An Investigation into the Relationship Between Adenine Modification and Cobamide Lower Base Synthesis

Biochemistry MSc, September 2022

Khethile Mhlongo

Total Word Count: 13800

I - Acknowledgements

Thank you, Dr. Andrew Lawrence, for your kind encouragement and guidance. Your mentorship has been invaluable and enabled me to complete the project. Thank you, Dr. Amira Abood, for your kind mentorship and advice. I would also like to thank all the members of the Lawrence Group for their support, friendship, and encouragement during my journey here, it has been a pleasure to work alongside you and learn from you.

Thank you kindly, Dr. Evelyne Deery for gifting the strains I used for this project as well as your precious advice along the way.

I would like to thank all my friends that I have made along the way on this journey, thank you for your support.

I would like to especially thank Makeda Ferguson for being so kind and being there for me when things became challenging.

Thank you so much to my sibling, Hlaluwakhe Mhlongo for your kindness and encouragement when I needed it most.

I would like to extend my deepest love and gratitude to my mother and father. Thank you for believing in me, thank you for your patience, kind words and unconditional love.

For my family, thank you for giving me the strength and encouragement when I needed it most. Thank you for giving me so much love.

In loving memory of Seapehi Molise,

may your positivity, kindness and strength live on through all those you have touched.

II – Abbreviations and Commonly Used Terms

Ade – Adenine as an adenyl lower ligand in a cobamide

Cobalamin – Vitamin B12, a complex organic molecule with DMB as the lower ligand

Cobamide (Cba) – an organic molecule like B12 with a lower ligand which may not be DMB

CobUST – an operon incorporated into the *E. coli* genome consisting of CobU, CobS and CobT

Corrinoid (Cbi) - a tetrapyrrole ring with cobalt as the central metal

MeAde – 2-methyladenine as a purinyl lower ligand in a cobamide

MeSAde – 2-methylsulfunyladenine as a purinyl lower ligand in a cobamide

III – Contents

List of Figures

List of Tables

IV – Abstract

Cobalamin, also known as vitamin B12 is a widely used cofactor in cellular metabolism but is only synthesized in a few species of prokaryotes. Interestingly, cobalamin in the large intestine is not bioavailable. But 25% of sequenced human intestinal bacteria have the capacity to produce corrinoids and 80% produce gene products that potentially require corrinoids, leading to some questions about the diversity of cobamides.

This project specifically takes interest in adenine and 2-methyladenine as lower ligands in cobamides. Which are present in some biomes such as the large intestine and are even produced by the flora which reside there. Even though they appear frequently in samples, little is known about the precursors to these lower ligands.

RlmN is a dual specificity methyl transferase which methylates adenine residues on tRNA and rRNA. In this project, we seek to investigate the effects of adenine supplementation and the deletion of this adenine methyltransferase on the lower ligands of cobamides synthesized by genetically modified *E. coli*.

This project demonstrated that the deletion of the *rlmN* gene was possible via CRISPR Cas9 gene deletion facilitated by homology directed repair with donor DNA. Additionally, it was demonstrated that the deletion may affect the growth rate of the deletion strain but is ultimately non-lethal.

Adenine supplementation appeared to impact the variety of lower ligand incorporated to form a cobamide in the absence of DMB. Without supplemented adenine, as an intracellular booster or externally with powder, 2-methyladenine appears to be the main preferred lower ligand to be incorporated into a cobamide and produces the

highest ratio of cobamides. Contrastingly, when supplemented with adenine, with an intracellular booster or externally, adenine becomes the main lower ligand of the cobamides produced by ratio. Interestingly, the addition of an intracellular adenine booster results in the loss of GDP cobinimide production. This project demonstrates the that the precursors for 2-methyladenine as a lower base are of intracellular origin, possibly from *rlmN* methylated t-RNA. Additionally, we have demonstrated that cobamides are preferentially produced when supplemented with adenine, even in the absence of DMB.

Chapter 1 – Introduction

1.1 What is Cobalamin (Vitamin B_{12})?

Vitamins are organic compounds which are important for normal growth and development but cannot be synthesized by the body, therefore they must be acquired through one's diet.[1] Cobalamin, also known as vitamin B12 is a widely used cofactor in cellular metabolism but is only synthesized in a few species of prokaryotes.[2] For humans with diets lacking in animal products, can result in a deficiency. As it is a component of a variety of essential pathways such as the formation of erythrocytes, the synthesis of DNA and the formation of the myelin sheath, this can result in diseases such as megaloblastic anemia and neurological symptoms such as lesions in the peripheral nerves and spinal cord. [1,3]

As a type of corrinoid, it is a fascinatingly complex molecule consisting of a β-axial (upper) ligand, a corrin ring containing cobalt and an α-axial (lower) ligand which is connected to the ring via a nucleotide loop as seen in Figure 1-1.[2] The lower ligand can be varied, with 5,6-dimethylbenzimidazole (DMB) being the only base which is the only form of cobamide used in eukaryotic metabolism. [4] An example of a cobalamin dependent enzyme in humans is methionine synthase, which converts L-homocysteine to L-methionine.[5] Methionine is a crucial amino acid which is coded for by the start codon AUG, thus its presence is required for protein synthesis.

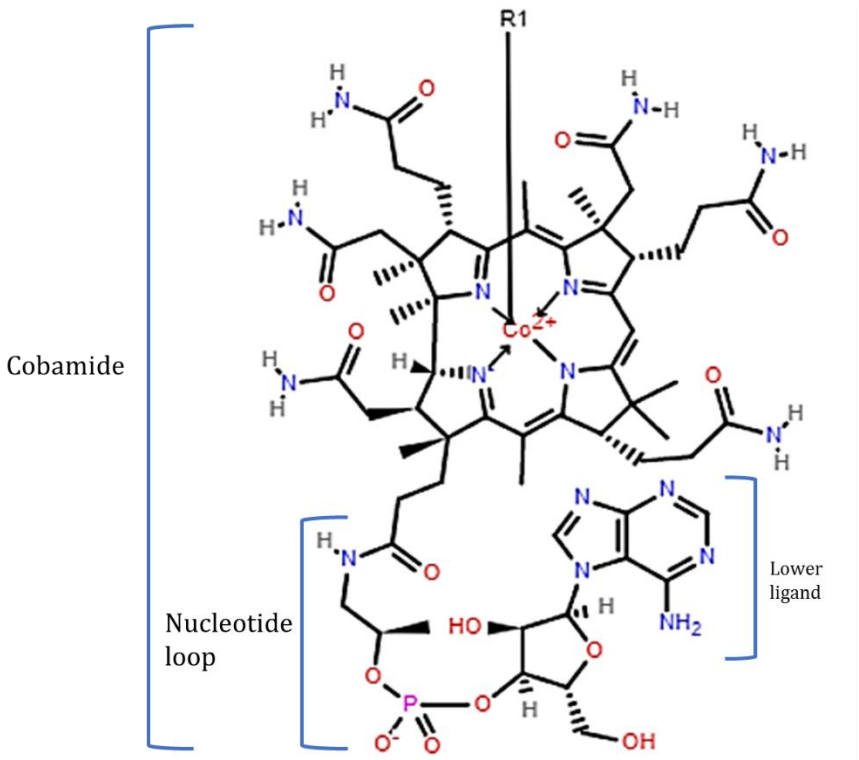


Figure 1-1: The chemical structure of pseudocobalamin or [MeAde] Cba, a cobamide with adenine as the lower ligand. The upper ligand (R1) can be varied. [Image generated in ChemSpider]

As a large natural product, it requires a complex network of around 30 enzymes to perform an elaborate series of reactions for its synthesis. There are two known independent biosynthetic pathways, the aerobic pathway, and the anaerobic pathway, though they both start with the same universal substrate, uroporphyrinogen III, they contrast in the timing of the step in which cobalt is inserted into the macrocycle, the requirement for molecular oxygen and the mechanism of ring contraction, where the C carbon at position C20 is extruded from the molecule.[6] Both pathways can be seen in figure 1-2.

In the aerobic pathway, cobalt is inserted after the formation of the precorrin ring and after its conversion to hydrogencobyrinic acid a,c-diamide. Contrastingly, in the

anaerobic pathway, cobalt is bound to sirohrydrochlorin and the corrin ring is formed after as seen in figure 1-2. Both of these result converge as they result in the formation of cob(II)yinic a,c-diamide.[7] Adenosylation takes place to form the upper ligand, after some steps, the lower ligand is added in a similar fashion for both pathways by CobU, CobS and CobT, resulting in the final product of coenzyme B12, a cobamide with an adenosine upper ligand replacing R1 and a DMB lower ligand, taking the form seen in figure 1-1. [6,8]

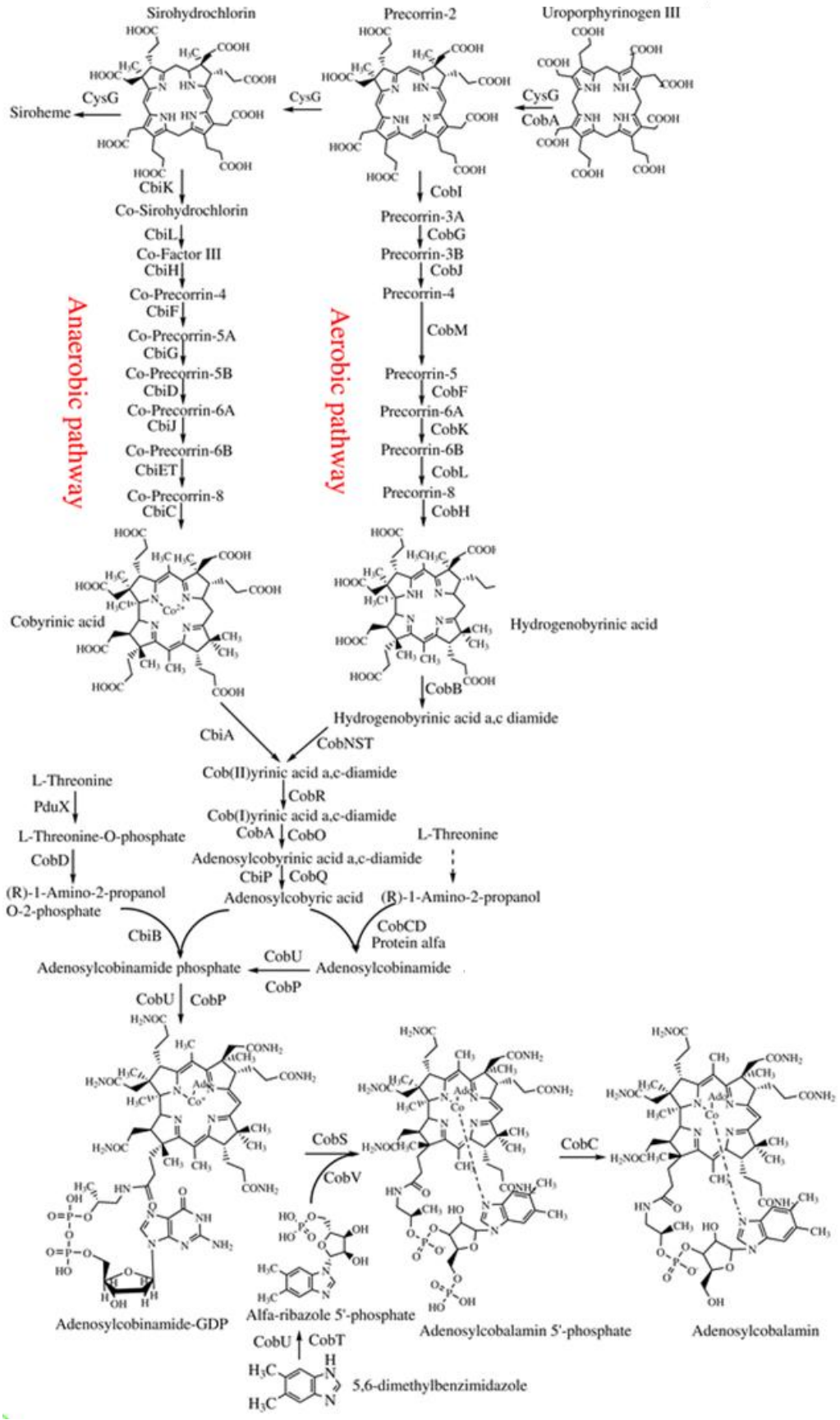
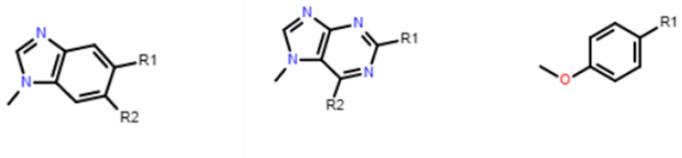

Sirohydrochlorin
Precorrin-2
Uroporphyrinogen III
CysG
Siroheme
CysG
CysG
CobA
CbiK
Co-Sirohydrochlorin
CobI
CbiL
Co-Factor III
Precorrin-3A
CobG
CbiH
Precorrin-3B
CobJ
Co-Precorrin-4
Precorrin-4
CbiF
Anaerobic pathway
Aerobic pathway
Co-Precorrin-5A
CobM
CbiG
Co-Precorrin-5B
Precorrin-5
CobF
CbiD
Co-Precorrin-6A
Precorrin-6A
CobK
CbiJ
Co-Precorrin-6B
Precorrin-6B
CobL
CbiET
Co-Precorrin-8
Precorrin-8
CobH
CbiC
Cobyrinic acid
Hydrogenobyrinic acid
CobB
Hydrogenobyrinic acid a,c diamide
CbiA
CobNST
Cob(II)yrinic acid a,c-diamide
L-Threonine
CobR
PduX
Cob(I)yrinic acid a,c-diamide
L-Threonine-O-phosphate
CobA
CobO
L-Threonine
CobD
Adenosylcobyrinic acid a,c-diamide
CbiP
CobQ
(R)-1-Amino-2-propanol
O-2-phosphate
Adenosylcobyric acid
(R)-1-Amino-2-propanol
CobCD
Protein alfa
CbiB
CobU
Adenosylcobinamide phosphate
Adenosylcobinamide
CobP
CobU
CobP
CobS
CobC
CobV
Adenosylcobinamide-GDP
Alfa-ribazole 5'-phosphate
Adenosylcobalamin 5'-phosphate
Adenosylcobalamin
CobU
CobT
5,6-dimethylbenzimidazole

Figure 1-2: The adenosyl cobalamin biosynthetic pathway, taken from Fang et. al[7]. The process of integrating the lower ligand can be observed in the lower part, with CobU adding a GDP group to the lower ligand of the Cbi. CobS incorporates the lower ligand containing DMB to form the precursor to cobalamin.

1.2 Cobamide Variations

In addition to their structural complexity as seen in figures 1-1 and 1-2, numerous analogues of B_{12} are found in nature. With variation in their upper and lower ligands, making them a diverse family of molecules. The main groups that can occupy the upper ligand (R1 in figure 1-1) are cyanide, methyl, adenosyl, and hydroxy groups, with the weakest bonds being that of the adenosyl and methyl groups, the cyanide upper ligand. The bond between the adenosyl group and the corrin ring is light sensitive thus, the molecule can convert to a hydroxy group when exposed to light, this property can be used to homogenise solutions with both types of upper ligand present.

The lower ligand for cobamides is varied to an even larger extent with over 15 known variants which can be differentiated through their structure classes, i.e., benzimidazoles, purines and phenolics as seen in figure 1-3.[4] These wonderfully diverse array of ligands is incorporated by the CobU, CobS and CobT enzymes. CobT converts benzimidazoles and other bases to a form of α-ribazole phosphate, which is then incorporated with adenosyl GDP cobinamide by CobS to produce cobalamin phosphate, as seen in figure 1-2.[9] CobT has a varied specificity in the classes of compound that can be incorporated into the alpha-ribazole, in the absence of DMB, they can incorporate adenine or 2-methyl adenine into a subsequent ribazole, this combined with substrate availability results in one aspect of cobamide diversity. [10]



Benzimidazole Purine Phenol

Figure 1-3: A depiction of the various cobamide lower ligands, the R1 regions can vary. [Images generated in ChemSpider]

This project specifically takes interest in adenine and 2-methyladenine as lower ligands in cobamides, as labelled in figure 1-1 with adenine as an example. Figure 1-4 depicts the formation of varying forms of α-ribazole phosphate from DMB, adenine and 2-methyladenine which form cobalamin, pseudo cobalamin ([Ade] Cba) and 2-methyladenyl cobamide ([MeAde] Cba) respectively.

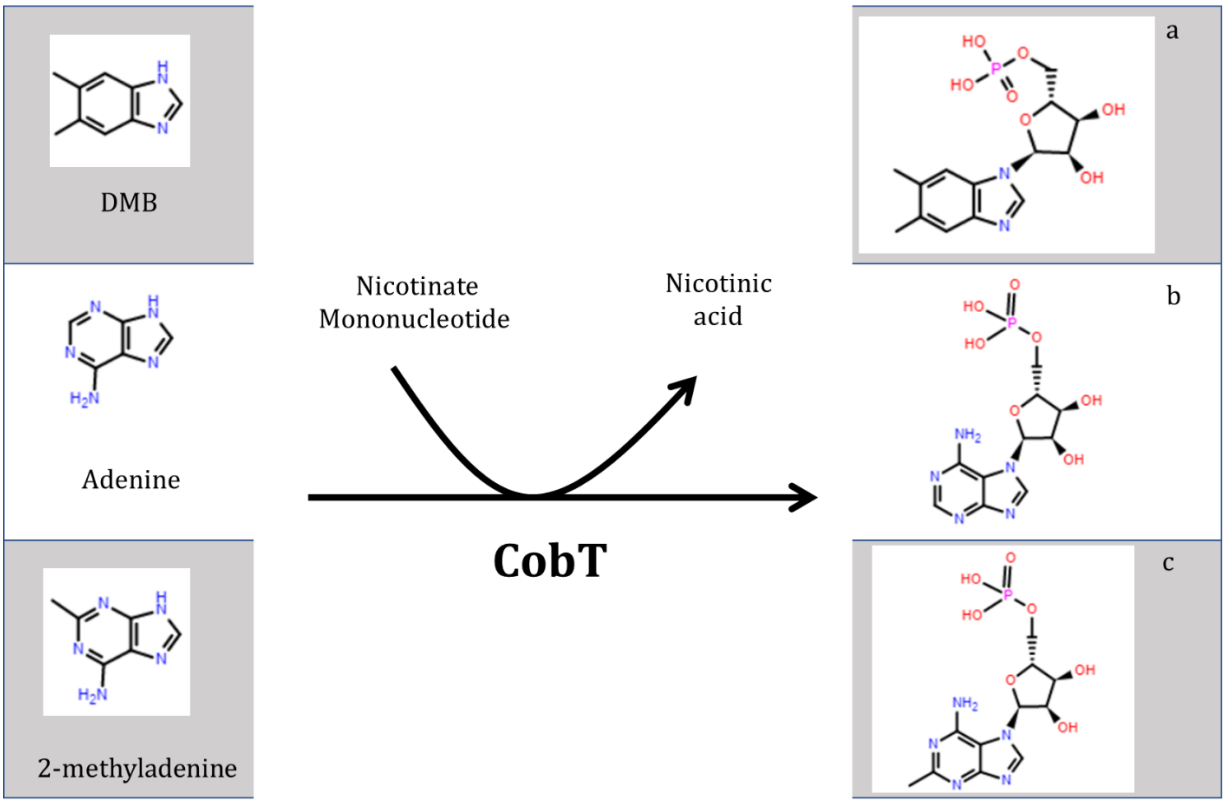


Figure 1-4 The formation of α-ribazole with is catalysed by CobT. The products on the right in the following order, a is the lower ligand to cobalamin (B_{12}), b is the lower ligand to [Ade]Cba, adenosyl cobamide (pseudo cobalamin), and c is the lower ligand to [MeAde] Cba, all in α-ribazole form.

