Development of a cell-free strategy for the directed evolution of enzymes for high-value natural products

Thesis for MSc by Research in Microbiology

School of Biosciences

University of Kent

Joshua Harman

2021

Declaration

Name: Joshua Harman

Degree: MSc-Microbiology

Title: Development of a cell-free strategy for directed evolution of enzymes for high-value natural products

I confirm that no part of this thesis had been submitted in support of an application for any degree or other qualification of the University of Kent, or any other University or institution of learning.

Joshua Harman

Abstract

Cell-free protein synthesis (CFPS) provides a robust method for protein production in a system in which all the contained energy can be channelled toward protein synthesis by removing the life goals of the cell. Herein, we sought to produce the haem dependant cytochrome P450 enzyme BM3 using a cell-free system as a proof of concept that complex cytochrome P450 enzymes can be studied using this technique and to provide a foundation for further study into directed evolution. To facilitate the production of BM3 in a cell-free reaction, we optimised a CFPS methodology by varying the reaction composition and expressed and characterised the BM3 protein using *in vivo* methods to act as a positive control. We observed several conditions that affect CFPS activity including, Mg2+, L-glutamate, maltodextrin, and D-ribose. However, while *E. coli in vivo* production of BM3 was strong, in contrast, the optimised *E. coli* CFPS system was unable to synthesise BM3. This could be due to several factors, including low promoter strength or low extract activity. In parallel, we also sought to produce a cell-free system for making haem the prosthetic group required for BM3 activity. Therefore, we expressed and characterised the enzymes for cell-free haem biosynthesis. Intriguingly, in an attempt to make haem b, instead, intermediates such as coproporphyrinogen III were formed. We suggest that either one of the enzymes was limiting (HemH - ferrochelatase), or a co-factor was absent in the reaction. In this research we were unable to synthesise BM3 using CFPS, this is most likely due to the low activity of the extract, by further refining the crude extract production method and the composition of the cell-free reaction it may be possible to produce BM3.

Table of content

List of Figures

List of Tables

Abbreviations

A_{260}	Absorbance at 260 nm
A_{280}	Absorbance at 280 nm
ALA	Delta-aminolevulinic acid
APS	Ammonium persulfate
BM3	Cytochrome P450 BM3
cAMP	Adenosine 3′,5′-cyclic monophosphate
CFPS	Cell-Free Protein Synthesis
CoA	Coenzyme A
CPR	NADPH-Cytochrome P450 Reductase
dNTP	Deoxynucleotide
dNTPs	Deoxyribonucleotide triphosphate
FAD	Flavin adenine dinucleotide
FMN	Flavin mononucleotide
His_6-tag	Hexahistidine tag
LB	Luria-Bertani
mRNA	Messenger Ribonucleic acid
NADPH	Nicotinamide adenine dinucleotide phosphate
P450	Cytochrome P450
PEG	Polyethylene Glycol
PEP	Phosphoenolpyruvate
PGA	3-phosphoglyceric acid
PTMs	Post-translation modifications
PURE	Purified recombinant element
SOC	Stable outgrowth medium
T7RNAP	T7 RNA polymerase
TAE	Tris-acetated-EDTA
TBST	Tris-buffered saline-Tween®-20
TEMED	Tetramethylethylenediamine
tRNA	Transfer ribonucleic acid
v/v	volume/volume
w/v	weight/volume
w/w	Weight/weight

Acknowledgement

Chapter 1 Introduction

1.1 Cell-Free Protein Synthesis

1.1.1 An introduction to cell-free protein synthesis (CFPS)

Since its development as a technique to determine the link between mRNA and protein production by decoding the genetic code (Nirenberg and Matthaei, 1961), CFPS has developed into a powerful tool in biochemistry with a wide variety of applications; from the production of antibodies (Yin *et al*., 2012) and site-specific antibody-drug conjugates (Zimmerman *et al*., 2011); to the understanding of biological systems (Hodgman and Jewett, 2012), the detection of cancer cells (Culler *et al*., 2010) and more. CFPS systems consist of crude extracts, prepared from the cell line of choice combined with the other necessary substrates of protein synthesis, such as amino acids, DNA, cofactors, salts, nucleotides and energy substrates. When combined, protein synthesis occurs resulting in the production of the target protein (Dondapati *et al*., 2020).

One of the appeals of CFPS systems is the lack of cellular membrane and a functional genome which removes the confliction between the cellular regulation (Carlson *et al*., 2014); furthermore, the in vitro approach removes the metabolic and cytotoxic burdens associated with large-scale protein production via *in vivo* methods (Focke *et al*., 2016). Some cell-free systems have a higher protein titre as a result of the fact that all the energy situated in the system is purely channelled towards protein production. CFPS allows direct control over reaction components as it is an open system; this allows the manipulation of the environment of protein synthesis and facilitates proper protein folding due to the distance between ribosomes (Rosenblum and Cooperman, 2014). Compared to *in vivo* protein synthesis, CFPS can be scaled for large quantities of protein production, in some instances up to 100 L (Hodgman and Jewett, 2012). Linear DNA can be used as the template for protein production in cell-free reactions; this eliminates the need for cloning steps in plasmid-based methods; as well as the transformation steps to incorporate the plasmid into the bacterial genome (Ayoubi-Joshaghani *et al.*, 2020). However, CFPS systems often show a high degree of variability between CFPS reaction and crude extract batches, this variation is attributed to the small reaction volumes, sensitive reagents, complicated protocols, and differences in techniques of individuals (Cole *et al*., 2019). A comparison of the workflow for i*n vivo* and i*n vitro* protein production systems can be observed in figure 1.1.
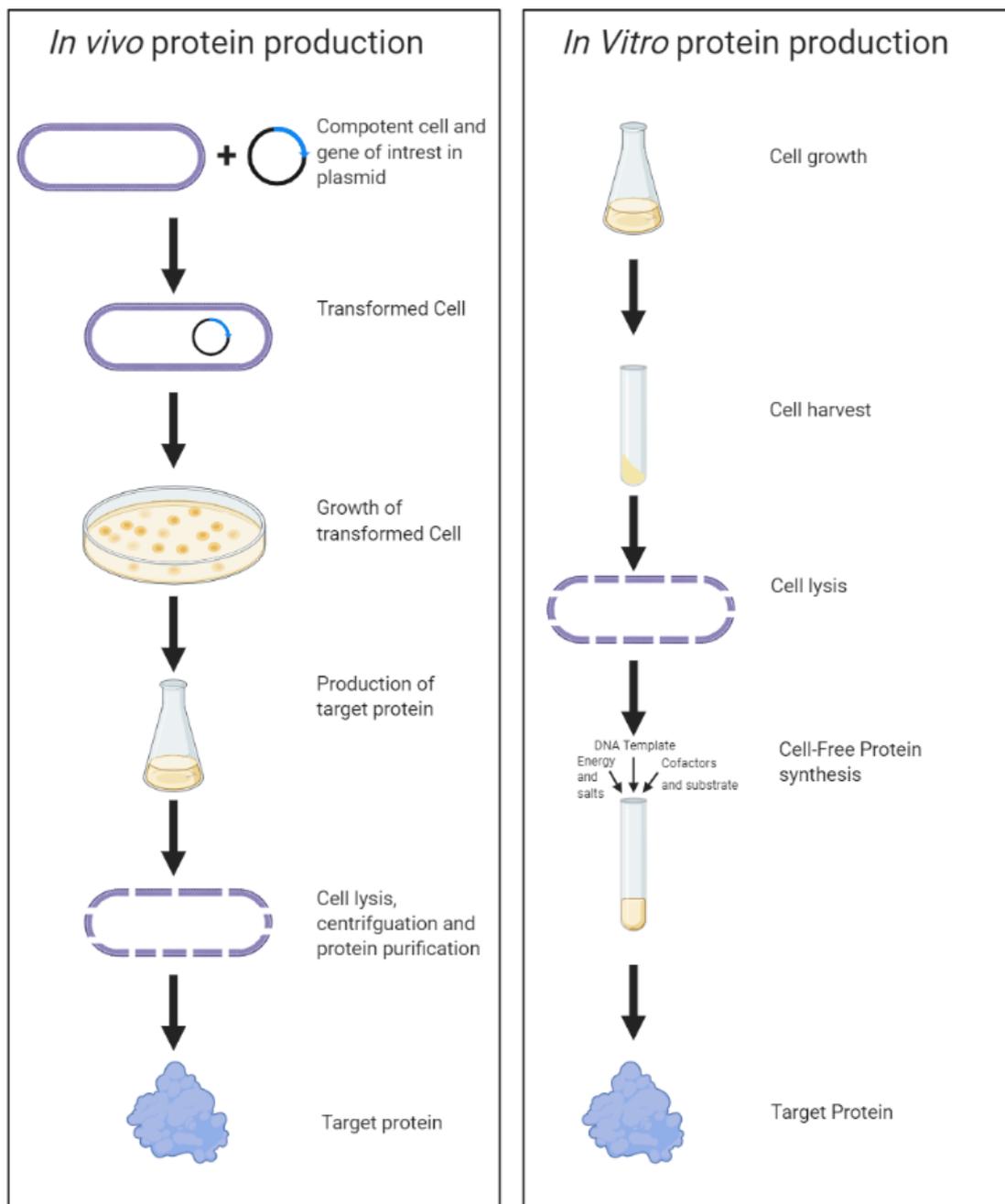
Figure 1.1 A comparison of *in vivo* and *in vitro* systems. A comparison of the standard workflow for *in vivo* (left) and *in vitro* (right) protein synthesis. CFPS provides advantages over *in vivo* protein synthesis such as the absence of a cell membrane removing a barrier, no genomic DNA removing genetic regulation and the fact that the system acts like a chemical reactor, you feed substrate and the product is produced. (Figure created with information from (Gregorio *et al*., 2019).

1.1.2 Construction of a CFPS systems

CPFS systems require several defined elements. A CFPS system reaction mixture can be divided up into four constituent parts; 1) cell extracts including the transcription and translation machinery required for protein synthesis, 2) A system for ATP regeneration, 3) Template DNA for protein-encoding and 4) substrates for protein synthesis (Lee and Kim. 2018). The standard component of a CFPS system and the role they play in the reaction can be observed in table 1.1.

Crude extract production

One of the major aspects of CFPS is crude cellular extracts which provide the system with the components necessary for transcription and translation (Katzen *et al.*, 2005). Several cell lines are viable for use in CFPS systems; these include but are not limited to *Escherichia coli* (Kigawa *et al.*, 2004), *Clostridium autoethanogenum* (Krüger *et al.*, 2020), insect cells (Carlson *et al.*, 2012), wheat germ (Guillaume *et al.*, 2019) and many more. The initial steps for the creation of a functioning CFPS are parallel regardless of the starting organism. This starting point involves the growth of the chosen cell line from which the transcription and translation machinery is to be extracted; followed by the lysis of the cells while preserving ribosomal functionality (Gregorio *et al.*, 2019). *E. coli* is the favoured cell line for CFPS due to several factors; *E. coli* extracts have been shown to produce a high protein yield when used in CFPS systems (Caschera *et al.*, 2014). The high yield of protein synthesis seen in *E. coli* extracts is due to the unparalleled fast growth kinetics and rapid doubling time means that a high number of total cells in the culture are capable of producing protein (Rosano *et al.*, 2014). Another beneficial factor in the use of *E. coli* is that the species is easy to cultivate in large quantities using low-cost media (Carlson *et al.*, 2012). The *E. coli* CFPS system also offers the lowest reaction cost when compared to other cell lines. *E. coli* cell-free are efficient due to their intrinsic metabolic ability to regenerate ATP from low-cost substrates, which provides energy for long-lasting (>8 hours) and high-level protein synthesis Swartz, 2006). The cell extracts contain upwards of 400 proteins, with ~108 dedicated to transcription-translation, and the remainder containing many enzymes from central metabolism and other factors that contribute to gene expression, including but not limited to folding chaperone, RNA degradation, electron transport chain (ETC) and nucleotide/amino acid biogenesis. We know the ETC is functional since the reactions are oxygen-dependent and catalyse oxidative phosphorylation from glucose catabolism (Vilkhovoy, et al., 2018). This reduces the need for the use of expensive energy-

providing molecules, although high-energy substrates such as 3-phosphoglycerate and pyruvate provide more efficient ATP regeneration schemes.

While *E. coli* is the favoured organism for CFPS, recent strides have been made in establishing high-yield CFPS systems using *Vibrio natriegens* (Des Soye *et al*., 2018). *V. natriegens* is a marine bacterium that is currently the fastest-growing non-pathogenic bacteria that has been observed with an observed doubling time of 9.4 minutes (Hoffart *et al*., 2017). This rapid doubling rate coupled with the fact that *V. natriegens* has been shown to grow to a notably higher cell density than *E. coli* makes it a prospective alternative chassis organism to *E. coli for* use in biotechnology (Hoff *et al.*, 2020). The ultra-fast growth rate of *V. natriegens* has been linked to high rates of protein production (Aiyar *et al.,* 2002) and metabolic efficiency (Hoffart *et al*., 2017) suggests the possibility of applying the organism in CFPS as an expression system (Wiegand *et al.,* 2018). Research has shown that is possible to produce high protein yields when using a crude extract derived from *V. natriegens* with Wiegand *et al* (2018) achieving >260 µg/mL, in this work, it was noted that *V. natriegens* required refined culturing methods to achieve the highest protein yield, with the greatest yield observed in LB media supplemented with V2 salts (204 mM NaCl, 4.2 mM KCl, 23.14 mM $MgCl_2$ Wiegand *et al* 2018). It was also noted that the most active crude extracts were produced from mid-logarithmic growth stage cells. Des Soye *et al* (2018) achieved 1.6 g/L of protein in CFPS reactions when using *V. natriegens* and noted that this system did not require codon optimisation but exhibited a lower transcription and translation efficacy than those which used *E. coli.*

Energy regeneration system

When preparing a CFPS reaction it is of fundamental importance to consider the metabolic pathways that will be providing energy to the system as this will have a direct bearing on the efficiency of the system. Protein translation represents the process with the highest energy demand in any CFPS system; with the need for 5 ATP molecules total per amino acid addition (Noireaux and Liu, 2020). This total is determined from the 2 ATP equivalents needed for amino acylation, during this process the carboxyl group of the backbone of the amino acid is covalently linked to the α-phosphate of the ATP molecule, releasing inorganic pyrophosphate, creating a 5' aminoacyl adenylate intermediate, in the second step of this reaction the generated

a 5' aminoacyl adenylate intermediate binds to a tRNA molecule (Pang et al.,2014). N-Formylmethionine is the starting residue in bacterial protein synthesis and is delivered to the ribosome by tRNA-fMet. Translation initiation requires 1 ATP (Sherman et al., 1985). 2-3 ATP molecules are used during the elongation of the polypeptide these are required for ribosome movement, the activation of translation factors (Neelagandan et al., 2022) and is required for the activation of a carboxy terminus causing a free peptide to be covalently bound elongating the polypeptide chain (Dall and Brandstetter, 2015) In the final step of protein synthesis one ATP molecule is used for the activation of the release factor during termination (Ray-Soni et al., 2016). With a total ATP cost of 5 per amino acid bond, it can be determined that the production of 1 mg/mL of an average-sized protein of 220 amino acids would require 44 mM of ATP (Calhoun and Swartz, 2007). There is a diverse range of secondary energy sources that can be added to a cell-free reaction to produce the required ATP. The simplest type of energy source is compounds that contain a high-energy phosphate bond so that ATP can be regenerated directly, such as in the case of PEP or creatine phosphate (Swarts, 2006). It is important to note that compounds containing high-energy phosphate bonds are unstable in cell-free reactions and can be rapidly degraded by non-specific phosphatases found in the cell extract (Shen *et al*., 1998). One strategy to prevent the breakdown of compounds is the periodic addition of PEP to the reaction to allow the regeneration of ATP; this method prolongs the period of protein synthesis and increases the production of recombinant proteins; while this method lessens the breakdown of PEP the accumulation of inorganic phosphates inhibits protein synthesis (Kim and Swartz, 2000). Research by Calhoun *et al* (2005) and Kim *et al* (2007) have shown that both glucose and fructose-1,6-bisphosphate respectively, are cheap and efficient energy sources for the regeneration of ATP that do not cause the accrue of inorganic phosphates. Another possible energy source is glutamate which has been shown to produce reducing agents such as NADH; which plays a key role in the generation of ATP by oxidative phosphorylation (Jewett *et al*., 2008).

DNA template

A cell-free system requires template DNA to allow the encoding of the desired protein, this is typically provided in the form of PCR-amplified DNA (Lee and Kim. 2018). PCR products or other forms of linear DNA are prone to rapid degradation due to the activity of exonucleases found in cell extracts; mRNA present in the cell extract is also prone to digestion (Brow and Burgess 1991). The combination of these factors means that CFPS systems using PCR

products generally produce a lower protein yield when compared to whole-cell systems (Lee and Kim. 2018). Efforts have been made to improve the efficiency and stability of DNA template CFPS systems. One approach is the use of GamS proteins which are introduced to the system to prevent the breakdown of DNA by exonucleases (Sitaraman *et al*., 2004). It is also possible to stabilise DNA via the introduction of a χ-site, this prevents the rapid degradation of the linear DNA by binding the RecBCD, the major nuclease responsible for the degradation of linear DNA, which subsequently enhances the output of cell-free systems (Marshall *et al.*, 2017). Alternatively it is possible to stabilise the mRNA generated during the reaction by preparing a cell extract from *E. coli* strains which are RNase E-defective, reducing the endonuclease digestion of mRNA. The presence of T7 terminator-overhang sequences in an mRNA sequence resulted in continuous protein synthesis when using ribonucleases E-defective cell extracts, which was speculated to be because of enhanced stability in the transcribed mRNA (Hahn and Kim 2006).

Substrates and co-factors

For a successful cell-free reaction several substrates and reagents are required to allow effective protein synthesis. Common reagents include T7RNAP, HEPES buffer, Mg^{2+} and K^+ salts, 20 amino acids, nucleotides, ATP, GTP, UTP, folic acid, NAD, coenzyme A, oxalic acid, putrescine, and spermidine (Levine *et al*., 2019; Lee and Kim. 2018).

Table 1.1 Components of an *E. coli* CFPS system and the role they play in the reaction.

Component	Role in a cell-free reaction
Crude extract	Provides the machinery for transcription and translation (Gregorio *et al.*, 2019).
T7RNAP	T7RNA polymerase catalyses the formation of RNA from the DNA template and contributes to an elevated level of protein synthesis. T7RNA polymerase is one of the most efficient polymerases; in part, due to the fact, the enzyme can complete a transcription cycle without requiring additional protein factors (McManus *et al*., 2019).

DNA template	Allows the encoding of the desired protein (Lee and Kim. 2018).
Potassium glutamate	Provides a source of potassium ions for the reaction. These ions balance the charges from nucleic acid phosphate groups and other ionic compounds in the system; preservation of the ionic balance is essential for several protein-nucleic acid interactions (Jewett and Swartz, 2004). Glutamate can produce reducing equivalents such as NADH, which plays a role in the generation of ATP via oxidative phosphorylation; making glutamate a key energy source in cell-free reactions (Jewett et al., 2008).
Magnesium glutamate	Provides a source of magnesium ions for the reaction; The decrease in magnesium ions in the reaction due to the accumulation of inorganic phosphates causes pre-mature stoppage of protein synthesis. The supplementation of magnesium ions has been shown to elongate the reaction time (Kim and Kim, 2009). Glutamate can produce reducing equivalents such as NADH, which plays a role in the generation of ATP via oxidative phosphorylation; making glutamate a key energy source in cell-free reactions (Jewett et al., 2008).
NTP's	Play a crucial role in translation and are used in transcription as substrates for the synthesis of RNA. NTP's couple together with systems transcription and translation components (Siegal-Gaskins et al., 2014).
Amino Acids	Monomers are used for the creation of the desired protein. After activation peptide bonds form between each amino acid elongating the peptide

	until the desired protein is synthesised. (Whittaker, 2014).
HEPES buffer	Used to maintain the physiological pH of the cell-free system (Rolf *et al*., 2019).
tRNA	Transport amino acids to the ribosome during protein synthesis. The addition of tRNA to the reaction supplements the concentration present in the cell extract; this increase in concentration means more amino acids are available for translation (Rolf *et al*., 2019).
CoA	A cofactor that plays a significant role in the generation of ATP via oxidative phosphorylation using glutamate (Cai *et al*., 2015).
NAD	A cofactor that plays a significant role in the generation of ATP via oxidative phosphorylation using glutamate (Cai *et al*., 2015).
Folinic acid	A precursor for formylmethione synthesis. Formylmethione is an amino acid derivative required for the initiation of prokaryotic protein synthesis (Cai *et al*., 2015).
Spermidine	Provides a source of ions for the reaction. These ions balance the charges from nucleic acid phosphate groups and other ionic compounds in the system; preservation of the ionic balance is essential for several protein-nucleic acid interactions (Jewett and Swartz, 2004).
3-PGA	Plays a role in energy regeneration. 3-PGA is viable as a secondary energy source and has been shown to yield the best protein production when compared to other buffers (Shin and Noireaux, 2010).

40% PEG (w/v)	PEG is used to maintain stable messenger levels and may mimic macromolecular crowding effects in the cell-free reaction. PEG is also used to simulate the viscosity of the cell cytoplasm (Rolf *et al*., 2019).

1.1.3 Other factors to consider when establishing a CFPS systems

One important consideration for CFPS is reaction temperature, which can differ greatly, depending on the platform used, cell line, or the protein being produced. Typically for *E. coli-based* systems, the optimum temperature ranges from 30°C to 37°C, however, the temperature differs depending on the time of incubation (Levine *et al*., 2019). The total time from preparation is also a factor to consider when designing a system; the use of *E. coli* is preferred as it requires the least time to prepare; a factor that can be attributed to its quick growth rate and the ease of extract preparation (Gregorio *et al*., 2019). The method of cell lysis is an area in which optimisation can be achieved; it has been shown that the use of sonication for cell lysis can produce a high number of extracts in a shorter period than the use of a high-pressure impinge homogenizer; while the use of sonication still achieves a high protein titre (Kwon and Jewett., 2015). It is possible to increase the efficiency of CFPS via the addition of supplements to the reaction; these supplements can improve components of the reaction such as the energy systems or transcription and translation (Dopp *et al.*, 2019).

1.1.4 Applications of CFPS

PTMs- CFPS is a valuable system for the study of PTMs; which have emerged as a critical area of study due to the role they play in a variety of cellular functions (Lothrop *et al*., 2013). The use of CFPS has allowed the study of PTMs through the generation of site-specific protein post-translation modification by the genetic code expansion strategy, which involves the genetic incorporation of modified amino acids into the target area of proteins (Chen *et al*., 2018). For example, Oza et al (2015) sought to tackle the limitations in studying serine phosphorylation, these being the inability to produce phosphoproteins at high purity and yields; by applying CFPS. Their cell-free system featured three key components, these being, a highly active crude extract produced using a gnomically recoded Release Factor 1-deficient strain (*E. coli* C321.ΔA) that also lacks the Sep-specific phosphatase SerB. The expression of an

improved Sep-OTS during the growth of the *E. coli* strain to enrich the lysate with phosphoryl-tRNA synthetase, tRNASep and EF-Sep, which is necessary for phosphoprotein synthesis, and finally integrating the components in a CFPS system which had previously been shown to mimic the *E. coli* cytoplasmic environment. Another example can be is Lysine acetylation, which is one of the most well-studied PTMs due to the crucial role it plays in DNA replication, work by Venkat et al (2019) incorporated the non-deacetylatable analogue Nε-thio-acetyllysine into proteins by combining flexizyme and the PURE system. Flexizymes are capable of changing a wide variety of non-natural amino acids on tRNAs allowing the reprograming of the genetic code by reassigning the codons that are generally assigned to natural amino acids to non-natural residues. In this Cell-free system, the flexizyme plays a role in incorporating the analogues into the protein.

High-throughput screening- CFPS has become the gold standard approach for high-throughput functional research (Chong, 2014); in particular, the use of linear DNA expression templates is a practical tool for high-throughput studies and protein screening (Schinn, *et al*, 2016) for determining protein functions (Gregorio *et al.*, 2019). *E. coli* is the most explored high-throughput platform to screen antibody mutant libraries for antibody engineering; it is consequently possible to scale up the production of the best mutants using the same platform to allow industrial production (Yin *et al.*, 2012). *E. coli* platforms have also been used for the identification of gene products involved in metabolic systems which result in protein accumulation and folding through the application of high-throughput functional genomics (Woodrow and Swartz, 2007).

Antibodies- Cell-free systems have been implemented to allow the production of functional antibodies, which represent the potential for a faster method of manufacturing when compared to *in vivo* systems which may have the disadvantages of protein insolubility, degradation of the protein inside the cell, or toxicity of the protein which may be detrimental to cell growth. A key advantage of cell-free systems is the open reaction environment that allows the conditions to be optimised directly, by the addition of chemicals or alteration of physical parameters that may inhibit cell growth, or in the case of substrates, may be restricted by the cell transport mechanism. This direct manipulation of the reaction can allow rapid design cycles to optimise reaction conditions. *E. coli* systems have been able to produce large yields of full-length, correctly folded, and assembled antibodies (Yin *et al*., 2012).

Large proteins - CFPS reactions can be performed using a batch of semi-continuous synthesis. Batch synthesise, has all the reactants necessary for protein production added to the reactor at the start of the reaction. Semi-continuous reactions involve a dialysis device to provide a short period (~10 hours) of substrate feed and removal of inhibitory waste products. Using *E. coli* systems, it has been possible to synthesis active proteins above 100 kDa; the production of such large proteins has been found to overwhelm *in vivo* systems leading to lower quantity protein expression (Rosenblum *et al*., 2014). The capability of CFPS systems to synthesise large proteins assures that the target protein in this research, which is above 100 kDa will be synthesisable using this approach.

Metalloproteins- Metalloproteins are known to be difficult to synthesis *in vivo* due to several factors such as low yields, insolubility, and poor metal cofactor assembly. The use of *E. coli* cell-free systems has allowed the synthesis of soluble and active metalloproteins, through the manipulation of reaction conditions and the use of chemical additives. For example, the production of Multicopper oxidases, which have potential biotechnological applications but the production of which is impaired by low expression when using *in vivo* expression methods. As an alternative, *E. coli* based CFPS has been shown as a viable platform for target protein production with yields as high as 1.2 mg/mL, more than 95% of the protein being soluble, and the ability to scale the reaction from 15 µL to 100 without a loss in productivity or solubility (Li *et al*., 2015).

Non-natural amino acid incorporation- The incorporation of toxic noncanonical amino acids into proteins can result in an enlarged scope of functionality, however, the incorporation of these analogue amino acids often inhibits cell growth as they modify endogenous proteins, limiting protein production. CFPS is an applicable method for the incorporation of toxic amino acids, making it a potent tool for further investigation into amino acid toxicity (Worst *et al.,* 2016).

Enzyme pathway studies- CFPS has been incorporated for the study of enzymes and multi-step biosynthesis pathways. By applying cell-free methods it is possible to produce an enzyme pathway, which subsequently allows for optimisation. This includes testing enzymes for the forward design of cellular systems, such as in the case of the production of biofuels (Karim and Jewett, 2018). Another pathway explored in CFPS involves the biosynthesis of glycan

structures. Glycosylation plays an important role in cellular functions and provides proteins with additional properties such as increasing stability and providing protection against denaturation (Barel and Charbit,. 2017). However, constructing precise pathways to study and engineer glycan structures on proteins remains difficult. By combining several CFPS-synthesised glycosyltransferases installed onto a protein target by an N-glycosyltransferase it is possible to assemble glycosylation pathways. This was demonstrated by producing 37 different protein glycosylation pathways (Kightlinger *et al*., 2019)

1.2 Cytochrome 450 enzymes and Cytochrome P450 BM3

1.2.1 Cytochrome 450 enzymes and their catalytic cycle

P450 enzymes are a superfamily of oxygenases found in all domains of life and are known for their catalytic versatility and dependency on a haem iron centre for activity (Nelson, 2006; Gilardi *et al*., 2017). The haem iron is co-ordinated by a cysteine residue in its thiolate form which is considered to be the active form of the enzyme; in this state, the enzyme typically has its major absorbance at roughly 418 nm (Miles *et al.*, 1992). Many P450 enzymes catalyse the reduction of oxygen molecules, binding them to their ferrous haem iron centre, and ultimately the formation of an oxygenated substrate and water produced from the reduction of the oxygen atom (Denisov *et al*., 2005). P450 enzyme activity can produce several outcomes, depending on the enzyme or substrate involved. P450 enzymes catalyse reactions such as hydroxylation, epoxidation, isomerization, aromatic hydroxylation and oxidation of alcohols and aldehydes, and oxidative C-C cleavage (Bernhardt, 2006. Therefore, P450 enzymes play a diverse role in various physiological reactions, including the metabolism of drugs and xenobiotics (Estabrook, 2003), the formation of steroids (Munro *et al*., 2018) and vitamin D synthesis in mammals (Ortiz de Montellano, 2015). In plants, P450 enzymes are involved in chemical defence mechanisms (Mizutani and Sato, 2011). P450 enzymes play an important part in the adaptation and survival of plants demonstrated through their role in the biosynthesis of antibiotics (Cupp-Vickery, 1995) and the ability to use various compounds as metabolism energy sources (Girvan and Munro, 2016).

Most P450 enzymes require an NADPH (Nicotinamide adenine dinucleotide phosphate)-driven redox system to properly function in living cells. In bacterial-derived P450 enzymes, this redox system involves the initial transfer of two electrons from NADPH to a FAD (flavin adenine dinucleotide)-binding ferredoxin reductase and then subsequently to an iron-sulphur cluster binding protein, which successively transfers two electrons to the P450 enzyme to initiate the P450 catalytic cycle (Munro *et al*., 2018). The arrival of the electrons at the P450 enzyme causes the reduction of the haem iron group, first to its ferrous state which allows oxygen binding and then further reducing the ferric-superoxo species formed to a reactive ferric-peroxo sate. In the following reaction step, the ferric-peroxo species undergoes two protonation steps: first forming a ferric-hydroperoxo species (compound 0). Compound 0 is further protonated and dehydrated to form a ferryl-oxo porphyrin radical cation species (compound I). Compound I is fundamental to the insertion of oxygen in P450 catalysed

reactions and is considered the major oxidant in the P450 reaction and extracts hydrogen from the substrate to produce a ferryl-hydroxo species (compound II). The substrate underdoes decarboxylation through one-electron oxidation caused by compound II generating a carbocation. This carbocation readily eliminates CO_2 with the formation of a double bond generating a terminal alkene; this occurs simultaneously with the protonation of the haem group, restoring the P450 enzyme to its resting state (Rittle and Green, 2010; Belcher *et al*., 2014; Munro *et al*., 2018). The catalytic cycle of Cytochrome P450 enzymes can be seen in figure 1.2

Some P450 enzyme catalysis can occur without the need for external electrons, this substitute pathway is instead driven by hydrogen peroxide or organic surrogates that are also able to react with the P450 enzyme-producing haem iron-oxo species. This alternative pathway, known as the 'peroxide shunt' is typically less efficient when compared to the standard catalytic cycle and can lead to haem modification and the loss of enzyme activity through modifications to the amino acids comprising the active site; these factors mean this method is rarely used where high yields of oxidised products are required (Munro *et al*., 2018).
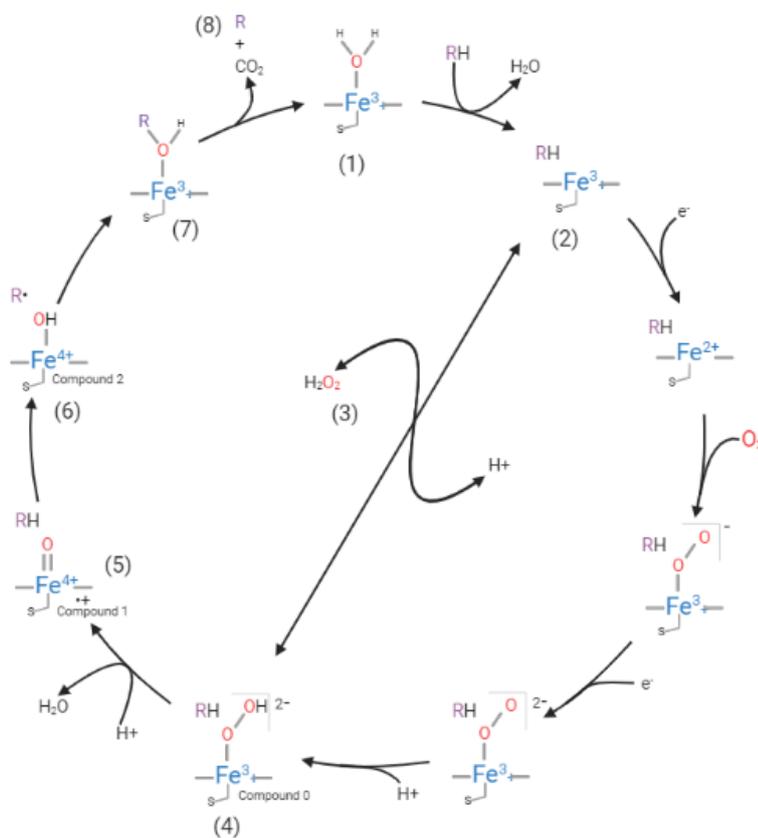
Figure 1.2 The cytochrome P450 catalytic cycle. The P450 enzyme (1) is in resting form, the haem iron is in a low-spin state and co-ordinated cysteine and water. (2) The substrate binds to the enzyme displacing the water molecule and converting the haem iron to a high-spin form. (3) Peroxygenase P450 enzymes use the peroxide shunt pathway which removes the need for external electrons; the interaction with hydrogen peroxide converts the substrate-bound enzyme to ferric-hydroperoxo species compound 0 (4). (5) Compound 0 is protonated and dehydrated resulting in compound 1- ferryl-oxo porphyrin radical cation species. (6) Compound 1 is protonated producing a ferryl-hydroxo species (compound 2). (7) The substrate is decarboxylated due to oxidation of the substrate caused by compound 2 generating a carbocation. (8) The carbocation eliminates CO_2 with the formation of a C_α -C_β double bond generating a terminal alkene; this occurs concurrently with the protonation of the haem iron restoring it to its resting state. (Figure created with information from Rittle and Green, 2010; Belcher *et al*., 2014; Munro *et al*., 2018).

