WARWICK
THE UNIVERSITY OF WARWICK

Heterologous repressors for controlled expression of recombinant proteins in *Escherichia coli*

By

Daniel Taylor

A thesis submitted in partial fulfilment of the requirements for the degree of:

Doctor of Philosophy in Synthetic Biology

School of Life Sciences, University of Warwick

Supervisors: Dr Christophe Corre & Dr Sara Kalvala

August 2022

This page intentionally left blank.

Table of Contents

List of Figures

Acknowledgements

Firstly, I would like to thank my supervisors, Dr Christophe Corre and Dr Sara Kalvala, for their encouragement, advice, and endless patience.

Secondly, I would like to thank the Corre group, in particular Dr Patrick Capel and Dr Nazia Auckloo for their sage advice and friendship.

Thirdly, I would like to thank my fellow members of the dearly departed Synthetic Biology Centre for Doctoral Training, namely Dr Cansu Küey, Dr Charlotte Gruender, Dr Patrick Capel (again), Jack Weaver, Laurence Legon, Dr Marlene Rothe and Muna Fuyal, in whose company countless quantities of tea and coffee were consumed, with particular thanks to my long-suffering housemates and friends Dr Alan Reed and Dr Ciaran Guy.

Fourthly, I would like to thank my family, particularly Mum and Dad and my sister Lucie, for all their support and encouragement, both during my PhD project and in all the many years of education that led me there, and for doing their best to follow my fumbling explanations in the absence of a scientific career of their own.

Fifthly, I would like to thank Vilma Wilke, who pointed me towards the farthest shore of human knowledge and gave me an oar, and Joseph Meller, who signalled out of the mists ahead to tell me where the rocks were.

Finally, I would like to thank the EPSRC and BBSRC for funding the SynBioCDT, and by extension my PhD project and my existence during that time.

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by me and has not been submitted in any previous application for a degree.

The research presented here was carried out by the author, unless otherwise noted.

No parts of this thesis have yet been published by the author.

Abstract

MmfR is a transcriptional repressor protein from *Streptomyces coelicolor* which regulates the production of the antibiotic methylenomycin. MmfR is chemically induced by methylenomycin furans (MMFs), a family of stable, low-molecular-weight diffusible molecules. MmfR has previously been proven to function when produced in *Escherichia coli*.

In the work described in this thesis, the use of MmfR as a transcriptional regulator for recombinant protein production in *E. coli* was explored, in tandem with its binding site and inducer, as an alternative to the traditional LacI/LacO/IPTG inducible expression system. It was necessary to establish whether MmfR could be expressed constitutively by *E. coli*, and whether it could be as effective a repressor as LacI for biotechnological applications.

A recombinant protein production construct was designed *de novo* to use MmfR as a transcriptional repressor, and to express GFP as a reporter. The construct was assembled, transformed into *E.* coli, and observed using time-resolved fluorescence. The construct was directly compared to a parallel construct that instead used LacI as its transcriptional repressor.

Experimental data showed *E. coli* to be capable of growing effectively whilst constitutively expressing MmfR, and MmfR to be able to repress transcription of the reporter gene when the latter was regulated by MmfR's MARE target sequence. However, MmfR was also observed to be leakier than LacI, less able to repress GFP expression when not exposed to its MMF inducer.

Computational modelling of the expression system in COPASI identified several approaches that could be taken to reduce the expression system's overall metabolic burden on the host cell. The MmfR expression system shows promise as an alternative to LacI-based expression, but further work is required to establish it as a viable alternative for large-scale recombinant protein production.

List of Abbreviations

BGC – Biosynthetic gene cluster

DMSO – Dimethyl sulfoxide

DNA – Deoxyribonucleic acid

EDTA – Ethylenediaminetetraacetic acid

GFP – Green fluorescent protein

IPTG - Isopropyl β-D-1-thiogalactopyranoside

LB – Lysogeny broth

LBA – Lysogeny broth agar

MARE – Methylenomycin active response element

MMF – Methylenomycin furan

PCR – Polymerase chain reaction

RNA – Ribonucleic acid

TFTR – TetR family transcriptional repressor

Chapter 1 – Introduction

1.1 Project Outline

Recombinant gene expression allows specific proteins to be produced in quantities that would otherwise not be possible, purified with relative ease, and processed for downstream applications. Recombinant proteins are widely used in research and in industry. Recombinant proteins are typically produced by monocultures of host cells – usually either bacteria (Chen, 2012; Rosano & Ceccarelli, 2014a), yeasts (Baeshen et al., 2014; Cregg, Cereghino, Shi, & Higgins, 2000), or specific insect or mammalian cells (Andersen & Krummen, 2002; Assenberg, Wan, Geisse, & Mayr, 2013) – though they can also be produced in plants (Boyhan & Daniell, 2011), or even animals (Tokareva, Michalczechen-Lacerda, Rech, & Kaplan, 2013).