1.3 Corrinoid Diversity in the Gut Microbiome

Like most vitamins humans consume, most of the cobalamin obtained through the diet is absorbed in the small intestine, therefore cobalamin in the large intestine is not bioavailable.[3] Despite this, 25% of sequenced human intestinal bacteria have the capacity to produce corrinoids and 80% produce gene products that potentially require corrinoids, most commonly methionine synthase suggesting the possibility of corrinoid sharing in this microbiome. [11]

Cobamides are present in the large intestine and are even produced by the flora which reside there, with cobalamin (DMB lower base) only making up under 2% of the

corrinoid content in human faecal samples. There are other forms of corrinoids present such as [2-MeAde] Cba, which is one of the most frequently occurring and abundant form of corrinoid at over 60% in this context appearing ubiquitously in samples. Other Cba analogues such as [Ade]Cba and [MeSAde]Cba occur at about 12% and 15% on average in samples, not nearly as frequent as [2-Me] Cba. However, little is known about the precursors to these lower ligands.[11]

1.4 Adenine Methylation

Adenine is one of four nucleotide bases in DNA and RNA, the coding and transcription materials in cells. Its methylation is often correlated to gene expression and regulation in cells, such as the methylation of tRNA and rRNA facilitates translational accuracy. In addition to this, 2-methyladenine is also one of a variety of lower bases attached to a cobamamide in the synthesis of cobalamin as it can be incorporated into an α-ribazole by CobT, making it a versatile molecule in cellular metabolism. [4,10]

The addition of adenine as the lower ligand renders cobalamin unusable as a cofactor for eukaryotes, earning it the name pseudo cobalamin. In the absence of DMB, adenine can be readily used as a substitute in the biosynthesis of cobalamin in *S. enterica*. [12] This form of cobamide could be utilized by the prokaryote to perform a corrinoid dependent reaction, demonstrating that a corrinoid with an alpha axial ligand other than DMB can serve as valid cofactor in at least one organism. However, when DMB was reintroduced, the synthesis of cobalamin with it as the lower base resumed, suggesting a preference for cobalamin formation. [12]

A family of enzymes that may be capable of methylating adenine on nucleic acids are the radical S-adenosylmethionine (SAM) superfamily of enzymes. The member that this

project particularly takes interest in is a dual-specificity RNA methyltransferase (rlmN), a methyl transferase that plays a role in modulating protein synthesis. This enzyme modifies an adenine base of the ribosomal RNA (rRNA) of the 23S ribosomal subunit during the ribosomal assembly, specifically A2503, converting it to 2-methyladenine. [13]

The methylation of the 23S ribosomal subunit usually takes place early in the process of ribosome assembly, with the process being highly conserved in bacteria. The modification of nucleotides in 23S ribosomal subunits is important for the formation of fully functional ribosomes, but the absence of this modification is not fatal to the cell. However, the A2503 site is an important site for the binding of translational RNA (tRNA) as it is in the binding pocket for the A-site, which facilitates the reading of the stop UAG codon, and this could increase the susceptibility of the *E. coli* to peptidyl transferase antibiotics.[13]

1.5 Clustered Regularly Spaced Palindromic Repeats (CRISPR)

The Cas9 protein from *S. pyogenes* is a popular choice for scientists for gene modification, its origins are as a form of defence against phage infection and subsequent plasmid transfer in bacteria. As an endonuclease, it performs a double strand cleavage on DNA, a process that has been harnessed into a versatile and practical tool for gene editing experiments. [14]

The CRISPR Cas9 system for gene editing requires two components, the Cas9 protein which interacts with the single strand guide RNA (sgRNA) to form the system that performs an accurate double strand cleavage on the target sequence of genomic DNA. To locate the desired cut site, a protospacer adjacent motif (PAM) sequence is needed, this varies among the types of Cas proteins. In *S. pyogenes* Cas9, the sequence is 5'-NGG-

3’ with N being any nucleotide base. In nature, this serves the additional purpose of preventing potential cuts in the bacterium’s own genome as the spacers within the CRISPR array of *S. pyogenes* is 5’-GTT-3’.

The Cas9 protein contains two active sites for nuclease activity known as the HNH and RuvC domains, which cleave the lagging and reading strands respectively. This double stranded cut takes place three base pairs preceding the PAM sequence on the 5’ strand, this can be observed in figure 1-5.[14]

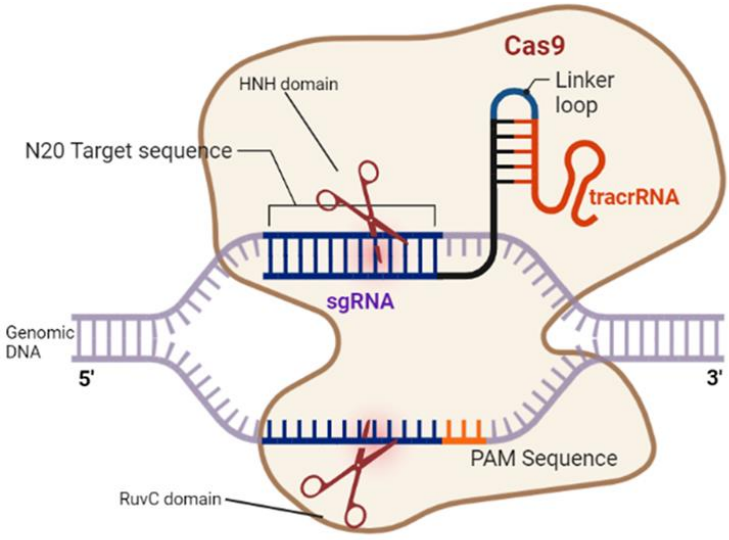


Figure 1-5: a depiction of the cut sites made by Cas9 for gene editing [an edited template from BioRender]. This figure illustrates the double lobed nature of the Cas9 protein, each domain contains a cleavage site specific to the sgRNA region. The HNH region cleaves the site complimentary to the sgRNA and the RuvC cleaves the identical site.

The cut can be repaired through non-homologous end joining or homology directed repair (HDR), the latter having more applications in experimental use due to its versatility and precision. HDR can be used to insert and delete sequences for gene

modification, this requires donor DNA to be used a template for repair in the form of single stranded oligonucleotides or, as in this project, a benign plasmid with the desired sequence. [14,15]

1.6 The Bacterial Species and Strains Used

In this project, *E. coli* strains which were genetically engineered to produce cobamides were used, they are herein referred to as ED671 and ED674. These contain most of the genes required for de novo synthesis of cobalamin as seen in figure 1-2 and were incorporated into the *E. coli* genome by Dr. E. Deery who also gifted the modified CRISPR system used in this project. The main difference between ED671 and ED674 is that ED671 contains an intracellular adenine booster in the form of AMP nucleosidase which produces free adenine in the cell by breaking down AMP into adenine and ribose phosphate, their respective genomes are detailed in table 2-1.

1.7 Project Aims

In this project, we will be investigating the role of *rlmN* and its correlation to the production of [2-MeAde] Cba. This will be achieved by using CRISPR to delete the *rlmN* gene in two strains of *E. coli* able to produce cobamides (ED671 and ED674). The cobamamides produced from the resulting and wild type strains will be extracted, purified, and analysed using high pressure liquid chromatography coupled with mass spectrometry.

Chapter 2 - Materials and methods

2.1 Media for Bacterial Work

Most of the media and components used were sourced from Melford except yeast extract which was sourced from Oxoid.

Lysogeny Broth (LB)

10g Tryptone

5g Yeast extract

5g NaCl

The components were dissolved in 1L of distilled water and autoclaved. To make LB for plates, 15g of agar powder was added before autoclaving.

Super optimal broth with catabolite repression (SOC)

2g Tryptone

0.5g Yeast extract

1mL (1M) NaCl

0.25mL (1M) KCL

The components were added to 97mL of distilled water and this solution was autoclaved. The two following components were made up and filter sterilized before 1mL of each was added to make the final solution.

(2M) glucose

(2M) Mg^{2+} solution

The 2 M Mg^{2+} solution consists of 2 g $MgCl_2 \cdot 6H_2O$ and 2.5 g $MgSO_4 \cdot 7H2O$ made up to 10mL with distilled water.

Supplemented Minimal Media for B12 producing strains

M9 Salts 10x

60g Na_2HPO_4

30g KH_2PO_4

10g NH_4Cl

These components were made up into a 1L solution with distilled water, this was then autoclaved.

HMO 10 Trace Elements Solution

500 mL Distilled and deionised H_2O

4 g Na-EDTA · $2H_2O$

0.395 g $CaSO_4$ ·$2H_2O$

0.15 g $ZnSO_4$ · $7H_2O$

3.7 g (NH4)2Fe [II]$(SO_4)_2$ · $6H_2O$

0.1 g $MnSO_4$ · H_2O

0.048 g $CuSO_4$ · 5 H_2O

0.02 g Na_2MoO_4 · 2 H_2O

0.02 g $Na2SeO_4$

0.5 g Citric Acid

These components were added to water while stirring, resulting in a pale green solution. The solution was stored at room temperature away from light and was filter sterilized directly before use.

$CoCl_2$ Solution

This solution was made up to 25 mg/mL with distilled water and autoclaved.

$MgSO_4$ Solution

This solution contained 2.4647 g of $MgSO_4 \cdot 7H_2O$ was dissolved in 10mL of H_2O to make 1 M and filter sterilized.

Final solution, 1 L

10 g Yeast extract

100 mL 10 x M9 salts

10 mL 50% glycerol

2 mL 1 M $MgSO_4$

2 mL $CoCl_2$, 6H2O 25 mg/ml

10 mL HMO10 trace elements

The yeast extract powder was dissolved in 900 mL of distilled water and autoclaved in a baffled flask covered with a spongy bung and foil. The 50% glycerol solution was made up and autoclaved separately before added to the final solution with the rest of the components.

Isopropyl beta-D-1-thiogalactopyronoside (IPTG)

This solution was made up to 1 M (0.238 g/mL) with distilled water and filter sterilized.

Arabinose Solution

This solution was made up to 1 M (0.105 g/mL) with distilled water and filter sterilized.

Classic Competent Cell Solution

This was made up of 0.1 M $CaCl_2$, 15% glycerol (v/v) and ddH_2O. This was filter sterilised through a 0.2 μm syringe and chilled in ice slurry before use.

TB solution

100 mL ddH_2O

302 mg $CaCl_2$

1.864 g KCl

1.088 g $MnCl_2 \cdot 4H_2O$

50 mL of ddH_2O, $CaCl_2$, and KCl were mixed, and the pH was adjusted to 6.7 with KOH. $MnCl_2$ and the rest of the water was added to give a final volume of 100mL. The solution was filter sterilised and chilled in ice slurry before use.

Antibiotics Used

Table 2-1: A record of the antibiotics used in this project.

Antibiotic	Stock Concentration	Working Concentration
Ampicillin	100 mg/mL	100 mg/L
Chloramphenicol	34 mg/mL (in ethanol)	34 mg/L
Kanamycin	25 mg/mL	25 mg/L
Spectinomycin	50 mg/mL	50 mg/L
Tetracycline	10 mg/mL	10 µg/L

All antibiotics are prepared in distilled and deionised water (ddH_2O) unless stated otherwise, they are also filter sterilized.

Bacterial strains Used In this Project

Table 2-2: A record of the bacterial strains used in this project.

Strain	Genotype	Phenotype/Description	Origin
JM109	endA1, recA1, gyrA96, thi, hsdR17 (rk–, mk+), relA1, supE44, Δ(lac-proAB), [F´ traD36, proAB, laqIqZΔM15].		Promega
BL2 (DE3) pLysS	F–, ompT, hsdSB (rB–, mB–), dcm, gal, λ(DE3), pLysS, Cm^r	Expresses protein under the control of T7 promoter	Novagen
DH10β	F– mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ (ara-leu)7697 galU galK λ– rpsL(StrR) nupG		New England BioLabs
ED674	BGEC043 with (T7P)-AIG*JFMKLHBWNSTQJDE C*FPUVE*647 integrated in the fim operon + with the ΔlacZYAΩ(T7RNAP) clone P3. + with (Bmei 708-cobWNST-btuR-cobR-cobB) integrated with CRISPR in 5' of the E. coli cobUST to form a longer operon	Produces adenyl cobamide	Dr. E. Deery
ED671	BGEC043 with (T7P)-AIG*JFMKLHBWNSTQJDE C*FPUVE*647 integrated in the fim operon + with the ΔlacZYAΩ(T7RNAP) clone P3. + with (Bmei 708-cobWNST-btuR-cobR-cobB) (Se Amn-PTH0960) integrated with CRISPR in 5' of the E. coli cobUST to form a longer operon	Produces adenyl cobamide and has an adenine booster	Dr. E. Deery
ED674 (RlmN deletion)		Is like ED674, however, the gene R*lmN* is deleted via CRISPR	This project

2.2 Methods for Bacterial Work

Aseptic Technique

All media was sterilized with an autoclave at 121 °C for 15 minutes, the equipment used for this was also processed in the same manner or sterile disposable plastic. A 0.2 μm filter attached to a syringe was used for filter sterilization. All sterile work takes place next to a blue flame.

Preparation of Starter Culture

One colony is selected from an LB plate to inoculate 5 mL of liquid LB with 0.2% (w/v) glucose and the working concentration of respective antibiotics, this culture is incubated overnight at 28 – 37 °C, 180RPM.

Preparation of Glycerol Stock

Sterile glycerol was prepared as 80% (v/v) solution with ddH_2O and autoclaved. In a cryotube, 700 μL of starter culture was added followed by 300 μL of glycerol solution, this was stored at -80 °C until required.

Preparation of Bacterial Culture on Agar Plate

An LB plate was prepared with heated LB agar solution. While the solution is hand-hot but still liquid, 0.2% (w/v) glucose and a working concentration of respective antibiotics are added, this solution is then poured into sterile petri dishes and left to solidify while covered. Bacteria is streaked on this plate with a sterile loop from glycerol stock.

Classic Competent Cells

Starter cultures of the bacteria required were made up the day before. 40mL of liquid LB with respective antibiotics in a glass flask was inoculated with 1% (v/v) of starter culture. This was incubated at 28 - 37°C, 180 RPM until OD_{600} reached 0.3 – 0.6. The culture was then transferred to a falcon tube, chilled in ice slurry for 10 minutes followed by centrifugation at 4,000 RPM for 10 minutes. The supernatant was discarded, and the pellet was resuspended in 20 mL of competent cell solution, this was kept in the ice slurry for a further 20 minutes. The contents of the falcon tube were centrifuged under the same conditions again, the supernatant was discarded, and the pellet was suspended in 2 mL of ice-cold competent cell solution. This was aliquoted in sterile 1.5 mL Eppendorf tubes as 50 µL and stored at -80 °C until further use.

Bacterial Transformation for Classic Competent Cells

Competent cells were thawed in ice slurry, 3 – 5 µL (up to 10% v/v) of desired plasmid or ligation mixture was added. This was kept on ice for 15 minutes followed by 1 minute in a 42 °C heat block. The mixture was kept on ice again for 2 minutes and 200 µL of sterile SOC media was added. This was incubated at 28 – 37 °C for 30 minutes to 1 hour before being plated on LB agar with 0.2% (w/v) glucose and respective antibiotics, it was incubated overnight at the same temperature.

Ultra-competent Cells (from updated CRISPR protocol)

25 mL of LB with 0.2% (w/v) glucose and respective antibiotics were inoculated with 1% (v/v) starter culture in a glass flask, this was incubated at 28 °C, 150 RPM until the OD_{600} reached 0.2. 250 µL of 10mM arabinose solution (final concentration) is added and the culture is incubated under the same conditions for a further 30 minutes. After

this, it is chilled in ice slurry for 15 minutes and centrifuges at 4000 RPM for 10 minutes at 4°C.

The supernatant was discarded, and the pellet is suspended in 5 mL of ice-cold TB solution, this incubated on ice for 25 minutes. This solution is centrifuged again under the same conditions, the supernatant was discarded, and the pellet is resuspended in 1 mL of TB solution, 0.25 mL of sterile glycerol is added as well before it is stored at -80 °C in 80 µL aliquots in sterile 1.5 mL Eppendorf tubes.

Measuring the Growth of Cobamide Producing Strains

LB agar plates with 0.2% (w/v) glucose were inoculated with the strains of interest and incubated at 28 °C overnight. One colony was used to inoculate an overnight starter culture in sterile LB with 0.2% (w/v) glucose.

The following day supplemented minimal yeast extract media was made up with the sterile method as mentioned in before. The overnight culture was diluted to 0.1 OD_{600} in the minimal newly made media. In Grenier 24 well tissue culture plates, 750 µL of the diluted overnight culture and 750 µL of additional sterile media were added to each plate, making a total of 1.5 mL culture. This was incubated at 28 ˚C double orbital shaking for 24 hours in a Fluorostar Omega shaker which took an OD_{600} measurement every 8 minutes from each well.

B12 Producing Strain Growth and B12 Extraction

B12 Producing Strain Growth

An LB agar plate was streaked with the culture of interest and incubated overnight (about 16 hours) at 28 ˚C. A 5 mL overnight starter culture was prepared from the LB

plate prepared from before this was incubated at 28 ˚C, 150 RPM. 400 mL total volume of sterile supplemented minimal media was prepared and inoculated with 1% (v/v) of overnight starter culture and incubated under the same conditions. When the OD_{600} reached 1, 0.4 mM (final concentration) IPTG was added, and the incubation was continued overnight (16 hours) under the same conditions. For adenine supplemented conditions, 0.1 g/L of adenine powder was added to the culture when IPTG was added.

Pellet Preparation for B12 Extraction

The following day, the culture containing cobalamin were centrifuged at 4000 RPM for 20 minutes, the supernatant was discarded, and the pellet was resuspended in 40mL of lower salt buffer (20 mM HEPES pH 7.5, 100 mM NaCl and 5 mM imidazole). This was boiled in a water bath for 15 minutes and centrifuged at 18,000RPM for 20 minutes. The pellet was discarded, and the supernatant kept for cobalamin purification.

Protein Growth and Pellet Harvesting

Making BL21 Star (DE3) pLysS pET14b-with gene of interest

A 5mL starter culture of BL21 Star (DE3) pLysS was inoculated in LB with chloramphenicol (34 mg/L) and glucose (0.2% w/v). This was used make competent cells and pET14b was transformed into BL21. The LB plate with chloramphenicol (34 mg/L), ampicillin (100 mg/L) and glucose (0.2% w/v) containing the newly made strain was incubated at 37˚C.

The following day, a 10mL starter culture was made by inoculating LB with chloramphenicol (34mg/L), ampicillin (100mg/L) and glucose (0.2% w/v) by adding on colony from the LB plate before. This was incubated at 37˚C overnight (about 16 hours).

A sterile 3 L flask containing 1 L of inoculating LB with chloramphenicol (34 mg/L), ampicillin (100 mg/L) was inoculated with 1% v/v of overnight starter culture and incubated at 37 ˚C, 150 RPM until the OD_{600} reached 0.6. 0.4 mM IPTG was added, and the culture was incubated at room temperature overnight.

Harvesting Pellet for Protein Isolation

The cells were centrifuged at 4000 RPM for 20 minutes at 4 ˚C, the resulting supernatant was discarded, and the pellet is suspended in 30mL of binding buffer (500 mM NaCl, 5 mM imidazole, 20 mM pH 7.5 HEPES) and stored at -20 ˚C or directly used for protein purification.

2.3 Molecular Biology Work

Obtaining the primers for the gene deletion process

A total of 8 primers were ordered for this experiment. They were synthesised by Integrated DNA Technologies listed in Table 3.

Table 2-3: A record of the primers used in this project.

Primer name	Sequence 5' - 3'
RlmN_del_5'flank_SalI_F	ACGTGTCGACAATCCGCGCCATCTCGTACG
N20_RlmN_F	CTAGTACGTTGTTCAGCAGGTTTTAGA
RlmN_del_3'flank_BamHI_R	CACCGCTCAGGCCGATAACCGGATCCTAG
N20_RlmN_R	CCGCTGCTCAACCTGAACAACGT
Checking_Deletion_Rev	GCGTGGACGTTGTAGCC
Checking_Deletion_Fwd	GCTGGTGATGGATGACGT
RlmN_del_5'flank_R	GACCACAGGCAGCATGCTAT
RlmN_del_3'flank_F	GGCCGGAATAGCATGCTGC

Miniprep

The Qiagen QIprep Spin Miniprep kit was used to isolate plasmids for this project.

Firstly, RNase solution and LyseBlue were added to Buffer P1. Pure ethanol was added to buffer PE. To prepare bacterial culture for the miniprep, 5mL of starter culture was incubated the day before. This was kept on ice slurry for 15 minutes before centrifugation at 4000RPM for 10 minutes. The pellet was resuspended in 250 µL of Buffer P1 and transferred to a 1.5 mL Eppendorf tube. 250 µL of Buffer P2 was added and the tube was inverted 4 – 6 times followed by the addition of 350µL of Buffer N3 and inversion again. The tube was centrifuged at 13,000RPM for 10 minutes.

800µL of the supernatant was added to the QIprep spin column on another Eppendorf tube and centrifuged at 13,000 RPM for 30 – 60 seconds, the flow through was

discarded. 0.5mL Buffer PB and 0.75 mL of Buffer PE were added to the column respectively and centrifuged at 13,000 RPM for 30 – 60 seconds between each addition. The column was again to remove any residual buffer.

To elute DNA from the column, it was transferred to a clean 1.5 mL Eppendorf tube, 30µL of sterile ddH_2O was added to the column and left to stand for 5 minutes followed by centrifuged at 13,000 RPM for 60 seconds. The eluent was stored at -20˚C until ready for use.

Agarose Gel and Electrophoresis

1% agarose (w/v) was made in TAE buffer to a final volume of 50mL and heated until dissolved. 1:10,0000 (5µL) of SYBR safe dye was added to the solution when it cooled to about 60˚C, it was gently swirled until homogenised and poured into a mould with 12 slots and was then left to solidify away from light.