Haem Biosynthesis

Haem plays a paramount role in the activity of P450 enzymes, for this reason, it is fundamental that a haem synthesis pathway is present during the generation of the P450 enzymes. In bacteria, haem is synthesized in an eight-enzyme mediated reaction that begins with the synthesis of δ-aminolevulinic acid (ALA) from either the condensation of succinyl-CoA and glycine in a reaction catalysed by ALA synthetase; in this reaction, the carboxyl carbon of the glycine molecule is lost, and the reaming components combined with CO_2 (O'Brian and Mark, 2002), or the conversion of glutamate to ALA in a three-enzyme pathway. The initial step of the conversion of glutamate to ALA involves the activation of glutamate by Glutamate-tRNA synthetase, a reaction in which glutamate is ligated to tRNA, this reaction requires ATP and magnesium. Next glutamate is reduced in a reaction that requires a reduced pyridine nucleotide, this is catalysed by Glutamate-tRNA reductase, and the product of this reaction is Glutamate-1-semialdehyde. In the final reaction step, the position of the nitrogen and oxo atoms of the reduced five-carbon intermediate are interchanged to form ALA, this reaction is catalysed by Glutamate-1-semialdehyde aminotransferase (Beale., 2007). The next step of the reaction pathway involves the condensation of two molecules of ALA by the enzyme ALA dehydrates (Hem B), producing a single molecule of porphobilinogen (Warren *et al*., 1998). 4 molecules of porphobilinogen are polymerised by the enzyme porphobilinogen deaminase (Hem C) producing one molecule of 1-hydroxymethylbilane and ammonia which is a by-product; 1-hydroxymethylbilane is then cyclized to form uroporphyrinogen III by uroporphyrinogen III synthase (Hem D). The four-acetate residues of the uroporphyrinogen III molecule are decarboxylated to methyl groups by uroporphyrinogen decarboxylase (Hem E) producing coproporphyrinogen III, with the release of four CO_2 groups (O'Brian and Mark, 2002). In the next step of the reaction coproporphyrinogen III is converted to coproporphyrinogen III oxidase (Hem F), this reaction produces 2 molecules of CO_2 as by-products (Dailey, 2001). The penultimate step of haem synthesis involves the production of protoporphyrinogen IX via the 6-electron oxidation of Protoporphyrinogen catalysed by Protoporphyrinogen IX Oxidase (Hem G), this reaction produces 6 hydrogen molecules as a by-product (Dailey, 2001). The terminal step of haem synthesis is catalysed by ferrochelatase (Hem H), this reaction involves the insertion of a ferrous iron into protoporphyrin IX producing protohaem (O'Brian and Mark, 2002).
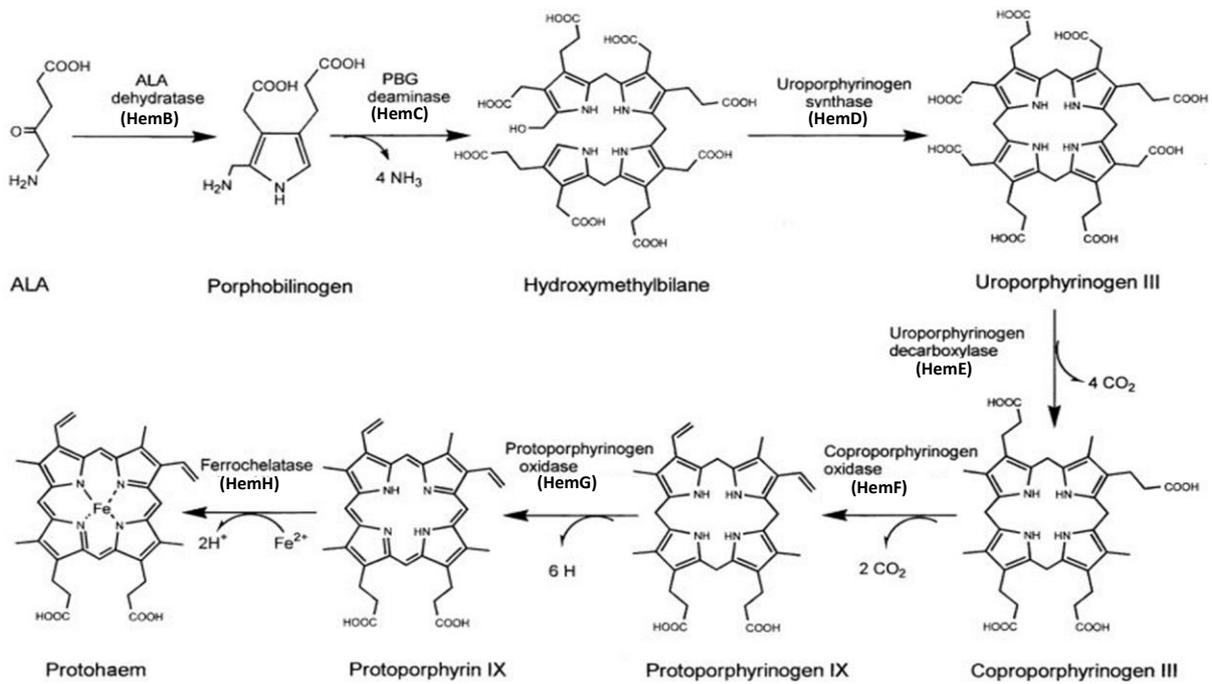
Figure 1.3 an overview of the enzymatic steps of the Haem synthesis pathway. Haem is synthesized in an enzyme-mediated reaction pathway that begins with the conversion of ALA to Porphobilinogen. In successive steps, porphobilinogen is converted into Hydroxymethylbilane, Uroporphyrinogen III, Coproporphyrinogen III, Protoporphyrinogen IX, Protoporphyrin IX, and finally protohaem. Each step in the pathway is catalysed by a specific enzyme (Adapted from Panek *et al*., 2002).

1.2.2 Cytochrome P450 BM3 from *Bacillus megaterium*

BM3 is a protein isolated from the Gram-positive, rod-shaped, mainly aerobic spore-forming bacteria *Bacillus megaterium* (Vary *et al.*, 2007). BM3 contains a reductase and P450 domains within a single 119 kDa fusion protein, as such, it is classed as a structurally self-sufficient enzyme. Analysis of the amino acid sequence of BM3 determined that the enzyme was formed because of the fusion of a P450 (N-terminal) gene and a CRP (NADPH-Cytochrome P450 Reductase) (C-terminal) gene and lacks a membrane anchor region (Ruettinger *et al.*, 1989); this information led to a new classification system, prokaryotic class II P450 systems (Munro *et al.*, 2002). BM3 has been observed to have some of the highest monooxygenase activity among all P450 enzymes detailed in the literature, with the highest catalytic rate reported for the enzyme being around 285 s^{-1} when arachidonic acid is used as a substrate (Michael *et al.*, 1999). BM3 catalyses the hydroxylation of fatty acids which possess an alkyl chain of lengths ranging from 12 to 20 carbon atoms (Girvan *et al.*, 2004), typically at the ω-1, ω-2, and ω-3 positions (Jeffreys *et al.*, 2019).

1.2.3 The structure of cytochrome P450 BM3

As mentioned, BM3 is a fusion between a fatty-acid binding P450 (Haem) domain that is roughly 55 kDa and a CPR that is roughly 65 kDa; these two units are linked through a flexible interdomain linker region. The CPR domain is responsible for the binding of NADPH and passes electrons through its FAD and FMN cofactors and then to the haem iron located in the P450 domain of the enzyme (Daff *et al.*, 1997). The transfer of these electrons triggers the catalytic cycle as described above. While this information is well documented, the exact structure of BM3 is still unknown as a crystal structure of full-length BM3 is not available in either its monomeric or catalytically active state (Jeffreys *et al.*, 2020), however, the structure of the complex between the haem and FMN- binding domains have been detailed. Analysis has shown that the asymmetric units consist of two haem-domain molecules and one flavin domain; the crystal structure reveals that this domain consists of a central 5-stranded parallel β-sheet surrounded by four α-helices (Sevrioukova *et al.*, 1999). The crystal structure (2X7Y) of the NADPH binding domain of BM3 has also been elucidated (Joyce *et al.*, 2012). Recently, important protein interaction sites corresponding to haem-CPR domain interactions at the dimeric interface have been identified using hydrogen-deuterium exchange mass spectrometry, which further illuminates the structure of a catalytic mechanism of BM3 (Jeffreys *et al.*, 2020).
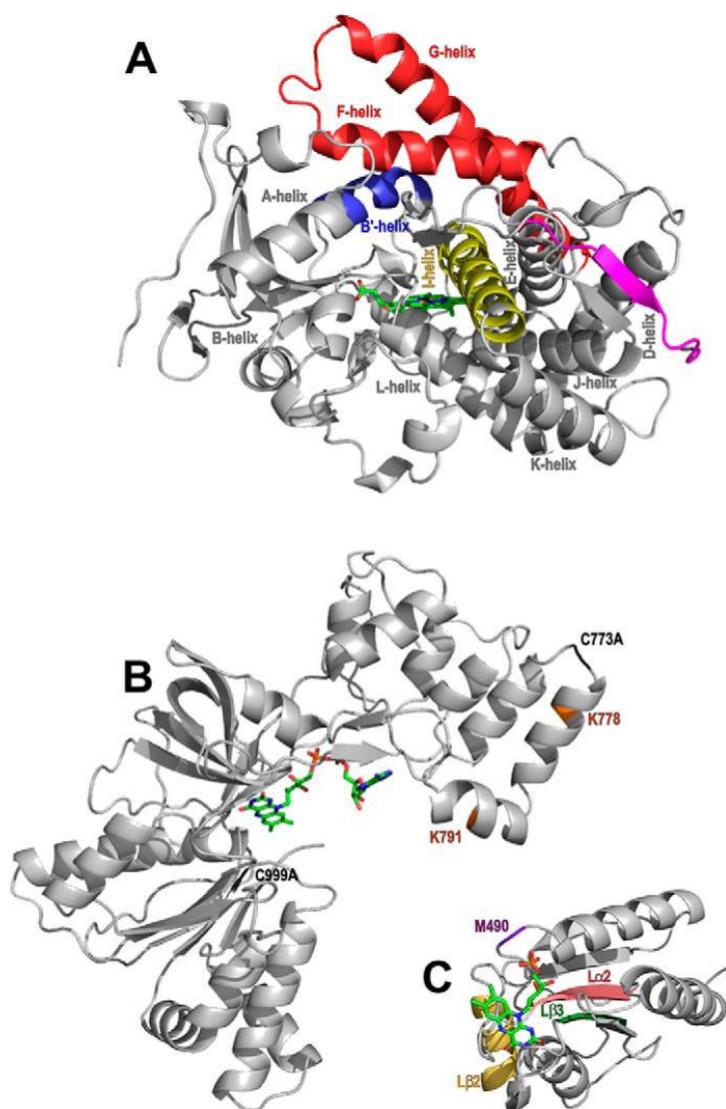
Figure 1.4 The structure and important features of P450 BM3 and its domains. The 1BU7 structure represents the catalytic heme domain of the P450 BM3 enzyme. The CRP domain of the enzyme consists of an FMN-binding domain and an FAD/NADPH-binding domain. Structures 1BVY and 4DQK show the structures of these subdomains. The heme domain of the P450 BM3 enzyme contains a heme prosthetic group shown in green which is responsible for substrate binding and oxidation of the bound substrate. The P450 fold contains highly flexible components, these being the I-helix shown in yellow and the F/G-helices shown in red. Other regions visualised in this figure are the mobile B'-helix shown in blue and the C terminus of the heme domain shown in magenta. B and C represents other selected P450 α helices. The reductase activity of the enzyme at the CRP domain is fuelled by electrons from NADPH. The

CRP domain is made up of two sub domains: the FAD/NADPH-binding domain shown in B and the FMN-binding domain shown in C. B, for successful crystallization of the FAD domain, a two amino acid mutation is required (C773A and C999A) these mutations are shown in light green these mutations also inhibit FAD domain dimer formation. Two lysine residues (Lys-778 and Lys-791) are coloured brown and are located close to the CPR dimeric interface. C shows regions of the FMN-binding (flavodoxin) domain which have been identified to play an important role in electron transfer to the heme prosthetic group, these are shown in purple, wheat, dark green and pink. (Figure and legend taken from Jeffreys *et al*., 2020).

1.2.4 Applications of cytochrome P450 BM3

The high reactivity and proficiency of BM3 coupled with the fact that the enzyme is self-sufficient (only requiring NADPH and substrate for activity) make it a prime candidate for use in the hydroxylation of many un-activated C-H bonds (Shoji, 2019) or the oxidation of several unnatural substrates. One issue with the use of BM3 to catalyse non-native substrates is that the enzyme displays a high level of substrate specificity (Carmichael *et al.*, 2001) however, the use of protein engineering or directed evolution to alter substrate specificity has yielded positive results; such as BM3 enzymes capable of hydroxylating octane at the 2-position to form S-2-octanol (Peters *et al.*, 2005), BM3 enzymes that can use testosterone as a substrate to produce 2β- and 15β- alcohols (Kille *et al.*, 2011) or enhancing the ability of BM3 to convert *p*-nitrophenoxydodecanoic acid to *p*-nitrophenol (Wong *et al.*, 2004).

BM3 mutants that have been engineered to metabolize drug-like compounds have been successfully used to activate the widely used anticancer drugs cyclophosphamide and ifosfamide. Cyclophosphamide and ifosfamide are administered as inactive prodrugs and require P450 enzyme activation to trigger the anticancer effects. The use of BM3 has yielded the fastest reported hydroxylation rate for both cyclophosphamide and ifosfamide making it a potential candidate for use in gene-directed enzyme product therapies (Vredenburg *et al.*, 2014). Similarly, engineering of BM3 through mutations of proximal ligand and other key active site residues has produced a BM3 variant with an estimated rate of 1000 turnovers per minute which can also be used under aerobic conditions capable of the enantioselective synthesis of the antidepressant levomilnacipran (Wang *et al.*, 2014).

1.3 Aims of this thesis

In this project, we aimed to question whether enzyme engineering can be facilitated using cell-free systems, which are potentially applicable to high-throughput screening. As an initial proof of concept, we sought to see if the well-characterized enzyme cytochrome P450 BM3 can be studied directly in CFPS. The production of an active BM3 enzyme will require the co-synthesis of haem b within the reaction; to show feasibility initial work was performed to create two plasmids that contain the genetic information for the enzymes of the haem synthesis pathway. This project has three main aims:

1.   CFPS of Cytochrome P450 BM3

The genomic DNA for *B. megaterium* DSM319 was used to create a pET15b expression plasmid which encodes the *bm3* gene. This plasmid will be used for T7-dependant expression of an N-terminally His_6-tagged BM3 protein. To achieve the highest yield of synthesis for the BM3 protein, an *E. coli* BL2 DE3 pLysS and *E. coli* Rossetta (DE3) pLysS crude extracts containing T7RNAP will be produced an optimised. Production of soluble BM3 will be determined via denaturing gel electrophoresis and Western blotting. Once produced, the activity of the synthesised BM3 enzymes will be tested directly from the CFPS samples.

2.   Haem biosynthesis

Free haem b is required to produce soluble and correctly folded BM3, to achieve this in a CFPS system the enzymes of the haem synthesis pathway herein referred to using the shorthand's, Hem A, B, C, D, E, F, G and H were cloned into two separate plasmids, these being HemBCD and HemEFGH. These plasmids will be transformed into *E. coli* BL21 DE3 pLysS to produce crude-extracts active for CFPS with T7RNAP and containing the haem b biosynthetic machinery. The activity of the enzymes will be determined via spectrometric assays.

Chapter 2 Materials and Methods

2.1 Materials

Table 2.1 Chemicals used in this study.

Chemical name	Supplier	Product code
1kb DNA Ladder	Promega	G571A
3-PGA	Sigma-Aldrich	P8877
5 x Green GoTaq® Flexi Buffer	Promega	M891A
AatII restriction enzyme	Thermo-Fisher	ER0991
cAMP	Sigma-Aldrich	A2252
Agar high gel strength, powder	Melford	A20020-500.0
Acetic Acid	Alfa Aesar	A/0400/PB17
Acetone, $\geq$ 99%	Alfa Aesar	A/0560/PC17
Agarose optimised grade	Melford	A20090-500.0
Ampliflu red	Sigma-Aldrich	90101
AseI Restriction Enzyme	New England Bio Labs	R0526S
NAD	Sigma-Aldrich	N6522
BamHI restriction enzyme	Promega	R602A
BSA	Sigma-Aldrich	A7906
BsaI restriction enzyme	New England Biolabs	R3535S
Calcium chloride anhydrous pills	Fluorochem	044727
Coomassie® Brilliant blue G 250	Merck	1.15444.0025
CoA sodium salt hydrate	Sigma-Aldrich	C4780
Cutsmart® buffer	New England Biolabs	B7204
D-glucose 6-phosphate dipotassium salt hydrate	Sigma-Aldrich	G73537
D-glucose	Fluorochem	094759
D-lactose monohydrate	Sigma-Aldrich	61345
dNTP	New England Biolabs	N0447S

DTT	Melford	D11000-25.0
DSMO, 100%	Thermo-Fisher	00607175
EDTA tetrasodium salt hydrate	Fluorochem	239931
FLAsH-EDT tag		
Ethanol (absolute, 99.8+ %)	Alfa Aesar	E/0650DF/17
Folinic acid	Sigma-Aldrich	F7878
Gel Loading Dye, Purple (6 x)	New England Biolabs	B7024S
Glycerol, 99%	Alfa Aesar	G/0650/17
Glycine, ≥99%	Alfa Aesar	G/0800/060
HEPES	Melford	H75030-1000.0
HisProbe-HRP Working Solution	Thermo-Fisher	15168
Imidazole	Fluorochem	021690
IPTG	Melford	156000-10.0
Iron (III) chloride hexahydrate	Fluorochem	10025-77-1
L-glutamic acid monopotassium salt monohydrate	Alfa Aesar	A17232
L-Glutamic acid hemimagnesium salt tetrahydrate	Sigma-Aldrich	49605
Laemmli SDS sample buffer, reducing (4 x)	Alfa Aesar	J60015
L-Alanine	Sigma-Aldrich	A7469
L-Arginine	Sigma-Aldrich	A5006
L-Asparagine	Sigma-Aldrich	A0884
L-Aspartic acid	Sigma-Aldrich	A8949
Lauric acid	Sigma-Aldrich	8.05333.0100
L-Cysteine	Sigma-Aldrich	C7352
L-Glutamic acid	Sigma-Aldrich	G8415
L-Glutamine	Sigma-Aldrich	G3126

L-Histidine	Sigma-Aldrich	H8000
L-Isoleucine	Sigma-Aldrich	I2752
L-Leucine	Sigma-Aldrich	L8000
L-Lysine	Sigma-Aldrich	L5501
L-Methionine	Sigma-Aldrich	M9625
L-Phenylalanine	Sigma-Aldrich	P2126
L-Proline	Sigma-Aldrich	P0380
L-Serine	Sigma-Aldrich	S4311
L-Threonine	Sigma-Aldrich	T8625
L-Tryptophan	Sigma-Aldrich	T0254
L-Tyrosine	Sigma-Aldrich	T3754
L-Valine	Sigma-Aldrich	V0500
$MgCl_2$, 25 mM	Promega	A351B
NADPH	Sigma-Aldrich	N5130
NEB ® SOC	New England Biolabs	B9035S
NdeI restriction enzyme	Thermo-Fisher	FD0583
Nickel (II) sulphate hexahydrate	Alpha Aesar	N/2000/50
NTPs solutions	Thermo-Fisher	R0481
PageRuler ™ Plus prestained Broad Range Protein Ladder	Thermo-Fisher	26619
PageRuler ™ Unstained Broad Range Protein Ladder	Thermo-Fisher	26630
PEG-8000	Promega	V3011
Potassium hydrogen phosphate	Alfa Aesar	A11321
Potassium hydroxide	Alfa Aesar	P/5600/53
Potassium phosphate monobasic	Fluorochem	094578
Pre-stained protein MW marker	Thermo-Fisher	26612
SDS	Melford	L22010-100.0

Sodium Chloride	Melford	S23020-1000.0
Spermidine	Sigma-Aldrich	85558-1G
SpeI restriction enzyme	Promega	R659A
SureCast™ acrylamide (40%)	Thermo Fisher	HC2040
SureCast™ Ammonium Persulfate (APS)	Thermo Fisher	HC2005
SureCast™ Resolving buffer	Thermo Fisher	HC2212
SureCast™ Stacking buffer	Thermo Fisher	HC2115
TEMED	Thermo Fisher	HC2006
SYBR™ safe DNA gel stain	Thermo-Fisher	S33102
T4 DNA Ligase	Promega	M180A
T4 DNA Ligase 10 x Buffer	Promega	C126A
TCEP		
Triss (hydroxymethyl) methylamine	Alfa Aesar	T/P630/60
tRNA	Sigma-Aldrich	10109541001
Tryptone, granulated	Melford	T60065-2000.0
Yeast Extract	Oxoid	LP0021
XbaI restriction enzyme	Promega	R618A

Table 2.2 Commercial kits used in this study.

Commercial Kit	Supplier	Product code
DNA Clean & Concentrator™ -5	Zymo Research	D4004
Gibson Assembly® Cloning Kit	NEB	E5510S
SuperSignal™ West HisProbe™ KIT	Thermo-Fisher	15168
Qiagen® Plasmid Plus Midi Kit	Qiagen	12945
QIAprep® Spin Miniprep Kit	Qiagen	27106
QIAquick® Gel Extraction Kit	Qiagen	28706

2.2 Equipment list

Table 2.3 Equipment used in this study.

Equipment	Supplier
12 place Microcentrifuge	SLS Lab Basics
Avanti™ J-30I centrifuge	Beckman Coulter
Benchtop Incubated Shakers, Incu-Shake MIDI	SciQuip
Cary® 50 UV-Vis Spectrophotometer	Varian Medical
CLARIOstar Plus microplate reader	BMG Labtech
Eppendorf 5417R Refrigerated Centrifuge	Merck
Eppendorf® 5800 Centrifuge Model 5810R	Merck
Eppendorf Concentrator plus	Eppendorf
FLUOstar Omega microplate reader	BMG Labtech
Innova 4300 incubator	New Brunswick Scientific
Jouan GR4.22Centrifuge	Thermo-Fisher
MIR-153 incubator	Sanyo
NanoDrop 2000c Spectrophotometer	Thermo-fisher
Plate Spinner Centrifuge	SLS flowgeg
VCX 130 ultrasonic processor	SONICS

2.3 Media and Buffers

2.3.1 Media and buffer sterilisation
All Media and buffer were sterilised by autoclaving or by filter sterilising using filters of a 0.2 µm pore size.

2.3.2 Media
All media described in table 2.4 were prepared by the addition of the listed components to the desired volume with ddH$_2$O and mixed well. Media was sterilised by autoclaving. For solid media, 1.5% (w/w) agar was added to the solution before autoclaving.

Table 2.4 Media Recipes.

Media	Media recipe and preparation
2YT	Tryptone 16 g/L, Yeast extract 10 g/L, Sodium Chloride 5 g/L
2 x YTPG media	Tryptone 16 g/L, Yeast extract 10 g/L, Sodium Chloride 5 g/L, Potassium phosphate dibasic 7 g/L, potassium phosphate monobasic 3 g/L. pH of the solution was adjusted to 7.2 using 5 M potassium hydroxide. D-Glucose was added to give a final concentration of 0.1 M in 1 litre.
Auto induction media	Solution 1: Sodium chloride 5 g/L, Tryptone 20 g/L, Yeast extract 5 g/L, Dipotassium phosphate 7 g/L, Potassium dihydrogen phosphate 3 g/L. the pH of the solution was adjusted to 7.2 using 5 M potassium hydroxide. Sugar solution: D-glucose 12.5 g/L, D-lactose 50 g/L, Glycerol 150 mL/L. the solution was filter sterilised.

	Solutions combined to give the final concentrations listed above.
Cell-Free Autoinduction media	Solution 1: Sodium chloride 5 g/L, Tryptone 20 g/L, Yeast extract 5 g/L, Dipotassium phosphate 14 g/L, Potassium dihydrogen phosphate 6 g/L. the pH of the solution was adjusted to 7.2 using 5 M potassium hydroxide. Sugar solution: D-glucose 0.5 g/L, D-lactose 4 g/L, 80% glycerol 7.5 mL/L Solutions combined to give the final concentrations listed above.
LB	Tryptone 10 g/L, Yeast extract 5 g/L, Sodium Chloride 5 g/L
Mini-Prep LB	Tryptone 10 g/L, Yeast extract 5 g/L, Sodium Chloride 10 g/L

Table 2.5 Antibiotic supplements.

Antibiotic	Supplier	Stock concentration	Working concentration
Ampicillin	Melford	100 mg/mL in ddH$_2$O	100 µg/mL
Carbenicillin	Melford	100 mg/mL in 50% Ethanol (v/v)	100 µg/l
Chloramphenicol	Melford	35 mg/mL in 70% Ethanol (v/v)	35 µg/l

Antibiotic supplements were stored at -20°C. Antibiotics were mixed with warmed agar before plates were poured and left to solidify for 30 minutes, before storage at 4°C.

2.3.3 Buffer and solution recipes

All solutions used for these experiments were prepared using ddH$_2$O.

Table 2.6 Buffer and solution recipes.

Buffer category	Buffer name	Buffer composition
Purifying His-Tag proteins	Binding Buffer (BB)	20 mM Tris pH adjusted to 8.0 using hydrochloric acid, 500 mM sodium chloride, 5 mM Imidazole
	Wash Buffer I (WBI)	20 mM Tris pH adjusted to 8.0 using hydrochloric acid, 500 mM sodium chloride, 30 mM Imidazole
	Wash Buffer II (WBII)	20 mM Tris pH adjusted to 8.0 using hydrochloric acid, 500 mM sodium chloride, 70 mM Imidazole
	Elution Buffer (EB)	20 mM Tris pH adjusted to 8.0 using hydrochloric acid, 500 mM sodium chloride, 1 M Imidazole
	Strip Buffer (SB)	20 mM Tris pH adjusted to 8.0 using hydrochloric acid, 500 mM sodium chloride, 100 mM EDTA
	Desalting buffer	50 mM Tris pH adjusted to 8.0 using hydrochloric acid, 400 mM sodium chloride, 10% glycerol V/v
Cell-free protein synthesis	S30 Buffer	14 mM magnesium acetate. 60 mM potassium acetate, 10 mM Tris-acetate and 2 mM DTT. DTT was added 15 minutes before use of buffer.
	S30 A Buffer	14 mM magnesium-glutamate, 60 mM potassium-glutamate, 50 mM Tris-HCl and 2 mM DTT. DTT was added 15 minutes before use of buffer.

	S30 B Buffer	14 mM magnesium-glutamate, 60 mM potassium-glutamate, 5 mM Tris-HCl and 2 mM DTT. DTT was added 15 minutes before use of buffer.
	Amino Acid solution	6 mM L-Alanine, 6 mM L-Arginine, 6 mM L-Asparagine, 6 mM L-Aspartic acid, 6 mM L-Cysteine, 6 mM L-Glutamic acid, 6 mM L-Glutamine, 6 mM Glycine, 6 mM L-Histidine, 6 mM L-Isoleucine, 5 mM L-Leucine, 6 mM L-Lysine, 6 mM L-Methionine, 6 mM L-Phenylalanine, 6 mM L-Proline, 6 mM L-Serine 6, 6 mM L-threonine, 6 mM L-Tryptophan, 6 mM L-Tyrosine, 6 mM L-Valine
	NTPs solution	25 mM ATP, 25 mM GTP, 12.5 mM CTP, 12.5 mM UTP.
	14 x Energy solution	700 mM HEPES buffer pH adjusted to 8, 2.8 mg/mL tRNA, 3.64 mM CoA, 4.62 mM NAD, 10.5 mM cAMP, 0.95 mM Folinic acid, 14 mM Spermidine, 420 mM 3-PGA
Western Blot	Transfer Buffer	25 mM Tris, 192 mM glycine, 20% methanol (v/v)
	TBST	25 mM Tris, 0.15 M sodium chloride, 0.05% Tween® -20 Detergent
	BSA/TBST	25 mg/mL BSA in TBST
	HisProbe-HRP Working Solution	1 µg/mL HisProbe-HRP in TBST
	SuperSignal West Pico Substrate Working Solution	1-part Luminol/Enhancer Solution to 1-part Stable Peroxide Solution

Denaturing gel electrophoresis	Resolving gel 12%	2.4 mL SureCast acrylamide (40%), 2 mL SureCast Resolving buffer, 3.42 mL distilled water, 80 µL 10% APS, 80 µL 10% SDS, 8 µL TEMED
	Stacking gel 4%	0.3 mL SureCast acrylamide (40%), 0.75 mL SureCast Stacking buffer, 1.89 mL distilled water, 30 µL 10% APS, 30 µL 10% SDS, 3 µL TEMED
	1 x SDS running buffer	25 mM Tris base, 192 mM glycine, 0.1% SDS
Agarose gel	1 x TAE buffer	40 mM Tris base, 20 mM acetic acid, 1 mM EDTA
	Agarose gel	0.25 g agarose, 25 mL TAE buffer, 2 µL SYBR safe
Enzyme activity assay	Phosphate Buffer	0.2 M potassium dihydrogen phosphate, 0.2 M disodium hydrogen phosphate in a ratio that gives a pH of 8
Estimation of protein concentration	Bradford's reagent	50 mg of Coomassie Blue G250 dissolved in 50 mL of methanol. 100 mL of 85% phosphoric acid. ddH$_2$O 1 L. Filtered to remove precipitates.

2.4 Microbial strains and plasmids

2.4.1 Bacterial strains
Bacterial strains (Table 2.7) were kindly provided by Dr Simon Moore (University of Kent)

Table 2.7 Bacterial strains.

Name	Description	Reference/Source
Escherichia coli		
BL21 DE3 pLysS	B F$^-$ *ompT gal dcm lon hsdS$_B$(r_B^- m_B^-)* λ (DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB$^+$*]$_{K-12}$($λ^S$)	Moore Laboratory
DH10β	Δ*(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14- ϕ80*d*lacZΔM15 recA1 relA1 endA1 nupG rpsL* (Str^R) *rph spoT1 Δ(mrr-hsdRMS-mcrBC)*	NEB
Rosetta (DE3) pLysS	B F$^-$ *ompT gal dcm lon? hsdS$_B$(r_B^- m_B^-)* λ (DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB$^+$*]$_{K-12}$($λ^S$) pLysSRARE[*T7p20 ileX argU thrU tyrU glyT thrT argW metT leuW proL ori*$_{p15A}$](Cm^R)	Moore Laboratory

2.4.2 Plasmids
Plasmids (Table 2.8) were kindly provided by Dr Simon Moore (University of Kent) or prepared in this work.

Table 2.8 List of plasmids.

Plasmid	Plasmid details	Antibiotics resistance	Origins
pET15b-*bm3*	T7 promoter, N-terminal His-tag, *bm3* protein, tetracysteine tag.	Ampicillin	Dr Moore
pET15b-*hemBCD*	T7 promoter, N-terminal His-tag, *hemb* protein *hemc* protein, *hemd*protein.	Ampicillin	Dr Moore

pET15b-*hemEFGH*	T7 promoter, N-terminal His-tag, *heme* protein *hemf* protein, *hemg* protein, *hemh* protein.	Ampicillin	Dr Moore
pET15b-*mScarlet*	T7 promoter, N-terminal His-tag, *mScarlet* red fluorescence protein, tetracysteine tag.	Ampicillin	Dr Moore
pTU1-SP44-*mScarlet*-Flga_CCXXCC_dBroccoli_B0015	SP44 promoter, N-terminal His-tag, *mScarlet* red fluorescence protein, tetracysteine tag.	Ampicillin	Dr Moore
pTU1-A-SP44-*bm3*-His6	SP44 promoter, N-terminal His-tag, *bm3 protein*, tetracysteine tag.	Ampicillin	This work

2.5 Microbiological techniques

2.5.1 Microbial growth and storage conditions

Unless otherwise stated, bacterial strains were grown aerobically overnight on LB agar media supplemented with appropriate antibiotics at 37°C.

2.5.2 Liquid cultures

Unless otherwise stated, overnight aerobic liquid cultures were prepared by adding a single colony from an LB agar plate to 5 mL of liquid LB media supplemented with the appropriate antibiotic. Cultures were grown overnight at 37°C and 200 rpm.

2.5.3 Spectrometric determination of cell density

For routine estimation of cell concentrations, the OD_{600} was measured to estimate the cell density of bacterial strains using the assumption that an OD_{600} of 1 equates to 1×10^9 cells per mL.

2.5.4 Glycerol stocks

Bacterial strains were stored by combining 650 µL liquid overnight culture with 350 µL of 87.5% (v/v) glycerol to give a final glycerol concentration of 30% (v/v). Samples were appropriately labelled kept on ice for 15 minutes and stored at –80°C.

2.6 Molecular biology techniques

2.6.1 Determination of DNA concentration

The concentration and purity of DNA concentration were determined by measuring the absorbance at 260 nm (A_{260}) and 280 nm (A_{280}). In a pure DNA solution, an A_{260} of 1 unit is equivalent to a concentration of dsDNA of 50 µg per mL. The purity of a DNA solution is determined by the ratio of A_{260} to A_{280}; a value between 1.8 and 2.0 indicates that the DNA can be considered pure. This process was performed using a Nanodrop 2000.

2.6.2 Mini-prep DNA purification

Were performed according to manufacturer's (Qiagen instructions.

2.6.3 Midi-Prep DNA purification

Were performed according to manufacturer's (Qiagen) instructions.

2.6.4 Zymoclean Gel DNA Recovery

Were performed according to manufacturer's (Zymo Research) instructions.

2.6.5 Agarose gel electrophoresis

Agarose gel electrophoresis was used for the analytical separation of DNA fragments. A prepared agarose gel was loaded into a tank and submerged in 1 x TAE buffer. The DNA sample was combined with a DNA loading dye (Gel Loading Dye, Purple (6 x)), and loaded into the agarose gel, an appropriate DNA ladder (1kb DNA Ladder) was also loaded. The gels were run at a constant voltage of 100 V for 30 minutes. After running, the DNA was detected under blue light.

2.6.6 PCR amplification of DNA

Oligonucleotide primers were designed for each DNA fragment of Interest to allow the amplification of the DNA sequence by PCR. Sequences coding for restriction endonucleases were inserted via these primers at either end of the DNA fragment. Oligonucleotide primers are listed in table 2.9, with the restriction endonuclease sequences coloured in red. Oligonucleotides were designed and obtained from Integrated DNA Technologies. Primers were designed to incorporate one or more restriction sites for successive cloning steps. To allow

the incorporation of a C-terminal His6-tag, the stop codon of the reverse primer was removed, and additional bases were added to keep the gene in frame.

Table 2.9 Oligonucleotides for PCR. Primary enzyme restriction sites are coloured red.

Primer Name	Sequence (5'-3')	Restriction site
BM3_Asel_F	CACATTAATATGACAATTAAAGAAATGCCTCAG	AseI
BM3_His_F	TAACTTTAAGAAGGAGATATACACAATGGGCAGCAGCCATCATC	
BM3_His_R	CTTGTCGTCGTCGTCCTTGTAGTCGTTACCCAGCCCACACGTC	
BBa_G00100 VF2	TGCCACCTGACGTCTAAGAA	NdeI cut site upstream of primer site.

For the amplification of the target DNA, PCR reactions of a total volume of 25 µL were prepared. PCR reactions used or GoTaq Polymerase. The 5 x Green GoTaq® Flexi Buffer, 25 mM MgCl and GoTaq® G2 flexi DNA polymerase was obtained from Promega. The composition of a GoTaq polymerase reaction can be seen in table 2.10.

Table 2.10 Composition of a GoTaq polymerase PCR reaction.

Component	Volume
5 x Green GoTaq® Flexi Buffer	5 µL
25 mM MgCl	2 µL
10 mM dNTPs	1 µL
10 µM forward primer	1.25 µL
10 µM reverse primer mix	1.25 µL
GoTaq® G2 flexi DNA Polymerase	0.125 µL
DNA (diluted to 0.1-10 ng/µL)	0.5 µL
DSMO	3% of the total volume
ddH2O	Fill to 25 µL

For standard PCR amplification using GoTaq polymerase, the following protocol was used (Table 2.12),

Table 2.12 PCR protocol for standard amplification using GoTaq DNA polymerase.

Process step	Temperature	Time	Number of cycles
Denaturation	95°C	2 minutes	1
Denaturation	95°C	30 seconds	30
Annealing	55°C	30 seconds	30
Elongation	72°C	1 minute per KB	30
Final Elongation	72°C	10 minutes	1
Cooling	8°C	Indefinite	1

2.6.7 Digestion of plasmid DNA with restriction enzymes

Digestion of plasmid DNA was performed using restriction enzymes purchased from Promega or Thermo-Fisher. Reaction conditions such as buffer, temperature, incubation time, and concentration of DNA and enzyme were determined from the manufacturer's instructions. Digested DNA was separated by agarose gel electrophoresis or Zymolclean gel DNA recovery.

2.6.8 DNA gel extraction

Were performed according to manufacturer's instructions.

2.6.9 Ligation of DNA- direct cloning

The ligation of purified DNA fragments with compatible ends was achieved using Promega Rapid Ligation buffer and T4 DNA ligase. The general ligation reaction consisted of a 15 µL mix which was prepared by combining the insert (0.5-1 µg) with vector (50-250 ng) in the 2 x ligation buffer (Promega) and 1-5 µL of T4 DNA ligase. A control ligation was prepared by creating a mix with the vector and no insert. If more colonies were present for the vector plus insert ligation, than the control, it was assumed that the ligation was successful.

2.6.10 Gibson Assembly Cloning

For the assembly of 2-3 fragments, PCR fragments of a concentration of 0.02- 0.5 pmols up to a volume no greater than 20% of the overall reaction were combined with 10 µL of Gibson Assembly Master Mix (2 x). The total volume of the reaction was made up to 20 µL with ddH$_2$O. The samples were incubated in a thermocycler at 50°C for 15 minutes and stored on ice before transformation into competent *E. coli* cells.

