bos taurus	2.80E-05
ΔbacA O/N 1 µM B12	2	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase (metE)	M. tuberculosis (strain ATCC 25618 / H37Rv)	1.40E-23
ΔbacA O/N 1 µM B12	3	Chaperone protein DnaK	M. tuberculosis (strain ATCC 25618 / H37Rv)	8.70E-12
ΔbacA O/N 1 µM B12	4	30S ribosomal protein S1	M. tuberculosis (strain ATCC 25618 / H37Rv)	4.40E-11
ΔbacA O/N 1 µM B12	5	Did not run	Did not run	Did not run
ΔbacA O/N 1 µM B12	6	Adenosylhomocysteinase	M. tuberculosis (strain ATCC 25618 / H37Rv)	4.40E-06
ΔbacA O/N 1 µM B12	7	Citrate synthase 1/ Elongation factor Tu	M. tuberculosis (strain ATCC 25618 / H37Rv)	4.40E-12 / 3.50E-08
ΔbacA O/N 1 µM B12	8	Keratin, type II cytoskeletal 1	Homo sapiens	0.63
ΔbacA O/N 1 µM B12	9	Electron transfer flavoprotein subunit alpha	M. tuberculosis (strain ATCC 25618 / H37Rv)	2.80E-10
ΔbacA O/N 1 µM B12	10	Electron transfer flavoprotein subunit beta	M. tuberculosis (strain ATCC 25618 / H37Rv)	5.50E-11
ΔbacA O/N urea	1	Chaperone protein DnaK	M. tuberculosis (strain ATCC 25618 / H37Rv)	1.70E-20
ΔbacA O/N urea	2	Glutamine--fructose-6-phosphate aminotransferase [isomerizing]	M. tuberculosis (strain ATCC 25618 / H37Rv)	0.0069
ΔbacA O/N urea	3	3-oxoacyl-[acyl-carrier-protein] synthase 2	M. tuberculosis (strain ATCC 25618 / H37Rv)	3.50E-05
ΔbacA O/N urea	4	Adenosine kinase	M. tuberculosis (strain ATCC 25618 / H37Rv)	0.0065
ΔbacA O/N urea	5	Electron transfer flavoprotein subunit alpha	M. tuberculosis (strain ATCC 25618 / H37Rv)	5.50E-14
ΔbacA O/N urea	6	Electron transfer flavoprotein subunit beta	M. tuberculosis (strain ATCC 25618 / H37Rv)	6.90E-10
ΔbacA O/N urea	7	Probable enoyl-CoA hydratase echA8	M. tuberculosis (strain ATCC 25618 / H37Rv)	15
ΔbacA O/N urea	8	30S ribosomal protein S4	M. tuberculosis (strain ATCC 25618 / H37Rv)	2.20E-09
WT urea	1	Vitamin B12-binding protein	Escherichia coli (strain K12)	2.80E-08
H37Rv urea	1	30S ribosomal protein S1	M. tuberculosis (strain ATCC 25618 / H37Rv)	6.90E-11
H37Rv urea	2	Vitamin B12-binding protein	Escherichia coli (strain K12)	0.14
H37Rv urea	3	30S ribosomal protein S3	M. tuberculosis (strain ATCC 25618 / H37Rv)	0.01
H37Rv urea	4	Vitamin B12-binding protein	Escherichia coli (strain K12)	2.20E-06

C.8: MetE and MetH amino acid sequences and their alignment

>MetH

MTAADKHLYDTDLLDVLSQRVMVGDGAMGTQLQAADLTLDDFRGLEGCNEILNET
RPDVLETIHRNYFEAGADAVETNTFGCNLSNLGDYDIADRIRDLSQKGTAIARRVAD