The dye used in this process is a 10x dye from Promega. The DNA ladder for this process is a 1 kb+ ladder from Promega as well and, electrophoresis was performed at 100 V for 25 minutes.

Gel Extraction

The Qiagen QIquick Gel extraction kit was used for this process. The desired band is first excised from the agarose gel and placed in a 3 mL Eppendorf tube. This is weighed and 3 volumes of Buffer QG are added per volume of gel for a maximum of 400 mg of gel, the tube is warmed at 50 ˚C for 10 minutes or until the agarose is fully dissolved, inverting every 2 – 3 minutes.

1 gel volume of pure isopropanol is added to the tube, and it was inverted until homogenised, this liquid was added to a QIquick spin column placed in a fresh Eppendorf tube. This was centrifuged for 1 minute at 13,000 RPM. 0.5 mL of Buffer QG was added to the column, and it was centrifuged again for 1 minute at 13,000 RPM. 0.73 mL of Buffer PE was added to the column, and it was centrifuged again for 1 minute at 13,000 RPM. The flowthrough was discarded between each step.

To remove residual buffer, the column was centrifuged again under the same conditions and the flowthrough discarded one more. The QIquick column was then transferred to a clean 1.5 mL microcentrifuge tube, 30 µL of autoclaved, distilled, deionised H_2O was added to the column instead of EB Buffer. The column was left to stand for 3 minutes and for 1 minute at 13,000 RPM, giving a yield of 30 µL of product.

Restriction Digest

The enzymes used for this process were mostly sourced from Promega. The constituents of the reactions are described in the table below.

Table 2-4: The components of a restriction digest

Components	Digest with 1 Enzyme (µl)	Digest with 2 Enzymes (µl)
Miniprep DNA	5	5
Buffer (x 10) [Selected based on the enzyme used]	1	1
Millipore water (autoclaved)	10	10
Restriction enzyme 1	1	0.5
Restriction enzyme 2	0	0.5

The reaction in conducted in clean microcentrifuge tubes and incubated at 37 ˚C for 1.5 hours. The Agarose Gel and Electrophoresis protocol was followed to analyse the size of the digest and the Gel Extraction protocol was followed to obtain the digested fragments.

Polymerase Chain Reaction (PCR) Amplification

The materials used for polymerase chain reaction (PCR) amplification in this project were NEB High-Fidelity Q5 Polymerase, Q5 High GC Enhancer and 5X Q5 Reaction Buffer. The genomic template was generated by obtaining one colony from an LB plate using the sterile method and suspending it in 10 µL of autoclaved H_2O.

Table 2-5: The components of a PCR reaction.

PCR Reaction Components	No additives (µL)	With DMSO (µL)	With DMSO and GC enhancer (µL)
Q5 reaction buffer 5x	5	5	5
DMSO	0	5	5
GC enhancer 5x	0	0	5
Forward primer @ 10 µM	2	2	2
Reverse primer @ 10 µM	2	2	2
dNTPS @ 2mM	5	5	5
Template	1	1	1
H_2O (distilled, deionised, autoclaved)	34.5	29.5	24.5
Q5 enzyme	0.5	0.5	0.5

PCR Protocol

This protocol was tailored to the expected length of the product in the elongation stage.

Table 2- 6: The PCR protocol used.

Cycles	Temperature (˚C)	Time (seconds)
1	96	120
30	96	30
	55	30
	72	60 seconds for 1,000 base pairs
1	72	420

The resulting product was stored at 4 ˚C on the PCR machine or -20 ˚C until use.

Annealing the N20 insert

The N20 primers were hybridized as they were heated to 95˚C for 30 seconds and cooled to 4˚C on a PCR block. This was done with the PCR reagent list with no Q5 enzyme or buffer.

pCas Plasmid Preparation

The pCas plasmid is #62225 on AdGene, it expresses a Cas 9 plasmid that is inducible with L-arabinose and is kanamycin resistant[15]. It was amplified in JM109 competent cells at 30˚C for 48 hours in an LB plate with kanamycin and glucose before being extracted.

Ligation

T4 ligase from NEB was used for this protocol, the vector and inserts for this protocol are obtained through the PCR amplification, miniprep and gel extraction protocols.

Table 2-7: The ligation reaction in this project.

Components	Starter Reaction (µL)	Adjustments, more insert
10x ligase buffer	2	2
Vector	2	4
Insert	2	1
T4 DNA Ligase	2	2
H_2O (distilled and deionised)	8	11

The reaction was left at room temperature for 2 hours or at 4 ˚C overnight. It was used in the Competent Cell Transformation protocol the following day.

Sequencing

The samples generated from the project were sent to GeneWiz for Sanger sequencing.

Protein Purification and Construction of Chelating Sepharose Column

Sonication

Sonication was used to lyse the cells in preparation for protein purification. The cells suspended in binding buffer were kept in an ice slurry throughout this process. The sonics Vibracell Ultrasonic processor was set to 50% amplitude, 30 second pulse for on and off for 5 minutes. After sonication the solution was centrifuged at 18,000 RPM at 4

˚C for 20 minutes and the supernatant, a cell lysate solution, was kept for the Chelating Sepharose column, the pellet was discarded.

Construction of Chelating Sepharose Column with Recombinant Protein

All buffer containing solutions were made with ddH_2O. HEPES buffer was adjusted to pH 7.5 with the addition of KOH 0.1 M.

Table 2-8: The buffers used for constructing a chelating Sepharose column.

Buffer Name	Buffer Components and Concentration
Binding Buffer	HEPES 0.02 M NaCl 0.5 M Imidazole 5 mM
Wash Buffer	HEPES 0.02 M NaCl 0.5 M Imidazole 0.06 M
Lower Salt Binding Buffer	HEPES 0.02 M NaCl 0.1 M Imidazole 5 mM
Charge Buffer	$NiSO_4$ 0.05 M

To construct the column, 3mL of chelating fast flow Sepharose was added to a column with wet filter paper at the base. A syringe with a needle in a bung was used to drip the

buffers into the Sepharose. The column was first charged with 20 mL charge buffer followed by 25 mL of binding buffer and subsequently, the cell lysate from the sonication step. The column was then washed with 25 mL of binding buffer followed by 10 mL of lower salt binding buffer as seen in figure 2-1.

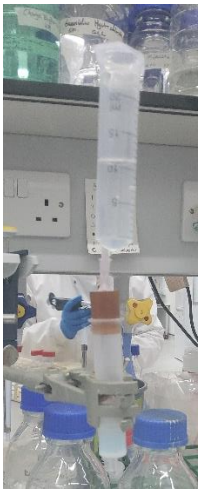


Figure 2-1: An image of Chelating Sepharose column with solvent reservoir charged with $NiSO_4$.

2.4 CRISPR Work

Plasmids Used

Table 2-9: A record of the plasmids used in this project.

Name	Description
pET14b	Vector for the expression of N-terminally 6xHis-tagged proteins modified to have an SpeI site.
pLysS	Expression of T7 lysozyme, T7 RNA polymerase inhibitor.
pTargetAfe	Vector containing spectinomycin resistance used to make pTargetN20.
pCas	Plasmid for the expression of Cas9 protein, contains kanamycin resistance vector.
pET14b-RecoveryED674	PCR product of RlmN flanking regions inserted into SalI and BamHI of pET14b.
pET14bBtuF	BtuF gene from *D. geothermalis* cloned into NdeI and SpeI of pET14b.
pTargetN20	N20 bases on RlmN cloned into NdeI and AfeI pTarget Afe.

Induction of the pCas Plasmid and Transformation of the pTargetN20 Plasmid

The resulting strain was used to inoculate 5mL overnight starter cultures with working concentrations of ampicillin and kanamycin. On the day of making the 'Ultra competent' cells, a 1:100 dilution in 25mL of LB was inoculated with the respective strains. These were grown for 4 hours, induced with 10mM L-arabinose, and incubated for a further 30 minutes to induce expression of the pCas plasmid. The cultures were then chilled in

an ice-water slurry for 15 minutes and made 'Ultra competent'. The procedure was like the CaCl2¬ method, but TB solution was used instead as seen in Chapter 2.2.

The procedure for plasmid transformation (pTargetN20) into the ED67x (pRecovery) (pCas) samples varied slightly from the transformation procedure in Chapter 2.2. 80 µL of competent cells. 3 µL of pTargetN20 and 0.8 µL of DMSO were incubated on ice for 5 minutes. This was followed by a minute of heat shock in a 42 ˚C water bath and further incubation on ice for 2 minutes. The cells were recovered in 200 µL of SOC at 28 ˚C for 1 hour and plated on an LB agar plate with kanamycin, spectinomycin and glucose.

Plasmid Curing

Four colonies were picked from the spectinomycin and kanamycin plates before and incubated as 5mL cultures in kanamycin (25mg/L) and tetracycline (10µg/L) overnight at 28 ˚C. and the resulting cultures were streaked on LB agar plates with kanamycin (25mg/L) and incubated overnight.

Selecting for the Deletion

A colony screening was performed for 20 of the resulting colonies on one of each of the following LB-agar plates with antibiotics, ampicillin (100mg/L), spectinomycin (50mg/L) and kanamycin (25mg/L).

2.5 Biochemical Work

B12 Purification

Buffers Used in the B12 Purification Process

All buffer containing solutions were made with ddH_2O. HEPES buffer was adjusted to pH 7.5 with the addition of KOH 0.1 M. Tris aminomethane (Tris) buffer was adjusted to pH 8.0 with 0.1M HCl.

Table 2-10: The buffers used for cobamide purification in this project.

Buffer Names	Components and Concentration
Lower Salt Buffer	HEPES 0.02 M NaCl 0.1 M Imidazole 5 mM
Lower Salt Binding Buffer	HEPES 0.02 M NaCl 0.1 M Imidazole 5 mM
Lower Salt Wash Buffer	HEPES 0.02 M NaCl 0.1 M Imidazole 0.06 M
Low Salt Buffer	HEPES 0.02 M NaCl 0.1 M
Elution Buffer	HEPES 0.02 M NaCl 0.1 M Guanidinium hydrochloride 6 M
Strip Buffer	Tris 0.02 M NaCl 0.5M Ethylenediaminetetraacetic acid (EDTA) 0.1M

Cobamide Purification Protocol using Affinity Chromatography

A charged column of Chelating Sepharose was used for this protocol to isolate cobamides with BtuF, cell extract from Section 2.2 contained the compounds to be extracted suspended in lower salt buffer was added to the column. The column was then washed with 25 mL of lower salt binding buffer and 10 mL of lower salt wash buffer respectively. 10mL of low salt buffer was added to the column as well. BtuF denaturation buffer is added to the column until it all colored solutions were eluted. BtuF is removed from the column by adding 30 mL of strip buffer or until the column is no longer blue, it is then washed with 50 mL water and can be reused.

Removal of Guanidinium Hydrochloride for HPLC Analysis

Reverse phase C18 fractionation was the preferred method to prepare the samples for LC-MS analysis by removing guanidine hydrochloride from the samples. LiChloroprep RP18 (25-10μM) sourced form Merk was used. It was activated with pure methanol and stored at room temperature until use. About 3mL of resin was added to a column and equilibrated with 5mL ddH_2O. The sample from Cobamide Purification Protocol was added and the column was washed again with 10mL ddH_2O. The sample was eluted with 50% methanol in ddH_2O until the column was white. To concentrate the samples and evaporate the methanol, they were dried in a vacuum centrifuge at 30˚C for up to 1 hour.

Reverse phase LC-MS analysis

All purified samples were dissolved in autoclaved double distilled water centrifuged at 50,000 RPM for 5 minutes and 40μL of sample was added to an HPLC vial. The mass spectrometry equipment used was attached to the HPLC machinery.

HPLC

The column used was Ace 5 AQ (2.1 x 250 mm; 5 µM; 100 Å) with the column oven set to 30 ˚C. Solvent A was 0.1% trifluoracetic acid (TFA) and solvent B was pure acetonitrile. The protocol used is described in the table and graph below.

Table 2-11: A description of the solvents used in the HPLC experiment; solvent A is 0.1% (v/v) TFA in water, and solvent B is acetonitrile.

Time (Minutes)	Solvent A (%)	Solvent B (%)
0	100	0
5	100	0
45	20	80
50	0	100
55	0	100
60	100	0
70	100	0

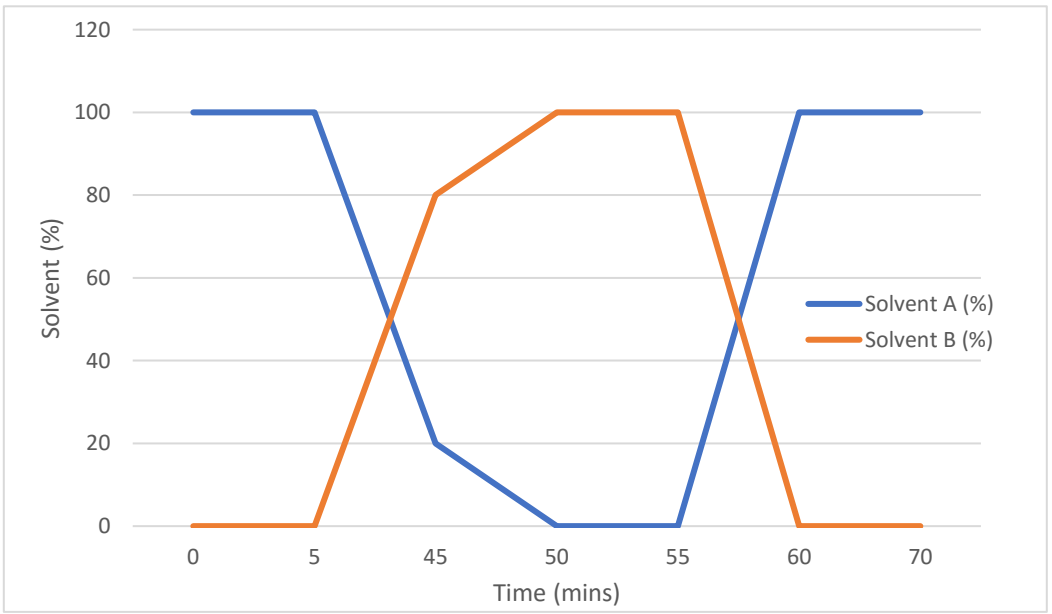


Figure 2-2: A graph depicting the ratio of solvents used in the HPLC experiment.

Mass Spectrometry

A Bruker micrOTOF II-MS was used to obtain electrospray ionisation mass spectrometry data (ESI-MS) using positive mode electrospray ionisation.

Chapter 3 – Results

3.1 – Molecular Biology Experiments

3.1.1 Preparing the Plasmids for the CRISPR Gene Deletion for *rlmN*

The purpose of the first molecular biology experiment was to prepare for the gene deletion process by preparing the plasmids required for the CRISPR process. These are the plasmids which express the Cas9 protein, supply the guide RNA and supply the donor DNA for homology directed repair. To tailor a CRISPR system for this project, these require synthesis from a template which was input in an online bioinformatics software Benchling.[16]

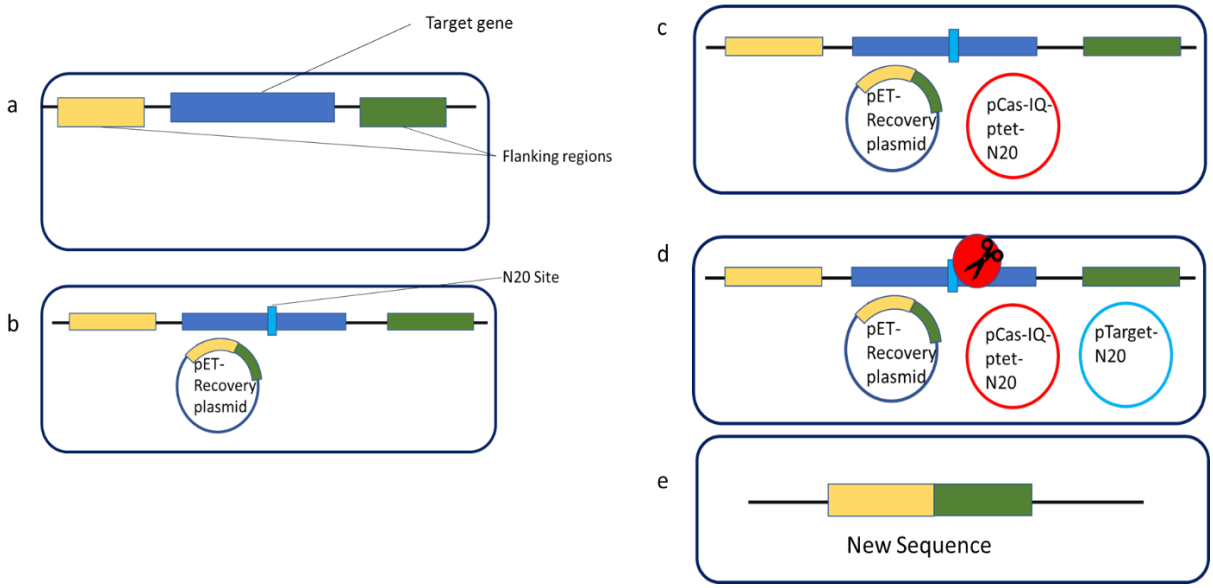


Figure 3-1: A depiction of the 2-plasmid system used in this project, the plasmids included in the system are pCas (red) and pTarget-N20 (light blue). The recovery plasmid (Dark blue) was tailored and synthesised as a part of the project.

The CRISPR system used was a two-plasmid system originating from Sheng Yang [15] which contains a Cas9 expressing plasmid and a NGG targeting plasmid (pTargetN20),

depicted in Figure 3-1. This system is capable of both gene insertion and deletion through homology directed repair (HDR). The pCas plasmid was modified by Dr. E. Deery, who replaced the IPTG promoter with the araBAD promoter, allowing the overproduction of the pCas when L-arabinose was added, this also increases efficiency as the cobamide biosynthetic genes in ED671 and ED674 are IPTG inducible. A recovery plasmid was also synthesized to facilitate HDR in this project and is described in the following subsection. In total, three plasmids were required for the CRISPR experiment to delete the *rlmN* gene, these were named pCas, pRecovery and pTargetN20.

Cloning the pRecovery Plasmid

The recovery plasmid (pRecovery) is a plasmid containing the donor DNA as an insert in a bacterial vector. To create a benign donor plasmid to reduce potential complications and increase efficiency, the promoter regions were removed. The donor DNA for this experiment consisted of DNA which was homologous to about 500 bp from both the flanking regions upstream and downstream of the *RlmN* gene. The two regions were amplified by PCR separately and spliced together, a technique referred to as overlap extension PCR to form the insert as seen in figure 3-2.

The pET14b vector used for this experiment was a gift from Dr. E. Deery, it contains an ampicillin resistant marker, and it was modified to contain an additional SpeI restriction site and tetracycline promoter site to aid in the curing of this plasmid later. The recovery plasmid (pRecovery) consists of the pET14b backbone cleaved at the SalI and BamHI restriction sites, removing the His tag site as it will not be used for protein expression.

The insert for this plasmid required a total of four primers. They were referred to as SalI-fwd, BamHI-rev, single strand homolog forward and single strand homolog reverse (refer to Chapter 2.3).

Designing the Primers

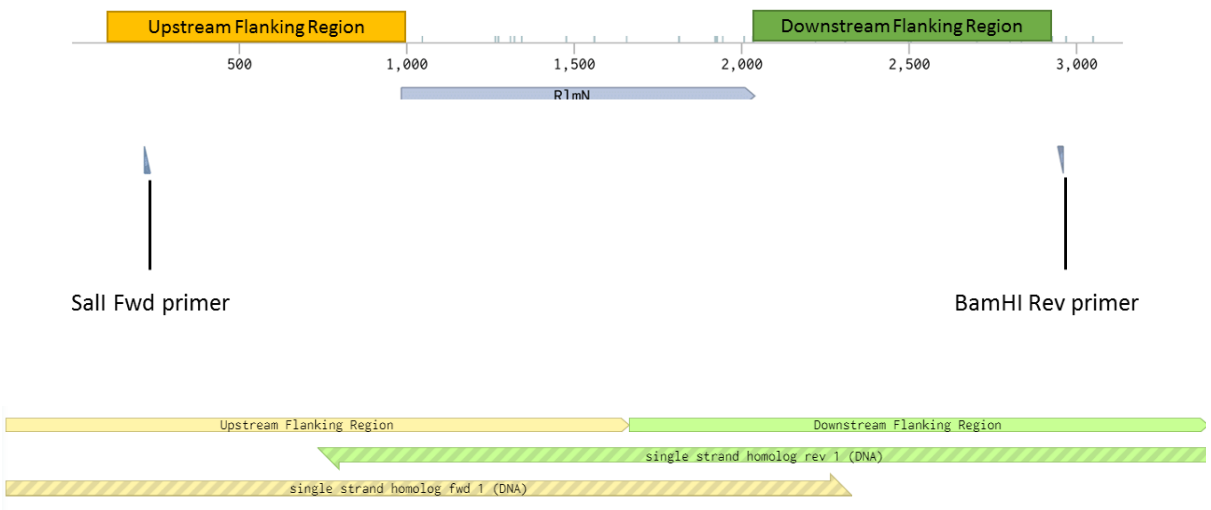


Figure 3-2: The template used for designing the single strand homolog primers and single strand homolog primers. The upstream flanking region is depicted as yellow and the downstream flanking region is depicted as green. The SalI Fwd and BamHI Rev primers are upstream and downstream of these regions respectively. The fusion of the upstream and downstream regions is illustrated in the lower portion of the diagram. Additionally, the single strand homolog rev primer (green striped) and the single strand homolog fwd primer (yellow striped) contain sequences from both regions.