2.6.11 Production of chemically competent cell

500 µL of an overnight culture was sub-cultured (1:100 dilution) into 50 mL of LB liquid medium supplemented with appropriate antibiotics, which was incubated aerobically at 200 rpm and 37°C until an OD$_{600}$ of 0.3 was reached. The culture was placed on ice for 15 minutes before being centrifuged at 4000 rpm for 10 minutes and 4°C. Following centrifugation, the supernatant was removed, and the pellet was re-suspended in 17 mL of ice-cold freshly filter-sterilised 0.1 M calcium chloride before being placed on ice for 30 minutes. The sample was centrifuged at 4000 rpm for 10 minutes and 4°C and the supernatant was discarded; the pellet was re-suspended in 2 mL of ice-cold 0.1 M calcium chloride, 15% glycerol. 100 µL aliquots of solution were added to sterile Eppendorf's and stored at -80°C.

2.6.12 Transformation of competent cells

Component cells were defrosted on ice for 10 minutes. 20 µL of competent cells were mixed with 2 µL of plasmid DNA; before being placed on ice for 20 minutes. The solution was heat-shocked at 42°C for 40 seconds before being placed on ice for 2 minutes. 100 µL of SOC broth was added to the solution which was then incubated at 37°C for 60 minutes. A 100 µL aliquot was spread plated onto LB agar with an appropriate antibiotic and incubated overnight at 37°C.

2.6.13 DNA sequencing

Following successful cloning confirmed by restriction enzymes digest analysis, all DNA prepared via PCR was screened to check for mutations. Sequencing was performed by Eurofins.

2.7 Protein biochemistry

2.7.1 Protein extraction

An overnight aerobic culture of BL21 DE3 pLysS was prepared by adding 5-10 colonies extracted from an LB agar media plate to 5 mL of liquid auto-induction media; this solution was incubated at 200 rpm and 37°C. Post incubation, the culture was centrifuged at 4000 rpm for 10 minutes at 4°C; the supernatant was discarded. The pellet was re-suspended in 500 µL of binding buffer. The solution was sonicated at a frequency of 20 kHz and 65% amplitude for 5 seconds on 5 seconds off cycles for a total on-time of 1 minute. The solution was centrifuged at 13000 rpm for 10 minutes at room temperature and the supernatant was discarded and the solution was stored at -20°C.

2.7.2 Purifying His-Tag proteins

Following protein extraction, 30 µL of the solution (1) was aliquoted and stored on ice. The total cell fraction was centrifuged at 14000 rpm for 10 minutes at 4°C. The supernatant was pooled and 30 µL of the solution (2) was aliquoted and stored on ice. The pellet was re-suspended in 10 mL of binding buffer and 30 µL of the solution (3) was aliquoted and stored on ice. The column was washed with 10 mL of strip buffer, followed by washing with 10 mL ddH$_2$0. 5 mL of 0.1 M nickel sulphate was applied to the column before washing with 10 mL ddH$_2$0 and then 10 mL of binding buffer. The supernatant was applied to the prepared column and the flow-through was collected and 30 µL of the solution (4) was aliquoted and stored on ice. 20 mL of binding buffer was applied to the column and the flow-through was collected and 30 µL of the solution (5) was aliquoted and stored on ice. 10 mL of wash buffer I was applied to the column and the flow-through was collected and 30 µL of the solution (6) was aliquoted and stored on ice. 10 mL of wash buffer II and the flow-through was collected and 30 µL of the solution (7) was aliquoted and stored on ice. 5 mL of eluting buffer was applied to the column and 1 mL fractions were collected (8), (9), (10), (11), (12), and (13) and stored on ice. Each of the numbered fractions was analysed by denaturing gel electrophoresis and stained with Coomassie Brilliant Blue G-250 stain.

2.7.3 Buffer exchange (desalting)

Purified protein was desalted in 30MTR dialysis tubing. The tubing was pre-equilibrated in a desalting buffer. Purified protein obtained via His-tag purification was pipetted into the dialysis

tubing; the tubbing was placed in 2 L of desalting buffer at 4°C and stirred overnight. After desalting purified protein was transferred to 1.5 mL Eppendorf and stored at –20°C.

2.7.4 A$_{280}$ protein concentration estimation

Protein concentrations were estimated using the method outlined in Warburg and Christian (1942). The absorbance at 280 nm was determined and the concentration of protein was estimated from the calculated extinction coefficient for the protein using the equation:

$A_{280} =$ concentration (M) x Extinction coefficient (M^{-1} cm^{-1}) x Path length (cm)

2.7.5 Bradford assay for estimation of protein concentration

A series of 20 µL dilutions of 1 mg/mL BSA were prepared to produce a concentration curve. Each 20 µL dilution was combined with 980 µL of 1 x Bradford dye and the absorbance was determined at 595 nm, water was used as a blank. A 20 µL dilution of protein was prepared so that the OD fell between 0.4 and 0.7, thus allowing the value to fall inside the range of the BSA standard curve. The absorbance of the protein sample was measured at 595 nm; this value was converted to concentration using the equation of the standard curve.

2.7.6 Cell-free protein synthesis

2.7.6.1 *E. coli* Crude Extract Preparation

To produce a crude extract, a Rosetta (DE3) pLysS *E. coli* strain was streaked onto 2YT agar media and incubated aerobically overnight at 37°C. An overnight aerobic culture was prepared by adding a single colony extracted from 2YT agar to 600 mL of Autoinduction media, this solution was incubated at 30°C and 150 rpm until an OD$_{600}$ of 1.5-2.0 was reached. The culture was centrifuged at 5000 x g for 12 minutes at 4°C and the supernatant was removed. 200 mL of S30A buffer was added to the pellet at 4°C, the bottles were shaken until the pellet was completely solubilized and then centrifuged at 5000 x g for 12 minutes at 4°C, the supernatant was removed, and this step was repeated. 40 mL of S30A buffer was added to the pellet and shaken to re-suspend at 4°C. The pellet was transferred to a pre-chilled falcon shaken to homogenise and centrifuged at 5000 x g for 8 minutes at 4°C, the supernatant was then discarded. The pellet was re-centrifuges at 5000 x g for 2 minutes at 4°C and residual supernatant was removed. The mass of the wet cell was recorded and 1.1 volumes of S30A buffer was added per 1 gram of wet cells and mixed thoroughly. The solution was centrifuged at 200 x g for 10 seconds to accumulate the cells at the bottom of the tube. The wet cell paste

was sonicated at a frequency of 20 kHz and 65% amplitude for 5 seconds on 5 seconds off cycles for a total on-time of 5 minutes. The extract was centrifuged at 12000 x g for 10 minutes at 4°C and the supernatant was transferred to a clean Eppendorf tube and incubated at 37°C for 80 minutes; the pellet was discarded. The extract was centrifuged at 12000 x g for 10 minutes at 4°C and the supernatant was transferred to dialysis tubing that had been pre-equilibrated in S30B buffer; the tubbing was placed in 2 L of S30B buffer at 4°C and stirred for 3 hours. After 3 hours, the extract was aliquoted into clean Eppendorf tubes and centrifuged at 12,000 x g for 10 minutes at 4°C. The supernatants were pooled and inverted 4-5 times to mix, then centrifuged at 1000 x g for 10-20 seconds at 4°C. A Bradford assay as described in 2.7.5 was performed to determine extract concentration and the extract was diluted with S30B buffer to a concentration of 30 mg/mL. 50 µL aliquots of extract were snap-frozen in dry ice and stored at -80°C. Method adapted from Sun *et al* (2013).

To produce a BL21 DE3 pLysS crude extract, a BL21 DE3 pLysS E. *coli* strain was streaked onto LB agar media with no antibiotics and incubated aerobically overnight at 37°C. A single colony was extracted from the LB plate and added to 30 mL of liquid LB media with no antibiotics and incubated aerobically overnight at 37°C and 250 rpm. The overnight culture was sub-cultured into 2 x YTPG at a 1:100 ratio and incubated at 37°C and 250 rpm. Once an OD_{600} of 0.3 was reached, IPTG was added to the solution at a final concentration of 400 µM and the solution was re-incubated until an OD_{600} of 3.0 was reached. The cells were harvested by centrifuging the culture at 5000 x g at 4°C for 15 minutes. Following centrifugation, the cells were washed; 200 mL of S30A was added to re-suspend the cells and then centrifuged at 5000 x g at 4°C for 10 minutes before the supernatant was discarded. The wash step was repeated for a total of three washes. 40 mL of buffer A was added to the pellet at 4°C. The pellet was transferred to a pre-chilled falcon shaken to homogenise and centrifuged at 5000 x g for 8 minutes at 4°C, the supernatant was then discarded. The pellet was re-centrifuges at 5000 x g for 2 minutes at 4°C and residual supernatant was removed. The wet mass of the cells was measured, and buffer A was added in a ratio of 1 mL to 1 g of cell mass. The cell suspension was sonicated at a frequency of 20 kHz and 65% amplitude for 5 seconds on 5 seconds off cycles for a total on-time of 5 minutes. The extract was centrifuged at 12000 x g for 10 minutes at 4°C and the supernatant was transferred to a clean Eppendorf tube and incubated at 37°C for 80 minutes; the pellet was discarded. A Bradford assay as described in 2.7.5 was performed to determine extract concentration. 50 µL aliquots of extract supernatant were snap-frozen in dry ice and stored at -80°C. Method adapted from Sun *et al* (2013).

Another method used to produce a BL21 DE3 pLysS crude extract involved the use of Cell-Free autoinduction media. In this method a BL21 DE3 pLysS (DE3) *E. coli* strain was streaked onto LB agar media with no antibiotics and incubated aerobically overnight at 37°C. A single colony from the overnight culture was inoculated into 400 mL of Cell-Free autoinduction media and incubated at 30°C and 200 rpm for roughly 15 hours, resulting in an OD_{600} of approximately 10. The overnight cultures were centrifuged at 5000 x g at 4°C for 10 minutes, the supernatant was discarded, and the pellet was transferred to a pre-chilled 50 mL falcon. The pellet was washed using 40 mL of S30 buffer by vortexing in 30-second cycles followed by a rest period on ice. Once the pellet was fully re-suspended, the solution was centrifuged at 5000 x g at 4°C for 10 minutes and the supernatant was discarded. The wash step was repeated a total of 3 times. The pellet was re-suspended in 1 mL of S30 buffer for each gram of wet cell mass, the pellet was re-suspended by vortexing in 30-second cycles followed by a rest period on ice. Following resuspension, the cell solution was sonicated at a frequency of 20 kHz and 50% amplitude for cycles of 5 seconds on and 10 seconds off for a total on-time of 3 minutes. Immediately after sonication 3.2 µL per 1 mL of pellet of DTT was added to the solution, which was mixed via inversion. Post sonication the solution was transferred to 1.5 mL Eppendorf tubes and centrifuged at 18,000 x g at 4°C for 10 minutes. The supernatant was pooled and aliquoted in 600 mL fractions before snap-freezing in dry ice and storage at -80°C. Method adapted from Levine *et al* (2019).

2.7.6.2 Preparation of Midi-prep DNA

DNA was amplified using the Midi-prep method detailed in 2.6.3 40 µL of 3 M potassium acetate and 400 µL of isopropanol were added to the DNA and mixed by inversion to cause precipitation before the sample was refrigerated at -20°C for 30 minutes. The sample was centrifuged at 13,000 rpm for 10 minutes at 4°C and the supernatant was discarded. 1 mL of 70% ethanol was added to the pellet before centrifugation at 13,000 rpm for 1 minute at 4°C, the supernatant was discarded, and the centrifugation step was repeated to remove any residual ethanol. The pellet was to completely dry in air and then re-dissolved in 400 µL of ddH$_2$O. The DNA concentration was measured via Nanodrop and diluted to 40 nM. DNA was stored at -20°C.

2.7.6.3 Preparation of Standard Master Mix

The standard Master Mix for the Cell-Free Protein Synthesis was prepared by combining the components listed in table 2.12. This produced a total volume of 91.61 µL of Master Mix; this volume is sufficient to run 3 Cell-Free Protein Synthesis (positive and negative controls and test reaction) reactions each using 24.75 µL.

Table 2.12 Cell-Free Protein Synthesis Standard Master Mix.

Order of component addition	Volume (µL)	Component	Final Concentration
1	40.3	30 mg/mL *E. coli* extract	10 mg/mL
2	1.5	0.5 M Mg-glutamate	6 mM
3	5.3	4 M K-glutamate	175 mM
4	6.6	NTPs 25 mM	1.5 mM
5	8.7	14 x Energy Solution	1 x
6	19.3	8.4 mM Amino Acids	1.33 mM
7	9.2	40% PEG-8000 (w/v)	3% (w/v)

2.7.6.4 Preparation of Cell-Free Protein Synthesis reaction

The Cell-Free reaction was performed by combining 24.75 µL of prepared Master Mix with 8.25 µL of plasmid 40 nM DNA (final DNA concentration is 10 nM). Samples were incubated for 6 hours at 28°C and 200 rpm. For non-fluorescent samples denaturing gel electrophoresis was performed as detailed in 2.7.9 to determine if protein production had occurred. For fluorescent proteins, fluorescence readings were recorded using a FLUOstar Omega microplate reader measuring fluorescence set at an excitation and emission maxima of 569 nm and 594 respectfully and a gain setting of 1000.

2.7.7 Enzyme Activity assays

2.7.7.1 BM3 activity assays

The activity of the synthesised BM3 enzyme at a concentration of 1 µM was tested by monitoring the turnover of 100 µM lauric acid (dissolved in 70% ethanol) indirectly through the oxidation of 50 µM NADPH; this was determined using a spectrophotometer measuring the absorbance at 340 nm over a period of time. All activity assays were performed in 0.2 M phosphate buffer pH 8.

2.7.7.2 BM3 activity assays at varying concentrations of lauric acid

A series of activity assays were performed to determine the optimum substrate concentration for the reaction; to achieve this the concentration of lauric acid was varied between the ranges of 100 µM to 800 µM. The substrate turnover for each reaction was measured indirectly by monitoring the oxidation of 50 µM NADPH; it was determined using a spectrophotometer measuring the absorbance at 340 nm over a period of time. The concentration of the BM3 enzyme in each reaction was 1 µM and all reactions were performed in 0.2 M phosphate buffer pH 8. The reactions were monitored using a FLUOstar Omega microplate reader. The pipetting function of the plate reader was used to trigger the reaction by adding a solution containing enzyme and buffer to a solution containing NADPH, lauric acid, and buffer, the plate reader was set up such that the required volume of enzyme solution would be added to each reaction well before briefly shaking the plate and then beginning absorbance readings, reaction wells were measured sequentially every minute for 68 minutes

2.7.7.2 Hem enzyme activity assays

The activity of the Hem pathway enzymes was tested to determine if when combined it was possible to synthesis haem. In these reactions, the 7 purified pathway enzymes were combined so that each had a final concentration of 1 mM. The purified enzyme mix was combined with varying concentrations of ALA and iron before incubation overnight at 37 °C. Following incubation, the excitation and emission of the reaction products were monitored using a CLARIOstar Plus microplate reader.

2.7.8 Denaturing gel electrophoresis

2.7.8.1 Denaturing gel electrophoresis gel preparation

Resolving gel 12% solution was poured into a mould and topped with 1 M isopropanol and left to set. Once set isopropanol was poured off. Stacking gel 4% solution was poured into the mould on top of the resolving buffer and left to set.

2.7.8.2 Preparation of CFPS samples for SDS page

For a 33 µL Cell-free sample 1 mL of ice-cold 100% acetone is added to the sample which is then stored at -20°C for 30 minutes. Post storage, the sample was centrifuged at 13,000 rpm and room temperature for 10 minutes, and the supernatant was discarded. The pellet underwent two rounds of washing with 1 mL of ice-cold 70 % acetone; in each washing step the sample was centrifuged at 13,000 rpm and room temperature for 10 minutes and the supernatant was discarded. 30 µL of ddH$_2$O was added to the washed pellet, before 10 µL of 4 x reducing Laemmli SDS sample buffer was added, and the samples were boiled at 100°C for 5 minutes.

2.7.8.3 Preparation of CFPS FLAsH-EDT tag analysis

For a 33 µL Cell-free sample 1 mL of ice-cold 100% acetone is added to the sample which is then stored at -20°C for 30 minutes. Post storage, the sample was centrifuged at 13,000 rpm and room temperature for 10 minutes, and the supernatant was discarded. The pellet underwent two rounds of washing with 1 mL of ice-cold 70 % acetone; in each washing step the sample was centrifuged at 13,000 rpm and room temperature for 10 minutes and the supernatant was discarded. 22 µL of ddH$_2$O was added to the washed pellet, before 10 µL of 4 x reducing Laemmli SDS sample buffer and 4 µL of TCEP was added, and the samples were boiled at 100°C for 5 minutes. 4 µL of 1 mM FLAsH-EDT was added to each sample before incubation at room temperature for 30 minutes. After incubation, the samples were centrifuged at 13,000 rpm for 10 minutes. For FLAsh-EDT tag analysis 5 µL of supernatant was added to the denaturing gel electrophoresis gel.

2.7.8.4 Running denaturing gel electrophoresis gels

Denaturing gel electrophoresis gels were loaded with the protein of interest at varying volumes; an appropriate ladder (PageRuler ™ Unstained Broad Range Protein Ladder) was also added. The gel was run at 200 V for 10 minutes while the sample migrated through the

stacking gel, before the voltage was switched to 150 V and run for around one hour or until the dye had reached the bottom of the resolving gel. Gels were stained with Coomassie Brilliant Blue G-250 stain.

2.7.9 Western Blotting

Western blots were performed using a SuperSignal™ West HisProbe™ KIT. Protein samples were run on denaturing gel electrophoresis gels as described in 2.7.9 with the exception being the use of an appropriate pre-stained ladder (PageRuler ™ Plus prestained Broad Ranged Protein Ladder). For Western blotting two replica denaturing gel electrophoresis gels are required for comparison. The filter paper, fibre pads, and gel are soaked in transfer buffer to allow saturation whilst the nitrocellulose membrane was equilibrated in methanol for 10 seconds, rinsed in ddH$_2$0, and soaked in transfer buffer. One of the fibre pads was placed in an open cassette holder and coated with transfer buffer to ensure complete saturation. Next, filter paper, gel, nitrocellulose membrane, filter paper, and the second fibre pad are placed into the cassette, each of these components was saturated with transfer buffer and all air bubbles were removed using a roller. The cassette was closed and loaded into a tank filled with transfer buffer, before being run at 100 V for 1 hour. For analysis of the Western blot, the nitrocellulose membrane was blocked with 10 mL BSA/TBST overnight at room temperature while shaking. The membrane was washed twice with 15 mL of TBST for 10 minutes. The blot was incubated with 10 mL HisProbe-HRP Working Solution for 1 hour at room temperature while shaking. The membrane was washed four times with 15 mL of TBST for 10 minutes. The blot was incubated with 7.5 mL of SuperSignal West Pico Substrate Working Solution and loaded onto a gel imaginer.

Chapter 3 Optimization of Cell-free Protein Synthesis

3.1 Optimization of Cell-free reaction

Variation in each batch of crude extract means that to facilitate the greatest yield of target protein it is necessary to refine the composition of a CFPS reaction. There are numerous variables in a cell-free reaction that can be modified to achieve greater protein production; as such, we optimised the CFPS of mScarlet to act as a reference protein (i.e., can be quantified easily) before further BM3 experiments. mScarlet is a fluorescent protein meaning it is possible to monitor the cell-free production by measuring relative fluorescence units (RFU) at excitation and emission maxima of 569 nm and 594. The amount of fluorescence produced is related to the concentration of protein. Fluorescence also allows the identification of properly folded protein, whereas a gel would only indicate the presence of produced protein.

In these experiments we sought to modify the standard CFPS reaction makeup as detailed in 2.14; we began by determining a baseline for mScarlet production in cell-free and generating a calibration curve from which we could estimate protein concentration from RFU values. We performed a series of reactions to identify which *E. coli* strain can be utilised to produce a crude extract with the highest yield of protein production; in this process, we also experimented with several production methods and concentrations to assess their impact on protein synthesis. Next, we sought to identify which promoter (constitutive or inducible) was best for strong protein synthesis, here we used both pTU1-SP44-*mScarlet* and pET15b-*mScarlet* as the source of plasmid DNA for the production of mScarlet plasmids of which can be seen in Figure 3.1 found in the appendix. We assessed the individual components of a cell-free reaction system, we aimed to determine the optimal concentration of magnesium and potassium glutamate for protein production testing them both separately and in tandem. Finally, the addition of D-ribose and maltodextrin in the reaction was evaluated, research by Garenne *et al*, (2021), showed that the addition of these chemicals to a CFPS reaction can cause a drastic increase in protein yield.

3.1.1 CFPS of mScarlet
Initial CFPS tests were performed to show the synthesis of mScarlet and determine a baseline for protein yield in Cell-Free. To test this an *E. coli* Rosetta (DE3) pLysS extract and pTU1-SP44-*mScarlet* were used to produce mScarlet, the reaction composition was detailed in table 2.12. This plasmid was chosen as it contained a constitutive promoter so the reaction would not require supplementation with T7RNAP. In addition, the plasmid-encoded the mScarlet protein, which is fluorescent to estimate relative protein production. To determine the

concentration of produced protein from the fluorescence value a calibration curve (figure 3.2 of the appendix) was prepared by combining known concentrations of purified mScarlet protein with *E. coli* Rosetta (DE3) pLysS crude extract and 700 mM HEPES buffer pH adjusted to 8; the samples were incubated at room temperature for 10 minutes before fluorescence readings were taken. The calibration curve produced from this data was used to estimate the mScarlet protein concentration produced in CFPS reactions. As the concertation of protein increased as did the RFU value.

Production of mScarlet was monitored over a 6-hour incubation period at 30°C using a FLUOstar Omega microplate reader measuring fluorescence every ten minutes at excitation and emission maxima of 569 nm and 594 respectfully and gain setting of 1000 (figure 3.3). In this test reaction, a negative control was prepared in which plasmid DNA was replaced with water. Over the 6 hours, an increase in fluorescence was observed beginning at the 50-minute mark, from here the increase in fluorescence is linear until the 200-minute mark, where the reaction begins to stall. During these readings the maximum fluorescence value recorded was 1400 RFU. Minimal background fluorescence (RFU = 50) was observed in the negative control reaction. Following the CFPS reaction to produce mScarlet denaturing gel electrophoresis was performed to positively confirm the synthesis of the protein (figure 3.4); to identify the production of mScarlet the CFPS negative control reaction was also loaded onto the gel (lane 2), along with dilution of purified mScarlet protein of an initial concentration of 181 nM (lanes 5, 6 and 7). When observing lane 3 of the gel a band can be observed at a similar size to the purified proteins loaded on lanes 5, 6 and 7 suggesting the production of the mScarlet protein.
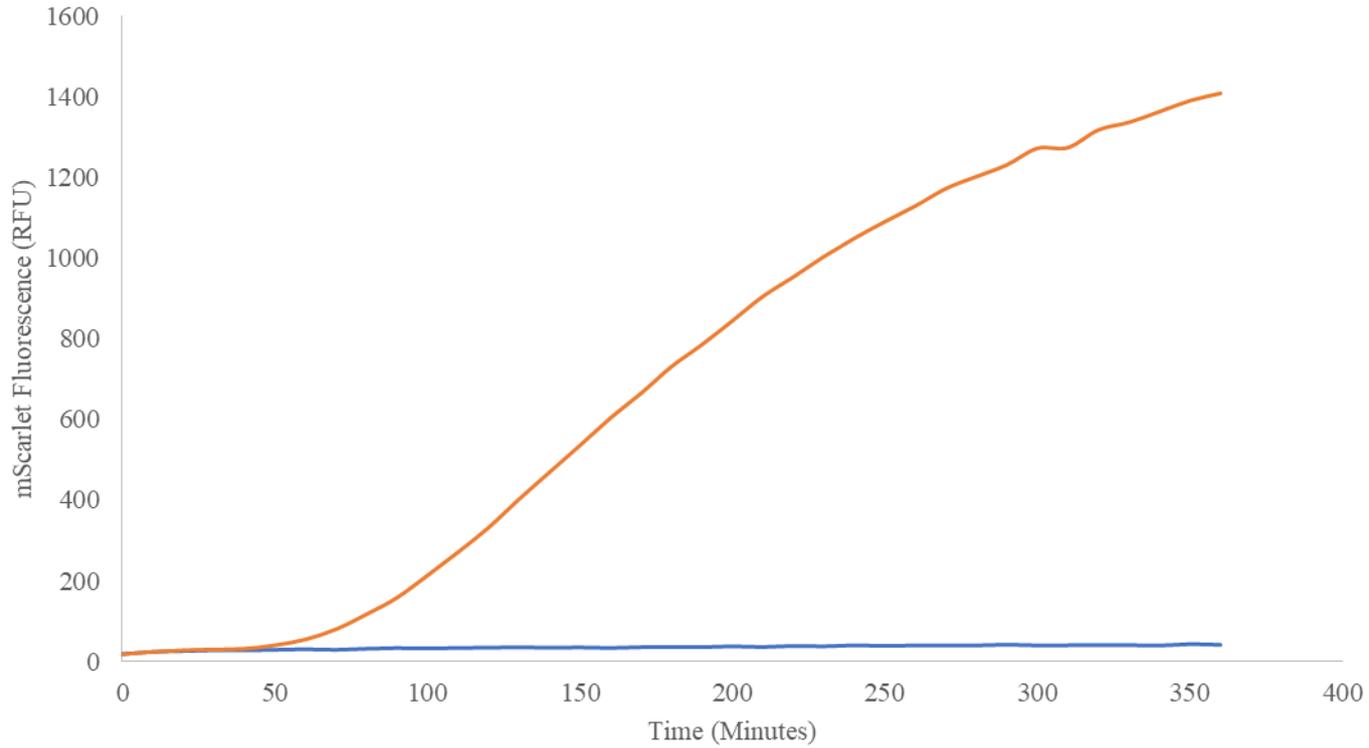
Figure 3.3 Production of mScarlet via CFPS in a time-course reaction. A CFPS reaction was performed using an *E. coli* Rosetta (DE3) pLysS crude extract and pTU1-SP44-*mScarlet-*Flga_CCXXCC_dBroccoli_B0015 (Redline) and negative control with no DNA (Blue line). The reaction was prepared in a 96-well plate and incubated in a FLUOstar plate reader at 30°C for 6 hours with fluorescence readings taken every ten minutes at an excitation and emission maxima of 569 nm and 594.
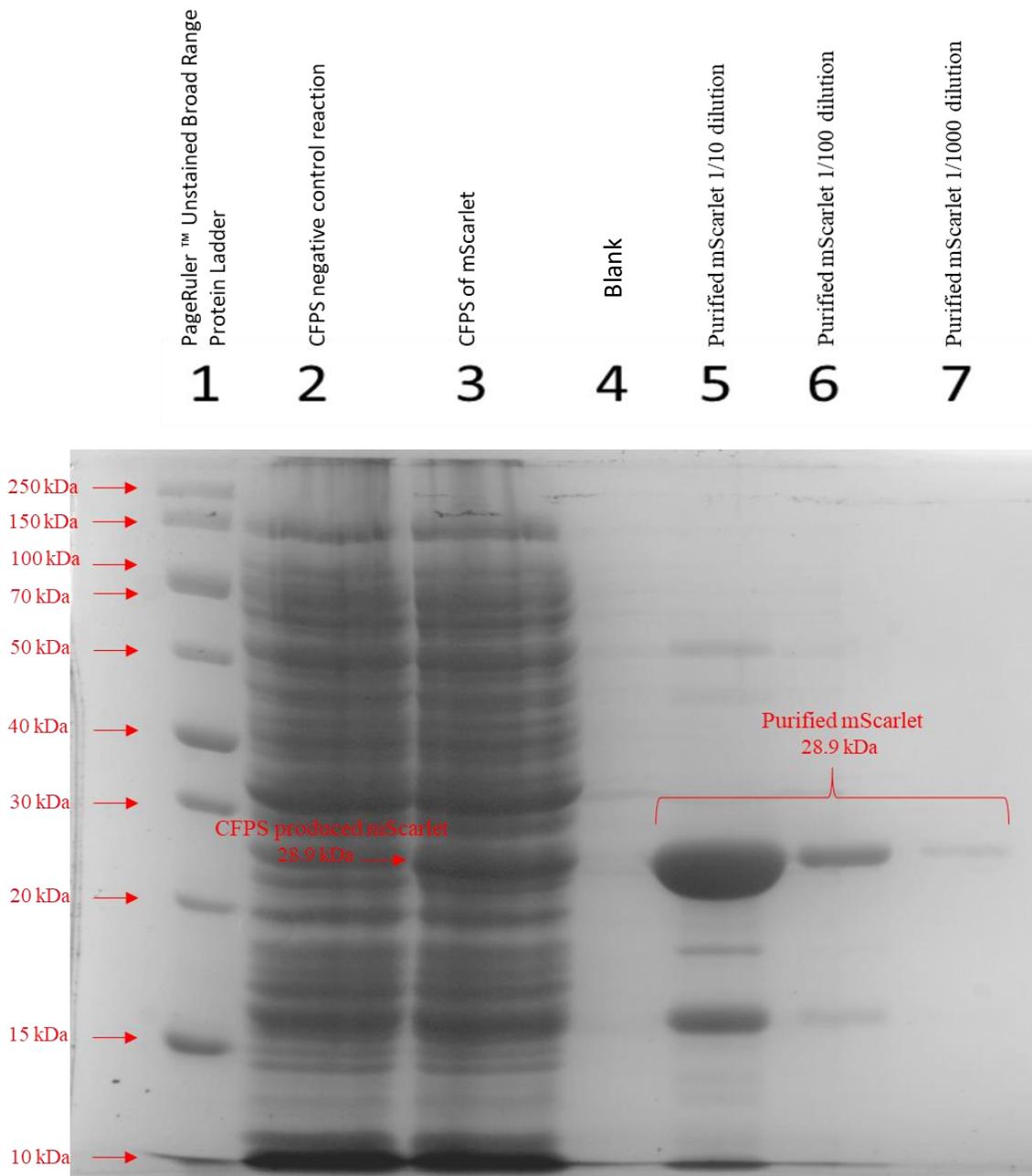
Figure 3.4 Cell-Free protein synthesis of mScarlet protein. Denaturing gel electrophoresis of mScarlet produced via CFPS using pET15b-*mScarlet* and an *E. coli* Rosetta (DE3) pLysS crude extract. Gel also shows dilutions of purified mScarlet (original concentration 181 nM). Lanes are as follows: 1- PageRuler ™ Unstained Broad Range Protein Ladder, 2- CFPS negative control reaction, 3- CFPS of mScarlet, 4- Blank, 5- Purified mScarlet 1/10 dilution, 6- Purified mScarlet 1/100 dilution, 7- Purified mScarlet 1/1000 dilution. The size of the mScarlet protein is 26.4 kDa without Histag which adds 2.5 kDa to protein size.

3.1.2 Determination of optimum *E. coli* strain for preparation of crude extract.

To determine which *E. coli* strain was optimal for protein production, crude extracts were prepared from both *E. coli* Rosetta (DE3) pLysS and BL21 DE3 pLysS; the activity of these extracts was compared to see which resulted in the greatest yield of protein. Rosetta (DE3) pLysS was chosen as it provides enhanced rates of protein production and has rare codons, as the *bm3* gene has a 44.8% G+C content. The codon pool of *E. coli* is ~52% GC. Therefore, we anticipated the addition of rare codons would aid the reaction. BL21 DE3 pLysS was chosen as a protein production host, due to its strong ability to produce recombinant proteins and the presence of the pLysS plasmid, which encodes three rare codons, these being AGA/AGG_{Arg}, AUA_{Ile} and CUA_{Leu} (Rosano and Ceccarelli, 2009). These reactions were performed as detailed in table 2.12 with the crude extract being the only source of variance, the final concentration of plasmid DNA in each reaction was 10 nM. The samples were incubated for 6 hours at 30°C and 200 rpm. The fluorescence data for these reactions can be seen in figure 3.5. In comparison, the prepared *E. coli* BL21 DE3 pLysS extract produced a greater fluorescence reading, which corresponds to a greater concentration of protein synthesized during the CFPS reaction When comparing the fluorescence values of the two extracts with the standard curve it can estimate that the BL21 DE3 pLysS extract produced 9.31 µM of mScarlet protein, whereas the Rosetta (DE3) pLysS extract produced 7.04 µM, this means that BL21 DE3 pLysS produced 1.3 times the concentration of protein. T-Test analysis comparing the two extracts yielded a p-value of 0.000262 which is statistically significant.
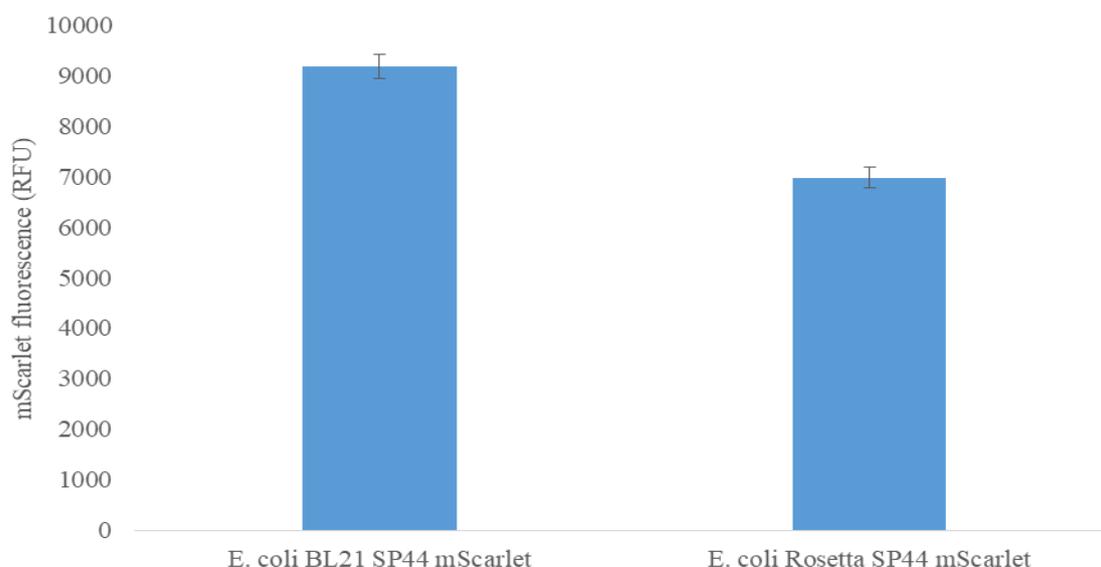
Figure 3.5 Determination of optimum *E. coli* strain for CFPS of mScarlet. Crude extracts were prepared using both *E. coli* Rosetta (DE3) pLysS and *E. coli* BL21 DE3 pLysS. Reactions were performed using the standard master mix solution detailed in table 2.12. Fluorescence readings were performed following 6 hours of incubation at 30ºC and 200 rpm. Three readings were taken for each reaction and an average fluorescence value was calculated; error bars represent one standard deviation. Samples were completed in triplicate with the mean RFU calculated, errors bars show one standard deviation.