ELGSPDRKRYVLGSMGPGTKLPTLGHTEYAVIRDAYTEAALGMLDGGADAILVETC
QDLLQLKAAVLGSRRAMTRAGRHIPVFAHVTVETTGTMLLGSEIGAALTAVEPLGV
DMIGLNCATGPAEMSEHLRHLSRHARIPVSVMPNAGLPVLGAKGAEYPLLPDELAE
ALAGFIAEFGLSLVGGCCGTTPAHIREVAAAVANIKRPERQVSYEPSVSSLYTAIPFAQ
DASVLVIGERTNANGSKGFREAMIAEDYQKCLDIAKDQTRDGAHLLDLCVDYVGRD
GVADMKALASRLATSSTLPIMLDSTETAVLQAGLEHLGGRCAINSVNYEDGDGPESR
FAKTMALVAEHGAAVVALTIDEEGQARTAQKKVEIAERLINDITGNWGVDESSILID
TLTFTIATGQEESRRDGIETIEAIRELKKRHPDVQTTLGLSNISFGLNPAARQVLNSVFL
HECQEAGLDSAIVHASKILPMNRIPEEQRNVALDLVYDRRREDYDPLQELMRLFEGV
SAASSKEDRLAELAGLPLFERLAQRIVDGERNGLDADLDEAMTQKPPLQIINEHLLA
GMKTVGELFGSGQMQLPFVLQSAEVMKAAVAYLEPHMERSDDDSGKGRIVLATVK
GDVHDIGKNLVDIILSNNGYEVVNIGIKQPIATILEVAEDKSADVVGMSGLLVKSTVV
MKENLEEMNTRGVAEKFPVLLGGAALTRSYVENDLAEIYQGEVHYARDAFEGLKL
MDTIMSAKRGEAPDENSPEAIKAREKEAERKARHQRSKRIAAQRKAAEEPVEVPERS
DVAADIEVPAPPFWGSRIVKGLAVADYTGLLDERALFLGQWGLRGQRGGEGPSYED
LVETEGRPRLRYWLDRLSTDGILAHAAVVYGYFPAVSEGNDIVVLTEPKPDAPVRY
RFHFPRQQRGRFLCIADFIRSRELAAERGEVDVLPFQLVTMGQPIADFANELFASNAY
RDYLEVHGIGVQLTEALAEYWHRRIREELKFSGDRAMAAEDPEAKEDYFKLGYRGA
RFAFGYGACPDLEDRAKMMALLEPERIGVTLSEELQLHPEQSTDAFVLHHPEAKYFN
V

>MetE
MTQPVRRQPFTATITGSPRIGPRRELKRATEGYWAGRTSRSELEAVAATLRRDTWSA
LAAAGLDSVPVNTFSYYDQMLDTAVLLGALPPRVSPVSDGLDRYFAAARGTDQIAP
LEMTKWFDTNYHYLVPEIGPSTTFTLHPGKVLAELKEALGQGIPARPVIIGPITFLLLS
KAVDGAGAPIERLEELVPVYSELLSLLADGGAQWVQFDEPALVTDLSPDAPALAEA
VYTALCSVSNRPAIYVATYFGDPGAALPALARTPVEAIGVDLVAGADTSVAGVPELA
GKTLVAGVVDGRNVWRTDLEAALGTLATLLGSAATVAVSTSCSTLHVPYSLEPETD
LDDALRSWLAFGAEKVREVVVLARALRDGHDAVADEIASSRAAIASRKRDPRLHNG

QIRARIEAIVASGAHRGNAAQRRASQDARLHLPPLPTTTIGSYPQTSAIRVARAALRA
GEIDEAEYVRRMRQEITEVIALQERLGLDVLVHGEPERNDMVQYFAEQLAGFFATQ
NGWVQSYGSRCVRPPILYGDVSRPRAMTVEWITYAQSLTDKPVKGMLTGPVTILAW
SFVRDDQPLADTANQVALAIRDETVDLQSAGIAVIQVDEPALRELLPLRRADQAEYL
RWAVGAFRLATSGVSDATQIHTHLCYSEFGEVIGAIADLDADVTSIEAARSHMEVLD
DLNAIGFANGVGPGVYDIHSPRVPSAEEMADSLRAALRAVPAERLWVNPDCGLKTR
NVDEVTASLHNMVAAAREVRAG

```
            1      10      20      30      40      50      60      70      80      90      100     110     120     130
            |-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------|
    met.H   MTAADKHLYDTDLLDVLSQRVMVGDGAMGTQLQAADLTLDDFRGLEGCNEILNETRPDVLETIHRNYFEAGADAVETNTFGCNLSNLGDYDIADRIRDLSQKGTAIARRVADELGSPDRKRYVLGSMGPG
    met.E
 Consensus  ..............................................................................................................................