The site for the SalI fwd primer was 361 base pairs upstream of the *rlmN* gene and the SalI restriction site was added to the 5' end of the primer. Additionally, the site for the BamHI rev primer was 791 bp down from the RlmN gene and BamHI restriction site was added to the 3' end of the primer each of which were 20 bp long.

The creation of the single strand homolog primers was more complex as they contained bases from the upstream and downstream region of the *rlmN* gene as seen in figure 3-2. The single strand homolog fwd primer was selected to have 19 bases, 14 from the

upstream region and the rest from the downstream region. The single strand homolog rev primer was a total of 20 bases long, 13 from the downstream region and 7 from the upstream region as seen in figure 3-2.

Amplifying the pRecovery Insert

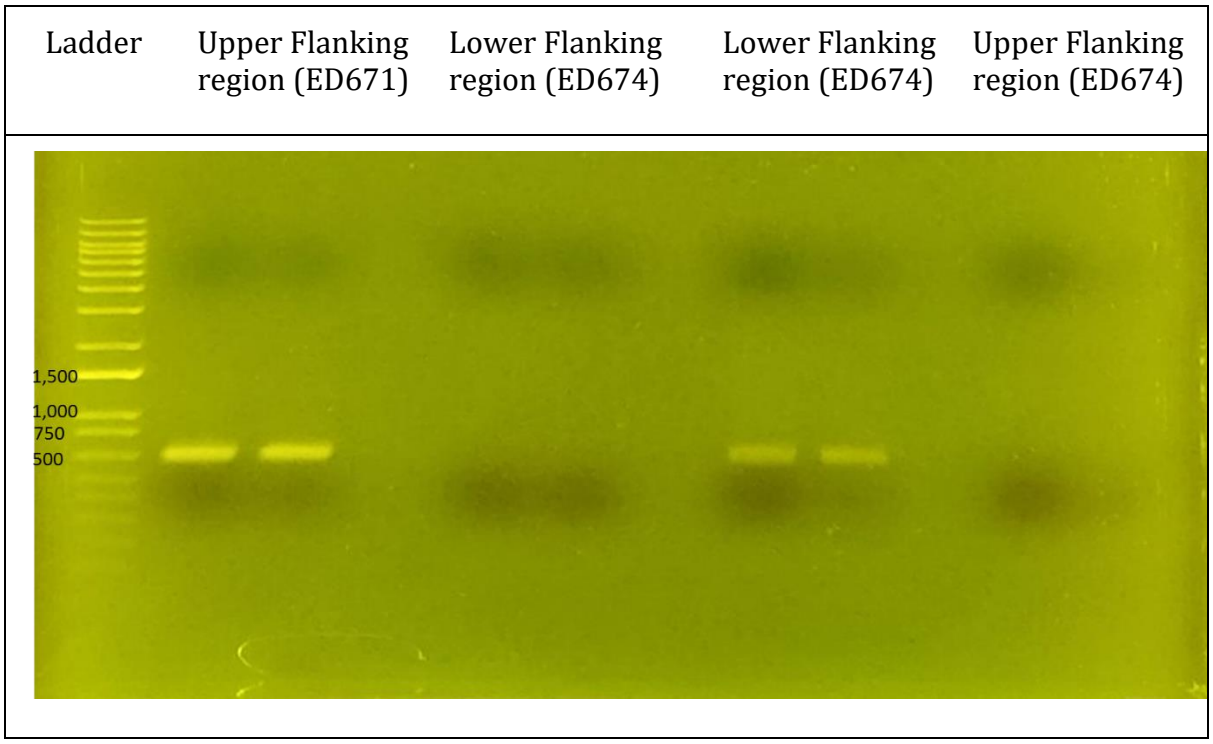
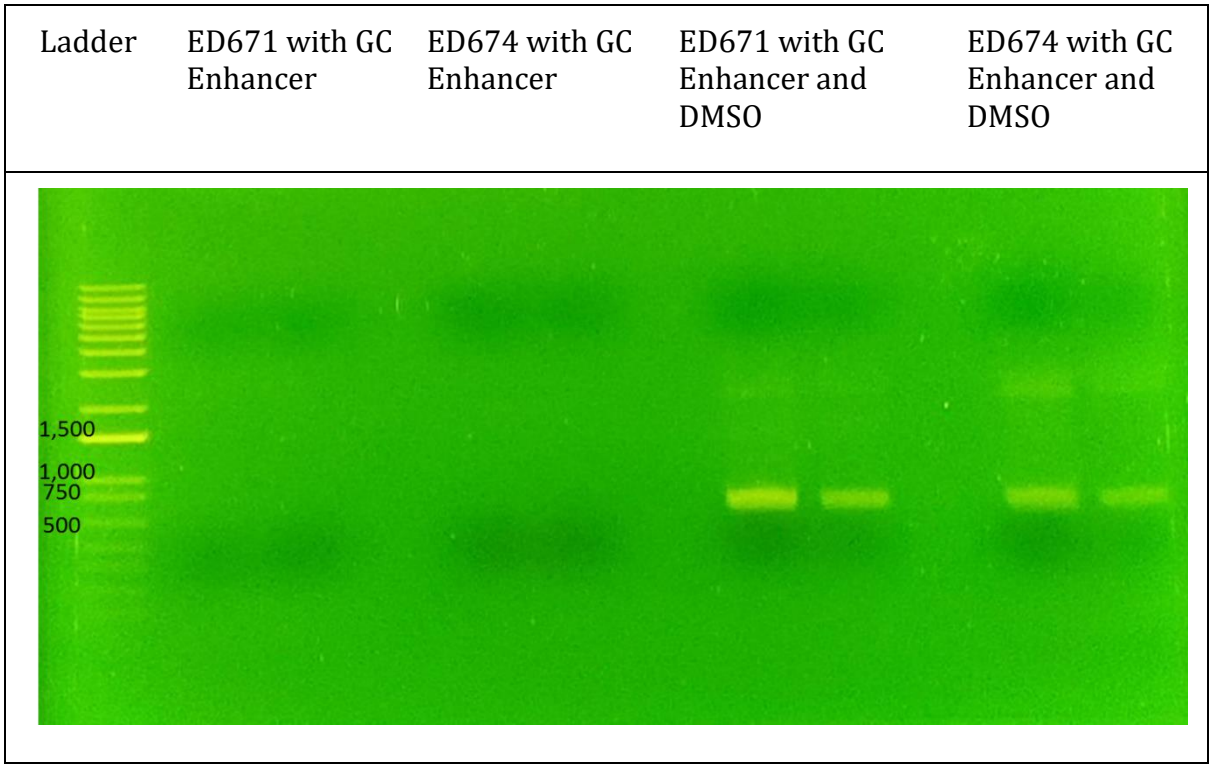


Figure 3-3: Image of the agarose gel containing PCR amplification fragments of the flanking regions in ED671 and ED674.



Figiure 3-4: Image of the PCR results for the lower flanking regions with DMSO.

For the PCR of the 5' flanking region, both amplifications from ED671 and ED674 resolved on the agarose gel at about 500 bp (figure 3-3), this is comparable the expected size of 623 bp as seen in figure 3-2. However, the 3' flanking regions required dimethyl sulfoxide (DMSO) and GC enhancer amplification as seen in the figure 3-3, where strong bands were observed around 700 bp, which matches the expected size of 751bp.

These bands were extracted and spliced together via overlap extension, this process involved amplifying the upstream and downstream regions separately as seen in figures 3-3 and 3-4. These fragments were then used as templates in an additional amplification process with the SalI Fwd and BamHI Rev primers. The homolog primers have regions which are complimentary to each other, therefore the resulting product is an insert with a SalI and BamHI restriction site on each end respectively. The resulting product was expected to be 1298 bp long, consisting of the upstream and downstream flanking

regions of *rlmN* as seen in figure 3-5. The agarose gel bands from the ED671 and ED674 strain were comparable to the expected size (figure 3-6). The excised bands and pET14b vector were digested with SalI and BamHI and ligated together to form the pRecovery plasmid (figure 3-8). This was confirmed via restriction digest (figure 3-7).

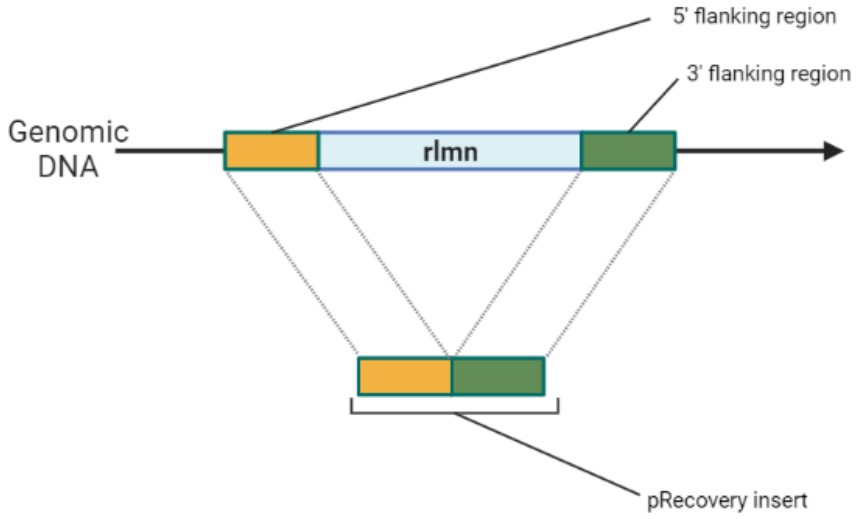


Figure 3-5: A depiction of how the pRecovery insert was designed.

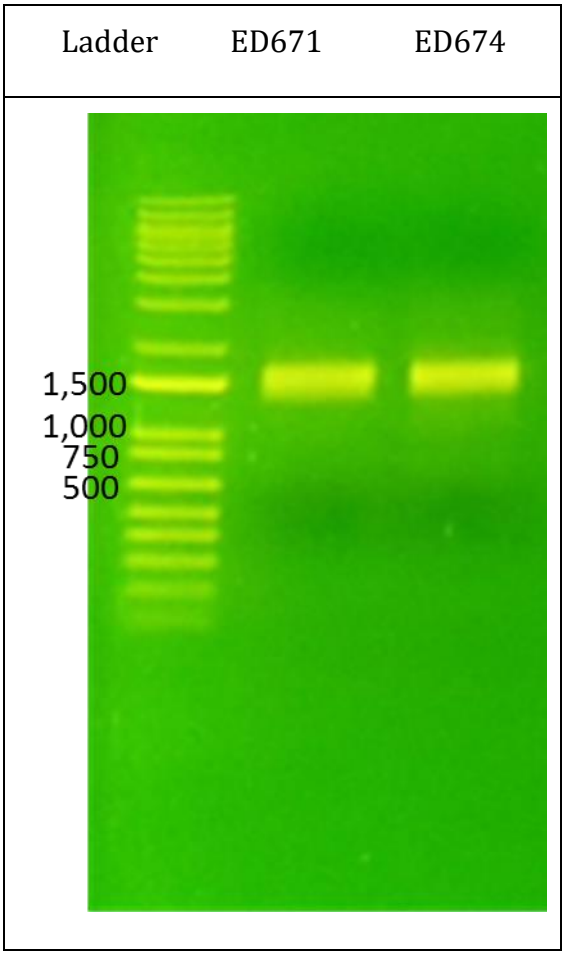


Figure 3-6: Image of the PCR results for the pRecovery inserts

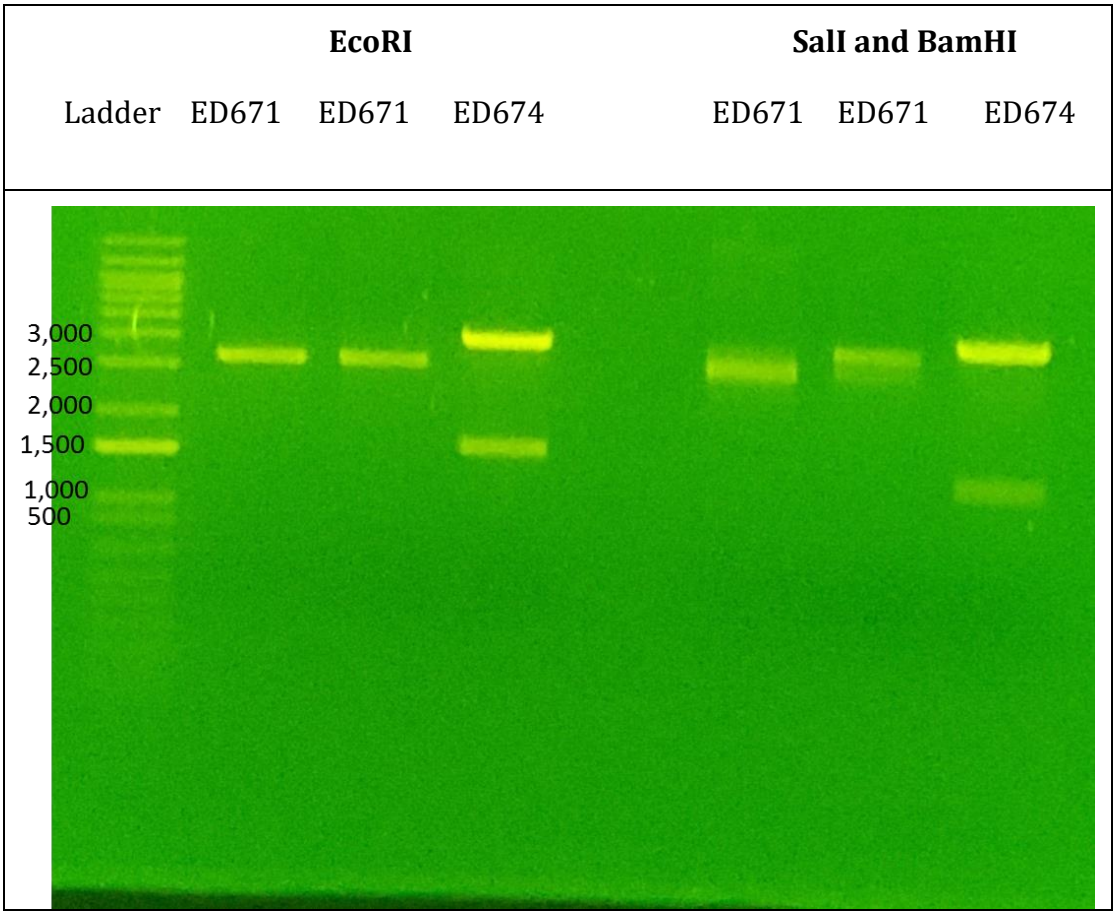


Figure 3-7: Image of the agarose gel results for the pRecovery double digests

In figure 3-7, ED671 samples showed only one band in both clones, suggesting that plasmid was not correct. However, the double digest of pET14b with the ED674 insert displayed 2 bands at the expected sizes with both digests, indicating a successful insertion, this was put forward as the pRecovery plasmid in this project.

rlmN deletion recovery DNA plasmid (5507 bp)

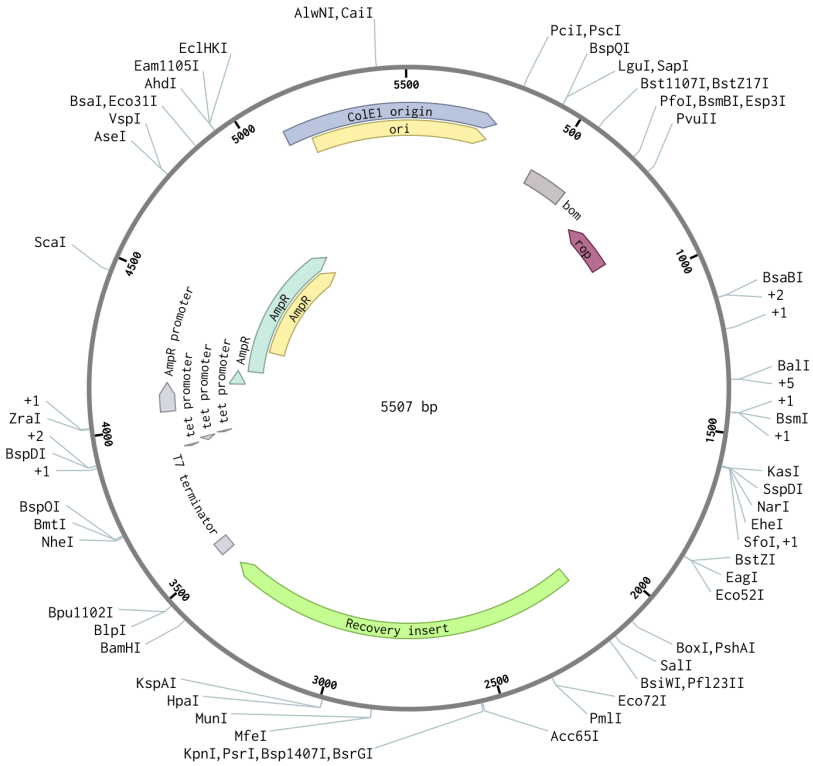


Figure 3-8: A 2D depiction of the pRecovery plasmid.

Cloning the pTarget-*rlmN* plasmid (pTarget N20)

The N20 sequence is meant to be transcribed into the target portion of the guide RNA as seen in figure 1-3, this contains the region that is complimentary to the NGG cut site in the genomic DNA. The location for the NGG sequence for the cut site of Cas9 was selected using Chopchop website, an online bioinformatics tool to select digest sites for various organisms. The target sequence ACGTTGTTCAGGTTGAGCAGCGG was selected in the NC_000913.3:2643721 on the + strand (figure 3-9). [17] This cut site is located near the centre of the *rlmN* gene and was used to design the N20 primers. The primers were designed so when hybridised together, they would create a blunt end and an overhanging end compatible with SpeI.

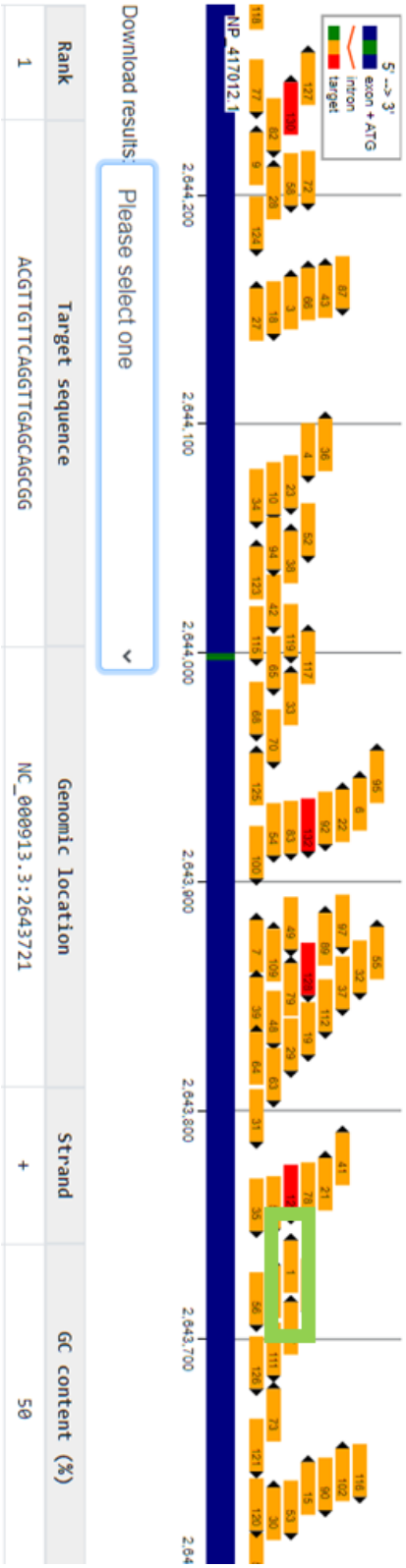


Figure 3-9: An image depicting the site selection of the NGG sequence (in green) on the CHOPCHOP genome editing website. The dark blue region is in and around the RlmN gene.

pTargetAfe contained a pTarget vector (AddGene #62226), spectinomycin resistance, modified to include AfeI site by Dr. E. Deery, (figure 3-10) as the backbone. It was digested with SpeI and AfeI and the reaction was run on an agarose gel. The large band was excised and extracted. This was ligated together with the hybridised N20 primers to form the pTargetN20 plasmid.