3.1.3 Determination of optimum method for preparation of *E. coli* BL21 DE3 pLysS crude extract

Once the optimum *E. coli* strain for crude extract production was determined, several methods were trialled to produce a crude extract that achieved the highest protein yield. Two main methods to produce an *E. coli* BL21 DE3 pLysS crude extract were tested, these being the preparation method detailed in Sun *et al* (2013) and the method outlined by Levine *et al* (2019). For the Sun *et al* (2013) method, the crude extract is prepared using an *E. coli* strain grown in 2 x YTPG to an optical density of 1.5-2.0 at 600 m, (corresponding to the mid-log growth phase). Once at the desired optical density the cells are pelleted at 5000 x g for 12 min at 4 ºC and the supernatant is removed before 200 mL of S30A buffer (10.88 g Mg-glutamate and 24.39 g K-glutamate, 50 ml 2M Tris, acetic acid (pH 7.7), water to 2 L add 4 ml 1 M DTT before use) is added and the pellet was shaken until completely solubilized followed by centrifugation at 5000 g for 12 minutes and 4 ºC, this process is repeated. 40 ml of S30A buffer is added and the solution is subjected to several centrifugation steps before S30A buffer is

added at a 1:1 ratio to pellet weight and the solution is vortexed until homogenous. The samples are sonicated for an on-cycle of 5 seconds and then followed by a 5-second off-cycle for a total on-time of 3 minutes at a frequency of 20 kHz and 65% amplitude to lyse the cells. After sonication, the supernatant is separated from the pellet via centrifugation and incubated at 200 rpm, 37 °C for 80 minutes. After incubation, the solution is centrifuged to remove any residual debris before being loaded into dialysis tubing and submerged in S30B buffer (10.88 g Mg-glutamate and 24.39 g K-glutamate, 50 ml 2M Tris, acetic acid (pH 7.7), water to 2 L and 2 ml 1 DTT added before use) stirring, at 4 °C for three hours. Post dialysis, the total protein concertation of the extract is quantified and stored in aliquots. The Levine *et al* (2019) method begins with growing an *E. coli* strain in LB media overnight and then sub-culturing to 2 x YTPG warmed to 37 °C supplemented with 0.4 M D-glucose to give a starting optical density of 0.1. Once an optical density of 0.6 is reached 1 mL of 1M Isopropyl β-d-1-thiogalactopyranoside is added and the cells are grown to an optical density of 3.0. The cells are then pelleted and subjected to several wash steps with S30A buffer (14 mM magnesium-glutamate, 60 mM potassium-glutamate, 50 mM Tris and 2 mM DTT. DTT was added 15 minutes before use of buffer.) before 1 mL of S30A buffer is added for each gram of cell mass pellet and the solution is vortexed until homogenous. The pellet is then lysed via sonication for 3 cycles of 45 seconds on and 59 seconds off at an amplitude of 50% delivering a total of 800-900 J of energy. Immediately after sonication, 4.5 µL 1 M DTT is added to the lysed cells which are subsequently centrifuged, with the supernatant removed and pooled and stored in aliquots.

For the extracts prepared using the Levine *et al* (2019) method the effect of a run-off reaction, which is not included in the preparation method was assessed to evaluate if any significant difference in protein production would be observed. A run-off reaction is a process of incubation of the extract post-cell lysis at 37°C for 80 minutes. The run-off reaction is performed to allow ribosomes to process leftover mRNA transcripts and exonucleases to degrade genomic DNA. In addition, metabolic reactions proceed to completion (Silverman *et al*., 2019). These metabolites are either left in the reaction or removed by dialysis in fresh S30 buffer.

Following the production of the varying crude extracts, CFPS reactions were performed using each extract. In these reactions, mScarlet was produced using pTU1-SP44-*mScarlet*-Flga_CCXXCC_dBroccoli_B0015 as the source of plasmid DNA and the standard cell-free

reaction as detailed in table 2.12. Each reaction was incubated for 6 hours at 30°C and 200 rpm. The fluorescence data for these reactions can be seen in figure 3.6.

When comparing the data for the crude extracts produced using the different methods a substantial difference can be observed when a run-off reaction is implemented as in Levine et al (2019) and omitted as in Sun et al (2013). When performing CFPS with the extract subjected to the 80-minute run-off reaction a greater fluorescence value is obtained, this value corresponds to an estimated protein concentration of 1.83 µM, while the reaction without a run-off yields an estimated protein concentration >0.098. T-Test analysis of the two extracts produced p-value <0.00001 is calculated; therefore, there is a statically significant difference.

When performing CFPS with a BL21 DE3 pLysS extract prepared using the Sun *et al* (2013) method an estimated protein concentration of 1.83 µM is achieved, when this value is compared to the Levine *et al*. method which produced an estimated concentration of 20 µM a value that is 11 times less than that produced by the Sun *et al* (2013) method. There is a substantially greater yield of protein from the Sun *et al* (2013) methodology for crude extract preparation. Using a T-Test to analyse the two extracts, a p-value of <0.00001 is calculated meaning there is a significant difference in fluorescence produced when using the Sun *et al* (2013) method.
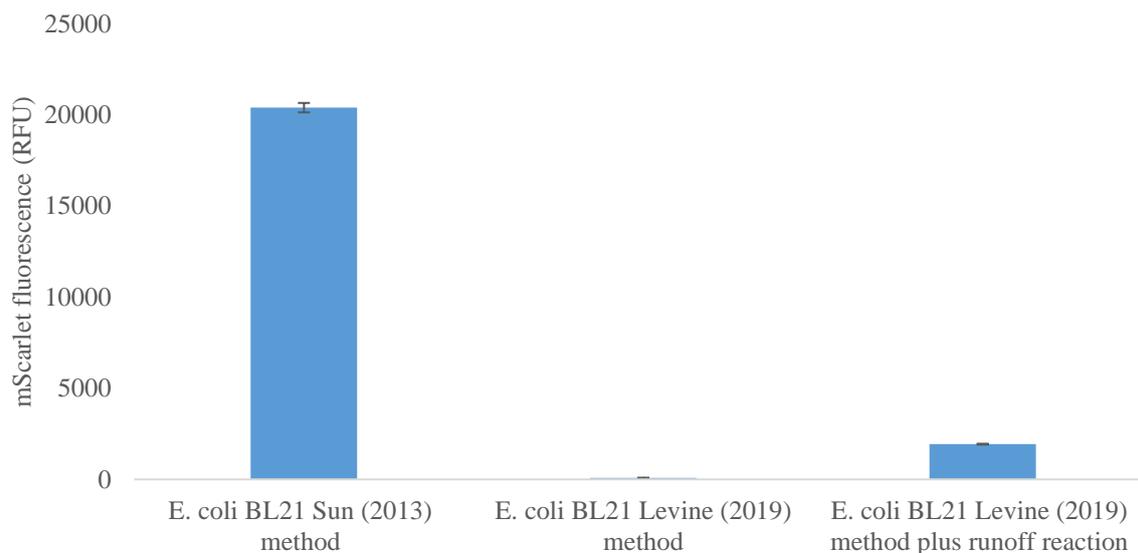
Figure 3.6 a comparison of *E. coli* BL21 DE3 pLysS crude extract preparation methods.
Varying crude extracts were prepared using the methods detailed by Sun *et al* (2013) and
Levine *et al*. (2019). For the Levine *et al* method an extract was prepared with and without a
run-off reaction. Postproduction a CFPS reaction was performed with each extract in which
mScarlet was synthesised. Fluorescence readings were performed following 6 hours of
incubation at 30°C and 200 rpm. Samples were completed in triplicate with the mean RFU
calculated, errors bars show one standard deviation.

3.1.4 Determination of optimum Plasmid vector for protein production.

Once the optimum *E. coli* strain to produce a crude extract was determined, the plasmid vector
which produced the greatest protein yield of mScarlet was determined. A vector containing the
constitutive promoter, this type of promoter does not require activation and expresses mRNA
at a constant rate. The other gene used was mScarlet (pTU1-SP44-*mScarlet*) was compared
with a vector T7RNAP-dependent promoter and *mScarlet* gene (pET15b-*mScarlet*). The
pTU1-SP44-*mScarlet* was chosen as it contains a strong constitutive promoter, pET15b-
mScarlet provides a T7 dependent inducible promoter, to provide, two options for protein
production to test for mScarlet synthesis. Both were compared to determine the optimal
strategy. Reactions were performed using the standard master mix detailed in table 2.12 and
using an *E. coli* BL21 DE3 pLysS crude extract. The samples were incubated for 6 hours at
30°C and 200 rpm. The fluorescence data for these reactions can be seen in figure 3.7.

Analysis of the fluorescence data yielded an estimated protein concentration of 9.31 µM when using pTU1-SP44-*mScarlet* as the source of DNA, and an estimated 3.05 µM when pET15b-*mScarlet* is used. This means that using pTU1-SP44-*mScarlet* yields 3 times the concentration of protein than pET15b-*mScarlet*. Using a T-Test to analyse the data for the two DNA sources, a p-value of < 0.0001 is calculated, meaning a statistically significant increase in protein production when using the pTU1-SP44-*mScarlet* plasmid.
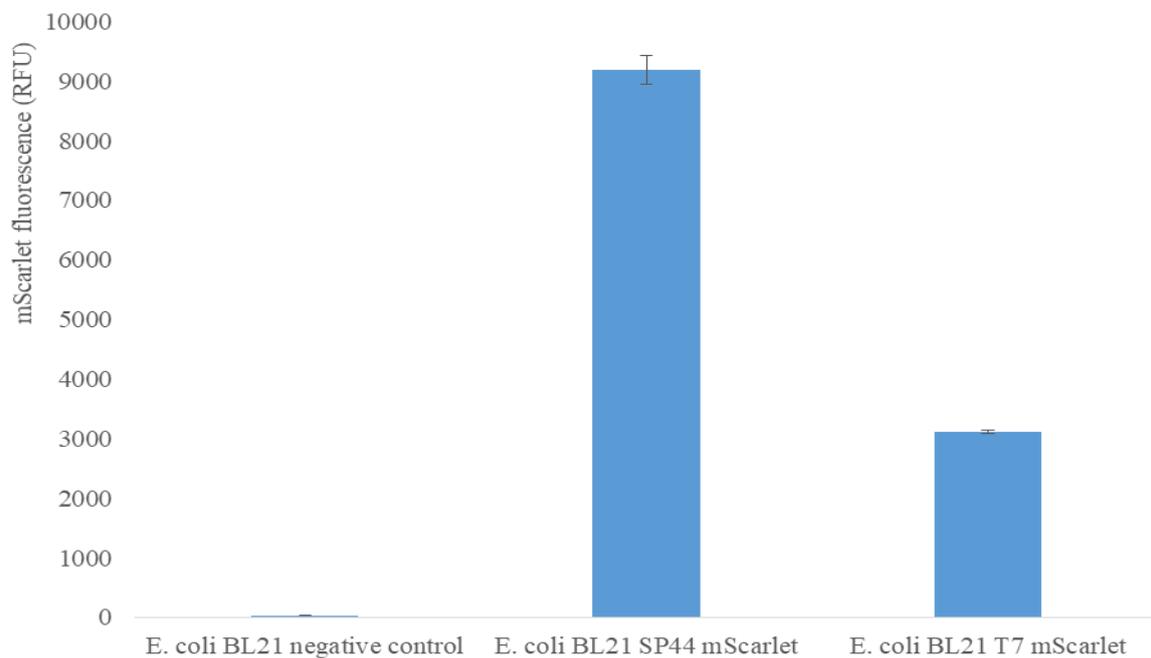
Figure 3.7 Determination of optimum plasmid for CFPS of mScarlet. Reactions were performed using the standard master mix solution detailed in table 2.12. Fluorescence readings were performed following 6 hours of incubation at 30°C and 200 rpm. Three readings were taken for each reaction and an average fluorescence value was calculated Samples were completed in triplicate with the mean RFU calculated, errors bars show one standard deviation.

3.1.5 Optimisation of crude extract concentration

To determine what concentration of crude extract achieved the maximum protein yield, a series of reactions were performed using varying crude extract concentrations, the crude extracts in this test were prepared using the method outlined by Sun *et al* (2013). In these reactions, the desired extract concentration was achieved by diluting the crude extract stock with S30 A

The concentration of the crude extract is measured in mg/mL which is an estimate of the total protein concentration. Once diluted to the desired concentration the crude extract was combined with the other components of the CFPS master mix as detailed in table 2.12 all of which were at their standard concentration. A total of six reactions were performed at different crude extract concentrations with one reaction having no extract present to act as a negative control. The final concentration of mScarlet plasmid DNA in each reaction was 10 nM. The samples were incubated for 6 hours at 30°C and 200 rpm. The fluorescence data for these reactions can be seen in figure 3.8 A

When analysing the fluorescence data produced the greatest fluorescence is observed at a crude extract concentration of 10.3 mg/mL, this fluorescence value corresponds to an estimated protein concentration of 0.92 µM. This value is 2.63 times larger when compared to the second-highest estimated protein yield concentration of 0.35 µM produced by the extract at a concentration of 7.72 mg/mL. T-Test statistical analysis for the two extract concentrations which yield the highest concentration of protein produced a p-value of <0.00001 which means there is a statistically significant increase in protein production when using an extract at a concentration of 10.3 mg/mL.

3.1.6 Optimisation of plasmid concentration

The optimum concentration of plasmid DNA which yielded maximum protein production was determined by performing a series of reactions at varying DNA concentrations; all other components of the reaction were kept the same as detailed in Table 2.12. The plasmid DNA at the desired concentration was added to an Eppendorf tube and dried using a vacuum centrifuge; before a CPFS reaction was performed using the standard reaction master mix detailed in table 2.12. The fluorescence data for these reactions can be seen in figure 3.8 B

The data shows that as DNA concentration increases so does the fluorescence reading. The greatest fluorescence reading was observed at a DNA concentration of 20 nM, this corresponded to an estimated protein concentration of 15.70 µM. We this value was compared to the other protein concentrations it was observed to be markedly greater, the second-highest protein production was observed at a DNA concentration of 15 nM which produced an estimated 12.55 µM of protein, a value that is 1.3 times lower than that seen at 20 nM of DNA. Using a T-Test to analyse the data for the DNA concentrations which yield the highest

concentration of protein, a p-value of 0.029751 is calculated, which means there is a statistically significant increase in protein production when the concentration of plasmid DNA is 20 nM. We also compared the concentration of protein produced at 20 nM with that produced at 10 nM, this being an estimated 7.75 µM which is two times lower than the value observed at 20 nM, is because 10 nM is the standard DNA concentration outlined in most CFPS protocols. A T-Test produced a p-value of 0.000507, meaning there is a statistically significant increase in protein production at 20 nM of DNA when compared to 10 nM.

3.1.7 Optimisation of magnesium and potassium glutamate concentration

The optimum concentrations of potassium-glutamate and magnesium-glutamate were determined by performing a series of reactions in which the concentrations were altered. In an initial test, the concentration of magnesium glutamate was varied by adding the desired concentration to sterile Eppendorf and drying down the solution using a vacuum centrifuge (figure 3.8 D). For varying concentrations of potassium glutamate, a series of dilutions was performed so that the desired potassium concentration would be 4 times the required value, meaning that when a volume of 8.25 µL (1/4 of the reaction volume) was added, the final reaction concentration would be that of the desired value (figure 3.8 E). Finally, magnesium glutamate and potassium glutamate concentrations were tested in tandem, in these reactions the desired concentration of magnesium-glutamate was added to a sterile Eppendorf and dried using a vacuum centrifuge, and the desired concentration of potassium glutamate was achieved by adding 8.25 µL (1/4 of the reaction volume) of solution four times the required value, which would result in a final reaction concentration at the desired value, the result of this can be seen in (figure 3.8 F). All other components of the reaction were kept the same as detailed in Table 2.12 with Mg-glutamate and K-glutamate replaced with ddH$_2$0. pTU1-SP44-*mScarlet*-Flga_CCXXCC_dBroccoli_B0015 DNA was added to each reaction and dried down in a vacuum centrifuge, the final concentration of plasmid DNA in each reaction was 10 nM. The samples were incubated for 6 hours at 30°C and 200 rpm.

When Magnesium glutamate concentrations are varied, the fluorescence data for these reactions can be seen in figure 3.8 C, no overall trend can be observed; the highest fluorescence reading was observed at a concentration of 10 mM, this RFU value corresponds to an estimated protein concentration of 0.92 µM. This value is 2.7 times greater than the second-highest estimated protein concentration of 0.34 µM observed at a magnesium glutamate concentration of 2 mM. A T-Test analysis of data for the magnesium glutamate concentrations which yield

the highest concentration of protein produced a p-value of 0.000206 is calculated which is statistically significant. The highest estimated protein yield value of 0.92 µM was also compared with the standard reaction magnesium glutamate concentration of 6 mM, which produced an estimated protein concentration of 0.07 µM, this value is 13.1 times smaller than the value estimated for a magnesium glutamate concentration of 10 mM. To determine if there was a significant difference in the values, a T-test was performed for the values of 10 mM and 6 mM, which produced a p-value of 0.000031, which is statistically significant.

As the concentration of potassium glutamate in the CFPS reaction increased so did the fluorescence values observed this is seen in figure 3.8 D, this increase reached its maximum at 175 mM where an estimated 0.8 µM of protein was produced, the fluorescence then began to decrease at the subsequent concentration of 200 mM where an estimated 0.8 µM of protein was produced. A T-test performed for the values for 175 mM and 200 mM, produced a p-value of 0.611575, which is statistically significant.

When magnesium glutamate and potassium glutamate levels were tested in tandem the greatest fluorescence readings were observed at higher magnesium glutamate concentrations, this is seen in figure 3.8 E. In each set of reactions, a magnesium glutamate concentration of 10 mM produced the greatest RFU value, except for in the reaction where potassium glutamate levels were 200 mM, here the magnesium glutamate concentration which produced the highest fluorescence reading was at 8 mM. The fact that the highest fluorescence was almost always observed at a magnesium concentration of 10 mM was consistent with the findings observed in the single magnesium test. When observing the data regarding potassium glutamate levels the highest readings are typically observed at a concentration of 175 mM, when comparing values that are at the same magnesium glutamate concentration. Except in the cases of the samples at concentrations of 100 mM potassium glutamate/ 8 mM magnesium glutamate, 175 mM potassium glutamate/ 8 mM magnesium glutamate, and 200 mM potassium glutamate/ 8 mM magnesium glutamate, here there is a drop in fluorescence when the potassium glutamate level increases from 100 mM to 175 mM and then an increase again as the value rises to 200 mM.

The greatest fluoresce is observed at 175 mM potassium glutamate/ 10 mM magnesium glutamate, this corresponds to an estimated protein concentration of 1.31 µM, these values also produced the highest protein concentration when tested independently. The second-highest

reading is seen at 200 mM potassium glutamate and 8 mM magnesium glutamate, here an estimated 0.97 µM of protein was produced. The two highest fluorescence readings were compared using a T-test, which resulted in a p-value of 0.174329, which is not statistically significant. The estimated protein concentration for the standard reaction at 175 mM potassium glutamate/ 6 mM magnesium glutamate, a value of 0.067 µM, was also compared with the highest estimated protein concentration. This estimated concentration was found to be 19.6 times less than that observed at 175 mM potassium glutamate/ 6 mM magnesium glutamate. A T-test, comparing the values produced a p-value of 0.003317, which is statistically significant.

A. Bar chart showing mScarlet fluorescence (RFU) versus Crude Extract concentration (0 mg/ml, 7.72 mg/ml, 10.3 mg/ml, 12.87 mg/ml, 18.03 mg/ml, 23.18 mg/ml, 28.3 mg/ml).

B. Bar chart showing mScarlet fluorescence (RFU) versus DNA concentration (0 nM, 1 nM, 2 nM, 5 nM, 10 nM, 15 nM, 20 nM).

C. Bar chart showing mScarlet fluorescence (RFU) versus Magnesium glutamate concentration (2 mM, 4 mM, 6 mM, 8 mM, 10 mM, 12 mM).

D.

E.
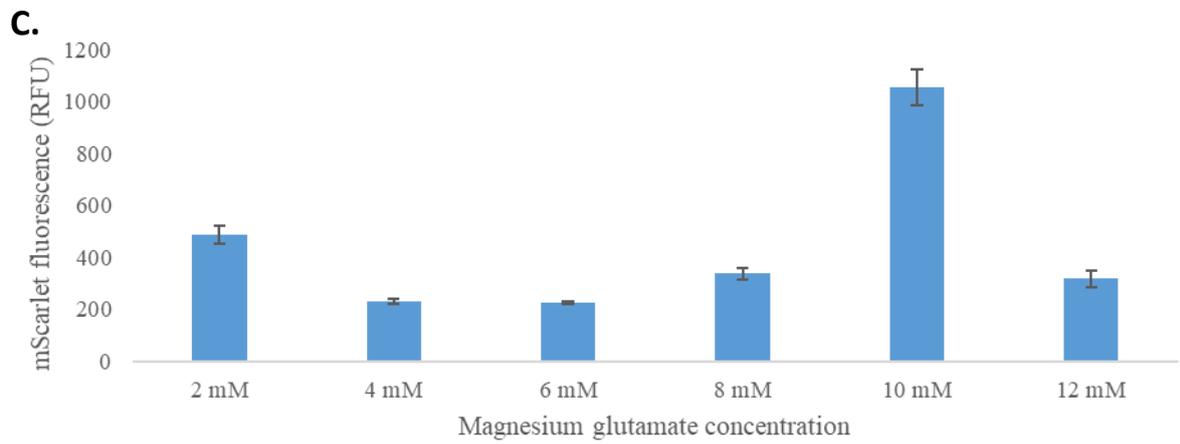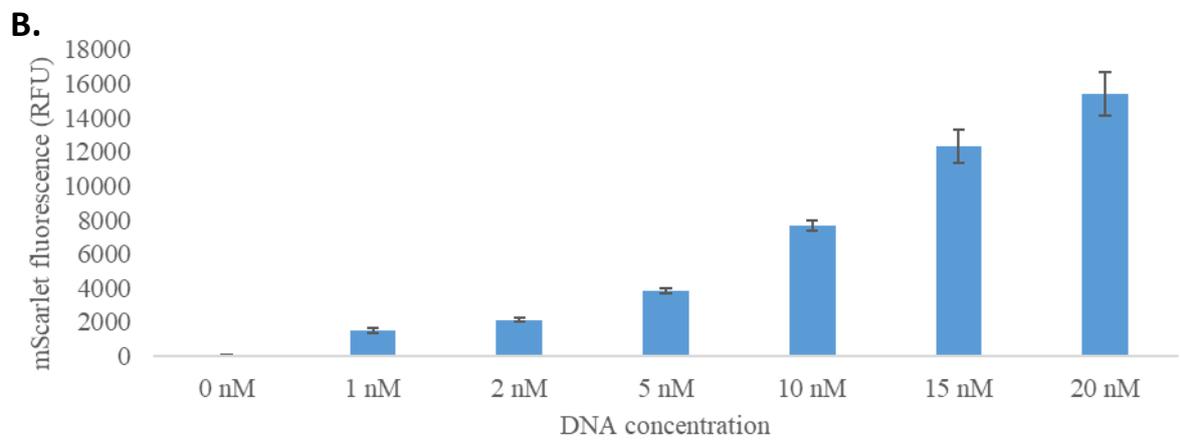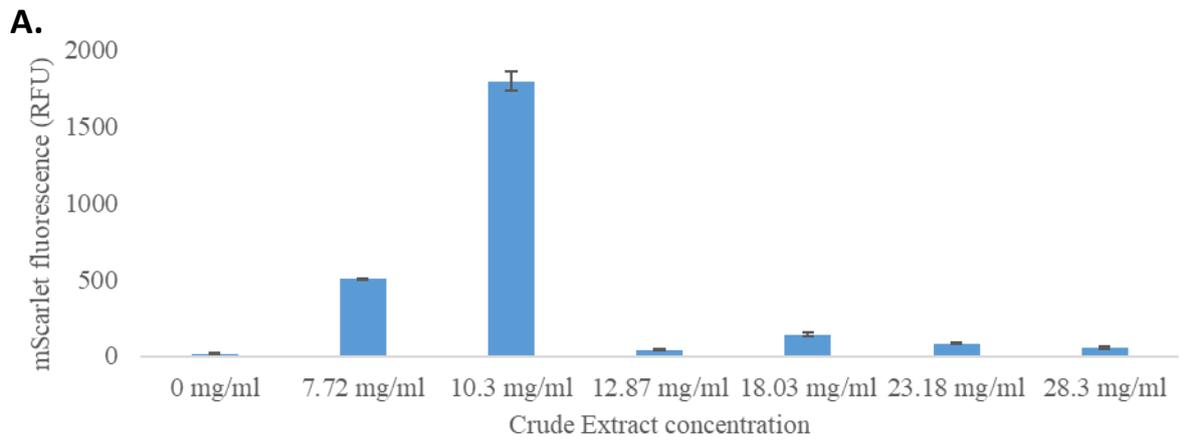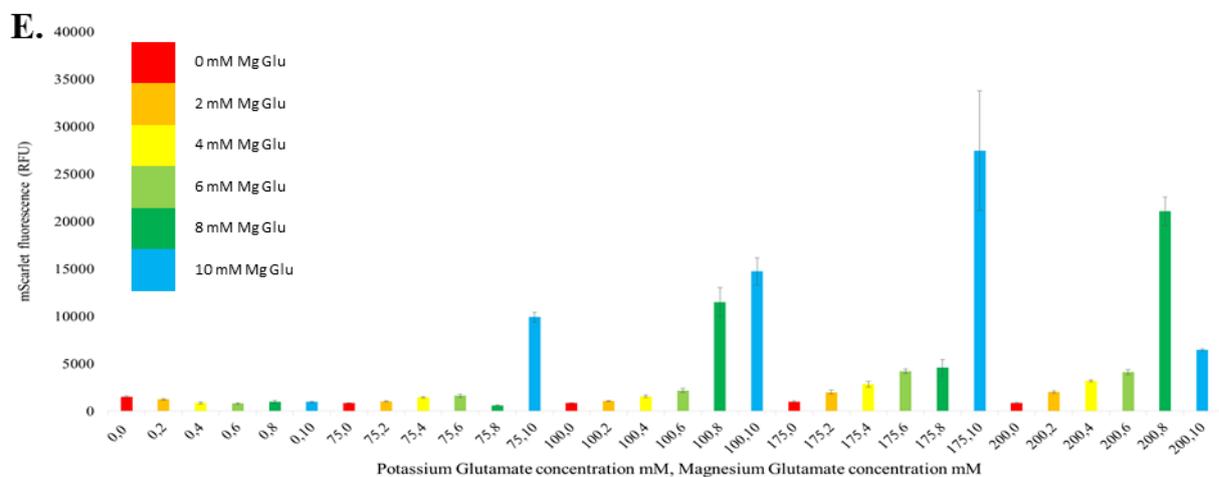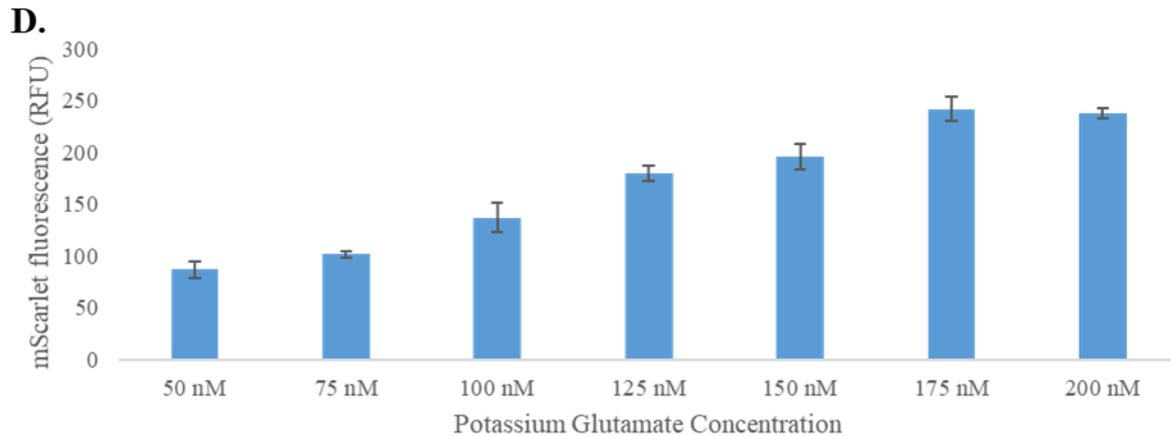
Figure 3.8 Optimisation of CFPS reaction for an _E. coli_ BL21 DE3 pLysS Crude extract.
A. The determination of the optimum crude extract protein concentration for the greatest production of mScarlet. **B.** The determination of the optimum DNA concentration for the greatest production of mScarlet. **C.** The determination of the optimum magnesium glutamate concentration for the greatest production of mScarlet. **D.** The determination of the optimum potassium glutamate concentration for the greatest production of mScarlet. **E.** The determination of the optimum potassium glutamate and magnesium glutamate concentration for the greatest production of mScarlet when both are at differing concentrations. For all reactions, Fluorescence readings were performed following six hours of incubation at 30°C and 200 rpm. Three readings were taken for each reaction and an average fluorescence value was calculated; samples were completed in triplicate with the mean RFU calculated, errors bars show one standard deviation.

3.1.8 The effect of D-ribose and maltodextrin on a CFPS reaction

Research by Garenne *et al*, (2021) has shown that the addition of D-ribose and maltodextrin to a cell-free reaction can greatly increase the protein yield. Maltodextrin has been shown to aid cell-free metabolic activity by indirectly stimulating the recycling of inorganic phosphates, which inhibit the reaction (Caschera and Noireaux, 2016). This happens through accelerating the regeneration of ATP from ADP and inorganic phosphate for glucose kinase activity. Garenne et al (2021) postulate that further supplementation of the reactions with D-ribose improves the regeneration of ATP, but the exact mechanism was not elucidated. To test this, we incorporated both D-ribose and maltodextrin into a CFPS reaction to synthesis mScarlet. In this test, the effect of D-ribose and maltodextrin addition were tested separately and together. The final concentration of D-ribose and maltodextrin in the CFPS reaction was 30 mM and 60 mM respectfully. The required volume of D-ribose and maltodextrin to achieve the desired final concentration was added to a sterile 2 mL Eppendorf tube and dried down using a vacuum centrifuge; once dried the other components of the cell-free reaction as detailed in table 2.12. For comparison, a reaction was performed using the standard CFPS reaction as detailed in table 2.12. pTU1-SP44-*mScarlet* was used as the source of DNA in these reactions at a concentration of 10 nM. The fluorescence data produced, seen in table 3.9 A, was used to estimate the protein yield for each reaction, these were as follows; Standard reaction as detailed in table 2.12 produced an estimated 4.1 µM of protein, reaction supplemented with just D-ribose produced an estimated 1.34 µM of protein, reaction supplemented with just maltodextrin produced an estimated 5.17 µM of protein, reaction supplemented with both D-ribose and maltodextrin produced an estimated 6.61 µM of protein.

When looking at the data a decrease in protein production can be observed when comparing the values recorded for the standard reaction mix and the reaction supplemented with just D-ribose, here a 3-fold decrease in protein production is seen when D-ribose is added. T-test analysis comparing the reactions yielded a p-value of 0.001042, which is statistically significant.

When comparing the data for the standard reaction and that supplemented with just maltodextrin a 1.2-fold increase in protein production can be seen in the samples where maltodextrin was added. T-test analysis comparing the reactions gave a p-value of 0.046042, which is statistically significant.

When comparing the data for the reaction supplemented with just maltodextrin and that supplemented with both D-ribose and maltodextrin a 1.6-fold increase in protein production can be seen in the samples where D-ribose and maltodextrin were added. T-test analysis of the two reactions produced a p-value of 0.004253, which is statistically significant.

Once the effect of supplementing the reaction with D-ribose and maltodextrin was determined, the optimum method for the incorporation of the supplements into a cell-free reaction was assessed. To test this D-ribose and maltodextrin were dried down in a vacuum centrifuge so that the final concentration of D-ribose was 30 mM and maltodextrin was 60 mM the reaction was then performed as detailed in table 2.12. The second method of incorporation tested was the addition of liquid D-ribose and maltodextrin, in this reaction the master mix in table 2.12 was altered to allow the addition of both supplements so that all components of the reaction were present in their desired concentration. To achieve this the volume of DNA in the reaction was adjusted such that the total volume for the reaction was still 33 µL, to ensure that the DNA was still at the correct concentration the standard 8.25 µL of DNA required to give a final reaction concentration of 10 nM was added to a sterile 2 mL Eppendorf tube and dried down using a vacuum centrifuge, the dried DNA was then re-suspended in a volume of ddH$_2$O which represented the standard 8.25 µL minus the volume used for the incorporation of D-ribose and maltodextrin. The fluorescence data for these reactions can be seen in figure 3.9 B

Using the fluorescence data produced it can be calculated that the method in which D-ribose and maltodextrin were dried down produced an estimated protein yield of 6.61 µM and when they were included in the master mix an estimated 5.29 µM of protein. This means that when the supplements are dried there is a 1.2 times increase in protein yield. T-test analysis of the values for the methods resulted in a p-value of 0.001423, which is statistically significant.
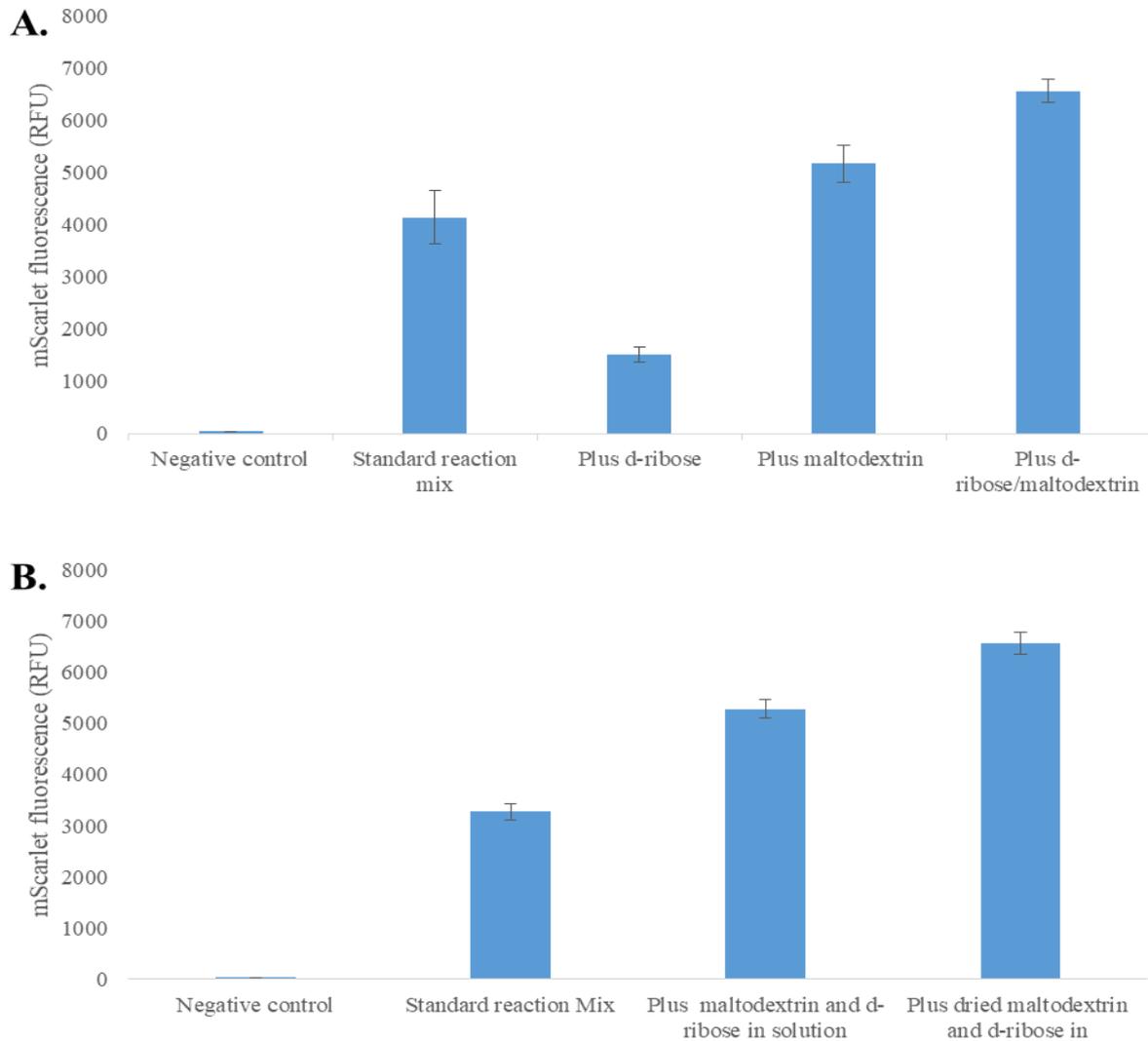
Figure 3.9 The effect of D-ribose and maltodextrin addition on the CFPS of mScarlet. A. Reactions were performed using the standard master mix solution detailed in table 2.12 and supplemented with just D-ribose, just maltodextrin, and both D-ribose and maltodextrin. Fluorescence readings were performed following 6 hours of incubation at 30°C and 200 rpm. Three readings were taken for each reaction and an average fluorescence value was calculated; error bars represent one standard deviation. **B.** The best method for the incorporation of D-ribose and maltodextrin into a CFPS reaction was determined by drying down both supplements for one sample and by implementing the supplements into the master mix solution detailed in table 2.12. Fluorescence readings were performed following 6 hours of incubation at 30°C and 200 rpm. Three readings were taken for each reaction and an average fluorescence value was calculated; error bars represent one standard deviation. Samples were completed in triplicate with the mean RFU calculated, errors bars show one standard deviation.