            131    140     150     160     170     180     190     200     210     220     230     240     250     260
            |-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------|
    met.H   TKLPTLGHTEYAVIRDAYTEAALGMLDGGADAILVETCQDLLQLKAAVLGSRRAMTRAGRHIPVFAHVTVETTGTMLLGSEIGAALTAVEPLGVDMIGLNCATGPAEMSEHLRHLSRHARIPVSVMPNAG
    met.E                               MTQPVRRQPFTATITGSPRIGPRRELKRATEGY--WAGRTSRSELEAVAATLRRDTHSALAAAGLDSVPVNTFSYYDQHLDTAVLLGALPPRYSPVSDGLD
 Consensus  ..................adailrrqcqdalqlgaariGpRRa$kRAgrgi..fAgrTsrsegealaaserraalsAlaaaGlDm!glNcasgpa#Ml#harhLgaharrpspVmdnad

            261    270     280     290     300     310     320     330     340     350     360     370     380     390
            |-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------|
    met.H   LPVLGAKGAEYPLLPDELAEALAGFIAEFGLSLVGGCCGTTPAHIREVAAAVANIKRPERQVSYEPSVSSLYTAIPFAQDASVLVIGERTNANGSKGFREAMIAEDYQKCLDIAKDQTRDGAHLLD-LCV
    met.E   RYFAAARGTDQ-IAPLEMTKWF-----DTNYHYLVPEIGPSTTFTLHPGKVLAELKEALGQGI--PARPVIIGPITF------LLLSKAVDGAGAPIERLEELVPVYSELLSLLADGGAQAVQFDEPALV
 Consensus  rpfaaArGa#q.iaPdE$aeal.....#fnlhllggccGpspahirepaaalA#iKraerQgi..ParpsiigaIpF......Lligeat#aaGakgeReaeiaedYqecLdiaaDqga#gaqld#.acV

            391    400     410     420     430     440     450     460     470     480     490     500     510     520
            |-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------|
    met.H   DYYGRDGVADMKALASRLATSSTLPIMLDSTETAYLQAGLEHLGGRCAINSVNYEDGDGPESRFAKTHALVAEHGAAVVALTIDEEGQARTAQKKVEIAERLINDITGNHGVDESSILIDTLTFTIATGQ
    met.E   TDLSPDAPALAEAVYTALCSVSNRPATYVATYFGDPGAALPALA-RTPVEAIGVDLVAGADTSVAGYPELA---GKTLVAGVYDGRNVART---DLEAALGTLATLLGSAATVAVSTSCSTLHVPYSLEP
 Consensus  ddlgrDapAdaeAlasaLassSnrPaildaTefadlqAaLeaLa.Rca!#a!ny#dgaGa#srfAgtmaLa...GaalVAgt!DernqaRT...dlEaAerliadilGnaatdasSilcdTLhfpialeq

            521    530     540     550     560     570     580     590     600     610     620     630     640     650
            |-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------|
    met.H   EESRRDGIETIEAIRELKKRHPDVQTTLGLSNISFGLNPAA--RQVLNSVFLHECQEAGLDSAIVHASKILPMNRIPEEQRNVALDLYYDRRREDYDPLQELMRLFE-GVSAASSKEDRLAELAGLP-LF
    met.E   ETDLDDALRSWLAFGAEKVREYVVLARALRDGHDAVADEIASSRAAIASRKRDPRLHNGQIRARIEAIVASGAHRGNAAQRRASQDARLHLPPLPTTTIGSYPQTSAIRVARAALRAGEIDEAEYVRRMR
 Consensus  EedrrDairsieAiraeKkRepdVqaragrdnhdaga#eaA..RaaiaSrkrderqeaGqdrAr!eAikalganRgnaaQRraaqDarldrrredtdpiqelmrlfa.rVaaAalradriaEaaglr.$r