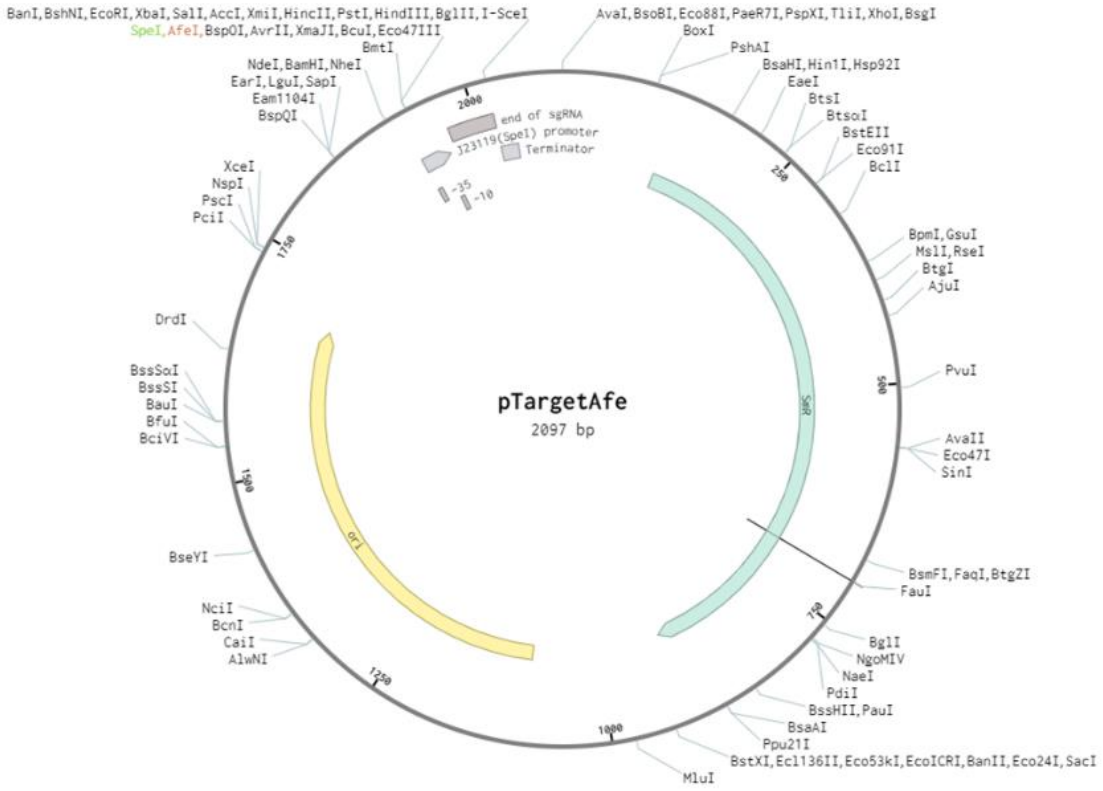


Figure 3-10: A depiction of the pAfe plasmid with the SpeI restriction site in green and the AfeI site in red.

To ensure the successful insertion of the N20 fragment in this vector, a digest with AfeI and SpeI was performed. A successful outcome would result in the loss of the Afe site. The resulting gel shows that the pTargetAfe with no insert is smaller than the newly formed pTargetN20 plasmid confirming the successful ligation of the fragments (figure

3-11). The sequence was also confirmed by DNA sequencing (Genewiz) as seen in figure 3-12).

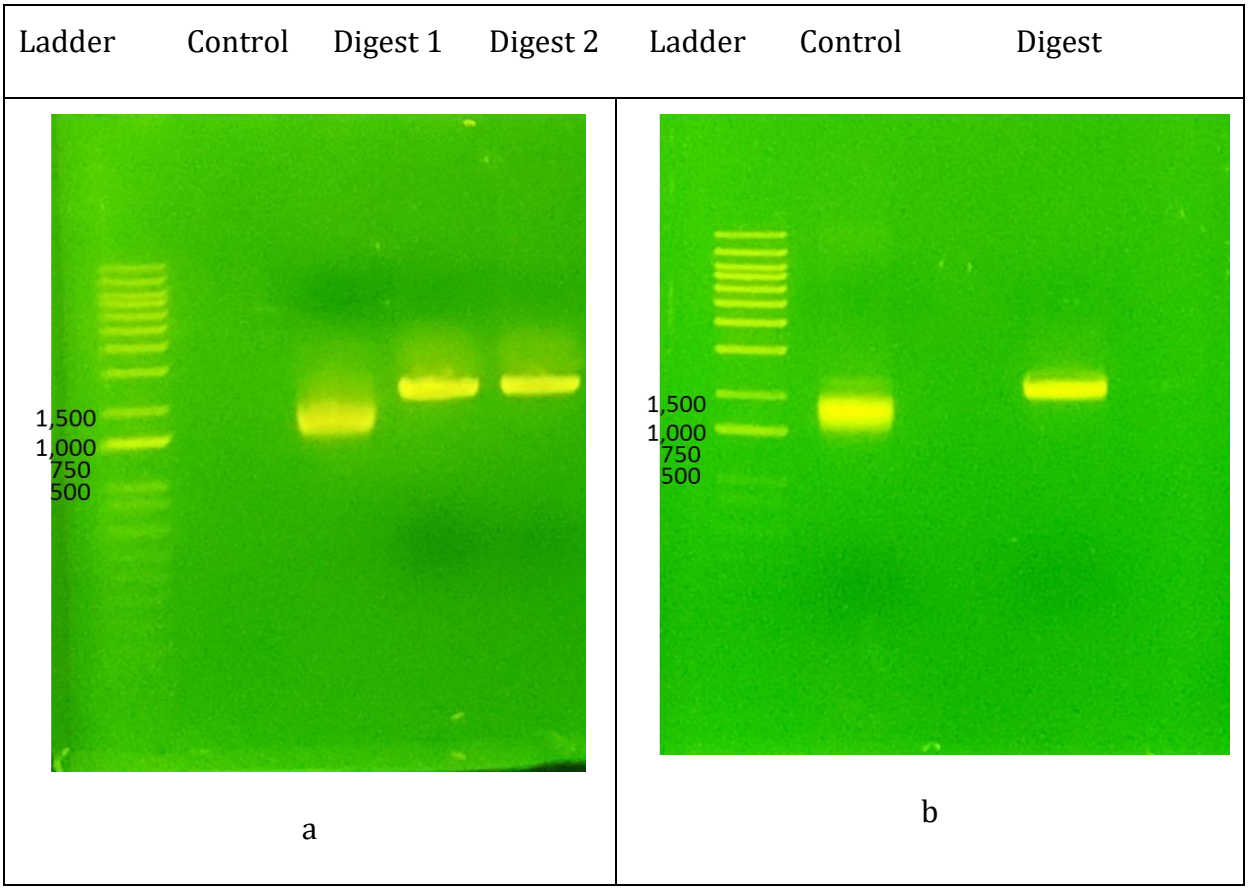


Figure 3-11: images of the agarose gels of a) pTargetN20 and b) pTargetAfe when digested with SpeI and Afe

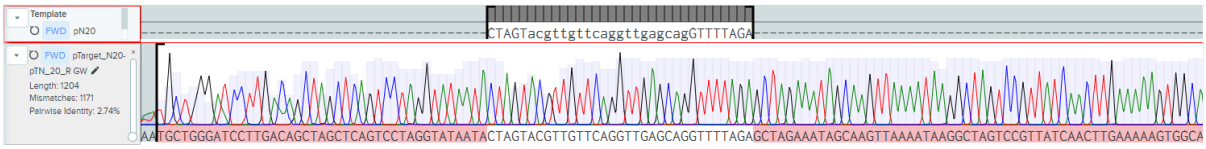


Figure 3-12: an image of the pairwise sequence alignment of the selected pN20 fragments (above) and sanger sequencing results of pTargetN20 (below)

3.1.2 CRISPR Results

The CRISPR process can be described in the following diagram (figure 3-13) with section *a* showing the starting product and *e* demonstrating the finished product. The aim of this process was to delete the *rlmN* gene leaving the upper and lower flanking regions intact.

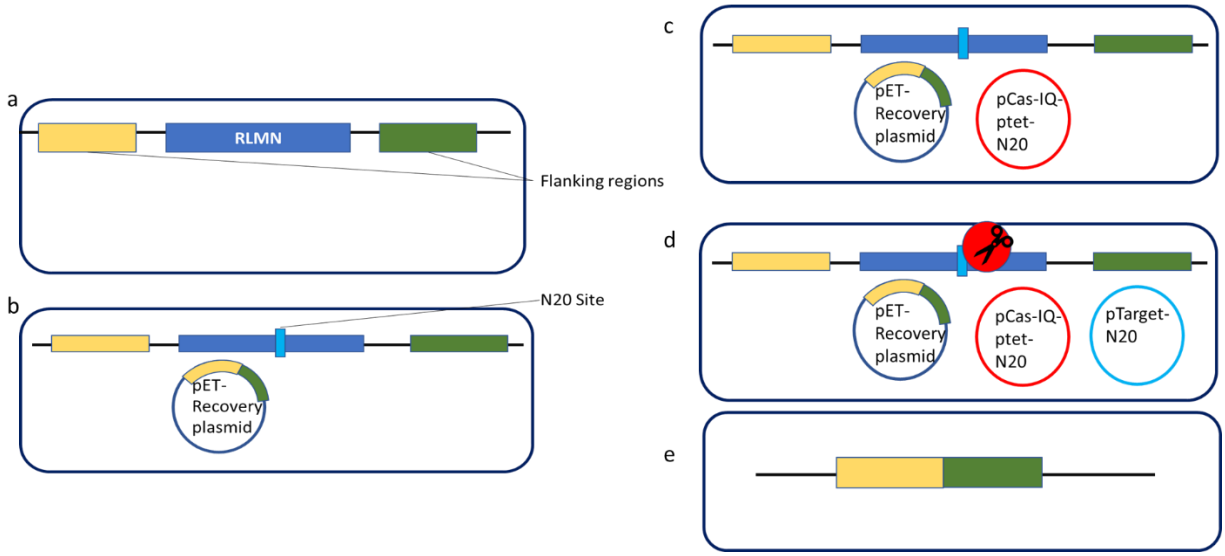


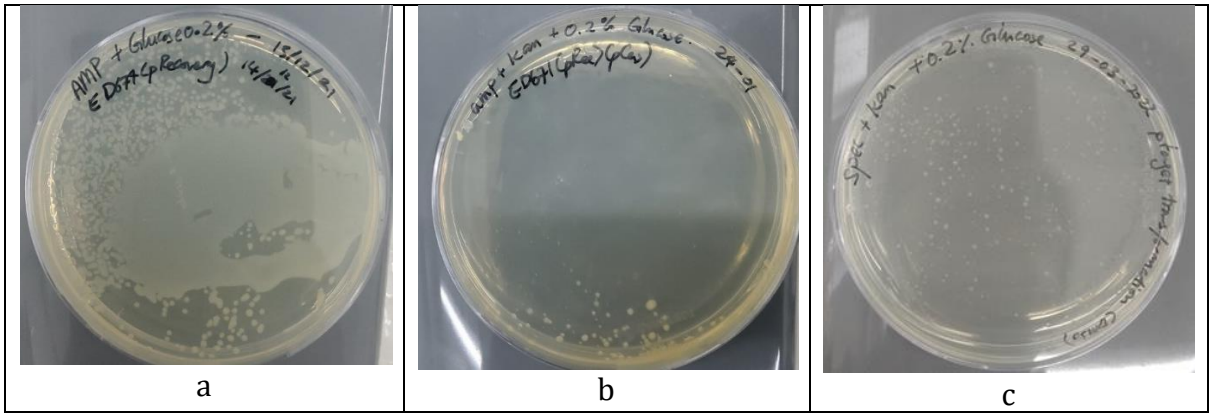
Figure 3-13: The CRISPR technique used in this project, the colour of the flanking regions referenced in figure 3-5 and the red circle in d references figure 1-3. This illustrates the process of deleting the rlmN gene (blue) and fusing the upstream and down stream flanking regions (yellow and green respectively).

Transforming Plasmids and Performing the Deletion of *rlmN using CRISPR*

The initial step of the process was transforming the pRecovery plasmid into chemically competent ED671 and ED674 cells. Transformants were selected for using the ampicillin marker, resulting in the strain ED67x (pRecovery) (figure 3-14 a). Single colonies were picked and made chemically competent and pCas-deltaIQ-Ptet-N20

(pCas) was introduced. Successful transformants were selected by ampicillin and kanamycin resistance, resulting in the ED674 (pRecovery) (pCas) strain (figure 3-14 b). However, this respective resistance was not demonstrated in the subsequent starter cultures made with ED671 (pRecovery) (pCas) suggesting a false positive result. At this step, ED671(pRecovery) candidates were deemed unsuccessful in pCas introduction, and the gene deletion process was halted here.

The most challenging step was making the strain ultracompetent to introduce the pN20 plasmid. Additionally, the strain CRISPR machinery must be expressed in the cell with the addition of L-arabinose beforehand. The pTargetN20 plasmid introduces guide RNA which is homologous to the NGG sequence which facilitates the Cas9 protein to a double cut in the genome and the pTarget plasmid in the pMB1 replicon. The genome is then repaired by the recombinase using the donor DNA from the pRecovery plasmid and the pTargetN20 plasmid should be lost with subsequent generations.



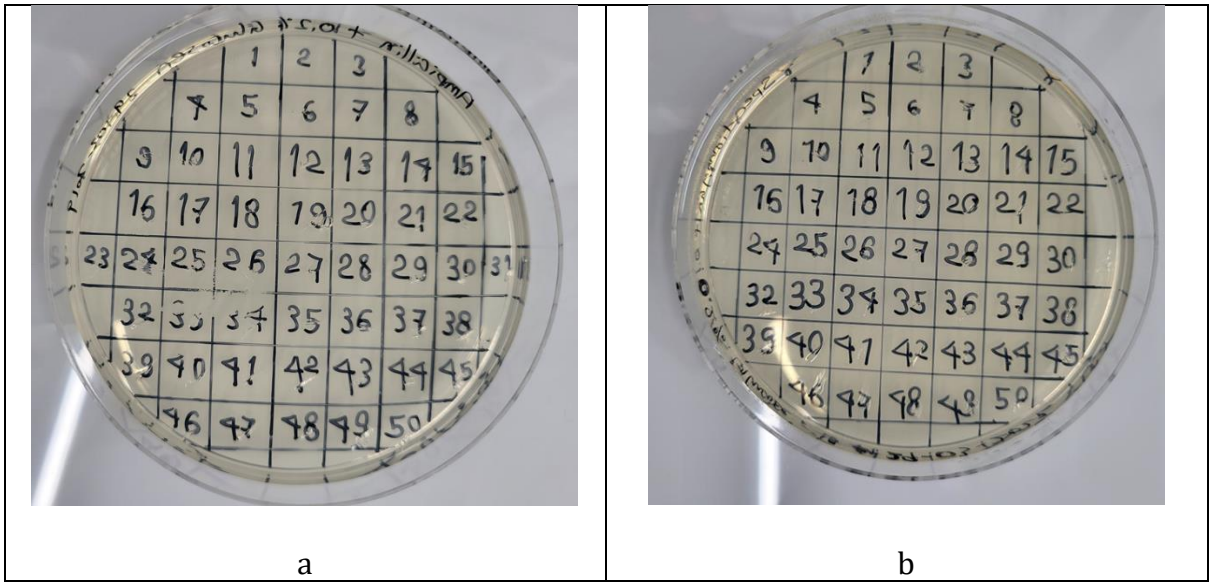
a b c

Figure 3-14: (a) ED674 with pRecovery growing on an ampicillin plate, (b) ED671 with pRecovery and pCas on an LB plate with ampicillin and kanamycin, (c) ED674 containing pCas and pTarget after pCas induction with L-arabinose on an LB plate with spectinomycin and kanamycin.

Plasmid Curing and Checking for *rlmN* Deletion

Once the deletion step is complete, the plasmids are removed. The cells are grown in liquid culture with LB and tetracycline overnight at 28 ˚C and streaked on kanamycin plates as mentioned in Chapter2. The curing process of introducing tetracycline to the culture is to induce the expression of an endonuclease which cures the pTargetN20 plasmid, the pCas plasmid is temperature sensitive, thus incubating the sample at 28 ˚C will remove this plasmid.

To confirm the loss of the pRecovery and pTargetN20 plasmids, a colony screening was performed for 50 of the resulting colonies on one of each of the following LB-agar plates with working concentration of the following antibiotics, ampicillin, spectinomycin and kanamycin. Figure 3-16 shows the colony selection plates with ampicillin and spectinomycin, no colonies grew despite an overnight incubation at 28˚C. Therefore, colonies which were both ampicillin and spectinomycin sensitive were selected from the kanamycin plate as seen in figures 3-15 and 3-16 for colony PCR using the RlmN_del_seq_Fwd and RlmN_del_seq_R primers. As the samples selected were sensitive to ampicillin and spectinomycin, they were considered free of the pRecovery and pTargetN20 plasmids.



a b

Figure 3-15: The colony selection plates a) ampicillin and b) spectinomycin to screen for plasmid curing of pRecovery and pTargetN20 respectively.

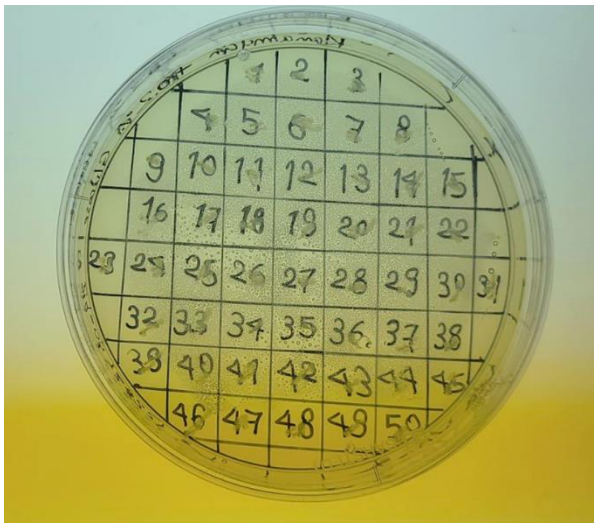


Figure 3-16: The colony selection plate with kanamycin to screen for pCas.

The pCas plasmid is a temperature sensitive plasmid, therefore the curing process was growing it at 42˚C for 6 hours in liquid LB followed by an overnight incubation at 37˚C on LB agar with glucose. The resulting colonies were used for cobamide production experiments.

To screen for the deletion of the *rlmN* gene, the Checking_Deletion_Fwd and Checking_Deletion_Rev primers were used, these contain sequences from the upstream and downstream regions of the entire CRISPR experiment, therefore, they will anneal with the template and produce a result that varies in size depending on the success of the deletion. The resulting PCR is visualised on a gel (figure 3-18). Two of the three colonies appeared to be correct with a band around 1,500 bp. This is congruent with the 1,555 bp loss of the *rlmN* gene from a selected area of about 3,000 bp (see figure 3-1). Colony 1 had an extra band around 3kb so tis strain was discarded. The PCR product was excised from the gel and sent for sequencing (GeneWiz) to confirm the deletion.