3.2 Discussion

The preparation of a highly active, cost and time-effective crude extract is fundamental when performing CFPS, in our research, we sort to outline the best method for crude extract production by comparing several methods and bacterial strains. Our research showed that the highest yield of protein was produced when using a crude extract derived from *E. coli BL21* which when compared to an extract produced using *E. coli* Rosetta (DE3) pLysS, yielded a significantly greater yield which was 1.3-fold higher. Typically in the literature, *E. coli* BL21 DE3 pLysS crude extracts prepared in the same manner as in this experiment have been observed to produce proteins yields of up to 0.75 mg/mL when using either *E. coli* BL21 DE3 pLysS or Rosetta (DE3) pLysS   (Sun *et al*., 2013), the highest concentration of protein produced in our experiments was estimated to be 20 µM which is 0.017424 mg/mL; compared to the yields observed in the literature the concentration of protein produced in our cell-free reactions is considerably lower.

It is also possible that the extracts lost activity over time or that repeated freeze-thawing of the aliquots leads to a reduction in activity. However, this is unlikely as the effect of freeze-thaw cycles on protein expression has been shown to have a negligible effect on protein expression (Levine *et al*., 2019).

A method for a 24-hour *E. coli* BL21 DE3 pLysS crude extract production was also trialled, in this method cell-free autoinduction media is used to sustain the cells once they reach the stationary growth phase and was designed so that this media could be inoculated with cells from a culture plate thus eliminating the need for time-consuming seed-culturing. This method was shown to produce >1 mg/mL of protein (Levine *et al., 2019).* When trialled in the lab two extracts were prepared, one following the method as detailed in the report and one with the inclusion of a run-off reaction. When observing the results of CFPS reactions for the extracts, no protein production was observed for the extract prepared following the method while a poor yield of protein was observed when the run-off reaction was performed. It has been shown that it is possible to create a functional *E. coli* BL21 DE3 pLysS crude extract that produces similar proteins yields without this step (Kim *et al*., 2014) and a decrease in activity has been observed as the run-off time increases (Kwon *et al*., 2015). As an *E. coli* BL21 DE3 pLysS extract does not require a run-off reaction it is unlikely that the omission of this step resulted in the production of an inactive extract.

The importance of producing a lysate that does not contain any cellular debris is emphasised by Kwon *et al* (2015). The presence of cellular debris in the final lysate affects the activity and can render the extract inactive. As the extracts show little or no activity it is possible that lysate contains cellular debris that has rendered them inactive.

It is possible that the production of protein was affected by the master mix used for the reaction; the master mix was prepared as detailed in Levine *et al* (2019) but was not supplemented with ammonium glutamate or phosphoenolpyruvate, both of which play a role in the regeneration of energy in the CFPS system, as they were not available. The reaction was also performed using the standard master mix detailed in table 2.12 but no protein synthesis was observed for either of the extracts. Another probable reason for the low activity of the extract was that the cells were harvested at an OD_{600} of 5.5 which is lower than the suggested harvest value of 10, by harvesting at a lower OD_{600} fewer cells are present in the extract which could result in cell lysate with a lower concentration of protein. However, Levine *et al* (2019) showed that a similar yield of protein was observed when the cells were harvested at an OD_{600} of 2.5 and 10.

In the method produced by Levine *et al* (2019) sonication is used to lyse the cells, this is similar to the method outlined by Sun *et al* (2013), however, the duration and number of on/off cycles differ in the Sun *et al* paper, the samples are sonicated for an on-cycle of 5 seconds and then followed by a 5-second off-cycle for a total on-time of 3 minutes at a frequency of 20 kHz and 65% amplitude. In the Levine *et al* method, the on-cycle lasts for 45 seconds and is followed by a 59-second off-cycle, the on-cycle is repeated 3 times for a total on-time of 2 minutes 15 seconds at a frequency of 20 kHz and amplitude of 50%. It is possible that the cell extract produced following the Levine *et al* method is not as active as detailed in the literature as the cells have not been completely lysed resulting in less intracellular material in the final lysate. Another possibility is that the prolonged-on cycles produce a large amount of heat which causes the proteins to denature.

In the pET system, target genes are cloned under a strong bacteriophage T7RNAP promoter, with expression induced by supplying a source of T7RNAP in the host cells. T7RNAp is typically extremely active and selective in that it can use almost all the cell's resources for target gene expression (Novagen, 2005). When looking at the expression for our pET vector genes, surprisingly lower expression is seen when compared to the SP44 vector, it is possible

that these low levels of protein expressed could result from low levels of tRNAP in the crude cell extract potentially restricting the yield of protein. This theory is supported by the fact that the SP44 system does not require the presence of T7RNAP for protein to be synthesised and higher yields are seen when using this system and the same batch of crude extract.

In this experiment the optimum concentration of plasmid DNA in each CFPS reaction was determined to be 20 nM, this equates to 578 ng/µL. The optimum concentration of DNA in our reactions is greater than that seen when compared to values in the literature, Wan, Kwon, and Jewett (2017), determined the optimum concentration to be 500 ng/µL, they also reported a decline in protein production as concentrations above this optimised value which the theorised may be a result of the consumption of nucleotide substrates. Takahashi *et al*, (2021) recorded the optimum protein yield at 270 ng/µL, while Zhang *et al* (2021) saw their greatest yield of protein at a similar DNA concentration of 300 ng/µL.

The use of a high concentration of crude extract has been shown to greatly reduce the efficiency of CFPS systems (Fujuwara and Nomura, 2013), meaning to be effective to produce proteins a cell-free system typically has a crude extract concentration of 10-20 mg/mL; this represents roughly 1/10 the intracellular concentration (Takahashi *et al*., 2021). This drastic difference in protein concentration means that a CFPS system produces an optimum protein yield which is less than 1 mg/mL per hour (Carlson *et al.,* 2012), when compared to the estimated 100 mg/mL per hour of protein produced by *E. coli* cells (Takahashi *et al*., 2021) is a substantial reduction in yield. In our experiment, the optimum value was determined to be ~ 10 mg/mL. Fujiwara and Doi (2016) also found that the optimum crude extract concentration for CFPS in their system was 10 mg/mL, while Kwon and Jewett (2015) observed that the greatest protein yield was obtained at an optimised crude extract concentration of 11 mg/mL. Takahashi *et al* (2021) have sort to develop a CFPS system capable of using high-concentration crude extracts in the hope of producing a protein yield more like that observed intracellularly. To achieve this, aim a high concentration of NTPs, which characteristically inhibits a CFPS reaction at lower crude extract concentrations (Nagaraj *et al.,* 2017), was used in conjunction with a high crude extract concentration and resulted in improved efficiency for the system. These findings suggest that it may be possible to increase the protein-producing potential of our system by increasing the crude extract and NTP concentrations in our system.

The observed optimum magnesium-glutamate concentration of 10 mM for a CFPS reaction is consistent with the values seen in the literature, earlier work by Jewett and Swartz (2004) showed an optimum magnesium ion concentration of 12 mM but more recent work by Li, *et al*., (2017) observed the greatest yield of protein production at a magnesium ion concentration of 10 mM in an *E. coli* CFPS system. Similarly, Levine *et al* (2019) also found that 10 mM of magnesium-glutamate is the optimal concentration for a CFPS reaction that uses an *E. coli*-derived crude extract. The calculated optimum potassium-glutamate in our reaction, determined to be 175 mM is considerably larger than the concentrations observed in the literature with an optimised concentration of 130 mM being reported by Levine *et al* (2019) and a similar value of 120 mM by Takahashi *et al* (2021). Sun *et al* (2013) suggested a larger range of potential concentrations of potassium-glutamate ranging from 40-160 mM, however, this represented an un-optimised concentration. As our reaction required a greater concentration of potassium-glutamate to generate the highest yield of protein it is possible that our crude extract contained a lower exogenous concentration of potassium ions when compared to those used by other research groups, this lower base concentration of ions would require greater supplementation of potassium ions to allow high yield protein synthesis.

The addition of maltodextrin as a secondary energy source in CFPS reactions is a viable option for protein synthesis and presents several advantages such as elevated levels of protein synthesis, no phosphate accumulation, and low cost, while the slow generation of glucose-1-phosphate followed by glycolysis supplies a homeostatic environment for CFPS (Wang & Zhang., 2009). Typically, 30-60 mM of maltodextrin is used to supplement CFPS reactions (Wang & Zhang, 2009, Garenne *et al*, 2021, Caschera and Noireaux 2013). It has been reported that the yield of protein produced via CFPS can be further increased by supplementing the reaction with D-ribose as well as maltodextrin (Garenne *et al*., 2021). A significant increase in protein production was observed when using both maltose and D-ribose it is possible that the addition of these two molecules as secondary energy sources in the reaction elongated to time over which protein synthesis could occur resulting in a greater protein yield.

When all the optimised conditions were combined in a single CFPS reaction no protein expression was observed, it is possible that when drying down some of the components in the reaction in the vacuum centrifuge crystals were formed, these crystals would affect the total volume of the reaction and subsequently the final concentration of the other components in the final reaction. Another plausible reason could be that the dried-down components of the

reaction were not fully re-suspended in the final reaction volume affecting the concentrations of these components. As the DNA in these reactions was dried down it is possible that this process was performed at too high a temperature leading to the denaturation of the plasmid, this would mean that no protein synthesis could occur as there is no coding information.

It is possible that the crude extract used when trailing the joint optimised conditions prevented the synthesis of protein as this was a different batch than that used during the optimisation steps. A high degree of variability is often seen between CFPS reaction and crude extract batches, this variation is attributed to the small reaction volumes, sensitive reagents, complicated protocols, and differences in techniques of individuals (Cole *et al*., 2019) this high variability is exemplified by Dopp *et al* (2019) who recorded a potential coefficient of variation at 97.3% between CFPS reactions. As previously mentioned, it is also possible that the presence of cellular debris in the cell lysate rendered the extract inactive.

Due to the difficulty in producing consistent results using the optimised conditions, it was decided that CFPS reactions would be performed using a crude extract produced using the Sun *et al* (2013) method, the standard reaction conditions detailed in table 2.12, adapted from Levine *et al* (2019) supplemented with maltodextrin and D-ribose as during testing the use of these factors in conjuncture produced the highest observed protein yield.

Chapter 4 BM3 characterisation and Cell-Free Protein synthesis

4.1 Characterisation of recombinant BM3

4.1.1 Characterisation of recombinant BM3

BM3 was expressed in *E. coli* using recombinant protein production methods this was done to allow the characterisation of the enzyme and provide a positive control to compare the cell-free experiments with and provide a baseline of protein activity. The pET15b-*bm3* plasmid, the map of which can be seen in figure 4.1 of the appendix (provided by Dr Moore), was pre-cloned with an N-terminal His6-tag under the control of a T7 promoter. Competent *E. coli* BL21 DE3 pLysS cells were transformed with the plasmid-encoded *bm3* gene sequence. The cells were grown to express the protein of interest which was purified via His_6-Tag purification. The fractions collected from the His_6-Tag purification were analysed by denaturing gel electrophoresis to determine if the BM3 protein was synthesised (figure 4.2). The purified BM3 protein was visible in lanes 9, 10, 11 and 12 this corresponds to the elution step of the protein purification, the bands are roughly 121.5 kDa which is larger than expected due to the addition of a His_6-tag following identification the protein was pooled, de-salted, and stored at -20°C. BM3 protein can be detected in all the lanes suggesting that at each of the wash steps a fraction of the produced protein is eluted, however, the greatest concertation is observed in lanes 9, 10, 11 and 12 where elution buffer was used. A second band is also visible in lanes 9, 10, 11 and 12 which is roughly 40 kDa, this suggests that a second protein was purified alongside the BM3 target protein.
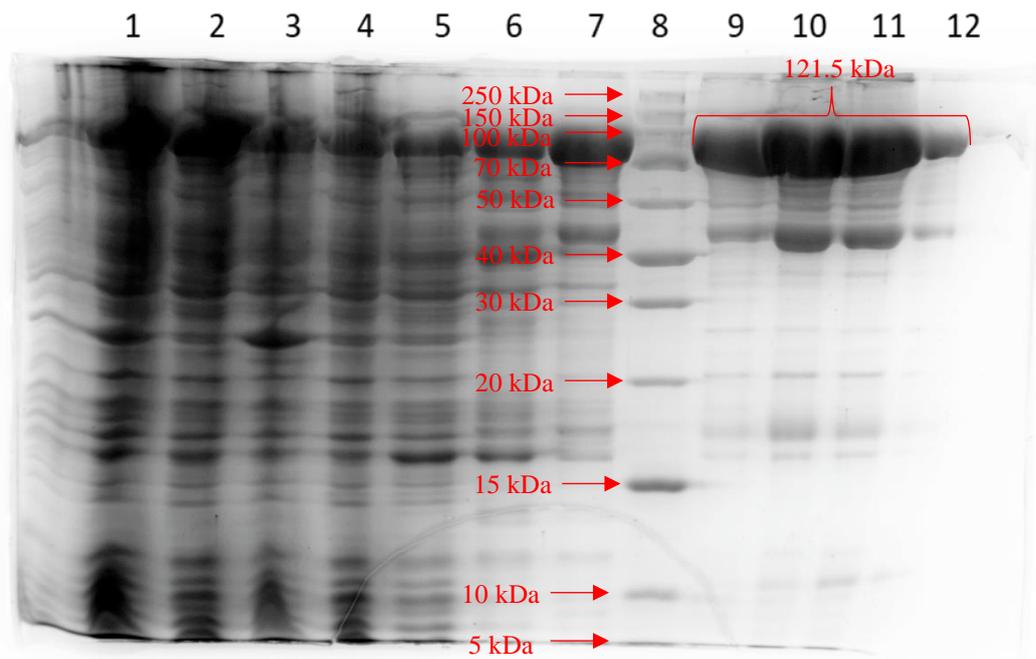
Figure 4.2 Purification of BM3. A denaturing gel electrophoresis of BM3 purification. Lanes are as follows: 1- crude extract, 2- soluble supernatant, 3- insoluble fraction, 4- supernatant flow-through, 5- binding buffer flow-through, 6- wash buffer I flow-through, 7- wash buffer II flow-through, 8- PageRuler ™ Unstained Broad Range Protein Ladder, 9, 10,11, 12- elution fractions. BM3 size is 119 kDa without Histag which adds 2.5 kDa to protein size.

4.1.2 BM3 activity assays

Next, we investigated the catalytic activity of the purified BM3. The activity of BM3 was determined by monitoring the monooxygenase activity of BM3 with respect to the substrate lauric acid measured indirectly through measuring the oxidation of NADPH. As NADPH is oxidised a fall in absorbance will be observed, this drop in absorbance can be used to interpret the turnover of lauric acid. The oxidation of NADPH was measured using a spectrophotometer measuring the absorbance at 340 nm, readings were taken over a period of three minutes with an absorbance reading taken every ten seconds (figure 4.3). Varying concentrations of BM3 were added to optimise the time course which allowed us to perform reactions in a reasonable time frame. This included 0.1 µM (figure 4.3), 1 µM (figure 4.3), and 10 µM BM3 (figure 4.3), combined with 100 µM lauric acid and 50 µM NADPH in 0.2 M phosphate buffer pH 8. A reduction in absorbance corresponds to the oxidation of NADPH, this is because BM3 catalyses the oxidation of NADPH when a substrate is present or absent, however, when a substrate is present the rate of oxidation is greater. A series of negative controls were also performed to determine the background oxidation of NADPH when no

substrate was present, this was necessary to determine if the inclusion of Lauric acid in the reaction increased the rate at which NADPH was oxidised. These are reactions without enzymes, NADPH, and lauric acid. At all enzyme concentrations, a decrease in absorbance is observed, and as the concentration of BM3 increases, the drop in absorbance becomes more pronounced which correlates to an increase in the rate of reaction.

BM3 was diluted across a range (0.1 - 10 µM) and initial activity was measured. 1 µM of BM3 provided a linear rate between the desired time intervals. Three negative controls were performed during this testing these being reactions with no substrate (figure 4.3 D), no enzyme (figure 4.3 E) and no NADPH (figure 4.3 F). In the reaction lacking BM3, the level of NADPH stays constant across the 3 minutes. However, a greater starting value is observed, this is likely due to enzyme activity between the period at which the enzyme is added, and the first absorbance reading is taken this suggests the enzyme is exhibiting pre-steady-state kinetics. In the reaction with no NADPH, a low level of absorbance is constant across the 3 minutes.

The activity of the purified BM3 enzyme was monitored at varying concentrations of the lauric acid substrate. In this series of reactions, the turnover of lauric acid was measured indirectly by measuring the oxidation of NADPH by monitoring the reduction in absorbance at 340 nm, this reduction corresponds to the oxidation of NADPH. The final concentration of BM3 purified enzyme used in the reactions was 1 µM and the final concentration of NADPH was 200 µM; the concentration of lauric acid ranged from 0 to 900 µM. Each reaction was performed in a 0.2 M phosphate buffer.
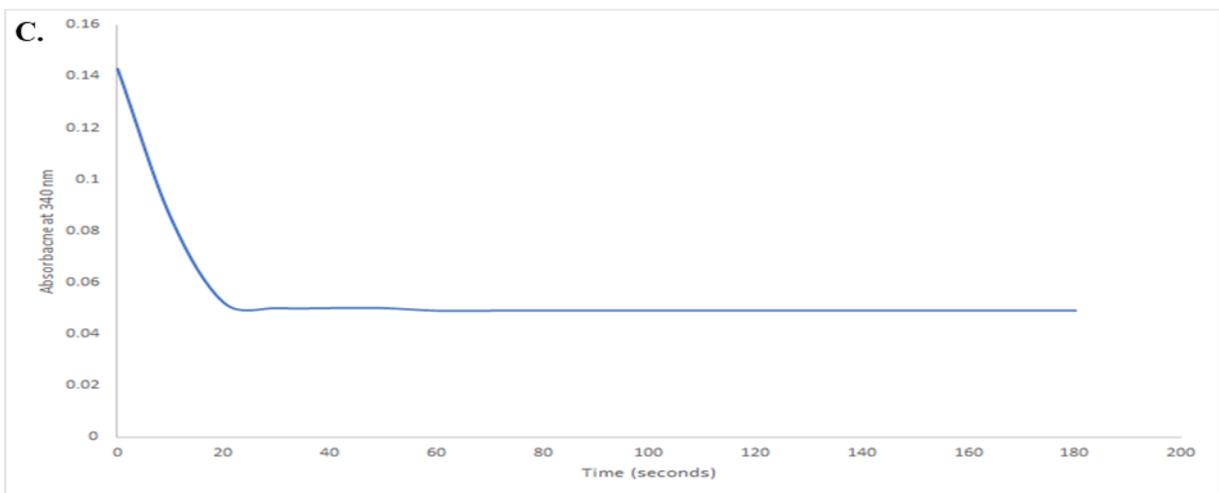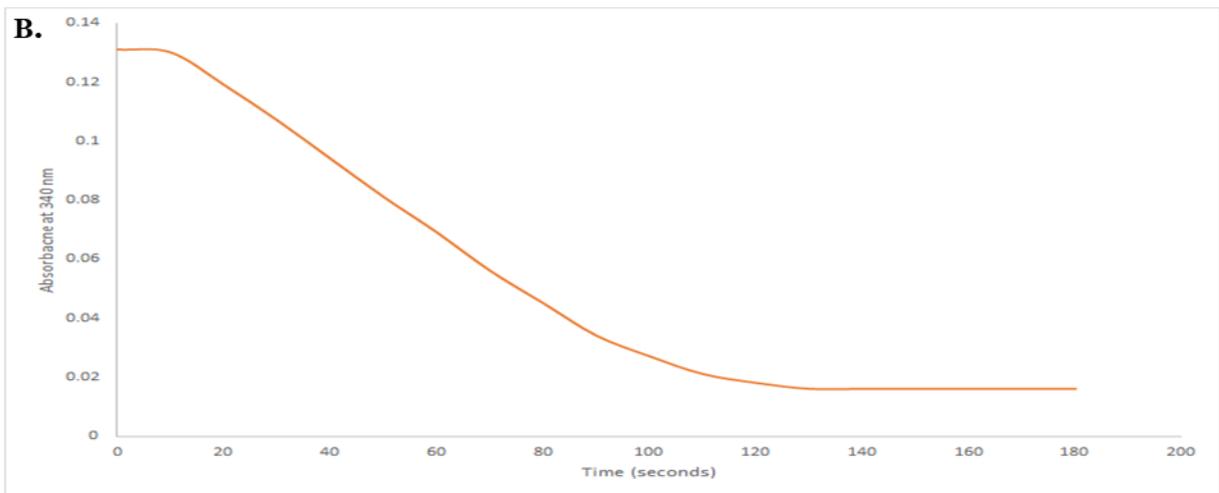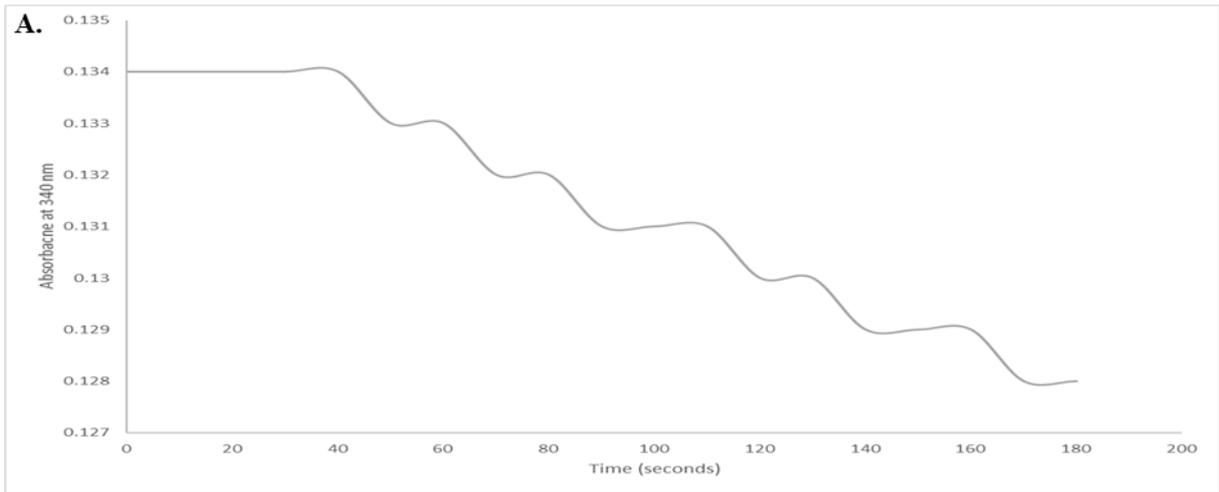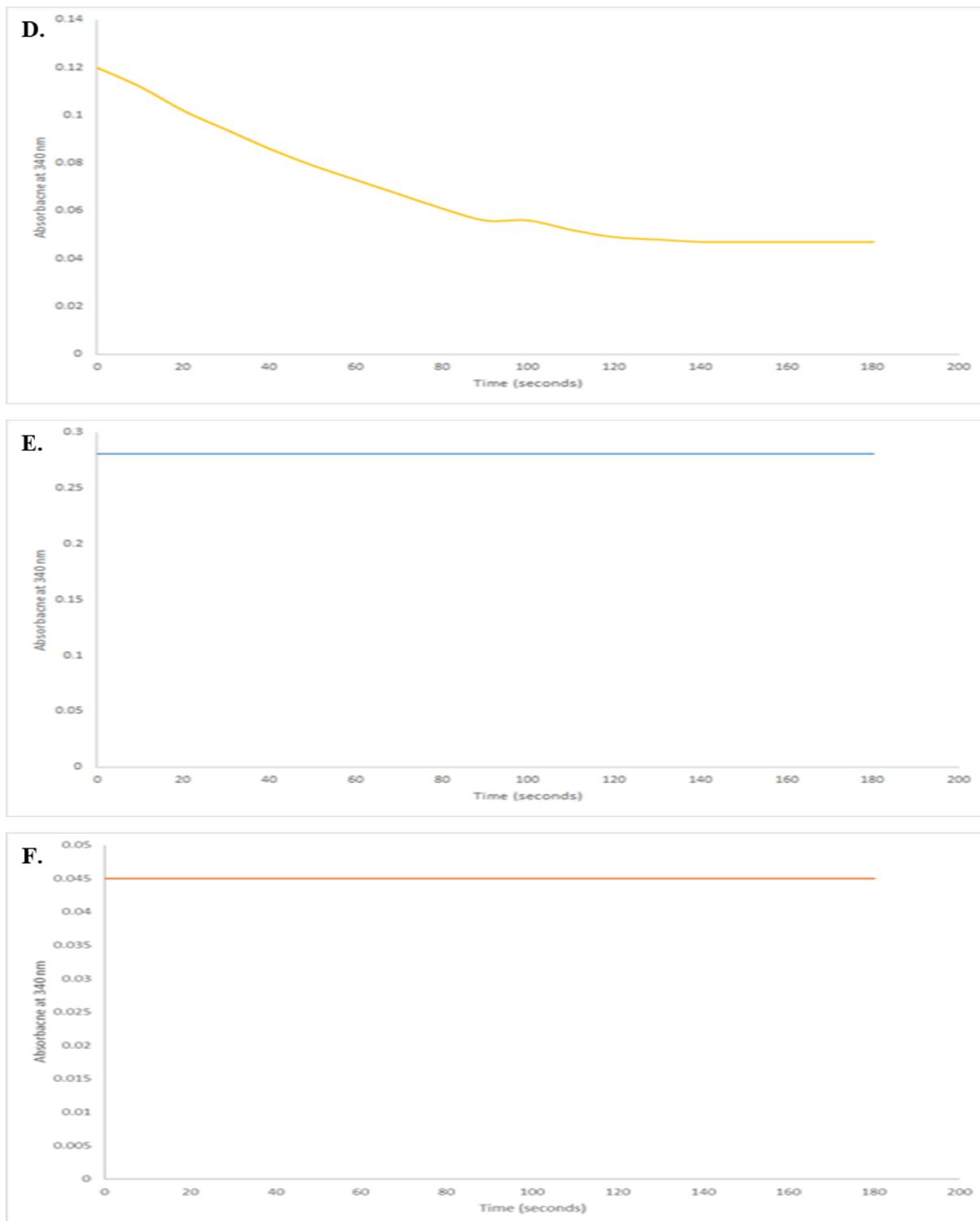
Figure 4.3 Characteristics of an activity assay for in vivo produced BM3 enzyme. Each reaction was monitored for 3 minutes with an absorbance reading at 340 nm taken every 10 seconds. All reactions were performed in 0.2 M phosphate buffer **A.** The change in absorbance of 50 µM NADPH at an enzyme concentration of 0.1 µM. **B.** The change in absorbance of 50 µM NADPH at an enzyme concentration of 1 µM. **C.** The change in absorbance of 50 µM

NADPH at an enzyme concentration of 10 µM. **D.** Negative control reaction showing the effect of no substrate; the reaction consists of 50 µM NADPH, 1 µM of BM3. **E.** Negative control reaction showing the effect of no BM3; the reaction consists of 50 µM NADPH, and 100 µM of lauric acid. **F.** Negative control reaction showing the effect of no NADPH; the reaction consists of 1 µM BM3 100 µM of lauric acid.

4.2 Synthesis of BM3 via CFPS

4.2.1 CFPS of BM3

Once the activity of *E. coli*-produced BM3 was confirmed, CFPS of the protein was performed using pET15b-*bm3* as the source of DNA. The reaction used an *E. coli* BL21 DE3 pLysS crude extract and was set up following the optimised conditions determined previously, the samples were incubated for 6 hours at 30°C and 200 rpm. Post incubation, denaturing gel electrophoresis (figure 4.4) and Western blot were performed to determine if the BM3 protein had been produced (figure 4.5). For comparison, mScarlet was combined with the CFPS reaction framework to act as a positive control the results of this can be seen in lane 2, however it is difficult to identify any observable bands. Dilutions of purified BM3 from *E. coli* expression of an original concentration of 0.61 mM for a total of three dilutions seen in lanes 5 (0.61mM), 6 (0.061mM) and 7 (0.0061mM) were also run on the denaturing gel electrophoresis gel. There was no visible production of BM3 in lane 3 of the gel.
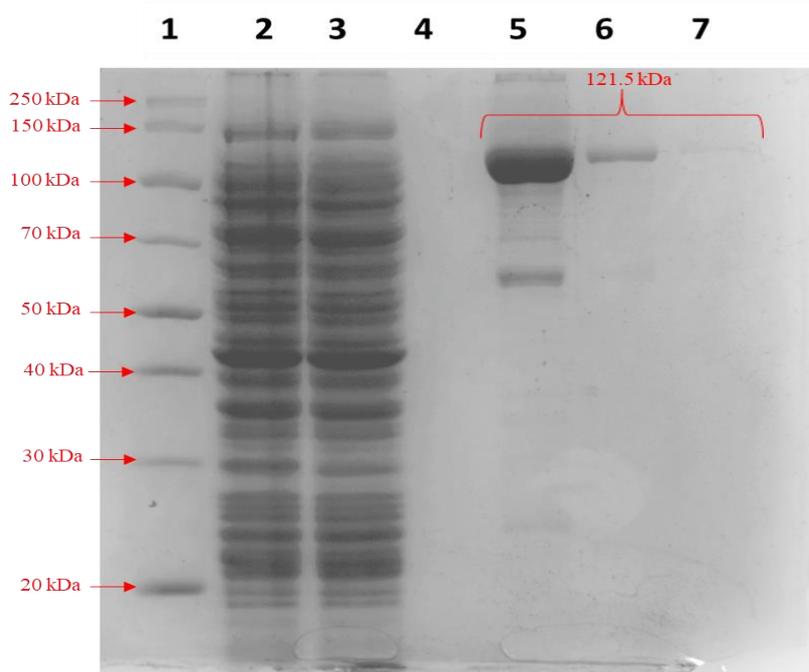

Figure 4.4 Denaturing gel electrophoresis gel of the CFPS of BM3 using pET15b-*bm3*. Lanes are as follows 1- PageRuler ™ Unstained Broad Range Protein Ladder, 2- CFPS negative control, 3- CFPS BM3, 4- Blank, 5- Purified BM3 1/10 dilution, 6- Purified BM3 1/100 dilution, 7- Purified BM3 1/1000 dilution. BM3 size is 119 kDa without Histag which adds 2.5 kDa to protein size.

The mScarlet protein produced as the positive control in the CFPS was also transferred to the Western blot, this would allow us to determine if the blot was successful as the CFPS-produced mScarlet protein would be His$_6$-tagged. Evidence of the transferred mScarlet can be observed in lane 2 of the blot. Purified BM3 protein at an original concentration of 0.61 mM was transferred to the blot at varying dilutions these can be seen in lanes 6 (0.61mM), 7 (0.061mM) and 8 (0.0061mM); this acted as a visual guide for BM3 production and would allow the estimation of the concentration of produced BM3. No evidence of BM3 production is viable in lane 3 of the blot suggesting that no protein was produced and the PageRuler ™ Plus prestained Broad Range Protein Ladder failed to transfer.
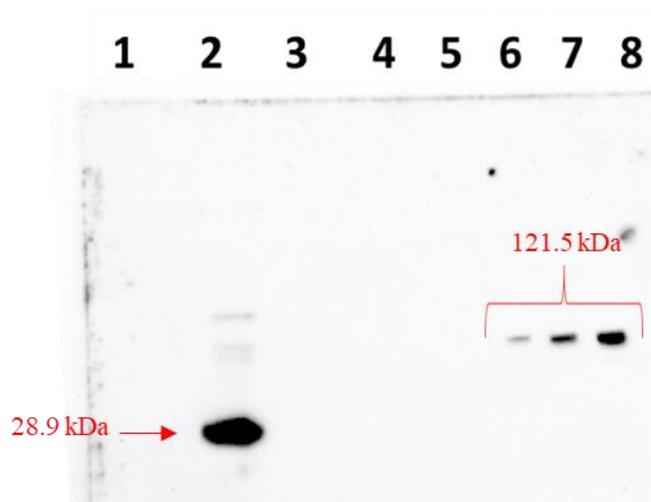
Figure 4.5 Western blot for the CFPS of BM3 using pET15b-*bm3*. A Western blot to determine if His-tagged BM3 has been produced via CFPS. mScarlet was also produced via CFPS as control and transferred to the blot. Purified BM3 (original concentration of 0.61 mM) was also transferred at varying dilutions. Lanes are as follows: 1- PageRuler ™ Plus prestained Broad Range Protein Ladder (not transferred to blot). 2- mScarlet CFPS control, 3- BM3 CFPS reaction, 4- blank, 5- blank, 6- Purified BM3 1/1000 dilution, 7- purified BM3 1/100 dilution, 8- purified BM3 1/10 dilution.

Initial CFPS reactions of BM3 using the pET15b-*bm3* as the source of DNA showed no production of the protein on denaturing gel electrophoresis gels and Western blots. To try and achieve the synthesis of the protein the encoded *bm3 gene* was cloned into the pTU1-SP44 vector using Gibson assembly. In this reaction a PCR-amplified version of the *bm3* gene was combined with a pTU1-SP44 plasmid, this backbone was selected as we found that it facilitated the highest protein yields during our optimisation experiments as seen in section 3.1.4. The map of the new plasmid can be seen in figure 4.6 of the appendix. A pre-cut pTU1-SP44 plasmid was used for the ligation, which also contained a cut mScarlet insert. This type of cloning is similar to the classic blue/white screening concept that relies upon a LacZ reporter (and 5-bromo-4-chloro-3-indolyl-β-d-galactopyranoside substrate) to indicate the presence of absence of insert. Re-ligation of the mScarlet-I gene yields pink colonies, after *E. coli* transformation and outgrowth. Alternatively, if the bm3 insert ligates, white colonies are expected. To provide a suitable negative control, a ligation was also set up lacking the insert, which is expected to produce only pink colonies. From our results, we observed, a greater proportion of the colonies were pink suggesting that the pTU1-SP44 vector more readily combined with the mScarlet insert than itself. The test plate showed (figure 4.7) a greater concentration of white colonies than pink colonies; these white colonies represent the ligation of the *bm3* insert with the pTU1-SP44 vector or the vector with itself. Four white colonies were selected from the test plate and plasmid DNA was isolated following the Qiagen MiniPrep protocol (see methods). When the ligated plasmid was digested with HindIII we expected to see 3 bands when agarose gel electrophoresis was performed, these being roughly 4.0 kb, 1.6 kb and 250 bp in size. To determine if the pTU1-SP44 vector and *bm3* insert had been correctly combined, a restriction enzyme digest and agarose gel (figure 4.8 of the appendix) were performed. The gel showed that the insert and vector had not correctly combined for 3 of the 4 tested colonies these being lanes 2, 3 and 4. While lane 5 suggested that the vector and insert may have correctly ligated with bands visible at the expected size of roughly 4.0 kb, and 1.6 kb. The smaller 250 bp band is not detectable on the band but this may be due to its small size. The MINI-prep for the possibly correct plasmid was re-transformed into DH10β competent cells and used to produce a MIDI-prep. To determine if the plasmid was correct it was sent for sequencing the data which can be seen in figure 4.11 of the appendix.
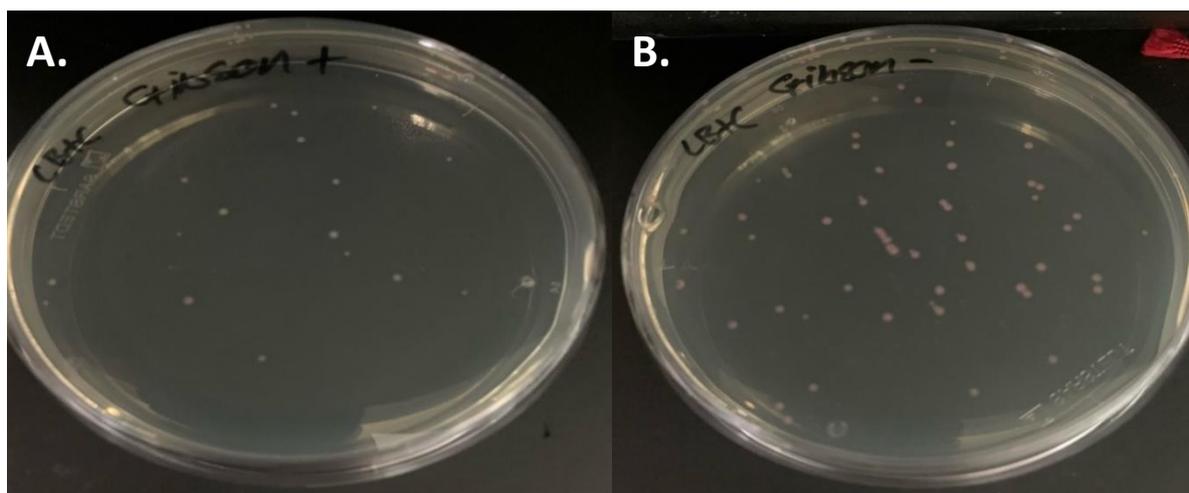
Figure 4.7 Plates of the transformed Gibson assembly plasmids. A. Plate for the transformation of a Gibson assembly plasmid containing a *bm3* insert, pTU1-SP44 vector, and mScarlet insert. White colonies on the plate (19 total) represent the ligation of the *bm3* insert with the pTU1-SP44 vector, while pink colonies (four total) represent the ligation of the mScarlet insert with the pTU1-SP44 vector. **B.** Plate for the transformation of a Gibson assembly only pTU1-SP44 vector and mScarlet insert, this reaction was used as a negative control. Pink colonies (67 total) represent the ligation of the mScarlet insert with the pTU1-SP44 vector.