            651    660     670     680     690     700     710     720     730     740     750     760     770     780
            |-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------|
    met.H   ERLAQRIVDGERNGLDADLDEAMTQKPPLQTINEHLLAGHKTVG---ELFGSGQMQLPFVLQSAEVHKAAVAYLEPHHERSDDDSGKGRIVLATVKGDVHDIGKNLVDIILSNNGYEVVNIGIKQPIATI
    met.E   QEITEVIALQERLGLDVLVHGEPERNDMVQYFAEQLAGFFATQNGWVQSYGSRCVRPPILYGDVSRPRAMTVEWITYAQSLTDKPVKGMLT-----GPVTILAWSFV-----RDDQPLADTANQVALAIR
 Consensus  #ria#rIadqERnGLDadldeamerndmlQiiaEqLaagmaTqn...#1%GSrcmrlPillqdaermrAataelepha#rldDdpgKGrit.....GdVhdiaknlV.....r#dqela#ianqqaiAir

            781    790     800     810     820     830     840     850     860     870     880     890     900     910
            |-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------|
    met.H   LEVAEDKSADVVGMSGLLVKSTVVMKENLEEMNTRGVAEKFPVLLGGALTRSYVENDLAEIYQGEVHYARDAFEGLKLMDTIMSAKRGEAPDENSPEAIKAREKEAERKARHQRSKRIAARQRKAAEEPV
    met.E   DETVDLQSAGIA-----VIQVDEPALRELLPLRRADQAEYLRWAVGAFRLATSGV-SDATQIH---THLCYSEF------GEVIGAIADLDADVTSIEAARSHMEVLDDLNAIGFANGVGPGVYDIHSP-
 Consensus  dEta#dqSAd!a.....l!qsdepalr#Lee$rradqAEklrwalGaaaLarSgV.nDaa#Ih...tHlardaF.......de!igAiadeaaDenSiEAarareeea#rlaahqranr!aaqrkaaeeP.

            911    920     930     940     950     960     970     980     990     1000    1010    1020    1030    1040
            |-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------|
    met.H   EVPERSDVAADIEVPAPPFWGSRIVKGLAVADYTGLLDERALFLGQWGLRGQRGGEGPSYEDLVETEGRPRLRYWLDRLSTDGILAHAAVVYGYFPAYSEGNDIVYLTEPKPDAPVRYRFHFPRQQRGRF
    met.E   RVPSAEEMADSLRAALRAVPAERLWVNPDCGLKTRNVDEVTASLHNMVAAAREVRAG
 Consensus  rVPeae#mAadiraaarafpaeRiwknlacadkTrnlDEraafLg#mgaaarrgraG.........

            1041   1050    1060    1070    1080    1090    1100    1110    1120    1130    1140    1150    1160    1170
            |-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------+-------|
    met.H   LCIADFIRSRELAAERGEVDVLPFQLVTMGQPIADFANELFASNAYRDYLEVHGIGVQLTEALAEYHHRRIREELKFSGDRAMAAEDPEAKEDYFKLGYRGARFAFGYGACPDLEDRAKMMALLEPERIG
    met.E
 Consensus  ..............................................................................................................................

            1171   1180    1190    1200
            |-------+-------+---------|
    met.H   VTLSEELQLHPEQSTDAFVLHHPEAKYFNV
    met.E
 Consensus  ..............................
```

Alignment produced using MultAlin