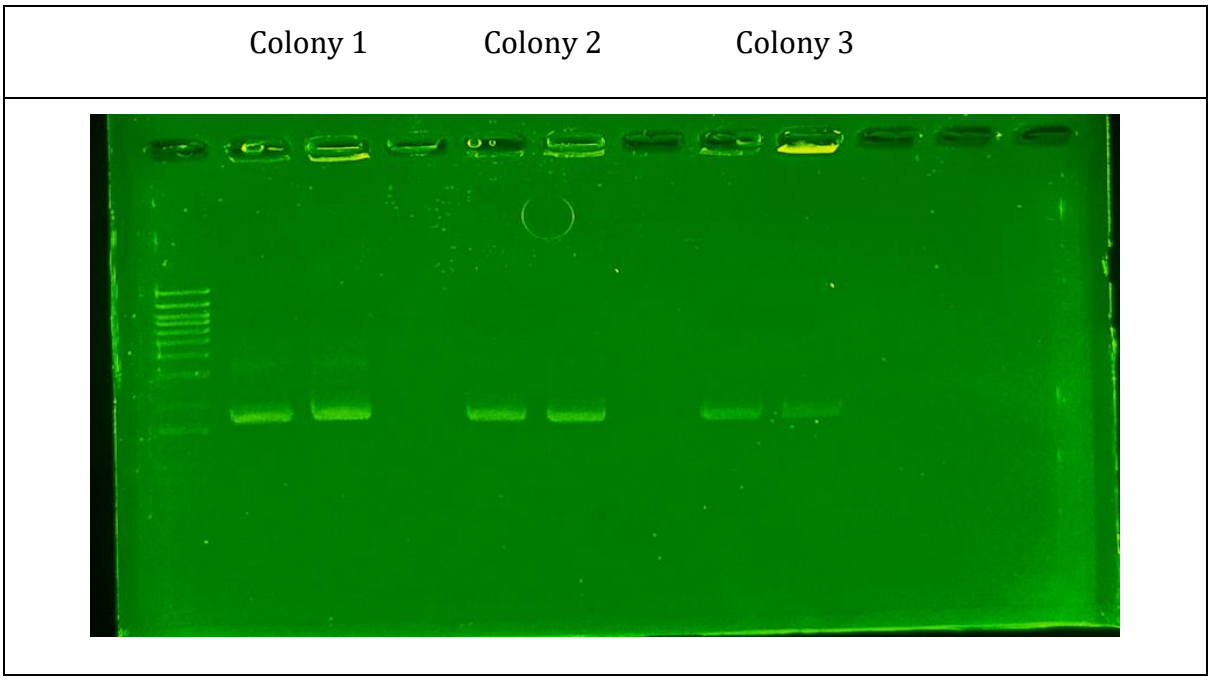


Figure 3-17: The agarose gel of the PCR to check for the deletion of rlmN.

3.2 Cobamide Producing Strains

The two adenosyl cobamide producing strains investigated in this project are ED674 and ED671. These are modified *E. coli* strains (gifted by Dr. E. Deery). ED674 is a BGC043 strain of *E. coli* modified with the genes from the biosynthetic pathway of

adenosyl cobalamin synthesis except for the synthesis of the lower ligand, as seen in Chapter 2.2. Therefore, in the absence of any lower base, GDP cobinamide would be the expected product. The cells contain some free adenine, thus adenine or 2-methyladenine would be the expected lower bases of the cobamides produced this strain. With the addition of 0.1 g/L adenine, pseudocobalamin ([Ade] Cba) would be the main product expected.

Comparably, ED671 also has identical components for the biosynthetic pathway added with the notable exception of an additional Se Amn-PTH0960 integrated to the 5' of the CobUST operon. The addition consists of an AMP nucleosidase which forms free adenine from AMP.[18] Therefore the expected product this strain would be more likely a [Ade] Cba or [MeAde] Cba.

Contrastingly, it is expected that the ED674 (RlmN deletion) strain will not produce 2-methyladenine containing cobalamin. However, GDP cobinamide and adenylyl cobamides are expected to appear with the latter especially in adenine supplemented cultures.

The main predictions in this project focus on the lower ligand because the adenosyl group on the beta axial ligand is bound to the main corrin ring via a light sensitive bond, prolonged exposure to light results in the conversion to a hydroxy group.

In addition, the gene of interest, *rlmN* does not appear to be essential to *E. coli*, despite its close working relationship with the ribosome and production of 2-methyladenine. However, its absence is correlated to a lower growth rate compared to wild type strains in *E. coli.* [19]

3.2.1 The Growth Rate of Cobamide Producing Cultures

The aim of this experiment was to compare the growth of the ED674*ΔrlmN* strain to ED671 and ED674. The growth process mirrored that of the larger volumes for cobamide production as described later in Chapter 2.2.

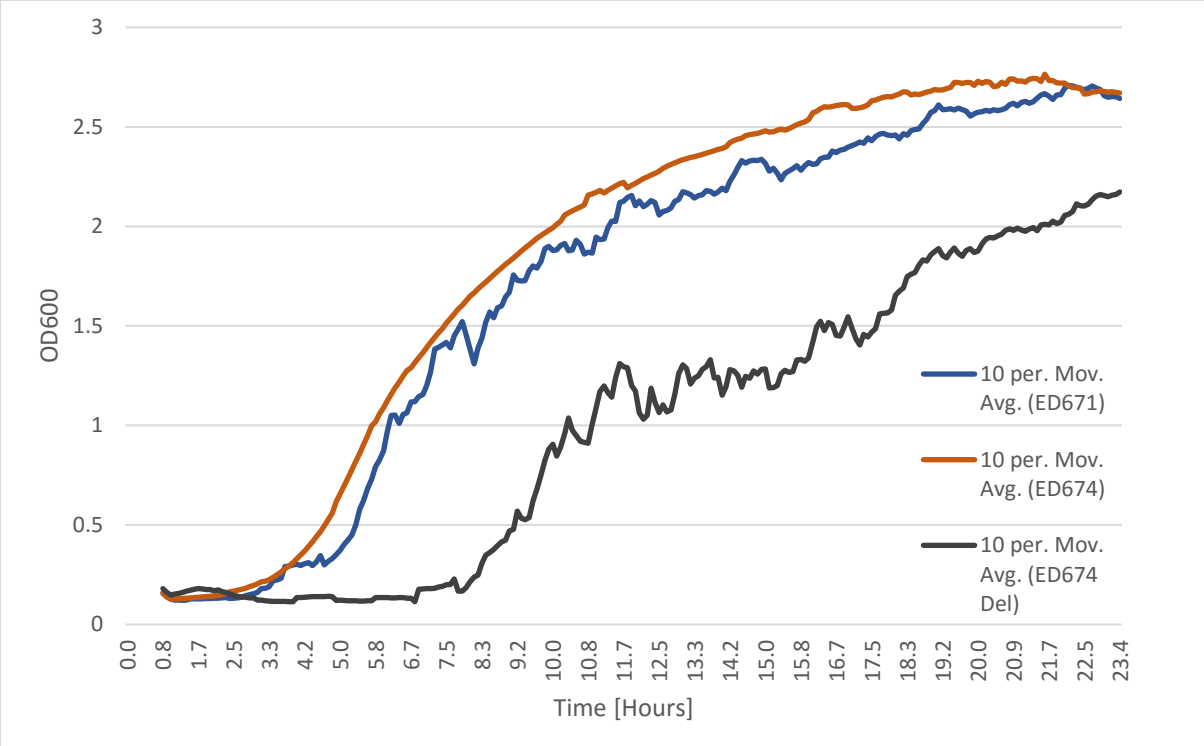


Figure 3-18: A growth curve comparing the growth rates between ED671, ED674 and ED674ΔrlmN

The growth curve of the three strains of interest were recorded over a 23-hour period (figure 3-18). ED671 and ED674 appear to have grown at similar rates and the ED674*ΔrlmN* strain noticeably slower with a long lag phase and an exponential phase lower than its unmodified counterparts. This supports the idea that while the *rlmN* deletion may not be fatal to the culture, it has a negative impact on its growth rate, therefore the deletion strain is most likely less competitive compared to ED674 as it grows at a reduced rate compared to the wild type.

The final OD for each of the strains is highest for ED674 and lowest for ED674*ΔrlmN* this could be due to the deletion. As the exponential phase for ED674*ΔrlmN* is not complete yet, the final reading may take longer. The function RlmN is related to protein synthesis, especially the modification of adenine residue on the small ribosomal subunit, a function that facilitates the reading of a stop codon.[13]

Table 3-1: the OD_{600} readings of the strains of interest after 24 hours.

	ED671	**ED674**	**ED674ΔrlmN**
Final OD_{600} (average of 3 biological repeats)	2.520	2.557	2.204

3.2.3 Growth of Cobamide Producing Cultures

The aim of this experiment was to assess the effect of the *rlmN* deletion and the addition of the AMP nucleosidase in the cobamide producing *E. coli* strains on the variety of cobamides produced. This consists of growing the cultures from glycerol stocks into starter cultures and inoculating supplemented minimal media, followed by the extraction and purification of cobamides.

The three cobamide producing cultures were grown in supplemented minimal media to maximize cobalamin production. The strains all contained the necessary genes to synthesize GDP cobinamide and CobUST operon which attaches the lower ligand to the corrin ring, but these strains do not possess the genes needed for DMB production.

Therefore, the most likely candidates for attachment are mostly free intracellular adenine and 2-methyladenine.

ED671 and ED674 contain most of the genes required to produce cobamides, they are all downstream of a T7 lac operon which requires IPTG to induce. Therefore, growth process consisted of three main steps, inoculation of the starter culture, inoculation of the minimal media and induction of the B12 producing genes over the course of three days as described in Chapter 2. This resulted in the collection of coloured supernatants in preparation for cobamide purification, especially after lysing cells through boiling.

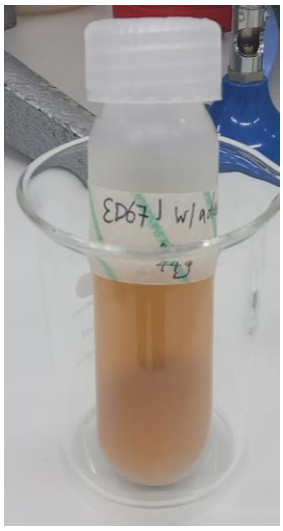


Figure 3-19: an image of a centrifuge tube containing the lyse cells ED671 culture after centrifugation at 18,000 RPM, the pink colour suggests the presence of corrinoids.

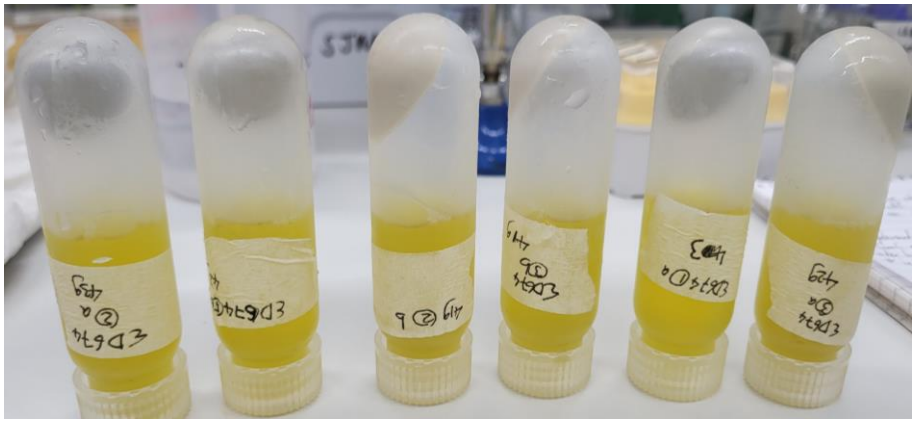


Figure 3-20: an image of centrifuge tubes containing culture ED674ΔrlmN after centrifugation at 18,000 RPM.

In this growth and extraction process, ED671 and ED674 produced pink supernatant (seen in figure 3-21), which is indicative of B12 production, this was kept for the purification step. However, the supernatant of the deletion strain was bright yellow as pictured in figure 3-20, it was also kept for purification.

3.2.3 Purification of Cobalamin

The aim of this experiment was to isolate and purify the cobamides present in the clarified cell extract step, this was done by immobilizing His_6-BtuF, an N-terminal his-tagged cobamide binding protein, onto a chelated Sepharose column charged with $NiSO_4$. The cobamides are bound to the column via the BtuF protein, were eluted by denaturing the BtuF protein using 6 M guanidinium hydrochloride. The pink colour in the columns and elution fractions indicated the presence of cobamides (figure 3-21). The cobamides were further purified and desalted over RP18 resin as described in Chapter 2.5, the resulting eluent contained 50% methanol which was subsequently removed by vacuum centrifugation.

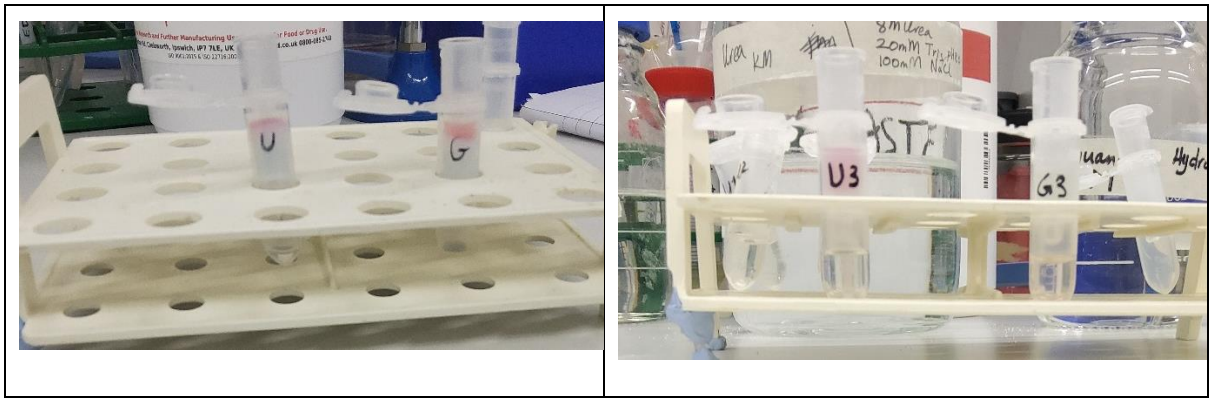


Figure 3-21: The depiction of chelating Sepharose columns with cobamides (pink), bound to the immobilised BtuF protein, they were then eluted, and the column was blue with pink flow through (right).

3.3 Analysis of Compounds Produced

The aim of this experiment was to ascertain the compounds produced by the bacterial strains of interest. The extracts of the strains ED671 and ED674 were analysed by HPLC-MS, to separate and identify the cobamides present. Unfortunately, the ED674*ΔrlmN* strain were yellow in colour (figure 3-19) and did not produce cell extract with compounds that were chelated by the BtuF column, suggesting that this strain did not produce a corrinoid that could be analysed within the scope of this project. The relative concentrations of the cobamides were estimated by extracted ion chromatograms (EIC), the area under the curve of mass spectrometry results were used to obtain this data with the manual integration function as seen in figure 3-4. While not quantitative, it does allow for the comparison of the relative abundance of the cobamides produced by the strains.

Expected Masses for the Compounds Produced

The three expected corrinoids are GDP Cbi, [Ade] Cba and [2MeAde] Cba. As the expected upper ligand is adenosine and the samples were not protected from light, both

adenosine and hydroxy upper ligands are expected. The anticipated m/z and molecular masses for these compounds can be found in the table below.

Table 3-2: the mass to charge ratio and molecular mass of the expected products in this project.

	[M+H]2+		**Molecular Mass**	
	adenosyl	hydroxy	adenosyl	hydroxy
GDP Cbi	720.76	707.75	1665.53	1414.5
[Ade] Cba	784	659.77	1567.63	1318.54
[2-MeAde] Cba	791	666.78	1581.64	1332.55

Manual Selection of Area Under Extracted Ion Chromatogram (EIC) curve

EIC data for each m/z value corresponding to the compounds of interest of each repeat was plotted, using the provided software from Bruker. The peaks were integrated and the value for the peak area was used to inform on the relative concentration of each compound.

Figure 3-22 shows an example from ED671 with the m/z of 784.83, suggesting it is [Ade] Cba. This was further confirmed with automated MS-MS, a process which allowed us to obtain the m/z of the lower ligand which will be discussed in the following section.

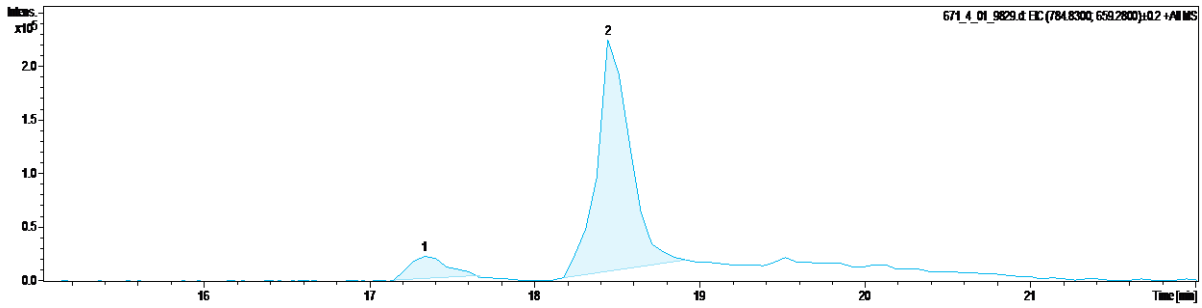


Figure 3-22: EIC for m/z 784.83 with manual integration of the peaks from a sample purified from ED671 without adenine supplementation. The blue area is the area included in the integration.

MS-MS Analysis of the Compounds Produced

The main compounds found were analysed by automatic MS-MS to confirm the masses of their constituents as they would occur through fragmentation. As the corrin ring is relatively stable, it was expected that most of the fragmentation would likely consist of the nucleotide loop and the lower base as described in the diagram in figure 1-1. Discernment as this level is required as the upper ligands appeared in varying forms in this project, giving varying masses for the same lower ligand in the initial LC-MS as seen in Section 3.3.

Table 3-3: the m/z r value of the expected lower base fragments of interest following MS-MS fragmentation.

Lower Ligand and Ribazole	m/z value ([M+H]2+)
GDP Cbi	-
[Ade]	348.07
[2-MeAde]	362.07

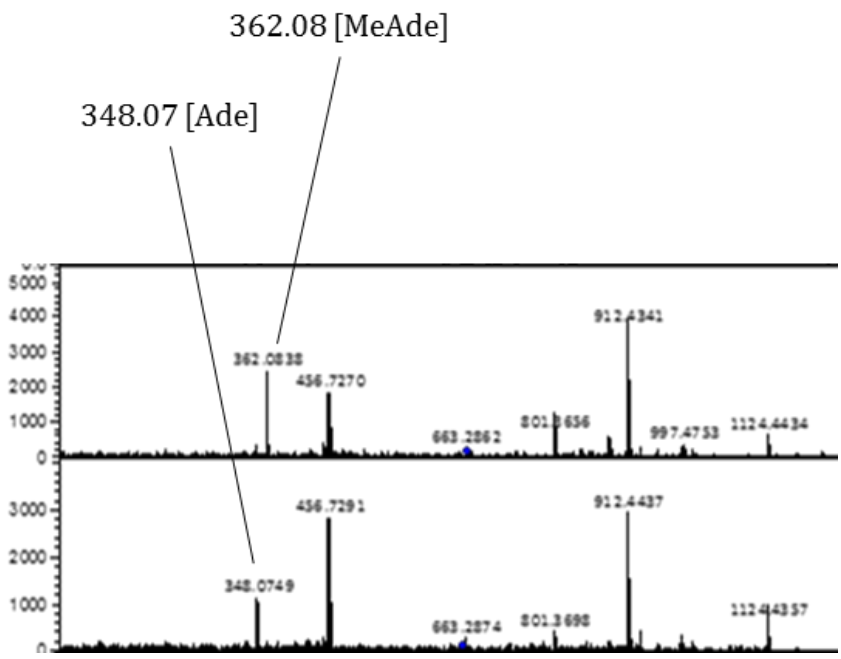


Figure 3-23: MSMS fragmentation revealing the varying lower ligands in ED671. The masses of the varying lower ligands are referred to in the image.

Compounds Produced by ED674

This strain was grown under adenine supplemented conditions for two repeats and without for three repeats, the EIC area for each cobamide variation was categorised by sample in figure 3-24. The main products from this strain were GDP cobinimide, [2MeAde] cobamide and [Ade] cobamide from both conditions, which all had a hydroxy upper ligand, this could be due to prolonged exposure to light.

A higher concentration of all expected cobamides was produced when the culture was not supplemented with adenine, with [2MeAde] most abundant. However, when supplemented with adenine, the concentrations of [Ade] Cba and GDP Cbi were higher than [2MeAde] Cba.