CFPS of pTU1-A-SP44-bm3

Following sequencing, pTU1-A-SP44-*bm3*-His$_6$ attempted to synthesise the BM3 target protein, as testing of the current batch of *E. coli* BL21 DE3 pLysS crude extract showed low activity, we verified this by carrying out the CFPS of mScarlet and comparing the concentration of protein produced with that seen in previous batches. As the production of a new *E. coli* BL21 DE3 pLysS, the crude extract would be time-consuming, the decision was made to attempt the CFPS of BM3 using a previously prepared *E. coli* Rosetta (DE3) pLysS crude extract that was known to be active. As the *E. coli* Rosetta (DE3) pLysS crude extract had not been optimised reactions were performed using the standard conditions detailed in table 2.12. However, as we had observed an increase in protein yield at higher DNA concentrations during our CFPS optimisation steps, we attempted the CFPS of BM3 using higher concentrations of pTU1-A-SP44-*bm3*-His$_6$, these being 20 nM 30 nm, and 40 nm respectfully. In these reactions the desired concentration of pTU1-A-SP44-*bm3*-His$_6$ DNA was added to a sterile 2 mL Eppendorf tube and dried down in a vacuum centrifuge; once dry the other components of the CFPS

reaction were added as detailed in table 2.12. To ensure that the CFPS reactions were working a positive control using mScarlet was used, and the composition of the reaction was as detailed in table 2.12; pTU1-SP44-*mScarlet* was used as the source of DNA in these reactions at a concentration of 10 nM. The CFPS product of this reaction was analysed using denaturing gel electrophoresis (Figure 4.9) in this gel purified BM3 protein is visible in lanes 5 and 6 but there is no detectable target protein produced via the CFPS reaction observed in lane 3. It is also noteworthy that the mScarlet control protein is not visible in lane 2, this suggests that it has either not been produced or has been produced at a very low concentration that is below the limit of the detection threshold.
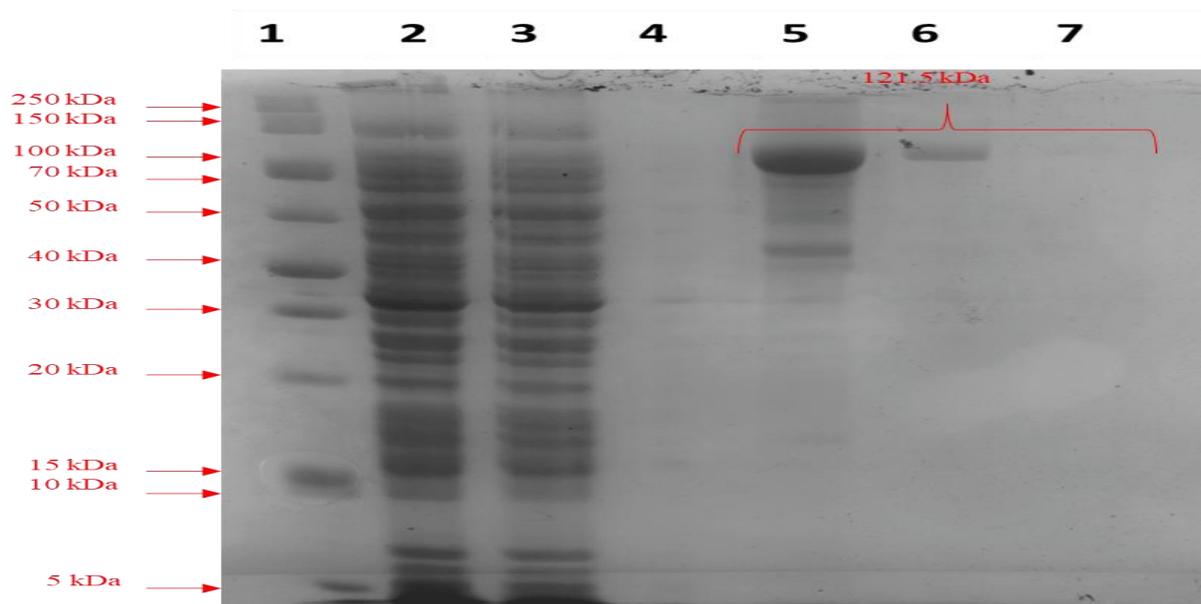
Figure 4.9 Denaturing gel electrophoresis of the CFPS of BM3 using pTU1-A-SP44-*bm3*-His₆. 1- PageRuler ™ Unstained Broad Range Protein Ladder, 2- CFPS negative control, 3- CFPS BM3, 4- Blank, 5- Purified BM3 1/10 dilution, 6- Purified BM3 1/100 dilution, 7- Purified BM3 1/1000 dilution. BM3 size is 119 kDa without Histag which adds 2.5 kDa to protein size.

The CFPS of BM3 was attempted using an *E. coli* Rosetta (DE3) pLysS crude extract and a reaction mixture supplemented with D-ribose and maltodextrin as this was found to increase the concertation of protein produced during our optimisation experiments. This experiment used varying concentrations of the pTU1-A-SP44-*bm3*-His$_6$ and pET15b-*bm3* plasmid. mScarlet was used as the positive control in this reaction. As the detection of the protein production via denaturing gel electrophoresis proved difficult an alternative method for protein identification was used. This involved incubation of the protein with a fluorescent fluorescein Arsenical Hairpin Binder-ethanedithiol (FLAsh-EDT$_2$) tag which binds to the tetracysteine sequence CysCysProGlyCysCys which was engineered into the BM3 target protein and the mScarlet control protein. Once bound the tag emits green-yellow fluorescence (528 nm). FLAsh-EDT$_2$, the method for which can be seen in 2.7.8.3 was used to identify the presence of the positive mScarlet control and the production of BM3 (figure 4.10). When viewed under blue light a band can be observed for the mScarlet positive control (lane 2), no bands can be observed for the CFPS of BM3 for either SP44 or pET-15b plasmid DNA sources (lanes 3-9) suggesting that no BM3 has been synthesized during the reaction.
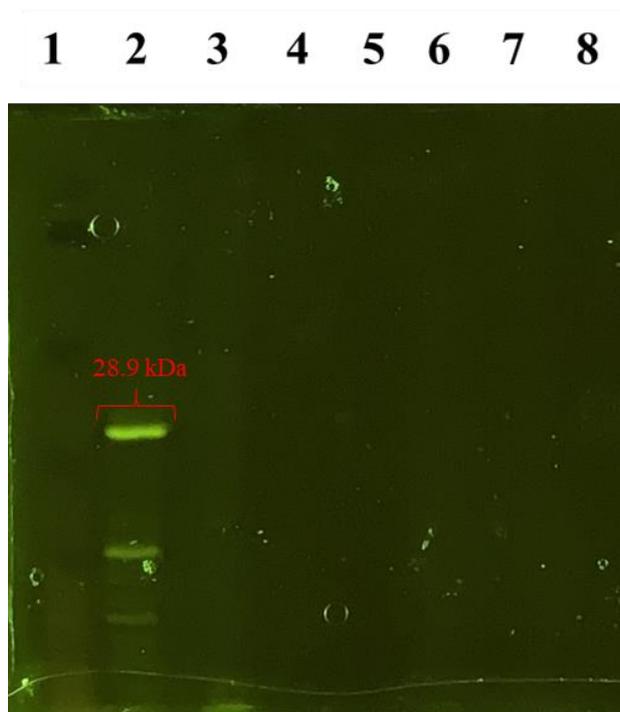
Figure 4.10 CFPS of pTU1-A-SP44-*bm3*-His₆ and pET15b-*bm3*. Denaturing gel electrophoresis analysis of the CFPS of BM3 using pTU1-A-SP44-*bm3*-His₆ and pET15b-*bm3* at varying concentrations as a source of DNA. CFPS of mScarlet was used as a positive control. CFPS samples were supplemented with D-ribose and maltodextrin and performed as detailed in table 2.12. Samples were incubated for 6 hours at 30°C and 200 rpm. Samples were prepared for denaturing gel electrophoresis and combined with denaturing gel electrophoresis loading buffer, TCEP, and FLAsH-EDT. The protein band on the gel were visualised under blue light. Lanes are as follows: 1- PageRuler ™ Plus prestained Broad Range Protein Ladder, 2- mScarlet CFPS, 3- pET15b-*bm3* 10 nM CFPS, 4- pET15b-*bm3* 20 nM CFPS, 5- pET15b-*bm3* 30 nM CFPS 6- pTU1-A-SP44-*bm3*-His₆ 10 nM CFPS, 7- pTU1-A-SP44-*bm3*-His₆ 20 nM CFPS, 8- pTU1-A-SP44-*bm3*-His₆ 30 nM CFPS.

4.3 Discussion

When observing the denaturing gel electrophoresis for the His_6-tag purification of the in vivo produced BM3 protein a second band can be observed, this suggests that the sample is not 100% pure and contains a contaminant, which is likely the result of non-specific protein binding during the purification process and may be due to a low ratio of the target protein to total protein in the sonicated cell pellet. The purified BM3 protein provides a positive control to compare to the cell-free reactions.

Through the performance of varying assays, the activity of the purified BM3 protein was confirmed. The rate at which NADPH was oxidised in the presence of the lauric acid substrate and the enzyme was greater than the rate of background oxidation when just the enzyme and NADPH were present. As expected, it was also possible to increase the rate of NADPH oxidation by increasing the concentration of lauric acid in the reaction.

It was not possible to produce an observable concentration of BM3 via CFPS when using the pET15b-*bm3* plasmid in CFPS. As the BM3 protein could be produced *in vivo* it suggests that the issue lies in the cell-free production of the protein.

It is also possible that the concentration of the DNA was not optimal, however, a series of reactions were performed using a range of plasmid concentrations, all of which yielded no observable protein production. The template design for the DNA may also affect protein synthesis, for example, the inclusion of rare codons at the beginning of the mRNA may compromise the initiation process for translation, this is because of the rarity of the cognate tRNAs for these rare codons (Wang *et al*., 2016). While the presence of rare codons may affect the CFPS reaction the use of Rosetta (DE3) pLysS should offset this because this strain supplements rare codons, however in these CFPS experiments no expression of the BM3 protein was observed suggesting that presence of rare codons is not a limiting factor for the reaction. A low concentration of T7RNAP in the CFPS reaction would inhibit the production of the target protein, to determine if this was a factor causing no protein expression during the production of the crude extracts the liquid cultures were supplemented with IPTG to facilitate the production of T7RNAP and the CFPS reactions were supplemented with purified T7RNAP, however, still, no protein was expressed in the cell-free reactions.

As it was not possible to express the target protein using the pET15 vector attempts were made to clone the *bm3* gene into the pTU1-SP44 vector as this was the more optimal vector during the optimisation stages. To produce the new vector Gibson assembly was employed to combine a pre-cut pTU1-SP44 vector which also contained the gene for mScarlet with PCR-amplified copies of the *bm3* gene. We made this decision as we found that pTU1-SP44 was the optimal plasmid for our cell-free system. When the initial Gibson reaction was performed, the plasmids produced were not the desired incorporation of pTU1-SP44 and *bm3*, instead, a greater proportion of pTU1-SP44 and mScarlet combination was observed, exemplified by a high number of pink colonies on the plate.

When the pTU1-SP44-*bm3* plasmid was used in a CFPS reaction to try a synthesise BM3 no expression of the protein was observed. It is feasible that the mutation in the produced pTU1-SP44-*bm3* plasmid prevented the expression of the desired protein, however as the protein was not expressed when using either of the vectors it is more likely that the issue is the protein itself, however, as CFPS is a viable option for the expression of difficult to synthesise proteins or those that are large it is unlikely the protein of interest in the issue. When observing the performed reactions and the low concentration of control protein produced the expression of the BM3 protein is likely limited by the low activity of the crude extract.

Chapter 5 Synthesis of active haem enzyme and their incorporation into a crude extract

In this chapter, we sought to produce active forms of the enzymes required for the biosynthesis of haem. As our target protein BM3 is a cytochrome P450 enzyme it requires a haem functional group for correct folding and activity, therefore our CFPS must contain the framework to produce haem so it can be incorporated into the enzyme during protein synthesis. *In vivo* produced BM3 is active due to the presence of haem in the *E. coli* BL21 DE3 pLysS cells used for protein expression, due to the presence of this endogenous haem it was postulated that when creating a crude extract there may be a sufficient concentration of haem in the extract to facilitate correct folding of the BM3 protein. We need both protein and haem for a functional enzyme, meaning that once the BM3 protein has been produced via CFPS haem must be present to allow correct protein folding. We were unable to merge these two separate objectives to complete the ultimate aim of the thesis.

Before the commencement of this project, two plasmids were prepared by Dr Moore, one containing the genes encoding Hem B, C, and D and the other coding for Hem E, F, G, and H. These plasmids were expressed in *E. coli* BL21 DE3 pLysS competent cells, and the proteins were purified. Once purified, denaturing gel electrophoresis gel was performed to determine if the 7 proteins had been purified. Activity assays were performed to determine whether the synthesised proteins were functional and if when combined with the ALA, the starting substrate in haem synthesis and Fe^{2+} it was possible to produce haem. Subsequently, we expressed the proteins in *E. coli* BL21 DE3 pLysS and used these cells to create a crude extract. This crude extract would provide the biosynthesis pathway required to produce haem in a CFPS reaction, this would provide the benefit of not having to supplement the reaction with purified versions of the enzyme.

5.1 *in vivo* synthesis of haem pathway enzymes and their characteristics

5.1.1 *in vivo* synthesis of the haem synthesis pathway enzymes

By producing the haem pathway enzymes *in vivo,* it is possible to confirm if active versions of the protein can be produced and identify the optimum reaction conditions for the pathway. The haem synthesis pathway enzymes were produced using in vivo methods to determine if they could be synthesised in correctly folded and active forms. To achieve this, the two plasmids HemBCD and HemEFGH (seen in figures 5.1 and 5.2 of the appendix) were separately transformed into competent *E. coli* BL21 DE3 pLysS cells. The cells were grown, and the proteins produced were extracted via His-Tag purification. The fractions collected from the immobilized metal ion affinity chromatography (IMAC) were analysed by denaturing gel electrophoresis to determine if the proteins of interest were synthesised. The expected molecular weight and the estimated molecular weight with the addition of the His$_6$- tag can be seen in table 5.1 of the appendix.

After the purification of proteins HemBCD, three bands were observed at around 37 kDa, 35 kDa, and 29 kDa these can be seen in lanes 9, 10, 11 and 12 of figure 5.3 A which corresponds to the wash steps. The expected sizes for the proteins were 35 kDa for HemB, 33 kDa for HemC and 27 kDa for HemD. The greatest concentration of protein can be observed in lane 10. These bands were larger than the expected size due to the addition of a His$_6$-tag which adds 2.5 kDa. The bands on the gel suggest that all three target proteins were produced and purified. The denaturing gel electrophoresis gel for the purification of proteins HemE, F, G, and H shows three clear bands at 41 kDa, 36 kDa, and 23 kDa these bands are visible in lanes 9, 10, 11 and 12 of figure 5.3 B which corresponds to the wash steps of the protein purification processes. The expected size for the proteins was 39 kDa for HemE, 34 kDa for HemF, 21 kDa for HemG and 36 kDa for HemH. Again, these bands are larger than expected as the purified protein has the addition of a Histag which adds 2.5 kDa to protein size. There may be a fourth band in between the bands of size 41 kDa and 36 kDa of lanes 9, 10, 11 and 12, if so, this band would be 37 kDa in size, it is not clear from the gel if this band is present. The three bands on the gel suggest that 3 of the 4 target proteins have been produced and purified these being HemE, HemF and HemG. HemH may be present but as it is similar in size to the other proteins its presence on the gel is not distinguishable.
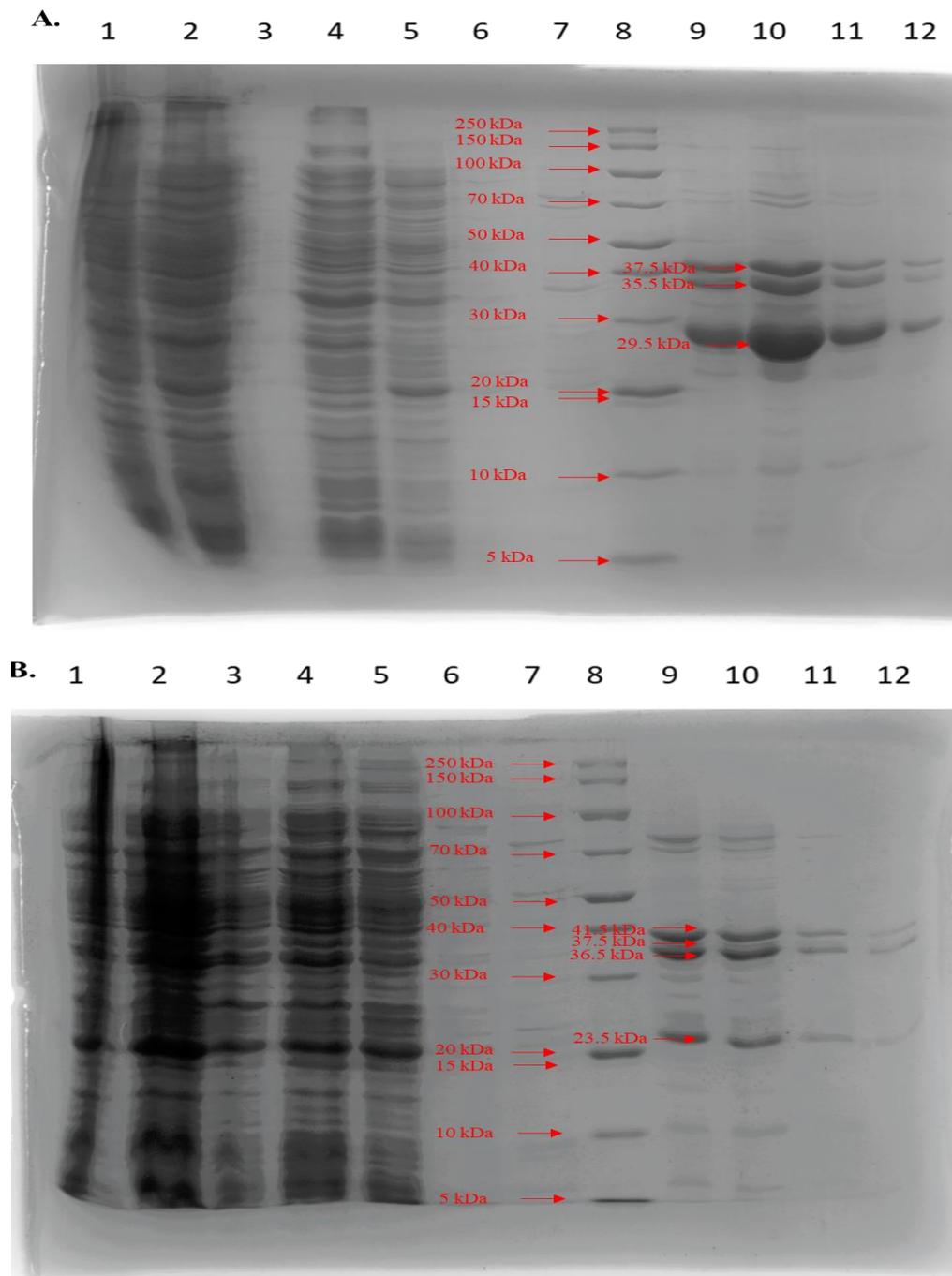
Figure 5.3 Purification of haem synthesis pathway proteins. A. Denaturing gel electrophoresis of HemBCD purification. **B.** Denaturing gel electrophoresis of HemEFGH purification**.** Lanes are as follows: 1- crude extract, 2- soluble supernatant, 3- insoluble fraction, 4- supernatant flow-through, 5- binding buffer flow-through, 6- wash buffer I flow-through, 7- wash buffer II flow-through, 8- PageRuler ™ Unstained Broad Range Protein Ladder, 9, 10,11, 12- elution fractions. Sizes of Haem pathway proteins are as follows HemB 35 kDa, HemC 33 kDa, HemD 27 kDa, HemE 39 kDa, HemF 34 kDa, HemG 21 kDa, and HemH 35 kDa. Sizes are without the addition of Histag which adds 2.5 kDa to protein size.

The genes for the haem biosynthesis proteins were individually expressed in competent *E. coli* BL21 DE3 pLysS cells as it was not possible to determine if all the HemEFGH proteins were present in the purified sample using denaturing gel electrophoresis. The transformed cells were subcultured into 10 mL of liquid autoinduction media. Once grown the cells were lysed and centrifuged to extract the proteins. Both the supernatant and the pellet for each of the Hem proteins were analysed using denaturing gel electrophoresis (figure 5.4) to determine if the proteins had been produced and if they were soluble in the supernatant.

When viewing the denaturing gel electrophoresis gel each protein of interest was observed in the insoluble fraction at a high concentration, this being lane 3 for HemE, lane 5 for HemF, lane 7 for HemG and lane 9 for HemH. These proteins were presumed to be inclusion bodies. In contrast, protein bands can only be seen in the supernatants for HemE (lane 2) and HemF (lane 4), and at a lower concentration than in the pellet. There was no detectable sign of proteins HemG (lane 6) and HemH lane 8). This suggests that either only proteins HemE and HemF are soluble or that the proteins HemG and HemH were extracted at such a low concentration that they cannot be detected on the gel.
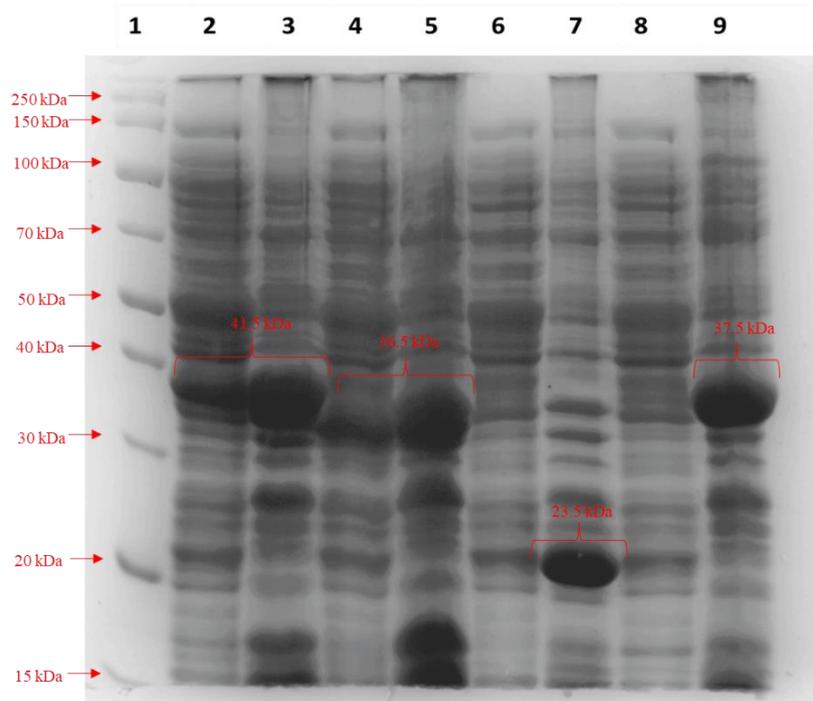
Figure 5.4 Single gene expression of Hem E, F, G, and H. Proteins Hem E, F, G, and H were individually transformed into competent *E. coli* BL21 DE3 pLysS cells and grown in autoinduction media. After sufficient growth, the cells were lysed by sonication and separated via centrifugation. The pellet and supernatant were analysed via denaturing gel electrophoresis to identify the presence of the haem synthesis pathway proteins. Lanes are as follows: 1- PageRuler ™ Unstained Broad Range Protein Ladder, 2- HemE supernatant, 3- HemE pellet, 4- HemF supernatant, 5- HemF pellet, 6- HemG supernatant, 7- HemG pellet, 8- HemH supernatant, 9- HemH pellet.

5.1.2 Haem synthesis pathway enzyme activity assays

Next, to determine if the synthesised haem synthesis pathway enzymes were active and capable of producing haem a series of enzyme assays were performed. In these assays, 1 µM final concentration of each enzyme was combined with ALA and Fe^{2+}, the precursors of haem biosynthesis. The concentration of ALA varied and ranged from 0 to 500 µM, and the concentration of iron in the reaction also varied ranging from 0 to 100 µM. A negative control reaction was performed in which no ALA was added; iron was still added in these reactions. Samples were incubated at $30^{o}C$ and 180 rpm overnight. Following incubation, the excitation and emission spectrums for the samples were measured so that we could attempt to identify the stage the reaction reached by comparing the spectrum peaks with those seen in the literature for compounds produced during haem biosynthesis. The Appearance of the samples under white light and blue light was recorded to determine if any pigmentation or fluorescence was observed. The Characteristics of fluorescent molecules in the haem synthesis pathway can be observed in table 5.2.

When looking at the sample with ALA concentrations of 50 µM, 100 µM, and 500 µM the same emission fluorescence peaks are observed, suggesting that all three reactions are creating the same final product (figure 5.6)

As the concentration increases the difference in spectrum seen at the differing iron concentrations decreases (Figures 5.5 and 5.6). In the sample with an ALA concentration of 500 µM (figure 5.6 C), the spectrum for the varying iron concentrations is almost entirely overlaid, with only a slight variation in fluorescence at the peaks. As the concentration of ALA increases so does the emission fluorescence of the sample.

A1.
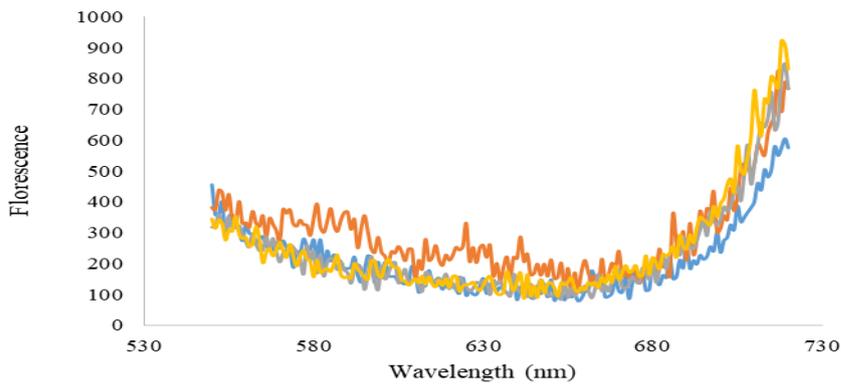
A2. **A3.**

Legend:
- 0 um ALA, 0 um Fe
- 0 um ALA, 10 um Fe
- 0 um ALA, 25 um Fe
- 0 um ALA, 100 um Fe

B1.

629 nm

589 nm
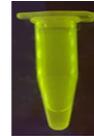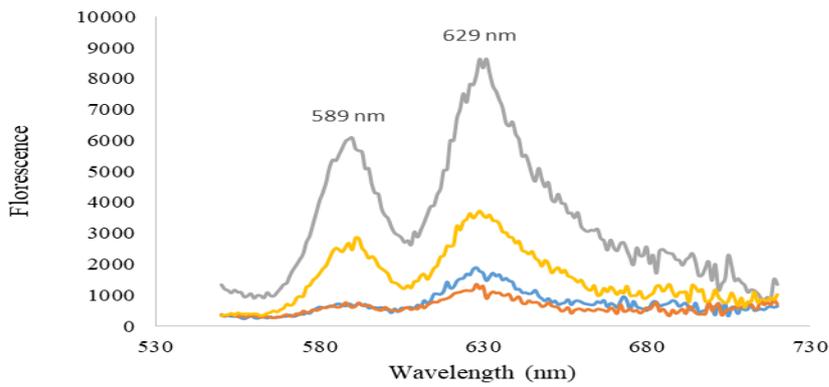
B2. **B3.**

Legend:
- 10 um ALA. 0 um Fe
- 10 um ALA. 10 um Fe
- 10 um ALA. 25 um Fe
- 10 um ALA, 100 um Fe

C1.

636 nm

672 nm

589 nm
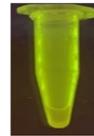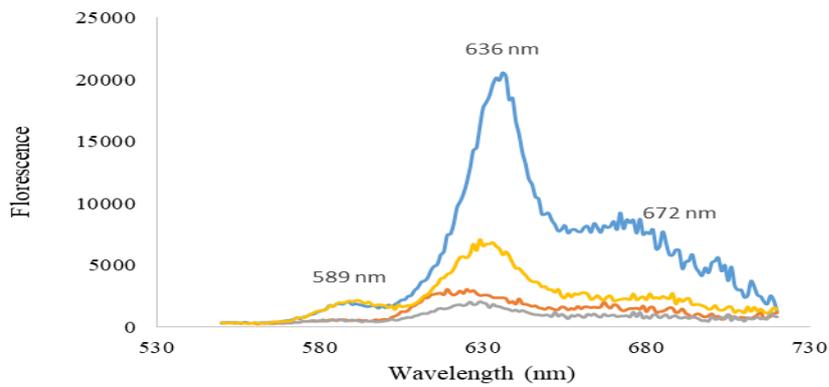
C2. **C3.**

Legend:
- 25 um ALA, 0 um Fe
- 25 um ALA, 10 um Fe
- 25 um ALA. 25 um Fe
- 25 um ALA. 100 um Fe
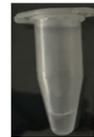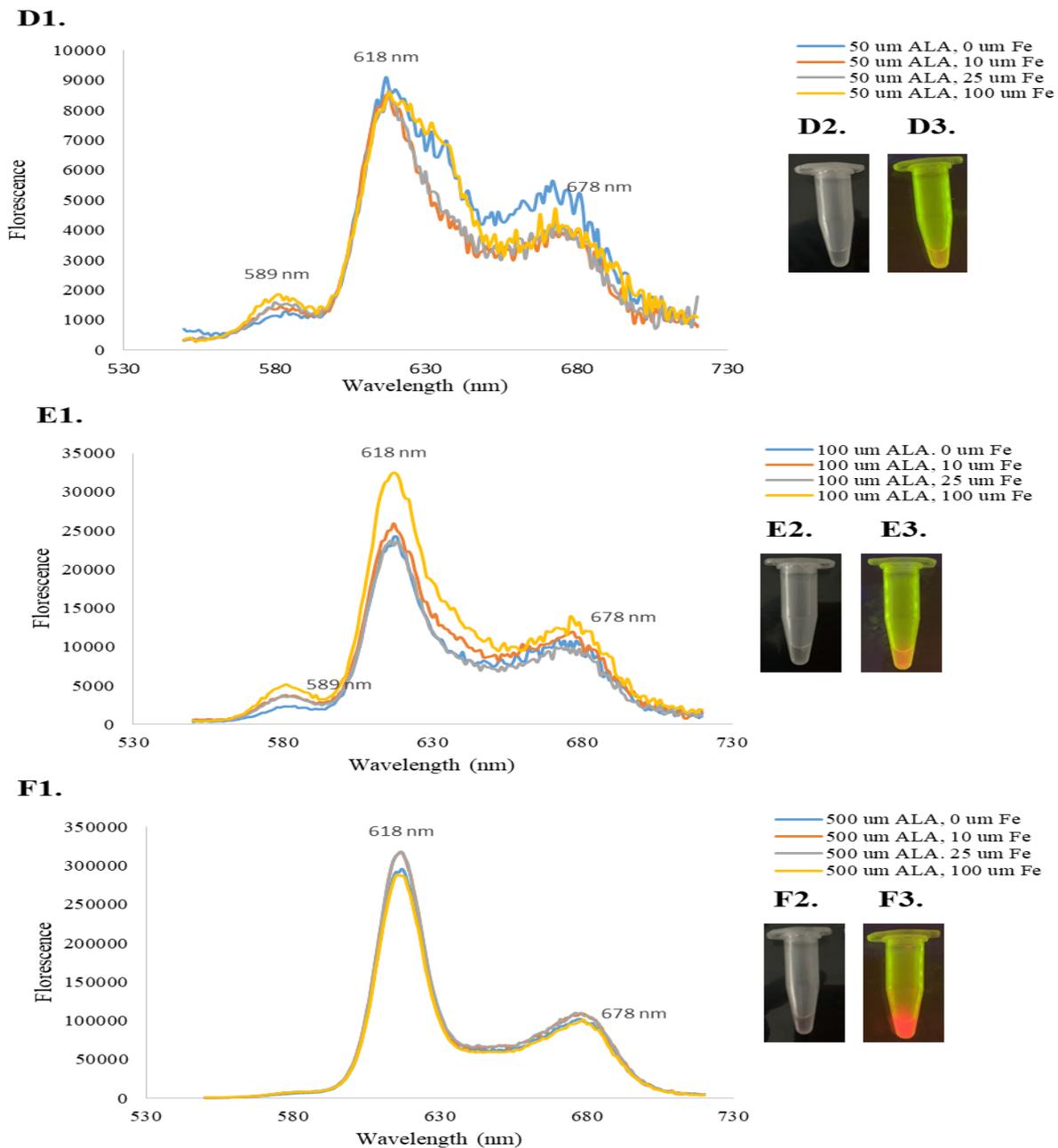
Figure 5.5 Characteristics of the products from an enzyme activity assay of in vivo produced Hem pathway enzymes at varying substrate and Iron concentrations. A1. Emission spectrum for ALA concentration of 0 µm and varying iron concentrations. **A2.** Sample for ALA concentration of 10 µm under visible light. **A3.** Sample for ALA concentration of 10 µm under blue light. **B1.** Emission spectrum for ALA concentration of 10 µm and varying iron concentrations. **B2.** Sample for ALA concentration of 10 µm under visible light. **B3.** Sample for ALA concentration of 10 µm under blue light. **C1.** Emission spectrum for ALA concentration of 25 µm and varying iron concentrations. **C2.** Sample for ALA concentration of 25 µm under visible light. **C3.** Sample for ALA concentration of 25 µm under

blue light. **D1.** Emission spectrum for ALA concentration of 50 µm and varying iron concentrations. **D2.** Sample for ALA concentration of 50 µm under visible light. **D3.** Sample for ALA concentration of 50 µm under blue light. **E1.** Emission spectrum for ALA concentration of 100 µm and varying iron concentrations. **E2.** Sample for ALA concentration of 100 µm under visible light. **E3.** Sample for ALA concentration of 100 µm under blue light. **F1.** Emission spectrum for ALA concentration of 500 µm and varying iron concentrations. **F2.** Sample for ALA concentration of 500 µm under visible light. **F3.** Sample for ALA concentration of 500 µm under blue light.

5.2 Incorporation of Haem pathway enzymes into a crude extract

5.2.1 Production of crude extract containing active haem enzymes

Once the activity of the haem synthesis pathway enzymes was confirmed the HemBCD and HemEFGH plasmids were transformed into *E. coli* BL21 DE3 pLysS, which was subsequently grown in 600 mL liquid autoinduction media cultures. The cells in the cultures were pelleted and used to create a crude extract, following the Levine *et al* (2019) method, this extract should, in theory, contain the haem synthesis pathway enzymes. To determine if the haem synthesis pathway enzymes were present in the extract a denaturing gel electrophoresis gel (not shown) was to run comparing the extract with the haem proteins expressed and a BL21 DE3 pLysS extract without haem protein expression. When the gel is viewed only one band is observed in the extract with the haem synthesis pathway enzymes expressed this is at around 20 kDa, this most closely corresponds to HemG which is expected to be 23.5 kDa in size when considering the addition of the His-tag. No other clear band can be observed.

To determine if the *E. coli.* BL21 DE3 pLysS extract with the haem synthesis pathway enzymes expressed was able to produce protein, a CFPS, the reaction was performed to produce mScarlet. The reaction was performed using the composition outlined in table 2.12. This sample was compared to a standard *E. coli.* BL21 DE3 pLysS extract and a negative control reaction where no DNA was added (Figure 5.7)

When looking at the fluorescence data produced for *E. coli.* BL21 DE3 pLysS extract with the haem synthesis pathway enzymes expressed an estimated 3.58 µM of protein was produced. T-test comparing the extracts provided a p-value of 0.000017 which is statistically significant. The extract with the haem synthesis pathway enzymes expressed was also compared with a standard *E. coli.* BL21 DE3 pLysS extract. The standard extract produced an estimated 9.19 µM of protein, a value that is 2.6 times greater than that of the extract with the haem synthesis pathway enzymes expressed. A T-test yielded a p-value of 0.000031 which means that significantly more protein is produced when using the standard crude extract.
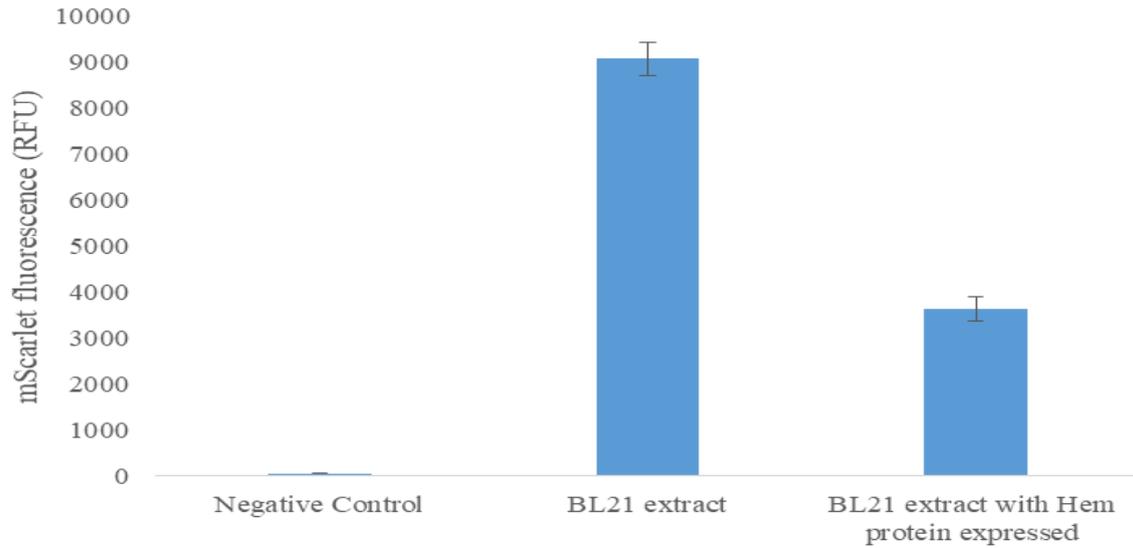
Figure 5.7 Determination of the activity of an *E. coli* BL21 DE3 pLysS crude extract with the Hem proteins expressed. Reactions were performed using the standard master mix solution detailed in table 2.12. Fluorescence readings were performed following 6 hours of incubation at 30°C and 200 rpm. Three readings were taken for each reaction and an average fluorescence value was calculated; samples were completed in triplicate with the mean RFU calculated, errors bars show one standard deviation.

5.3 Discussion

In this chapter we sort to synthesise active forms of the enzymes required for the biosynthesis of haem, when looking at the gels for the in vivo expression of the haem synthesis pathway proteins for the HemBCD plasmid three distinct bands can be observed and the expected size for the three proteins suggesting all three have been produced. In contrast, it is not as easy to determine if all four proteins of the HemEFGH plasmid have been made, this is due to proteins HemF and Hem H being similar in size, there appear to be three bands in close proximity on the gel, but this does not conclusively show that all proteins are present. To determine if the proteins are capable of being made each of the four genes in HemEFGH was expressed singularly, with the produced cell pellet sonicated to release the cellular components. When these cellular components and the re-constituted cellular pellet were analysed using denaturing gel electrophoresis the proteins of interest could be observed in each of the cellular pellets at their expected size. In comparison, only proteins HemE and HemF can be detected in the cellular components. The absence of proteins HemG and HemH suggest that these proteins are not soluble.

Reconstituting haem biosynthesis-
The data from the enzyme activity assays show several potential products are being produced. When comparing the emission data observed the reaction has likely progressed to the point of producing Coproporphyrinogen III which has been shown to have an emission spectrum with peaks at 618 nm and 680 nm (Seo *et al*., 2009); peaks can be observed at these values for reactions D1, E1 and F1. Fluorescence can also be seen in the samples for these reactions when viewed under blue light. When analysing the spectrums for the other reactions no correlation can be seen between the peaks and those seen in the other fluorescence compounds of the haem synthesis pathway. This is except for C1 which shows a peak at 636 nm and a broad peak at 672 nm, the broadness of the peak at 672 nm means that there is also a peak at around 690 nm which could potentially mean that Protoporphyrin IX which shows characteristic emission peaks at 630 nm and 690 nm was synthesised in this reaction.

The production of haem in *E. coli* is a heavily regulated process with several potential limiting steps, Pranawidjaja *et al* (2015) postulate that the intracellular accumulation of coproporphyrinogen I, caused by the spontaneous conversion of hydroxymethylbilane to coproporphyrinogen I is one of these bottleneck steps. The up-regulation of the genes which

code for HemD and HemF has been shown to cause the accumulation of ALA while the up-regulation of the other genes of the haem synthesis pathway proteins results in a decrease in ALA accumulation (Zhang *et al*., 2015) this suggests that it is important to find the correct balance between the enzyme concentrations to achieve the desired goal of haem synthesis.

When looking at the data from the activity assay the likely accumulation of coproporphyrinogen III suggests that this is not a limiting step in our reaction. Another documented limiting step is the limited conversion of protoporphyrin IX to haem (Pranawidjaja *et al* 2014), if in reaction C1 Protoporphyrin IX is produced it is possible this observable accumulation is due to this limited conversion, and it is also possible that there is haem present in this sample. Another possibility is that all of the required enzymes for the production of haem are not present in the reaction or that they are not at a sufficient concentration for the synthesis of haem.

From the denaturing gel electrophoresis gel of the crude extract containing the expressed haem synthesis pathway proteins only one band at 20 kDa, this band is similar in size to that expected for HemG, as only one band can be observed it stands to reason that only one of the haem synthesis pathway proteins is present in the extract at a detectable concentration. The absence of the other haem synthesis pathway proteins suggests that they were not overexpressed during the cellular growth stage of the crude extract preparation, to assess this further it would be necessary to perform a Western blot to examine if the proteins were present but at such low concentrations, they are not detectable via a denaturing gel electrophoresis gel. If the desired proteins are absent the growth conditions used during the crude extract preparation would require refinement to achieve the expression of all 7 haem synthesis pathway proteins.

When assessing the activity of the crude extract with the haem synthesis pathway proteins expressed the extract is active, however, the yield is significantly less than a standard *E. coli* crude extract. This significant decrease may be due to the depletion of cell resources during cell growth which would lower the availability of resources for the production of the target protein resulting in a lower yield. It is also possible that the low activity observed in this batch is caused by the batch being of low quality compared to previously prepared ones as variability between crude extract batches is well documented in the literature.

Chapter 6 General Discussion

6.1 Optimisation of a CFPS reaction and the synthesis of BM3

The BM3 protein provides an optimal P450 target enzyme for this study due to its high activity and the fact that it is well characterized. We sought to determine if it was possible to produce and study this enzyme using an optimised CFPS system.

In this work we were able to produce an optimised method for a CFPS reaction, this involved the incorporation of two supplements, these being maltodextrin and D-ribose as reported by Sun *et al* (2013), previously not used in the in-house methodology which was adapted from Levine et al (2019). This combined with a crude extract produced using the methods detailed in Sun *et al* (2013) and our standard reaction mix, resulted in the greatest yield of protein. By drying down both maltodextrin and D-ribose and then adding the other components used in our CFPS methodology, a statistically significant increase in protein yield was observed. Maltodextrin has been shown to aid cell-free metabolic activity by indirectly stimulating the recycling of inorganic phosphates, which inhibit the reaction (Caschera and Noireaux, 2016). This happens through accelerating the regeneration of ATP from ADP and inorganic phosphate for glucose kinase activity. Garenne et al (2021) postulate that further supplementation of the reactions with D-ribose improves the regeneration of ATP, but the exact mechanism was not elucidated.

While we were able to produce an optimised CFPS reaction methodology, we were unable to express BM3 using CFPS, it is most likely that this inability lies in the low activity of the prepared crude extract. To address this situation, it would be necessary to continue to refine the process for the construction of the *E. coli* crude extract with the goal being the production of a highly active crude extract similar to those observed in the literature. There is an exceptionally large number of variables to consider when trying to create a highly efficient CFPS system with varying approaches taken by several research groups, to maximise the protein yield of our extract it would be necessary to investigate the effects of a large number of variables have on our reaction system such as the addition of ammonium glutamate or phosphoenolpyruvate as in the Levine *et al* (2019) method both of which play a role in the regeneration of energy in the CFPS system, by further improving energy regeneration in the cell-free system it is possible to increase the duration of protein synthesis to optimise the yield of the target protein. By exploring the variables beneficial additives and reaction conditions could be discovered and implemented. It may also be beneficial to explore a cell-free system

that uses a non-*E. coli*-derived crude extract, such as one which uses *V. natriegens* as the cell line. *V. natriegens* is an appealing option due to its rapid growth rate and has previously been incorporated into CFPS systems, where yields of protein > 260 μg/mL were reproducibly produced, and the extract could be prepared by a single user in 1-2 days (Wiegand *et al.*, 2019).

Another important aspect to explore would be reducing the variability between reactions and crude extract batches, one method to do this would be to bulk-prepare crude extract to reduce variability between batches by eliminating the need for multiple preparations. Similarly, the preparation of CFPS reaction master mixes which could be aliquoted and frozen would reduce the potential for variation when creating the 14-x energy solution or the reaction mix. There is also the potential to use commercial CFPS kits such as PURE. PURE systems offer an alternative means for providing the *E. coli* transcription and translation machinery for a CFPS reaction; this comprises 36 enzymes directly involved in transcription and translation, highly purified 70S ribosomes, tRNA mixtures, 20 amino acids, four nucleoside triphosphates, and buffers (Shimizu *et al.*, 2005).

In contrast with crude extract systems, a PURE system lacks proteases and ribonucleases leading to an absence of side reactions resulting in undesired products in the cell extract (Kuruma and Ueda, 2015). The use of a PURE system allows for higher purity of the target protein and the reaction conditions are, to a degree, more easily manipulated than cell-extract-based CFPS systems (Ohashi *et al.,* 2010). The pure system also allows for easy purification post-translation for products that are affinity-tagged (Ohashi *et al.,* 2010), such as the His_6-tag of the target BM3 protein. This system would reduce the variation between extract batches and would remove the time-consuming steps involved in the preparation of crude extracts. While the PURE system represents a more efficient methodology it is vastly more expensive than the inexpensive crude extract route which also has the further advantage of being a scalable reaction producing high-volume protein yields (Jin *et al.*, 2019; Des Soye *et al.*, 2019). Pure systems are also less productive than crude-extract-based systems, producing around 100 μg/mL (Hillebrecht and Chong, 2008).

To achieve the greatest chance of producing BM3 in CFPS it would also be necessary to address the issues surrounding the cloning of the target plasmid DNA. For example, we attempted, Gibson assembly was not very successful, since only a single clone was sequence verified and found to contain mutations. This suggests the bm3 gene when combined with a constitutive

promoter is toxic to *E. coli* growth and selection for the stages. In addition, we were unable to determine why the pET15b-bm3 produced the protein in *E. coli* BL21 DE3 pLysS cells but was not actively produced in the cell-free system – possibly the extracts had low activity and were variable in performance since the mScarlet synthesis varied more considerably than we normally see in standard cell-free reactions. To this end, it would be necessary either perform a MAXI-prep of the pET15b-*bm3* vector with the aim being to greatly increase the DNA concentration and purity followed by the performance of CFPS reactions supplemented using purified T7RNAP to ensure the activation of the plasmid. Alternatively, efforts could be made to re-clone the BM3 gene using a polymerase less prone to mutation such as a proofreading enzyme like Q5 polymerase. However, attempts to clone the bm3 gene using Q5 polymerase were unsuccessful, the gene is likely toxic when constantly overexpressed with a constitutive promoter. There was no issue with the T7 promoter, which is silent in *E. coli* cloning strains, but Q5 polymerase is a highly efficient enzyme and the fact that no colonies were seen following the Gibson assembly, suggests it was too toxic. To overcome the issue of toxicity a bacterial expression system with tightly regulated gene expression and plasmid copy number such as that developed by Bowers et al (2004) can be used. Once cloned repetition of the Gibson reaction and sequencing would determine if a non-mutated plasmid was produced. Either of these plasmids in conjunction with a highly active crude extract represents the best chance for the cell-free production of BM3.

6.2 Haem biosynthesis

For the synthesis of an active P450 enzyme, a haem functional group is required. To ensure that haem is present in the CFPS reaction we expressed and purified the enzymes required for the biosynthesis of haem and then determine if it was possible to produce haem from its precursor compound ALA and Fe^{2+}. When comparing the emission data to values seen in the literature for the intermediate compounds of the haem biosynthesis pathway either coproporphyrinogen III or Protoporphyrin IX was likely produced, however, the emission data obtained does not provide conclusive evidence as to which stage of the haem synthesis pathway the performed reactions reach; through the use of mass spectroscopy it would be possible to confirm the compounds in the reaction samples and determine which stage of the reaction has been reached and if any of the samples contain haem.

Another consideration is the need to confirm that all the required enzymes for the synthesis of haem are present in the reaction or express each of the proteins individually and combine them following purification. Once combined, activity assays can be performed to determine which stage of the pathway is reached. From here optimisation can be performed so that haem is produced, this optimisation would involve variation in the concentrations of each enzyme and the addition of supplements. With the enzymes expressed individually, it would also be possible to perform individual activity assays to determine if the enzymes are active.

The production of an extract that contains all the enzymes required for haem synthesis is dependent on the ability to produce an active version of the enzyme. In this work, the extract created in which the enzymes required for haem synthesis were expressed showed evidence of only one of the enzymes being present. To induce the expression of the other enzymes it may be necessary to elongate the incubation time, however, this could lead to the cells no longer being in their logarithmic growth phase when harvested meaning that a lower concentration of ribose would potentially be present in the extract meaning it is less active. To address this the use of the cell-free autoinduction media used by Levine *et al* (2019) could lead to a crude extract grown over a long time that maintains high activity. It may also be beneficial to explore the use of *V. natriegens* as this may produce a crude extract where the haem proteins are readily expressed. *V. natriegens* is an appealing target for crude extract production due to its rapid growth rate and high protein yields. Another potential approach would be the addition of purified enzymes to a highly active crude extract or a pure system.

As the crude extract with the enzymes required for haem synthesis was significantly less active than a standard crude extract it is necessary to optimise the reactions to try and increase the protein yield, again there are several potential variables to change to try and achieve the highest possible protein yield.

References

Ahn, J.-H., Chu, H.-S., Kim, T.-W., Oh, I.-S., Choi, C.-Y., Hahn, G.-H., Park, C.-G. and Kim, D.-M. (2005). Cell-free synthesis of recombinant proteins from PCR-amplified genes at a comparable productivity to that of plasmid-based reactions. *Biochemical and Biophysical Research Communications*, 338(3), pp.1346–1352. doi:10.1016/j.bbrc.2005.10.094.

Aiyar, S.E., Gaal, T. and Gourse, R.L. (2002). rRNA Promoter Activity in the Fast-Growing Bacterium Vibrio natriegens. *Journal of Bacteriology*, 184(5), pp.1349–1358. doi:10.1128/jb.184.5.1349-1358.2002.

Arnold, F.H. and Volkov, A.A. (1999). Directed evolution of biocatalysts. *Current Opinion in Chemical Biology*, [online] 3(1), pp.54–59. doi:10.1016/s1367-5931(99)80010-6.

Ayoubi-Joshaghani, M.H., Dianat-Moghadam, H., Seidi, K., Jahanban-Esfahalan, A., Zare, P. and Jahanban-Esfahlan, R. (2020). Cell-free protein synthesis: The transition from batch reactions to minimal cells and microfluidic devices. *Biotechnology and Bioengineering*, 117(4), pp.1204–1229. doi:10.1002/bit.27248.

Barel, M. and Charbit, A. (2017) "Role of glycosylation/deglycolysation processes in francisella tularensis pathogenesis," *Frontiers in Cellular and Infection Microbiology*, 7. Available at: https://doi.org/10.3389/fcimb.2017.00071.

Beale, S.I. (2007). Biosynthesis of Hemes. *EcoSal Plus*, 2(2). doi:10.1128/ecosalplus.3.6.3.11.

Belcher, J., McLean, K.J., Matthews, S., Woodward, L.S., Fisher, K., Rigby, S.E.J., Nelson, D.R., Potts, D., Baynham, M.T., Parker, D.A., Leys, D. and Munro, A.W. (2014). Structure and Biochemical Properties of the Alkene Producing Cytochrome P450 OleTJE (CYP152L1) from the Jeotgalicoccus sp. 8456 Bacterium. *Journal of Biological Chemistry*, 289(10), pp.6535–6550. doi:10.1074/jbc.m113.527325.

Bernhardt, R. (2006). Cytochromes P450 as versatile biocatalysts. *Journal of Biotechnology*, 124(1), pp.128–145. doi:10.1016/j.jbiotec.2006.01.026.

Bowers LM, Lapoint K, Anthony L, Pluciennik A, Filutowicz M. Bacterial expression system with tightly regulated gene expression and plasmid copy number. Gene. 2004 Sep 29;340(1):11-8. doi: 10.1016/j.gene.2004.06.012. PMID: 15556290.

Cai, Q., Hanson, J.A., Steiner, A.R., Tran, C., Masikat, M.R., Chen, R., Zawada, J.F., Sato, A.K., Hallam, T.J. and Yin, G. (2015). A simplified and robust protocol for immunoglobulin expression in escherichia Coli cell-free protein synthesis systems. *Biotechnology Progress*, 31(3), pp.823–831. doi:10.1002/btpr.2082.

Calhoun, K.A. and Swartz, J.R. (2005). Energizing cell-free protein synthesis with glucose metabolism. *Biotechnology and Bioengineering*, [online] 90(5), pp.606–613. doi:10.1002/bit.20449.

Calhoun, K.A. and Swartz, J.R. (2007). Energy Systems for ATP Regeneration in Cell-Free Protein Synthesis Reactions. *In Vitro Transcription and Translation Protocols*, 375, pp.3–17. doi:10.1007/978-1-59745-388-2_1.

Carlson, E.D., Gan, R., Hodgman, C.E. and Jewett, M.C. (2012). Cell-Free Protein Synthesis: Applications Come of Age. *Biotechnology advances*, [online] 30(5), pp.1185–1194. doi:10.1016/j.biotechadv.2011.09.016.

Carmichael, A.B. and Wong, L.-L. (2001). Protein engineering of bacillus megateriumCYP102. *European Journal of Biochemistry*, 268(10), pp.3117–3125. doi:10.1046/j.1432-1327.2001.02212.x.

Caschera, F. and Noireaux, V. (2014). Synthesis of 2.3 mg/ml of protein with an all Escherichia coli cell-free transcription-translation system. *Biochimie*, 99, pp.162–168. doi:10.1016/j.biochi.2013.11.025.

Chartrain, M., Salmon, P.M., Robinson, D.K. and Buckland, B.C. (2000). Metabolic engineering and directed evolution for the production of pharmaceuticals. *Current Opinion in Biotechnology*, 11(2), pp.209–214. doi:10.1016/s0958-1669(00)00081-1.

Chen, H., Venkat, S., McGuire, P., Gan, Q. and Fan, C. (2018). Recent Development of Genetic Code Expansion for Posttranslational Modification Studies. *Molecules*, 23(7), p.1662. doi:10.3390/molecules23071662.

Chong, S. (2014). Overview of Cell-Free Protein Synthesis: Historic Landmarks, Commercial Systems, and Expanding Applications. *Current Protocols in Molecular Biology*, [online] 108(1). doi:10.1002/0471142727.mb1630s108.

Cobb, R.E., Chao, R. and Zhao, H. (2013). Directed evolution: Past, present, and future. *AIChE Journal*, [online] 59(5), pp.1432–1440. doi:10.1002/aic.13995.

Coco, W.M., Levinson, W.E., Crist, M.J., Hektor, H.J., Darzins, A., Pienkos, P.T., Squires, C.H. and Monticello, D.J. (2001). DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nature Biotechnology*, 19(4), pp.354–359. doi:10.1038/86744.

Cupp-Vickery, J.R. and Poulos, T.L. (1995). Structure of cytochrome P450eryF involved in erythromycin biosynthesis. *Nature Structural & Molecular Biology*, 2(2), pp.144–153. doi:10.1038/nsb0295-144.

DaCosta, R.S., Andersson, H. and Wilson, B.C. (2007). Molecular Fluorescence Excitation-Emission Matrices Relevant to Tissue Spectroscopy¶. *Photochemistry and Photobiology*, 78(4), pp.384–392. doi:10.1562/0031-8655(2003)0780384mfemrt2.0.co2.

Daff, S.N., Chapman, S.K., Turner, K.L., Holt, R.A., Govindaraj, S., Poulos, T.L. and Munro, A.W. (1997). Redox Control of the Catalytic Cycle of Flavocytochrome P-450 BM3. *Biochemistry*, 36(45), pp.13816–13823. doi:10.1021/bi971085s.

Dailey, H.A. (2002). Terminal steps of haem biosynthesis. *Biochemical Society Transactions*, 30(4), pp.590–595. doi:10.1042/bst0300590.

Dall, E. and Brandstetter, H. (2016). Structure and function of legumain in health and disease. *Biochimie*, 122, pp.126–150. doi:10.1016/j.biochi.2015.09.022.

David, G., Fogeron, M.-L., Montserret, R., Lecoq, L., Page, A., Delolme, F., Nassal, M. and Böckmann, A. (2019). Phosphorylation and Alternative Translation on Wheat Germ Cell-Free Protein Synthesis of the DHBV Large Envelope Protein. *Frontiers in Molecular Biosciences*, 6. doi:10.3389/fmolb.2019.00138.

Denisov, I.G., Makris, T.M., Sligar, S.G. and Schlichting, I. (2005). Structure and Chemistry of Cytochrome P450. *Chemical Reviews*, 105(6), pp.2253–2278. doi:10.1021/cr0307143.

Des Soye, B.J., Davidson, S.R., Weinstock, M.T., Gibson, D.G. and Jewett, M.C. (2018). Establishing a High-Yielding Cell-Free Protein Synthesis Platform Derived fromVibrio natriegens. *ACS Synthetic Biology*, 7(9), pp.2245–2255. doi:10.1021/acssynbio.8b00252.

Des Soye, B.J., Gerbasi, V.R., Thomas, P.M., Kelleher, N.L. and Jewett, M.C. (2019). A Highly Productive, One-Pot Cell-Free Protein Synthesis Platform Based on Genomically Recoded Escherichia coli. *Cell Chemical Biology*, 26(12), pp.1743-1754.e9. doi:10.1016/j.chembiol.2019.10.008.

Dondapati, S.K. *et al.* (2020) "Cell-free protein synthesis: A promising option for future drugdevelopment," *BioDrugs*, 34(3), pp. 327–348. Available at: https://doi.org/10.1007/s40259-020-00417-y.

Dondapati, S., Wüstenhagen, D., Strauch, E. and Kubick, S., 2017. Cell-free production of pore-forming toxins: Functional analysis of thermostable direct hemolysin fromVibrio parahaemolyticus. *Engineering in Life Sciences*, 18(2), pp.140-148.

Dopp, B.J.L., Tamiev, D.D. and Reuel, N.F. (2019). Cell-free supplement mixtures: Elucidating the history and biochemical utility of additives used to support in vitro protein synthesis in *E. coli* extract. *Biotechnology Advances*, 37(1), pp.246–258. doi:10.1016/j.biotechadv.2018.12.006.

Dopp, J.L., Jo, Y.R. and Reuel, N.F. (2019). Methods to reduce variability in *E. coli*-based cell-free protein expression experiments. *Synthetic and Systems Biotechnology*, [online] 4(4), pp.204–211. doi:10.1016/j.synbio.2019.10.003.

Estabrook, R.W. (2003). A PASSION FOR P450s (REMEMBRANCES OF THE EARLY HISTORY OF RESEARCH ON CYTOCHROME P450). *Drug Metabolism and Disposition*, 31(12), pp.1461–1473. doi:10.1124/dmd.31.12.1461.

Focke, P.J., Hein, C., Hoffmann, B., Matulef, K., Bernhard, F., Dötsch, V. and Valiyaveetil, F.I. (2016). Combining *in Vitro* Folding with Cell-Free Protein Synthesis for Membrane Protein Expression. *Biochemistry*, 55(30), pp.4212–4219. doi:10.1021/acs.biochem.6b00488.

Fujiwara, K. and Doi, N. (2016). Biochemical Preparation of Cell Extract for Cell-Free Protein Synthesis without Physical Disruption. *PLoS ONE*, [online] 11(4). doi:10.1371/journal.pone.0154614.

Fujiwara, K. and Nomura, S.M. (2013). Condensation of an Additive-Free Cell Extract to Mimic the Conditions of Live Cells. *PLoS ONE*, 8(1), p.e54155. doi:10.1371/journal.pone.0054155.

Garenne, D., Thompson, S., Brisson, A., Khakimzhan, A. and Noireaux, V. (2021). The all-*E. coli*TXTL toolbox 3.0: new capabilities of a cell-free synthetic biology platform. *Synthetic Biology*, 6(1). doi:10.1093/synbio/ysab017.

Gilardi, G. and Di Nardo, G. (2016). Heme iron centers in cytochrome P450: structure and catalytic activity. *Rendiconti Lincei*, 28(S1), pp.159–167. doi:10.1007/s12210-016-0565-z.

Girvan, H.M., Marshall, K.R., Lawson, R.J., Leys, D., Joyce, M.G., Clarkson, J., Smith, W.E., Cheesman, M.R. and Munro, A.W. (2004). Flavocytochrome P450 BM3 Mutant A264E Undergoes Substrate-dependent Formation of a Novel Heme Iron Ligand Set. *Journal of Biological Chemistry*, 279(22), pp.23274–23286. doi:10.1074/jbc.m401716200.

Girvan, H.M. and Munro, A.W. (2016). Applications of microbial cytochrome P450 enzymes in biotechnology and synthetic biology. *Current Opinion in Chemical Biology*, 31, pp.136–145. doi:10.1016/j.cbpa.2016.02.018.

Gregorio, N.E., Levine, M.Z. and Oza, J.P. (2019). A User's Guide to Cell-Free Protein Synthesis. *Methods and Protocols*, [online] 2(1). doi:10.3390/mps2010024.

Hahn, G.-H. and Kim, D.-M. (2006). Production of milligram quantities of recombinant proteins from PCR-amplified DNAs in a continuous-exchange cell-free protein synthesis system. *Analytical Biochemistry*, 355(1), pp.151–153. doi:10.1016/j.ab.2006.05.004.

Hendricksen, A., Hook, B. and Schagat, T. (2010). *RNase Contamination Happens Recombinant RNasin Inhibitor Can Safeguard Your Samples*. [online] www.promega.com. Available at: https://www.promega.co.uk/resources/pubhub/rnase-contamination-happens-recombinant-rnasin-inhibitor-can-safeguard-your-samples/ [Accessed 8 Jun. 2022].

Hillebrecht, J.R. and Chong, S. (2008). A comparative study of protein synthesis in in vitro systems: from the prokaryotic reconstituted to the eukaryotic extract-based. *BMC Biotechnology*, [online] 8, p.58. doi:10.1186/1472-6750-8-58.

Hodgman, C.E. and Jewett, M.C. (2012). Cell-free synthetic biology: Thinking outside the cell. *Metabolic Engineering*, 14(3), pp.261–269. doi:10.1016/j.ymben.2011.09.002.

Hoff, J., Daniel, B., Stukenberg, D., Thuronyi, B.W., Waldminghaus, T. and Fritz, G. (2020). Vibrio natriegens: an ultrafast-growing marine bacterium as emerging synthetic biology chassis. *Environmental Microbiology*. doi:10.1111/1462-2920.15128.

Hoffart, E., Grenz, S., Lange, J., Nitschel, R., Müller, F., Schwentner, A., Feith, A., Lenfers-Lücker, M., Takors, R. and Blombach, B. (2017). High Substrate Uptake Rates Empower Vibrio natriegens as Production Host for Industrial Biotechnology. *Applied and Environmental Microbiology*, 83(22). doi:10.1128/aem.01614-17.

Jeffreys, L.N., Pacholarz, K.J., Johannissen, L.O., Girvan, H.M., Barran, P.E., Voice, M.W. and Munro, A.W. (2020). Characterization of the structure and interactions of P450 BM3 using hybrid mass spectrometry approaches. *Journal of Biological Chemistry*, 295(22), pp.7595–7607. doi:10.1074/jbc.ra119.011630.

Jeffreys, L.N., Poddar, H., Golovanova, M., Levy, C.W., Girvan, H.M., McLean, K.J., Voice, M.W., Leys, D. and Munro, A.W. (2019). Novel insights into P450 BM3 interactions with FDA-approved antifungal azole drugs. *Scientific Reports*, [online] 9(1), p.1577. doi:10.1038/s41598-018-37330-y.

Jewett, M.C., Calhoun, K.A., Voloshin, A., Wuu, J.J. and Swartz, J.R. (2008). An integrated cell-free metabolic platform for protein production and synthetic biology. *Molecular Systems Biology*, 4(1), p.220. doi:10.1038/msb.2008.57.

Jewett, M.C. and Swartz, J.R. (2004). Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. *Biotechnology and Bioengineering*, 86(1), pp.19–26. doi:10.1002/bit.20026.

Jin, X., Kightlinger, W. and Hong, S.H. (2019). Optimizing Cell-Free Protein Synthesis for Increased Yield and Activity of Colicins. *Methods and Protocols*, 2(2), p.28. doi:10.3390/mps2020028.

Joyce, M.G., Ekanem, I.S., Roitel, O., Dunford, A.J., Neeli, R., Girvan, H.M., Baker, G.J., Curtis, R.A., Munro, A.W. and Leys, D. (2012). The crystal structure of the FAD/NADPH-binding domain of flavocytochrome   P450   BM3. *FEBS Journal*, 279(9), pp.1694–1706. doi:10.1111/j.1742-4658.2012.08544.x.

Joyce, M.G., Girvan, H.M., Munro, A.W. and Leys, D. (2004). A Single Mutation in Cytochrome P450 BM3 Induces the Conformational Rearrangement Seen upon Substrate Binding in the Wild-type Enzyme. *Journal of Biological Chemistry*, 279(22), pp.23287–23293. doi:10.1074/jbc.m401717200.

Karim, A.S. and Jewett, M.C. (2018). Cell-Free Synthetic Biology for Pathway Prototyping. *Methods in Enzymology*, pp.31–57. doi:10.1016/bs.mie.2018.04.029.

Katzen, F., Chang, G. and Kudlicki, W. (2005). The past, present and future of cell-free protein synthesis. *Trends in Biotechnology*, 23(3), pp.150–156. doi:10.1016/j.tibtech.2005.01.003.

Kigawa, T., Yabuki, T., Matsuda, N., Matsuda, T., Nakajima, R., Tanaka, A. and Yokoyama, S. (2004). Preparation of Escherichia coli cell extract for highly productive cell-free protein expression. *Journal of Structural and Functional Genomics*, 5(1/2), pp.63–68. doi:10.1023/b:jsfg.0000029204.57846.7d.

Kightlinger, W., Duncker, K.E., Ramesh, A., Thames, A.H., Natarajan, A., Stark, J.C., Yang, A., Lin, L., Mrksich, M., DeLisa, M.P. and Jewett, M.C. (2019). A cell-free biosynthesis platform for modular construction of protein glycosylation pathways. *Nature Communications*, 10(1). doi:10.1038/s41467-019-12024-9.

Kille, S., Zilly, F.E., Acevedo, J.P. and Reetz, M.T. (2011). Regio- and stereoselectivity of P450-catalysed hydroxylation of steroids controlled by laboratory evolution. *Nature Chemistry*, 3(9), pp.738–743. doi:10.1038/nchem.1113.

Kim, D.-M. . and Swartz, J.R. (2000). Prolonging Cell-Free Protein Synthesis by Selective Reagent Additions. *Biotechnology Progress*, 16(3), pp.385–390. doi:10.1021/bp000031y.

Kim, H.-C. and Kim, D.-M. (2009). Methods for energizing cell-free protein synthesis. *Journal of Bioscience and Bioengineering*, 108(1), pp.1–4. doi:10.1016/j.jbiosc.2009.02.007.

Kim, T., Keum, J., Oh, I., Choi, C., Kim, H. and Kim, D. (2007). An economical and highly productive cell-free protein synthesis system utilizing fructose-1,6-bisphosphate as an energy source. *Journal of Biotechnology*, [online] 130(4), pp.389–393. doi:10.1016/j.jbiotec.2007.05.002.

Kim, T.-W., Keum, J.-W., Oh, I.-S., Choi, C.-Y., Park, C.-G. and Kim, D.-M. (2006). Simple procedures for the construction of a robust and cost-effective cell-free protein synthesis system. *Journal of Biotechnology*, 126(4), pp.554–561. doi:10.1016/j.jbiotec.2006.05.014.

Krüger, A., Mueller, A.P., Rybnicky, G.A., Engle, N.L., Yang, Z.K., Tschaplinski, T.J., Simpson, S.D., Köpke, M. and Jewett, M.C. (2020). Development of a clostridia-based cell-free system for prototyping genetic parts and metabolic pathways. *Metabolic Engineering*, 62, pp.95–105. doi:10.1016/j.ymben.2020.06.004.

Kuruma, Y. and Ueda, T. (2015). The PURE system for the cell-free synthesis of membrane proteins. *Nature Protocols*, 10(9), pp.1328–1344. doi:10.1038/nprot.2015.082.

Kwon, Y.-C. and Jewett, M.C. (2015). High-throughput preparation methods of crude extract for robust cell-free protein synthesis. *Scientific Reports*, 5(1). doi:10.1038/srep08663.

Lee, K.-H. and Kim, D.-M. (2018). Recent advances in development of cell-free protein synthesis systems for fast and efficient production of recombinant proteins. *FEMS Microbiology Letters*, 365(17). doi:10.1093/femsle/fny174.