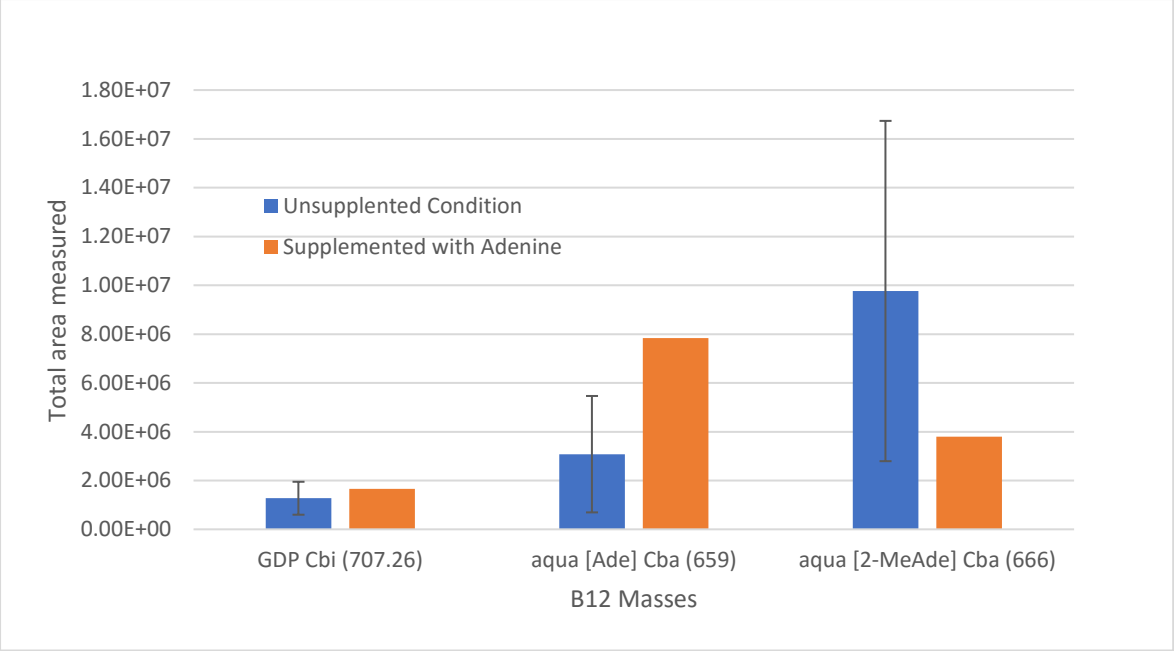


Figure 3-24: a graph comparing integrated peak areas for the cobamide varieties produced by ED674.The non-supplemented conditions contain the average of three repeats and the supplemented conditions contain the average of two repeats.

Compounds Produced by ED671

This strain was grown in three biological repeats for non supplemented conditions and two biological repeats with adenine supplementation. When supplemented with adenine, the total cobamide levels increased on average, the most abundant type was [Ade]Cba, which occurred at an over 40% higher concentration than when grown without adenine, with a large variance between the two adenine supplemented biological repeats (figure 3-25).

Interestingly, the concentration of [MeAde] Cba is lower than [Ade] Cba in both conditions, with adenine supplemented conditions producing about 40% less [MeAde] Cba compared to non supplemented conditions on average though this number is a small portion of the entire cobamide yield.

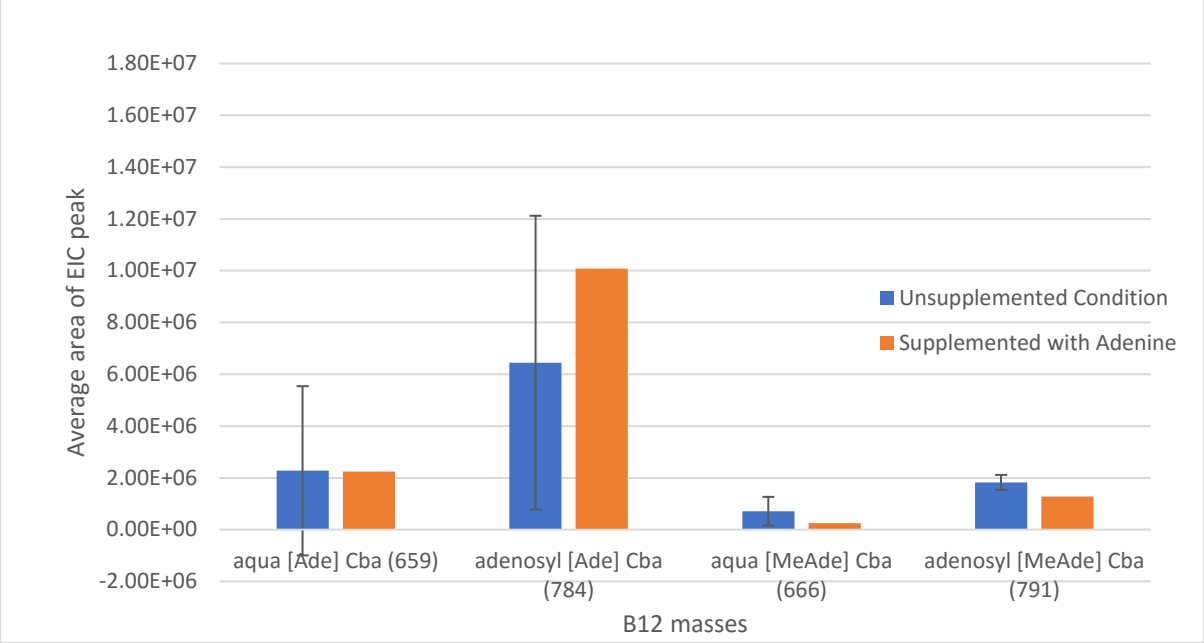


Figure 3-25: a graph comparing the integrated peak areas for the cobamide varieties produced by ED671. The non-supplemented conditions contain the average of three repeats and the supplemented conditions contain the average of two repeats.

Comparison of Cobamide Production

The graph below (figure 3-26) details the average cobamide production compared between both strains of interest under adenine supplemented and not supplemented conditions. Though [Ade]Cba and [MeAde]Cba appeared in all samples, their levels varied within the different strains and conditions. In both samples, the supplementation of adenine appeared to in the amount of total cobamide production in one repeat of ED671. An additional notable exception was GDP Cbi which increased by over 30%. The highest concentration of cobamide production had come form ED674 when grown without adenine followed by ED671 under the same condition and Cba production was higher in ED674 under both conditions.

[Ade]Cba production was highest by ED671 without adenine supplementation and was higher in both ED671 conditions compared to ED674. It was lowest from ED674 without adenine supplementation. Contrastingly, [MeAde]Cba production was highest in ED674 when not supplemented with adenine, about 84% higher than ED671 in the same condition. When comparing both strains, ED674 produced more [MeAde]Cba under both conditions compared to ED671.

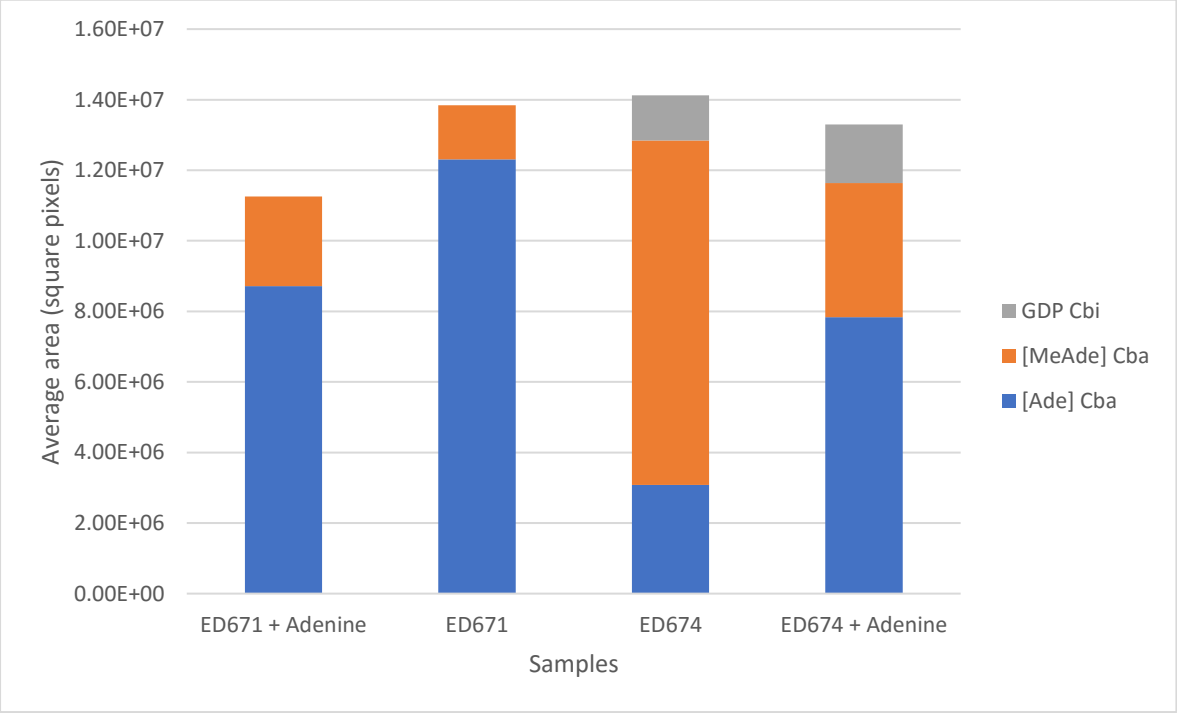


Figure 3-26: a graph comparing the integrated peak areas for the total cobamide varieties produced by both strains. The adenine supplemented (+Adenine) conditions were repeated three times and the ED671, ED674 conditions were repeated twice.

Chapter 4 – Conclusion and Discussion

4.1 Conclusions of the Project

rlmN Gene Deletion

The *rlmN* gene was successfully deleted using CRISPR. Overall, the newly made strain has impaired growth compared to the WT strain and unfortunately, the ability of the strain to produce cobamides was lost in the process, as observed in the bright yellow lysates obtained in Section 3.2.2.

Cobamide Lower Ligand Synthesis

Adenine supplementation appears to impact the variety of lower ligand incorporated to form a cobamide in the absence of DMB. In the absence of a form of supplemented adenine, as an intracellular booster or external addition, 2-methyladenine appears to be the main preferred lower ligand to be incorporated into a cobamide and produces the highest ratio of cobamides as observed in ED674 grown without additional adenine. Contrastingly, when supplemented with adenine, with an intracellular booster or externally, adenine becomes the main lower ligand of the cobamides produced by ratio. Interestingly, the addition of an intracellular adenine booster in the form of Amn results in the mitigation of GDP Cbi production, even though the total concentration of cobamides decreased. It would be interesting to evaluate the effect of the growth rate of the culture on the lower ligands observed as RNA breakdown from tRNA or rRNA. As the breakdown of these compounds is known to be higher at low growth rates and under conditions of stress.

4.2 Discussion

4.2.1 Gene Deletion Experiment

It is important to understand why the ED674*ΔrlmN* strain did not make cobamides which could be detected in this experiment. The Cas9 protein in this project creates double strand breaks (DSB) on the NGG site which can be repaired with non-homologous end joining (NHEJ) or HDR. The likelihood of the former is not as high in the laboratory K-12 strain from which ED674 is modified from. [20] Therefore, it is unlikely NHEJ contributed to the ED674*ΔrlmN* strain's lack of cobamide production. However, it is worth noting that the deletion of rlmN, though non-lethal, can result in UAG readthrough and slower growth. [13,19] It is worth noting that the readings were taken with minimal media and the effects of *rlmN* deletion could be more pronounced. Thus, it could be worth looking into the Cob operon added to the ED674 strain and whether they may have any UAG codons within the reading frame resulting in a malfunctional protein. As there are two other codons, UGA and UAA which also function as stop codons, a point mutation to convert UAG to UAA on respective Cob operon genes could result in cobamide production.

<u>Targeting rlmN and correlating which of its targets contribute to lower ligand formation</u>

RlmN is a bifunctional enzyme which methylates tRNA and rRNA, with two potential starting points for the presence of free 2-methyladenine in the cell, it is crucial to understand which of these contribute to lower ligand formation. It is important to assess if these are broken down in congruence with other events which affect the cell.

rRNA Degradation

RlmN methylates the A2503 residue on the 23S ribosomes, this methylation is correlated to translational accuracy, especially the reading of the UAG stop codon. [13] To estimate the number of potential free nucleotides generated by ribosomal break down, it is important to understand that ribosomes typically have stable RNA with two notable exceptional conditions, quality control and starvation conditions. Under quality control circumstances, rRNA can broken down due to its exposure as the ribosome does not protect it from RNase, additionally, under starvation conditions, rRNA can be broken down into nucleotides. [21] However, as ribosomal synthesis is an energetically costly affair for the cell, this process is highly regulated through quality control and repair [22], therefore the number of free nucleotides generated from ribosomes is highly varied between growths.

tRNA Degradation

Interestingly, RlmN also methylates a variety of tRNA, specifically on the A37 of $tRNA^{Arg}_{ACG}$, $tRNA^{Asp}_{GUG}$, $tRNA^{Gln}_{UUG}$, $tRNA^{Gln}_{CUG}$, $tRNA^{Glu}_{UUC}$, and $tRNA^{His}_{GUG}$, in *E. coli*, a process which also contributes to translational accuracy. [13] Additionally, this enzyme prefers certain residues over others, despite its wide specificity. This process is specific this residue and its neighboring residues, 35 is usually guanine of cytosine and 36 is usually thymine or cytosine. Under starvation conditions, without amino acids readily available, tRNA is also degraded rapidly suggesting that they are tightly regulated component of protein synthesis in the cell. [23]

In addition to their larger number compared to ribosomes, tRNA with different specificities occur and varying frequencies within the cell, therefore it is important to

use their amounts to estimate the potential amount of methylated adenosine available for the degradation process into nucleotides. The total number of tRNA molecules in question add up to over 14,000 on average in each cell as seen in the table below.[24] However, it is worth noting that these numbers can fluctuate with the state of the cell, they can be higher during the exponential growth phase of a culture and lower when free amino acids are scarce. [23]

Table 4-1: A table stating the numbers of varying tRNA in E. coli, sourced from Dong et. al. [24]

tRNA	Number of molecules per cell
$\mathbf{tRNA^{Arg}_{ACG}}$	4752
$\mathbf{tRNA^{Asp}_{GUG}}$	2396
$\mathbf{tRNA^{Gln}_{UUG}}$	764
$\mathbf{tRNA^{Gln}_{CUG}}$	881
$\mathbf{tRNA^{Glu}_{UUC}}$	4717
$\mathbf{tRNA^{His}_{GUG}}$	639

Nucleotide Degradation and Freeing Adenine

The orientation of 2-methyladenine within the α-ribazole differs from its orientation within a methylated nucleotide as seen in figure 4-1. Thus, the degradation of the nucleotide to the release a free methylated purine is necessary for its incorporation into the lower ligand, the structural differences are noted in the figure below. However, the process of nucleotide breakdown into purines is not currently known.

Figure 4-1: 2-methyladenine incorporated into alpha ribazole (left) and an adenine nucleotide (right)

4.2.2 Cobamide Producing Culture Experiments

Adenine Supplementation

Free adenine supplementation was selected because *E. coli* can import external sources of adenine using the *purP*, a high affinity adenine permease. This facilitates adenine import without using energy from the cell. [25]

Adenine (0.1 g/L) was added to the cultures when IPTG was added, at this point the OD_{600} of the culture was mostly above 1. Additionally, the solubility of adenine in water is 0.1 g/L, this could limit the effect of supplementation on cobamide producing cultures. This variance can be seen in how the average ED671 culture produces a lower variety of cobamides compared to ED674 when supplemented with an external source of adenine. Furthermore, adenine can be made soluble in a higher concentration (20 mg/mL) in a solution of 1M HCl. This can be filter sterilised, leading to a lower risk of experimental contamination, a likely useful tool in potential *ΔrlmN* strains as they have a lower growth rate as seen in Chapter 3.2. However, the impact on the pH of media must be assessed first.

4.2.3 Concluding Remarks

To investigate one aspect of cobamide lower ligand variation, we used two strains of cobamide producing *E. coli* to assess the effect of intracellular and extracellular adenine supplementation would have an effect on the cobamides produced. We observed that the supplementation of adenine would increase the ratio of [Ade] Cba produced by the cell and stopped the production of GDP cobinimide. We also managed to successfully delete the *rlmN* gene in a cobamide producing strain of *E. coli.* However, the newly made strain lost its cobamide producing capabilities. This suggests that 2-methyladenine is a potential cobamide lower ligand, however, the concentration of it under the conditions of this project could be lower than the amount of corrinoids produced by the cell and when adenine is available, it is incorporated as a lower ligand as well, [MeAde] Cba appears ubiquitously in human faecal samples, even with the supplementation of adenine. This displays an interesting parallel to the sample of cobamide varieties in the human faecal samples where [MeAde] Cba may not be produced in the highest concentration, but it appears most frequently of all the varieties investigated.[11] This could suggest that this pathway prefers a complete cobamide compared to GDP Cbi. Additionally, this data could also be applied to strain which naturally make and use [MeAde] such as *S. tyohimurium,* which also have the CobUST system of lower ligand synthesis and attachment to the corrinoid ring.[26,27]

A potential future direction of this work can go into the rRNA or specific tRNA varieties which could potentially provide substrate for the lower ligand. After which, the pathway which frees purines to be incorporated into an α-ribazole could potentially be characterised. An example of this would be to substitute A37 in all the respective tRNA which contain it for thymine and assess the cobamides produced by the resulting

mutants. If the resulting cobamides do not contain [MeAde] as a lower ligand, it could indicate that tRNA contains a precursor to the 2-methyladenine lower ligand of potential cobamides. This method will also prevent the potential readthrough errors that may have resulted in the lack of cobamides produced in the *ΔrlmN* strain in this experiment.

References

(1) Stabler, S. P., and Allen, R. H. (2004) Vitamin B12 deficiency as a worldwide problem. *Annu Rev Nutr*.

(2) Escalante-Semerena, J. C., and Warren, M. J. (2008) Biosynthesis and Use of Cobalamin (B 12). *EcoSal Plus 3*.

(3) Herrmann, W., and Obeid, R. (2012) Cobalamin deficiency. *Subcell Biochem 56*, 301–322.

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

(5) Mascarenhas, R., Gouda, H., Ruetz, M., and Banerjee, R. (2022) Human B 12-dependent enzymes: Methionine synthase and Methylmalonyl-CoA mutase. *Methods Enzymol 668*, 309–326.

(6) Warren, M., Schubert, H. L., and Escalante-Semerena, J. C. (2002) The biosynthesis of adenosylcobalamin (vitamin B-12).

(7) Fang, H., Kang, J., and Zhang, D. (2017) Microbial production of vitamin B12: A review and future perspectives. *Microb Cell Fact 16*.

(8) Escalante-Semerena, J. C., and Warren, M. J. (2008) Biosynthesis and Use of Cobalamin (B 12) . *EcoSal Plus 3*.

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

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

(11) Degnan, P. H., Taga, M. E., and Goodman, A. L. (2014) Vitamin B12 as a Modulator of Gut Microbial Ecology. *Cell Metab 20*, 769–778.

(12) Anderson, P. J., Lango, J., Carkeet, C., Britten, A., Kräutler, B., Hammock, B. D., and Roth, J. R. (2008) One pathway can incorporate either adenine or dimethylbenzimidazole as an α-axial ligand of B12 cofactors in Salmonella enterica. *J Bacteriol 190*, 1160–1171.

(13) Benítez-Páez, A., Villarroya, M., and Armengod, M. E. (2012) The Escherichia coli RlmN methyltransferase is a dual-specificity enzyme that modifies both rRNA and tRNA and controls translational accuracy. *RNA 18*, 1783–1795.

(14) Jiang, F., and Doudna, J. A. (2017, May 22) CRISPR-Cas9 Structures and Mechanisms. *Annu Rev Biophys*. Annual Reviews Inc.

(15) Jiang, Y., Chen, B., Duan, C., Sun, B., Yang, J., and Yang, S. (2015) Multigene editing in the Escherichia coli genome via the CRISPR-Cas9 system. *Appl Environ Microbiol 81*, 2506–2514.

(16) Benchling [Biology Software].

(17) Labun, K., Montague, T. G., Krause, M., Torres Cleuren, Y. N., Tjeldnes, H., and Valen, E. (2019) CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res 47*, W171–W174.

(18) Leung, H. B., and Schramm, V. L. (1980) Adenylate degradation in Escherichia coli. The role of AMP nucleosidase and properties of the purified enzyme. *Journal of Biological Chemistry 255*, 10867–10874.

(19) Toh, S.-M., Xiong, L., Bae, T., and Mankin, A. S. The methyltransferase YfgB/RlmN is responsible for modification of adenosine 2503 in 23S rRNA.

(20) Chayot, R., Montagne, B., Mazel, D., and Ricchetti, M. (2010) An end-joining repair mechanism in Escherichia coli. *Proc Natl Acad Sci U S A 107*, 2141–2146.

(21) Sulthana, S., Basturea, G. N., and Deutscher, M. P. (2016) Elucidation of pathways of ribosomal RNA degradation: an essential role for RNase E. *RNA 22*, 1163.

(22) Giuliano, M. G., and Engl, C. (2021) The Lifecycle of Ribosomal RNA in Bacteria. *RNA Damage and Repair* 27–51.

(23) Svenningsen, S. lo, Kongstad, M., Stenum, T. S., Muñoz-Gómez, A. J., and Sørensen, M. A. (2017) Transfer RNA is highly unstable during early amino acid starvation in Escherichia coli. *Nucleic Acids Res 45*, 793.

(24) Dong, H., Nilsson, L., and Kurland, C. G. (1996) Co-variation of tRNA Abundance and Codon Usage in Escherichia coli at Different Growth Rates. *J. Mol. Biol 260*, 649–663.

(25) Burton, K., and Su, M. (1994) Adenine transport in Escherichia coli. *Proceedings of the Royal Society B: Biological Sciences 255*, 153–157.