Lesley, S.A., Brow, M.A. and Burgess, R.R. (1991). Use of in vitro protein synthesis from polymerase chain reaction-generated templates to study interaction of Escherichia coli transcription factors with core RNA polymerase and for epitope mapping of monoclonal antibodies. *Journal of Biological Chemistry*, 266(4), pp.2632–2638. doi:10.1016/s0021-9258(18)52291-2.

Levine, M.Z., Gregorio, N.E., Jewett, M.C., Watts, K.R. and Oza, J.P. (2019). Escherichia coli-Based Cell-Free Protein Synthesis: Protocols for a robust, flexible, and accessible platform technology. *Journal of Visualized Experiments*, (144). doi:10.3791/58882.

Levine, M.Z., So, B., Mullin, A.C., Fanter, R., Dillard, K., Watts, K.R., La Frano, M.R. and Oza, J.P. (2020). Activation of Energy Metabolism through Growth Media Reformulation Enables a 24-Hour Workflow for Cell-Free Expression. *ACS Synthetic Biology*, 9(10), pp.2765–2774. doi:10.1021/acssynbio.0c00283.

Li, J., Lawton, T.J., Kostecki, J.S., Nisthal, A., Fang, J., Mayo, S.L., Rosenzweig, A.C. and Jewett, M.C. (2015). Cell-free protein synthesis enables high yielding synthesis of an active multicopper oxidase. *Biotechnology Journal*, 11(2), pp.212–218. doi:10.1002/biot.201500030.

Li, J., Wang, H., Kwon, Y.-C. and Jewett, M.C. (2017). Establishing a high yielding streptomyces-based cell-free protein synthesis system. *Biotechnology and Bioengineering*, 114(6), pp.1343–1353. doi:10.1002/bit.26253.

Liebeton, K., Zonta, A., Schimossek, K., Nardini, M., Lang, D., Dijkstra, B.W., Reetz, M.T. and Jaeger, K.-E. (2000). Directed evolution of an enantioselective lipase. *Chemistry & Biology*, 7(9), pp.709–718. doi:10.1016/s1074-5521(00)00015-6.

Lothrop, A.P., Torres, M.P. and Fuchs, S.M. (2013). Deciphering post-translational modification codes. *FEBS Letters*, 587(8), pp.1247–1257. doi:10.1016/j.febslet.2013.01.047.

Marshall, R., Maxwell, C.S., Collins, S.P., Beisel, C.L. and Noireaux, V. (2017). Short DNA containing χ sites enhances DNA stability and gene expression in *E. coli* cell-free transcription-translation systems. *Biotechnology and Bioengineering*, 114(9), pp.2137–2141. doi:10.1002/bit.26333.

Mathews, M.A.A. (2001). Crystal structure of human uroporphyrinogen III synthase. *The EMBO Journal*, 20(21), pp.5832–5839. doi:10.1093/emboj/20.21.5832.

McInerney, P., Adams, P. and Hadi, M.Z. (2014). Error Rate Comparison during Polymerase Chain Reaction by DNA Polymerase. *Molecular Biology International*, 2014, pp.1–8. doi:10.1155/2014/287430.

McManus, J.B., Emanuel, P.A., Murray, R.M. and Lux, M.W. (2019). A method for cost-effective and rapid characterization of engineered T7-based transcription factors by cell-free protein synthesis reveals insights into the regulation of T7 RNA polymerase-driven expression. *Archives of Biochemistry and Biophysics*, 674, p.108045. doi:10.1016/j.abb.2019.07.010.

Miles, J.S., Munro, A.W., Rospendowski, B.N., Smith, W.E., McKnight, J. and Thomson, A.J. (1992). Domains of the catalytically self-sufficient cytochrome P-450 BM-3. Genetic construction, overexpression, purification and spectroscopic characterization. *Biochemical Journal*, 288(2), pp.503–509. doi:10.1042/bj2880503.

Mizutani, M. and Sato, F. (2011). Unusual P450 reactions in plant secondary metabolism. *Archives of Biochemistry and Biophysics*, 507(1), pp.194–203. doi:10.1016/j.abb.2010.09.026.

Morlock, L.K., Böttcher, D. and Bornscheuer, U.T. (2017). Simultaneous detection of NADPH consumption and H_2O_2 production using the Ampliflu^{TM} Red assay for screening of P450 activities and uncoupling. *Applied Microbiology and Biotechnology*, 102(2), pp.985–994. doi:10.1007/s00253-017-8636-3.

Müller, D., Bayer, K. and Mattanovich, D. (2006). Potentials and limitations of prokaryotic and eukaryotic expression systems for recombinant protein production – a comparative view. *Microbial Cell Factories*, 5(Suppl 1), p.P61. doi:10.1186/1475-2859-5-s1-p61.

Munro, A.W., McLean, K.J., Grant, J.L. and Makris, T.M. (2018). Structure and function of the cytochrome P450 peroxygenase enzymes. *Biochemical Society Transactions*, 46(1), pp.183–196. doi:10.1042/bst20170218.

Nagaraj, V.H., Greene, J.M., Sengupta, A.M. and Sontag, E.D. (2017). Translation inhibition and resource balance in the TX-TL cell-free gene expression system. *Synthetic Biology*, 2(1). doi:10.1093/synbio/ysx005.

Nelson, D.R. (2004). Cytochrome P450 Nomenclature, 2004. *Cytochrome P450 Protocols*, pp.1–10. doi:10.1385/1-59259-998-2:1.

Neelagandan, N. et al. (2020) "What determines eukaryotic translation elongation: Recent molecular and quantitative analyses of protein synthesis," Open Biology, 10(12), p. 200292. Available at: https://doi.org/10.1098/rsob.200292.

Nirenberg, M.W. and Matthaei, J.H. (1961). The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proceedings of the National Academy of Sciences*, 47(10), pp.1588–1602. doi:10.1073/pnas.47.10.1588.

NOBLE, M.A., MILES, C.S., CHAPMAN, S.K., LYSEK, D.A., MACKAY, A.C., REID, G.A., HANZLIK, R.P. and MUNRO, A.W. (1999). Roles of key active-site residues in flavocytochrome P450 BM3. *Biochemical Journal*, 339(2), pp.371–379. doi:10.1042/bj3390371.

Ohashi, H., Kanamori, T., Shimizu, Y. and Ueda, T. (2010). A Highly Controllable Reconstituted Cell-Free System -a Breakthrough in Protein Synthesis Research. *Current Pharmaceutical Biotechnology*, 11(3), pp.267–271. doi:10.2174/138920110791111889.

Ortiz, P.R. (2005). *Cytochrome P450: structure, mechanism, and biochemistry*. New York: Kluwer Academic/Plenum Publishers.

Oza, J., Aerni, H., Pirman, N., Barber, K., ter Haar, C., Rogulina, S., Amrofell, M., Isaacs, F., Rinehart, J. and Jewett, M., 2015. Robust production of recombinant phosphoproteins using cell-free protein synthesis. *Nature Communications*, 6(1).

Panek, H. and O'Brian, M.R. (2002). A whole genome view of prokaryotic haem biosynthesis. *Microbiology*, 148(8), pp.2273–2282. doi:10.1099/00221287-148-8-2273.

Pang, Y.L., Poruri, K. and Martinis, S.A. (2014) "trna Synthetase: trna aminoacylation and beyond," WIREs RNA, 5(4), pp. 461–480. Available at: https://doi.org/10.1002/wrna.1224.

Peters, M.W., Meinhold, P., Glieder, A. and Arnold, F.H. (2003). Regio- and Enantioselective Alkane Hydroxylation with Engineered Cytochromes P450 BM-3. *Journal of the American Chemical Society*, 125(44), pp.13442–13450. doi:10.1021/ja0303790.

Pranawidjaja, S., Choi, S.-I., Lay, B.W. and Kim, P. (2015). Analysis of Heme Biosynthetic Pathways in a Recombinant Escherichia coli. *Journal of Microbiology and Biotechnology*, 25(6), pp.880–886. doi:10.4014/jmb.1411.11050.

Ramm, F., Dondapati, S.K., Thoring, L., Zemella, A., Wüstenhagen, D.A., Frentzel, H., Stech, M. and Kubick, S. (2020). Mammalian cell-free protein expression promotes the functional characterization of the tripartite non-hemolytic enterotoxin from Bacillus cereus. *Scientific Reports*, 10(1). doi:10.1038/s41598-020-59634-8.

Ray-Soni, A., Bellecourt, M.J. and Landick, R. (2016) "Mechanisms of bacterial transcription termination: All good things must end," *Annual Review of Biochemistry*, 85(1), pp. 319–347. Available at: https://doi.org/10.1146/annurev-biochem-060815-014844.

Rittle, J. and Green, M.T. (2010). Cytochrome P450 Compound I: Capture, Characterization, and C-H Bond Activation Kinetics. *Science*, 330(6006), pp.933–937. doi:10.1126/science.1193478.

Rolf, J., Rosenthal, K. and Lütz, S. (2019). Application of Cell-Free Protein Synthesis for Faster Biocatalyst Development. *Catalysts*, [online] 9(2), p.190. doi:10.3390/catal9020190.

Rosano, G. and Ceccarelli, E., 2009. Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted Escherichia coli strain. *Microbial Cell Factories*, 8(1), p.41.

Rosano, G.L. and Ceccarelli, E.A. (2014). Recombinant protein expression in Escherichia coli: Advances and challenges. *Frontiers in Microbiology*, 5(172). doi:10.3389/fmicb.2014.00172.

Rosenblum, G. and Cooperman, B.S. (2014). Engine out of the Chassis: Cell-Free Protein Synthesis and its Uses. *FEBS letters*, [online] 588(2), pp.261–268. doi:10.1016/j.febslet.2013.10.016.

Ruettinger, R.T., Wen, L.P. and Fulco, A.J. (1989). Coding Nucleotide, 5′ Regulatory, and Deduced Amino Acid Sequences of P-450BM-3, a Single Peptide Cytochrome P-450:NADPH-P-450 Reductase from Bacillus megaterium. *Journal of Biological Chemistry*, 264(19), pp.10987–10995. doi:10.1016/s0021-9258(18)60416-8.

Santos, G. de A., Dhoke, G.V., Davari, M.D., Ruff, A.J. and Schwaneberg, U. (2019). Directed Evolution of P450 BM3 towards Functionalization of Aromatic O-Heterocycles. *International Journal of Molecular Sciences*, 20(13), p.3353. doi:10.3390/ijms20133353.

Schinn, S.-M., Broadbent, A., Bradley, W.T. and Bundy, B.C. (2016). Protein synthesis directly from PCR: progress and applications of cell-free protein synthesis with linear DNA. *New Biotechnology*, 33(4), pp.480–487. doi:10.1016/j.nbt.2016.04.002.

Schmidt-Dannert, C. and Arnold, F.H. (1999). Directed evolution of industrial enzymes. *Trends in Biotechnology*, 17(4), pp.135–136. doi:10.1016/s0167-7799(98)01283-9.

Seng Wong, T., Arnold, F.H. and Schwaneberg, U. (2004). Laboratory evolution of cytochrome P450 BM-3 monooxygenase for organic cosolvents. *Biotechnology and Bioengineering*, 85(3), pp.351–358. doi:10.1002/bit.10896.

Seo, I., Tseng, S.H., Cula, G.O., Bargo, P.R. and Kollias, N. (2009). Fluorescence spectroscopy for endogenous porphyrins in human facial skin. *Photonic Therapeutics and Diagnostics V*. doi:10.1117/12.811913.

Sevrioukova, I.F., Li, H., Zhang, H., Peterson, J.A. and Poulos, T.L. (1999). Structure of a cytochrome P450–redox partner electron-transfer complex. *Proceedings of the National Academy of Sciences*, 96(5), pp.1863–1868. doi:10.1073/pnas.96.5.1863.

Sherman, F., Stewart, J.W. and Tsunasawa, S. (1985) "Methionine or not methionine at the beginning of a protein," BioEssays, 3(1), pp. 27–31. Available at: https://doi.org/10.1002/bies.950030108.

Shimizu, Y., Kanamori, T. and Ueda, T. (2005). Protein synthesis by pure translation systems. *Methods*, 36(3), pp.299–304. doi:10.1016/j.ymeth.2005.04.006.

Shin, J. and Noireaux, V. (2010). Efficient cell-free expression with the endogenous *E. coli* RNA polymerase and sigma factor 70. *Journal of Biological Engineering*, 4(1), p.8. doi:10.1186/1754-1611-4-8.

Shoji, O., Aiba, Y. and Watanabe, Y. (2019). Hoodwinking Cytochrome P450BM3 into Hydroxylating Non-Native Substrates by Exploiting Its Substrate Misrecognition. *Accounts of Chemical Research*, 52(4), pp.925–934. doi:10.1021/acs.accounts.8b00651.

Siegal-Gaskins, D., Tuza, Z.A., Kim, J., Noireaux, V. and Murray, R.M. (2014). Gene Circuit Performance Characterization and Resource Usage in a Cell-Free 'Breadboard'. *ACS Synthetic Biology*, 3(6), pp.416–425. doi:10.1021/sb400203p.

Silverman, A.D., Kelley-Loughnane, N., Lucks, J.B. and Jewett, M.C. (2018). Deconstructing Cell-Free Extract Preparation forin VitroActivation of Transcriptional Genetic Circuitry. *ACS Synthetic Biology*, 8(2), pp.403–414. doi:10.1021/acssynbio.8b00430.

Sitaraman, K., Esposito, D., Klarmann, G., Le Grice, S.F., Hartley, J.L. and Chatterjee, D.K. (2004). A novel cell-free protein synthesis system. *Journal of Biotechnology*, 110(3), pp.257–263. doi:10.1016/j.jbiotec.2004.02.014.

Spradlin, J., Lee, D., Mahadevan, S., Mahomed, M., Tang, L., Lam, Q., Colbert, A., Shafaat, O.S., Goodin, D., Kloos, M., Kato, M. and Cheruzel, L.E. (2016). Insights into an efficient light-driven hybrid P450 BM3 enzyme from crystallographic, spectroscopic and biochemical studies. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1864(12), pp.1732–1738. doi:10.1016/j.bbapap.2016.09.005.

Sun, Z.Z., Hayes, C.A., Shin, J., Caschera, F., Murray, R.M. and Noireaux, V. (2013). Protocols for Implementing an Escherichia coli Based TX-TL Cell-Free Expression System for Synthetic Biology. *Journal of Visualized Experiments*, (79). doi:10.3791/50762.

Surzycki, S. (2000). General Aspects of DNA Isolation and Purification. *Basic Techniques in Molecular Biology*, pp.1–32. doi:10.1007/978-3-642-56968-5_1.

Swartz, J. (2006). Developing cell-free biology for industrial applications. *Journal of Industrial Microbiology & Biotechnology*, 33(7), pp.476–485. doi:10.1007/s10295-006-0127-y.

Takahashi, K., Sato, G., Doi, N. and Fujiwara, K. (2021). A Relationship between NTP and Cell Extract Concentration for Cell-Free Protein Expression. *Life*, 11(3), p.237. doi:10.3390/life11030237.

Vary, P.S., Biedendieck, R., Fuerch, T., Meinhardt, F., Rohde, M., Deckwer, W.-D. and Jahn, D. (2007). Bacillus megaterium—from simple soil bacterium to industrial protein production host. *Applied Microbiology and Biotechnology*, 76(5), pp.957–967. doi:10.1007/s00253-007-1089-3.

Vredenburg, G., den Braver-Sewradj, S., van Vugt-Lussenburg, B.M.A., Vermeulen, N.P.E., Commandeur, J.N.M. and Vos, J.C. (2015). Activation of the anticancer drugs cyclophosphamide and ifosfamide by cytochrome P450 BM3 mutants. *Toxicology Letters*, [online] 232(1), pp.182–192. doi:10.1016/j.toxlet.2014.11.005.

Wang, Y., Li, C., Khan, Md.R.I., Wang, Y., Ruan, Y., Zhao, B., Zhang, B., Ma, X., Zhang, K., Zhao, X., Ye, G., Guo, X., Feng, G., He, L. and Ma, G. (2016). An Engineered Rare

Codon Device for Optimization of Metabolic Pathways. *Scientific Reports*, 6(1). doi:10.1038/srep20608.

Wang, Y. and Zhang, Y-H.P. (2009). Cell-free protein synthesis energized by slowly-metabolized maltodextrin. *BMC Biotechnology*, 9(1), p.58. doi:10.1186/1472-6750-9-58.

Wang, Z.J., Renata, H., Peck, N.E., Farwell, C.C., Coelho, P.S. and Arnold, F.H. (2014). Improved Cyclopropanation Activity of Histidine-Ligated Cytochrome P450 Enables the Enantioselective Formal Synthesis of Levomilnacipran. *Angewandte Chemie*, 126(26), pp.6928–6931. doi:10.1002/ange.201402809.

Warren, M.J., Cooper, J.B., Wood, S.P. and Shoolingin-Jordan, P.M. (1998). Lead poisoning, haem synthesis and 5-aminolaevulinic acid dehydratase. *Trends in Biochemical Sciences*, 23(6), pp.217–221. doi:10.1016/s0968-0004(98)01219-5.

Whittaker, J.W. (2012). Cell-free protein synthesis: the state of the art. *Biotechnology Letters*, 35(2), pp.143–152. doi:10.1007/s10529-012-1075-4.

Wiegand, D.J., Lee, H.H., Ostrov, N. and Church, G.M. (2018). Establishing a Cell-Free

Vibrio natriegens Expression System. *ACS Synthetic Biology*, 7(10), pp.2475–2479. doi:10.1021/acssynbio.8b00222.

Venkat S, Chen H, Gan Q, Fan C. The Application of Cell-Free Protein Synthesis in Genetic Code Expansion for Post-translational Modifications. Front Pharmacol. 2019 Mar 20; 10:248. doi: 10.3389/fphar.2019.00248. PMID: 30949051; PMCID: PMC6436179.

Vilkhovoy, M. et al. (2018) "Sequence specific modeling of *E. coli* cell-free protein synthesis," ACS Synthetic Biology, 7(8), pp. 1844–1857. Available at: https://doi.org/10.1021/acssynbio.7b00465.

Woodrow, K.A. and Swartz, J.R. (2007). A sequential expression system for high-throughput functional genomic analysis. *PROTEOMICS*, 7(21), pp.3870–3879. doi:10.1002/pmic.200700471.

Worst, E.G., Exner, M.P., De Simone, A., Schenkelberger, M., Noireaux, V., Budisa, N. and Ott, A. (2016). Residue-specific Incorporation of Noncanonical Amino Acids into Model

Proteins Using an Escherichia coli Cell-free Transcription-translation System. *Journal of Visualized Experiments*, (114). doi:10.3791/54273.

Yin, G., Garces, E.D., Yang, J., Zhang, J., Tran, C., Steiner, A.R., Roos, C., Bajad, S., Hudak, S., Penta, K., Zawada, J., Pollitt, S. and Murray, C.J. (2012). Aglycosylated antibodies and antibody fragments produced in a scalable in vitro transcription-translation system. *mAbs*, [online] 4(2), pp.217–225. doi:10.4161/mabs.4.2.19202.

Zhang, J., Kang, Z., Chen, J. and Du, G. (2015). Optimization of the heme biosynthesis pathway for the production of 5-aminolevulinic acid in Escherichia coli. *Scientific Reports*, 5(1). doi:10.1038/srep08584.

Zhang, L., Lin, X., Wang, T., Guo, W. and Lu, Y. (2021). Development and comparison of cell-free protein synthesis systems derived from typical bacterial chassis. *Bioresources and Bioprocessing*, [online] 8(1). doi:10.1186/s40643-021-00413-2.

Zimmerman, E.S., Heibeck, T.H., Gill, A., Li, X., Murray, C.J., Madlansacay, M.R., Tran, C., Uter, N.T., Yin, G., Rivers, P.J., Yam, A.Y., Wang, W.D., Steiner, A.R., Bajad, S.U., Penta, K., Yang, W., Hallam, T.J., Thanos, C.D. and Sato, A.K. (2014). Production of Site-Specific Antibody–Drug Conjugates Using Optimized Non-Natural Amino Acids in a Cell-Free Expression System. *Bioconjugate Chemistry*, 25(2), pp.351–361. doi:10.1021/bc400490

Appendix

Figure 3.1 Plasmid maps of pTU1-SP44-*mScarlet*-Flga_CCXXCC_dBroccoli_B0015 and pET15b-*mScarlet*. <mark>The top plasmid map shows pTU1-SP44-*mScarlet*-Flga_CCXXCC_dBroccoli_B0015, this plasmid has a T7 promoter, an N-terminal His-tag, a tetracysteine tag and is ampicillin resistant. The plasmid contains the gene which codes for the production of mScarlet red fluorescence protein. The bottom plasmid shows pET15b-*mScarlet*, this plasmid has an SP44 promoter, an N-terminal His-tag, a tetracysteine tag and is ampicillin resistant. The plasmid contains the gene which codes for the production of mScarlet red fluorescence protein</mark>

$$y = 968.89x + 160.93$$

Figure 3.2 Calibration curve for mScarlet concentration and fluorescence. The fluorescence of several samples of known mScarlet concentration was recorded and plotted to create a standard curve. Samples were prepared by combining purified mScarlet protein diluted to the desired concentration, crude extract, and 700 mM HEPES buffer pH adjusted to 8. The standard curve was used to estimate the concentration of protein produced via CFPS using the observed fluorescence reading.
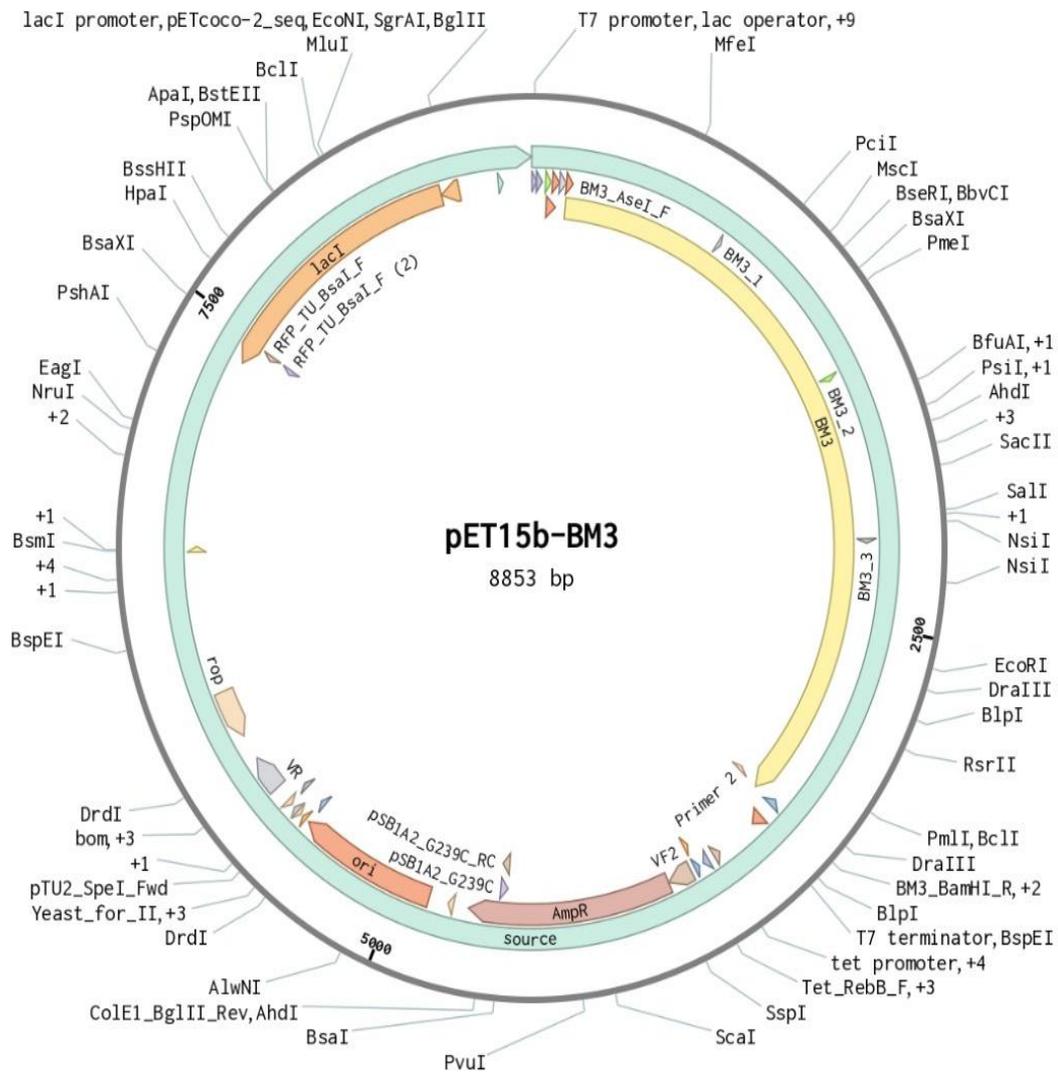
Figure 4.1 Plasmid map for pET15b-*bm3*. The Plasmid map for pET15b-*bm3,* this plasmid has an SP44 promoter, an N-terminal His-tag, a tetracysteine tag and is ampicillin resistant. The plasmid contains the gene which codes for the production of the BM3 protein.
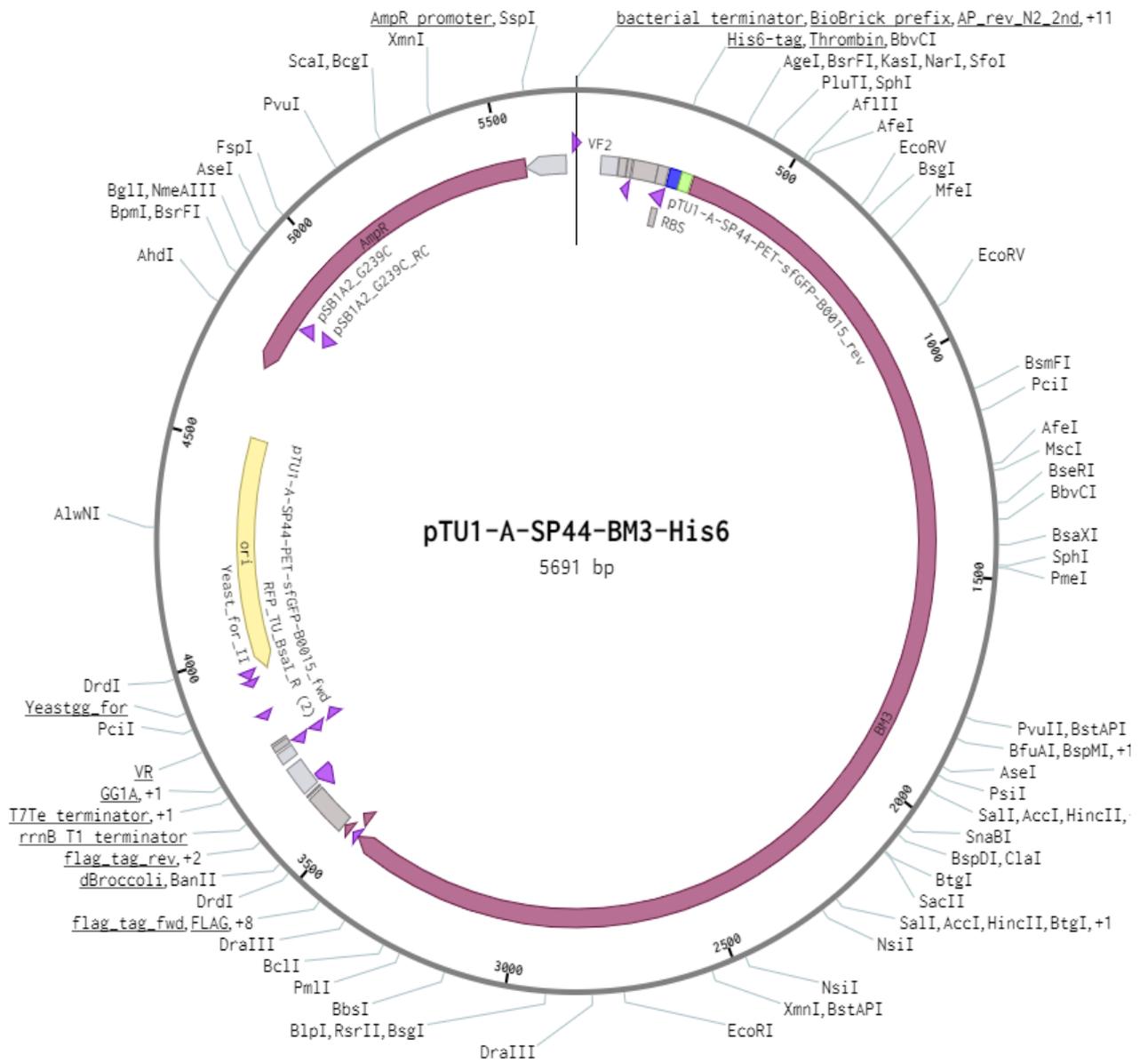
Figure 4.6 Plasmid map for pTU1-SP44-*bm3*-His₆. The Plasmid map for pTU1-SP44-*bm3*-His₆., this plasmid has a T7 promoter, an N-terminal His-tag, a tetracysteine tag and is ampicillin resistant. The plasmid contains the gene which codes for the production of the BM3 protein.
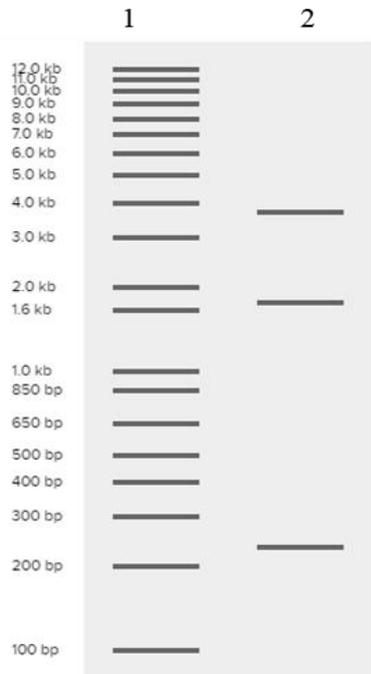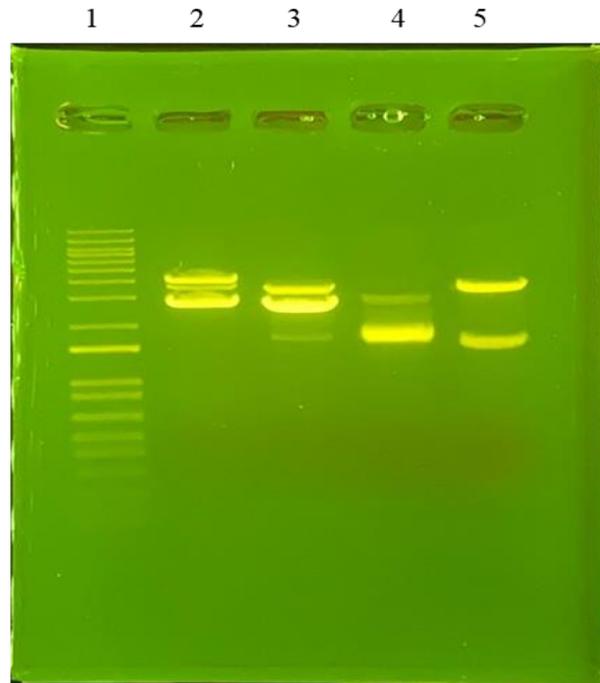
Figure 4.8 Digest of Gibson assembly plasmids. A. Virtual digest of pTU1-A-SP44-*bm3*-His6, using the restriction enzyme HindIII, lanes are as follows; 1- 1kb DNA Ladder, 2 expected bands. **B.** digest of pTU1-A-SP44-*bm3*-His$_6$, using the restriction enzyme HindIII, lanes are as follows; 1- 1kb DNA Ladder, 2- test plasmid 1, 3- test plasmid 3, 4- test plasmid 3, 5- test plasmid 4. Test plasmid 4 is believed to be the desired plasmid, and the expected band at $\approx$ 230 kb is too faint to be seen

Figure 4.11 Sequencing data for the pTU1-SP44-*bm3* plasmid.
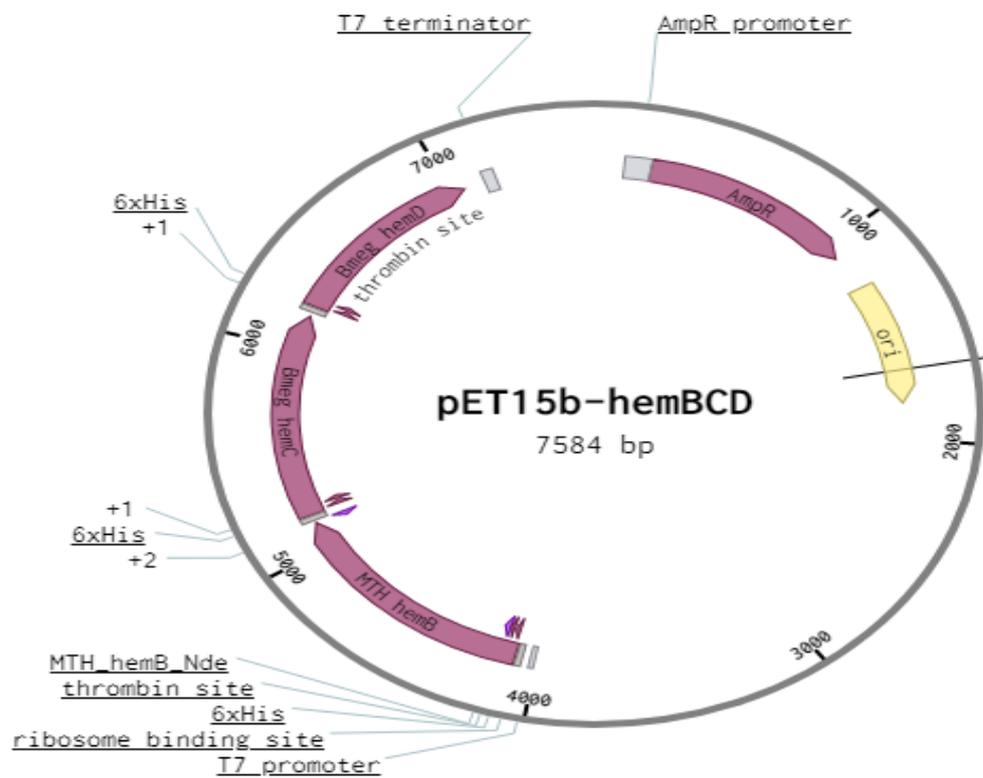
Figure 5.1 Plasmid map for pET15b-hemBCD. The Plasmid map for pET15b-hemBCD, this plasmid has an SP44 promoter, an N-terminal His-tag, a tetracysteine tag and is ampicillin resistant. The plasmid contains the gene which codes for the production of proteins hemB, hemC and hemD.

Figure 5.2 Plasmid map for pET15b-hemEFGH. The Plasmid map for pET15b-hemEFGH, this plasmid has an SP44 promoter, an N-terminal His-tag, a tetracysteine tag and is ampicillin resistant. The plasmid contains the gene which codes for the production of proteins hemE, hemF, hemG and hemH.

Table 5.1 Categorisation of haem synthesis pathway proteins molecular weight

Protein	Expected molecular weight	Estimated Observed molecular weight
HemB	35 kDa	37 kDa
HemC	33 kDa	35 kDa
HemD	27 kDa	29 kDa
HemE	39 kDa	41 kDa
HemF	34 kDa	36 kDa
HemG	21 kDa	23 kDa
HemH	35 kDa	37 kDa

The molecular weights seen in this table are taken from the Protein Data Bank.

Table 5.2 Characteristics of fluorescent molecules in the haem synthesis pathway

Compound	Excitation wavelength	Emission Wavelength
Uroporphyrinogen III	404 nm (Mathews *et al.*, 2001)	618 nm (Mathews *et al.*, 2001)
Coproporphyrinogen III	404 nm (Seo *et al.*, 2009)	618 nm and 680 nm (Seo *et al.*, 2009)
Protoporphyrin IX	405 nm and 442 nm (DaCosta *et al.*, 2003)	630 nm and 690 nm (DaCosta *et al.*, 2003)