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

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

Appendix

5.1 Sequencing Data for RlmN deletion

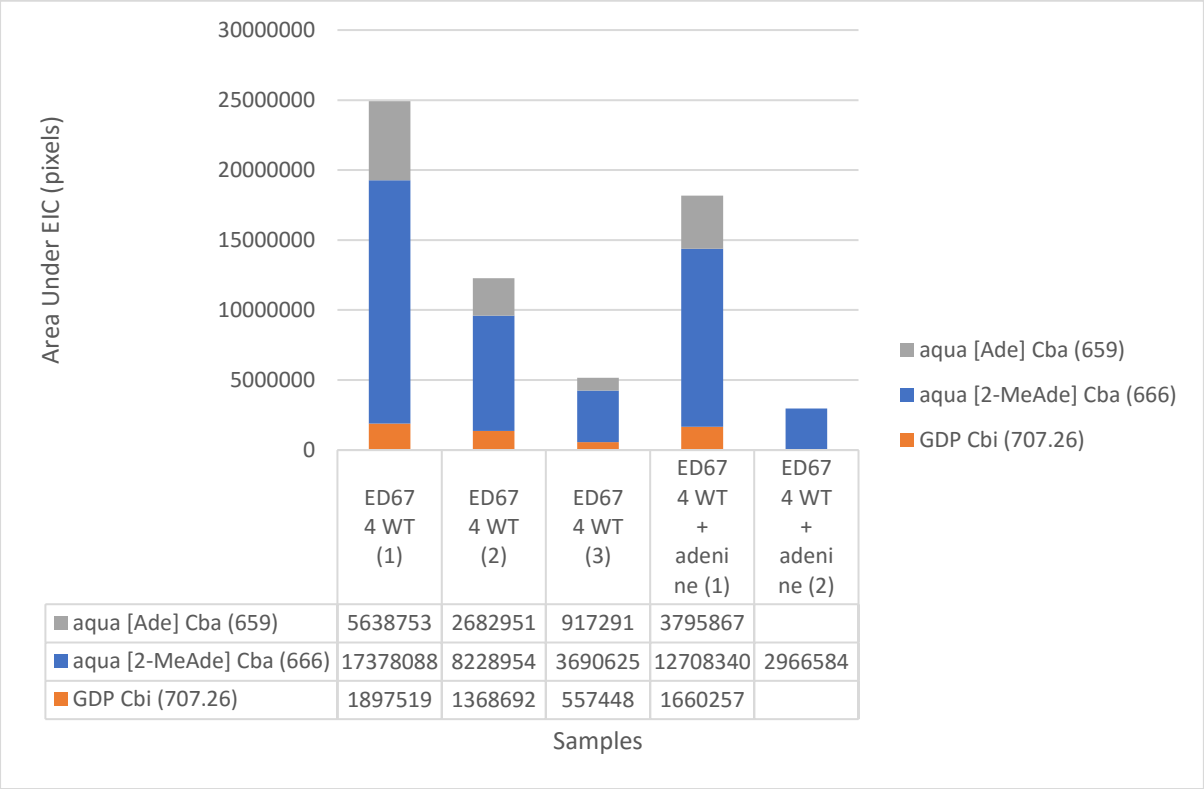
The first sample is the template consisting of the upper and lower flanking regions, the second sample is the sequencing data from the forward strand and the third is from the reverse strand.

```
            1                                                                                   82
RlmN dele... CCCGCCATACGGTCACGATGCCCAGCTCCCGCTGCAGCTGACAGGCGTTCGCGATGGCGCGATTATTAAACGTTTACCGGGC
RlmN_del_...  ---------------------------------------------------------------------------------
RlmN_del_... ---------------------------------------------------------------------------------

.....................................................................................................

            83                                                                                  164
RlmN dele... GCAGCGGAAGCGACTTTGCCGTTGCAATCAAGTGGAGGGGCAGGTGAACGCTGGTGGTTTCTGAATGGCGAACCGTTAACTG
RlmN_del_...  ---------------------------------------------------------------------------------
RlmN_del_... ---------------------------------------------------------------------------------

.....................................................................................................

            165                                                                                 246
RlmN dele... AACGCGGGCGCAACGTGACCCTGCATTTGACGGATAAAGGTGATTATCAATTGCTGGTGATGGATGACGTGGGGCAAATCGC
RlmN_del_...  ---------------------------------------------------------------------------------
RlmN_del_... ---------------------------------------------------------------------------------

.....................................................................................................

            247                                                                                 328
RlmN dele... GACAGTGAAATTTGTCATGCAATAGTCTGTTTTGTCCCTGATTGTTGCTAAAACTCATTTTATTTTAAAAAAATGTTACCTT
RlmN_del_... --AAGAGTGAATTGTCATGCAATAGTCTGTTTTGTCCCTGATTGTTGCTAAAACTCATTTTATTTTAAAAAAATGTTACCTT
RlmN_del_... ---------------------------------------------------------------------------------

.....................................................................................................

            329                                                                                 410
RlmN dele... CATCAATAGTCAACGGCCCTGTTGCTCATTATAATCCGCGCCATCTCGTACGCTGGTACAGACAACAACAGAACAATTTACA
RlmN_del_... CATCAATAGTCAACGGCCCTGTTGCTCATTATAATCCGCGCCATCTCGTACGCTGGTACAGACAACAACAGAACAATTTACA
RlmN_del_... ---------------------------------------------------------------------------------

.....................................................................................................

            411                                                                                 492
RlmN dele... GAGGTAAAAATGGCTATTGAACGTACTTTTTCCATCATCAAACCGAACGCGGTAGCAAAAAACGTCATTGGTAATATCTTTG
RlmN_del_... GAGGTAAAAATGGCTATTGAACGTACTTTTTCCATCATCAAACCGAACGCGGTAGCAAAAAACGTCATTGGTAATATCTTTG
RlmN_del_... ------------------------------------------------------------GGGANCNAAAACTCTTTTAAATTTT

.....................................................................................................

            493                                                                                 574
RlmN dele... CGCGCTTTGAAGCTGCAGGGTTCAAAATTGTTGGCACCAAAATGCTGCACCTGACCGTTGAACAGGCACGTGGCTTTTATGC
RlmN_del_... CGCGCTTTGAAGCTGCAGGGTTCAAAATTGTTGGCACCAAAATGCTGCACCTGACCGTTGAACAGGCACGTGGCTTTTATGC
RlmN_del_... TGCGGCCTTGAAAAGCCCGGGGGTTAAAATTTGGGCCCCAAAGTTNGCCCCCTGCCCCTGGAACCGGCCCGGGGCTTTTTNAG

.....................................................................................................
```

```
            575                                                                    656
RlmN dele... TGAACACGATGGAAAACCGTTCTTTGATGGTCTGGTTGAATTCATGACCTCTGGCCCGATCGTGGTTTCCGTGCTGGAAGGT
RlmN_del_... TGAACACGATGGAAAACCGTTCTTTGATGGTCTGGTTGAATTCATGACCTCTGGCCCGATCGTGGTTTCCGTGCTGGAAGGT
RlmN_del_... TGAACCCGAAGGAAAACCGTTTTTTGA-GGTCTGGTGGAATTCA-GGCCTTTGGCCCGATCGGGGTTTCCGTCC--GGAAGG

.....

            657                                                                    738
RlmN dele... GAAAACGCCGTTCAGCGTCACCGCGATCTGCTGGGCGCGACCAATCCGGCAAACGCACTGGCTGGTACTCTGCGCGCTGATT
RlmN_del_... GAAAACGCCGTTCAGCGTCACCGCGATCTGCTGGGCGCGACCAATCCGGCAAACGCACTGGCTGGTACTCTGCGCGCTGATT
RlmN_del_... GAAAACCCCGTTCAGGGTCACCGCGATTTGCTGGGCGCNACCATTCCGGCAAACGCACTGGCTGGTACTTTGCGCGCTGATT

.....
739                                                                         820
RlmN dele... ACGCTGACAGCCTGACCGAAAACGGTACCCACGGTTCTGATTCCGTCGAATCTGCCGCTCGCGAAATCGCTTATTTCTTTGG
RlmN_del_... ACGCTGACAGCCTGACCGAAAACGGTACCCACGGTTCTGATTCCGTCGAATCTGCCGCTCGCGAAATCGCTTATTTCTTTGG
RlmN_del_... ACGCTGACAGCCTGACCGAAAACGGTACCCACGGTTCTGATTCCGTCGAATCTGCCGCTCGCGAAATCGCTTATTTC-TTGG

.....

            821                                                                    902
RlmN dele... CGAAGGCGAAGTGTGCCCGCGCACCCGTTAATAATTTCGTAAATGCCGCGTGCAAACGTGGCATCCGTGCGCCAGAATTTGT
RlmN_del_... CGAAGGCGAAGTGTGCCCGCGCACCCGTTAATAATTTCGTAAATGCCGCGTGCAAACGTGGCATCCGTGCGCCAGAATTTGT
RlmN_del_... CGAAGGCGAAGTGTGCCCGCGCACCCGTTAATAATTTCGTAAATGCCGCGTGCAAACGTGGCATCCGTGCGCCAGAATTTGT

.....

            903                                                                    984
RlmN dele... ACAATGCAGCGCCCCCGGACGAGCAGCCGCTCACCGGGGCGTTTCTTTTTTCAACCCTCCAGGGGCCATAACGTGTAATAAC
RlmN_del_... ACAATGCAGCGCCCCCGGACGAGCAGCCGCTCACCGGGGCGTTTCTTTTTTCAACCCTCCAGGGGCCATAACGTGTAATAAC
RlmN_del_... ACAATGCAGCGCCCCCGGACGAGCAGCCGCTCACCGGGGCGTTTCTTTTTTCAACCCTCCAGGGGCCATAACGTGTAATAAC

.....

            985                                                                   1066
RlmN dele... GAGGCCGGAATAGCATTAGCCATTACGCCACGGTTACTTTATGTACCGTGGCGTAATGTTTTATGAATGAAGTATTAAGTTG
RlmN_del_... GAGGCCGGAATAGCATTAGCCATTACGCCACGGTTACTTTATGTACCGTGGCGTAATGTTTTATGAATGAAGTATTAAGTTG
RlmN_del_... GAGGCCGGAATAGCATTAGCCATTACGCCACGGTTACTTTATGTACCGTGGCGTAATGTTTTATGAATGAAGTATTAAGTTG

.....

            1067                                                                  1148
RlmN dele... CAATTGTCTGGTCAATATTAACGGTGCGTTTTTGCTGACTTTAAGGCAGTATGTAACGATGCAACAGTAAGTTAGCCTTAGT
RlmN_del_... CAATTGTCTGGTCAATATTAACGGTGCGTTTTTGCTGACTTTAAGGCAGTATGTAACGATGCAACAGTAAGTTAGCCTTAGT
RlmN_del_... CAATTGTCTGGTCAATATTAACGGTGCGTTTTTGCTGACTTTAAGGCAGTATGTAACGATGCAACAGTAAGTTAGCCTTAGT

.....

            1149                                                                  1230
RlmN dele... GAATGTGGGCTTTGTCACGAGCACACAGACGGTCTTATACTGTATGATAACGGTTAACTTAACGGATGTTTCGCGGTGTGGG
RlmN_del_... GAATGTGGGCTTTGTCACGAGCACACAGACGGTCTTATACTGTATGATAACGGTTAACTTAACGGATGTTTCGCGGTGTGGG
RlmN_del_... GAATGTGGGCTTTGTCACGAGCACACAGACGGTCTTATACTGTATGATAACGGTTAACTTAACGGATGTTTCGCGGTGTGGG
```

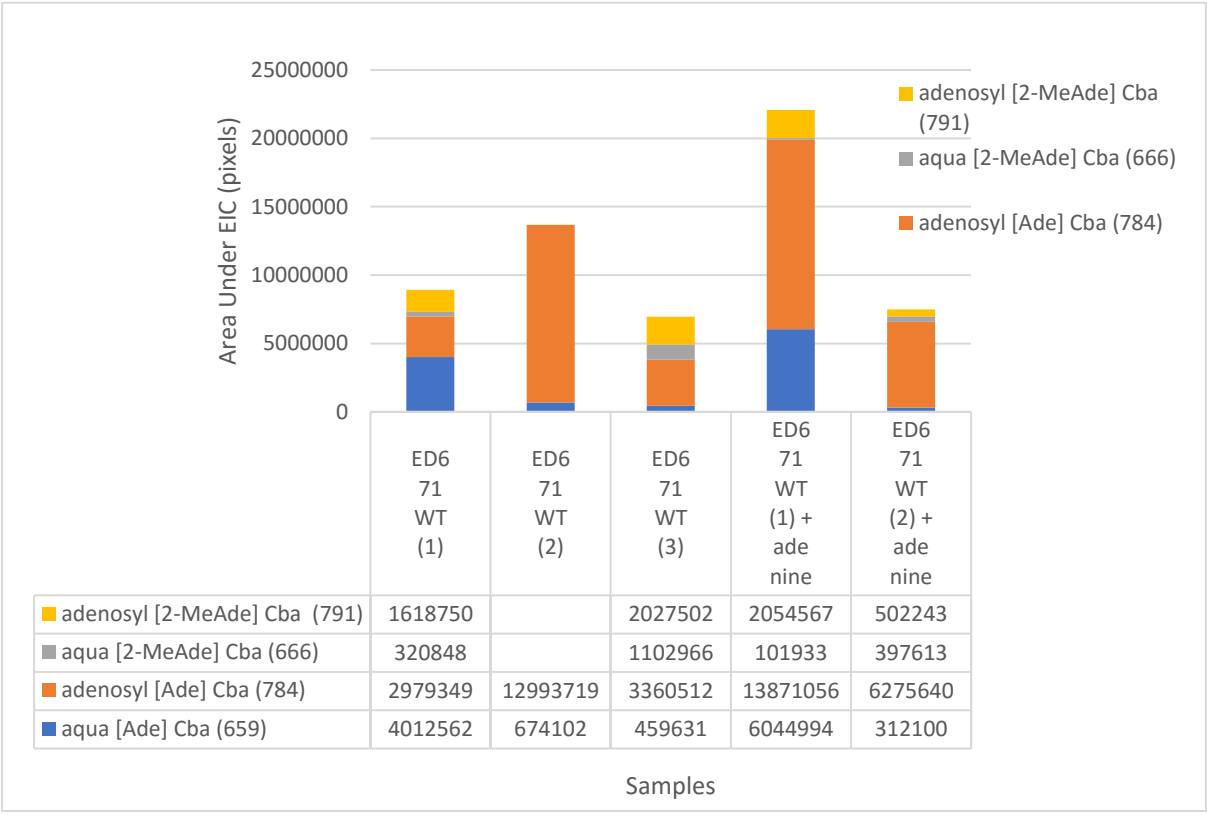
```
            1231                                                                                1312
RlmN dele... TGAGCATGATGGTTCACCGGCATCTCAATTCTCATTTAAACGTACCTGCAGCGAATGAATACTGAAGCCACGCACGACCAAA
RlmN_del_... TGAGCATGATGGTTCACCGGCATCTCAATTCTCATTTAAACGTACCTGCAGCGAATGAATACTGAAGCC-CCCCCGACCAAA
RlmN_del_... TGAGCATGATGGTTCACCGGCATCTCAATTCTCATTTAAACGTACCTGCAGCGAATGAATACTGAAGCCACGCACGACCAAA

.....................................................................................................

            1313                                                                                1394
RlmN dele... ATGAAGCACTTACTACCGGCGCTCGCCTGCGTAATGCTCGCGAACAACTAGGACTTAGTCAGCAGGCCGTTGCCGAGCGACT
RlmN_del_... ATGAAGCACTTACAACCGGGGCTCGCCTGCGTAAGGCNCCGAAAACTAGGANTTANTCGNCANCCGTTGCCAACGACTTGCC
RlmN_del_... ATGAAGCACTTACTACCGGCGCTCGCCTGCGTAATGCTCGCGAACAACTAGGACTTAGTCAGCAGGCCGTTGCCGAGCGACT

.....................................................................................................

            1395                                                                                1476
RlmN dele... TTGCCTGAAGGTTTCCACGGTACGCGACATTGAAGAAGATAAGGCACCCGCCGATCTTGCTTCAACATTCCTGCGCGGATAT
RlmN_del_... TGAAGGTTCCCCGGACCCCGATTNGAAAAAAAAANGGACCCGCCNAATTTGGTTAAAATTCCGGGCGGAATATTCGCCCTTAT
RlmN_del_... TTGCCTGAAGGTTTCCACGGTACGCGACATTGAAGAAGATAAGGCACCCGCCGATCTTGCTTCAACATTCCTGCGCGGATAT

.....................................................................................................
1477                                                                                        1558
RlmN dele... ATCCGCTCTTATGCGCGTCTGGTACATATTCCAGAAGAAGAACTGCTGCCAGGGCTGGAAAAGCAGGCTCCACTTCGGGCTG
RlmN_del_... GCCCCTTGGGANAATTTCCAAAAAAAAAATGNNGGCCGGGCNNGAAAAAGGGGCCCCTTTCGGTCTAAAATTTGCCCAAATGA
RlmN_del_... ATCCGCTCTTATGCGCGTCTGGTACATATTCCAGAAGAAGAACTGCTGCCAGGGCTGGAAAAGCAGGCTCCACTTCGGGCTG

.....................................................................................................

            1559                                                                                1640
RlmN dele... CAAAAGTTGCGCCGATGCAGAGTTTTTCCCTCGGTAAACGCCGCAAAAAACGCGACGGCTGGCTGATGACCTTCACTTGGCT
RlmN_del_...  AATTTTTCCCTGAAACCCCCAAA-------------------------------------------------------------
RlmN_del_... CAAAAGTTGCGCCGATGCAGAGTTTTTCCCTCGGTAAACGCCGCAAAAAACGCGACGGCTGGCTGATGACCTTCACTTGGCT

.....................................................................................................

            1641                                                                                1722
RlmN dele... GGTGTTGTTTGTGGTTATCGGCCTGAGCGGTGCCTGGTGGTGGCAAGACCGCAAAGCTCAGCAGGAAGAGATCACCACTATG
RlmN_del_...  -------------------------------------------------------------------------------
RlmN_del_... GGTGTTGTTTGTGGTTATCGGCCTGAGCGGTGCCTGGTGGTGGCAAGACCGCAAAGCTCAGCAGGAAGAGATCACCACTATG

.....................................................................................................

            1723                                                                                1804
RlmN dele... GCCGATCAATCTTCGGCGGAACTGAGCAGTAA-TAGCGAGCAGGGGCAGAGTGTTCCATTAAATACGTCGACAACTACAGAC
RlmN_del_...  -------------------------------------------------------------------------------
RlmN_del_... GCCGATCAATCTTCGGCGGAACTGAGCAGTAAGTAGCGAGCAGGGGCGAGAGTTCCAAA-----------------------

.....................................................................................................

            1805                                                                                1886
RlmN dele... CCGGCTACAACGTCCACGCCGCCAGCGTCTGTGGATACTACCGCAACCAACACGCAAACACCTGCCGTAACTGCGCCAGCAC
RlmN_del_...  -------------------------------------------------------------------------------
RlmN_del_... -------------------------------------------------------------------------------

.....................................................................................................

            1887                                                                                1968
RlmN dele... CAGCTGTTGATCCGCAACAGAATGCGGTTGTTTCGCCTTCGCAGGCAAATGTTGATACCGCCGCGACCCCGGCACCAACGGC
RlmN_del_...  -------------------------------------------------------------------------------
RlmN_del_... -------------------------------------------------------------------------------

.....................................................................................................

            1969                        2001
RlmN dele... AGCAACAACGCCAGATGGTGCTGCGCCGTTGCC
RlmN_del_... --------------------------------
RlmN_del_... --------------------------------

.....................................................................................................
```

5.2 Graphs Detailing the EIC Data for Cobamide Varieties Produced by ED674



	ED674 WT (1)	ED674 WT (2)	ED674 WT (3)	ED674 WT + adenine (1)	ED674 WT + adenine (2)
aqua [Ade] Cba (659)	5638753	2682951	917291	3795867	
aqua [2-MeAde] Cba (666)	17378088	8228954	3690625	12708340	2966584
GDP Cbi (707.26)	1897519	1368692	557448	1660257	

5.3 Graphs Detailing the EIC Data for Cobamide Varieties Produced by ED671



	ED671 WT (1)	ED671 WT (2)	ED671 WT (3)	ED671 WT (1) + adenine	ED671 WT (2) + adenine
adenosyl [2-MeAde] Cba (791)	1618750		2027502	2054567	502243
aqua [2-MeAde] Cba (666)	320848		1102966	101933	397613
adenosyl [Ade] Cba (784)	2979349	12993719	3360512	13871056	6275640
aqua [Ade] Cba (659)	4012562	674102	459631	6044994	312100