Commercial recombinant protein production systems typically include an induction system. Overproducing a particular protein places a metabolic stress on the host cell, which slows growth and creates a selection pressure for the host to mutate in a way that prevents the protein from being produced at all (Boo, Ellis, & Stan, 2019; Borkowski, Ceroni, Stan, & Ellis, 2016). It is typical for the protein's gene to be placed under the influence of a transcriptional repressor, which prevents production of the protein until a predetermined condition is met.

In previous research, a transcriptional repressor from *Streptomyces coelicolor*, MmfR, has been shown to function in *Escherichia coli*, an organism used very commonly as a host for recombinant protein production. The objective of this project was to further characterise MmfR in *E. coli*, to explore its potential as a regulator for recombinant protein production, and to develop such a production system for widespread use.

1.2 Recombinant Protein Production

1.2.1 DNA Recombination

Recombinant DNA technology is the cutting and splicing of genetic code from one organism to another. Pioneered in the 1970s, researchers were able to clone the human insulin gene into E. coli, allowing the bacterium itself to produce insulin that could later be harvested and processed into a viable treatment for type II diabetes (Goeddel et al., 1979). Restriction enzymes were used to cut both human and bacterial DNA at specific locations, and the relevant parts were chemically ligated together. Today, DNA can be modified far more easily and precisely thanks to the availability of more restriction enzymes and other cutting methods (such as CRISPR-Cas9), more methods of DNA assembly (such as Golden Gate assembly and Gibson assembly), and more methods of targeted mutagenesis (including site-directed mutagenesis), and greater access to chemical oligonucleotide and gene synthesis services.

As touched upon above, restriction enzymes allow DNA sequences to be reliably cut at specific points. Each restriction enzyme recognises a specific target sequence of nucleotides that may be as few as four nucleotides, or as many as eight or even more (Di Felice, Micheli, & Camilloni, 2019; Pingoud, 2001). Type II restriction enzymes, the category most often used in DNA recombination, cut DNA in a specific location relative to their recognition site. This location may be within the recognition site, or a fixed number of base pairs to one side, and the cut may be made at the same location on both strands (leaving a “blunt” end) or at different bases (leaving a “sticky” end) (Di Felice et al., 2019; Pingoud, 2001).

Most restriction enzymes can only cut at their specific recognition sites, but some other methods of cutting DNA allow the cut to be targeted to any theoretical nucleotide sequence. Zinc finger nucleases are restriction enzymes that consist of distinct DNA-binding and cleavage domains connected by a linker, the former of which can be engineered to bind to any specific nucleotide sequence (Carroll, 2011; Kim, Cha, & Chandrasegaran, 1996). Transcription activator-like effector nucleases (TALENs) use

the same nuclease domain but a more easily-tailored form of DNA-binding domain (Christian et al., 2010; Joung & Sander, 2013; T. Li et al., 2011; Mahfouz et al., 2011). However, CRISPR-Cas9 has supplanted both of these methods of targeted gene editing. Using a short guide RNA sequence, the bacterial Cas9 enzyme can be used to cut a DNA sequence at an exact location complementary to the matching guide RNA (Jinek et al., 2012).

Ligation of DNA fragments is typically done using ligase enzymes, such as T4 ligase, from the bacteriophage of the same name. Different ligases have different requirements, but ligation can generally be done between both blunt- and sticky-ended fragments. In the latter case, the sticky ends of adjacent fragments should be complementary to one another (Engler & Marillonnet, 2013).

PCR uses RNA polymerase enzymes, typically with high thermal stability (originally from *Thermus aquaticus*, a bacterium native to hot springs) to "amplify" existing DNA sequences at high speed, generating copies from a reaction mixture of free nucleotide building blocks (Saiki et al., 1988). Single-stranded DNA primers direct the polymerase to the ends of the region to be replicated, and can include modifications to the sequence, such as individual base modifications, or the addition of restriction enzyme sites to enable downstream assembly (Carter, 1986; Engler & Marillonnet, 2013).

Gene synthesis technology has made it vastly cheaper and easier to obtain entirely custom DNA sequences from tens to thousands of base pairs long, all of which may contain genes and other functional domains alongside any configuration of restriction enzyme sites and other tools for cloning. These sequences can then be combined into even longer sequences using one of several assembly methods, including Gibson assembly and Golden Gate assembly.

In Gibson assembly, the component DNA sequences are designed with overlaps that are 50-500 base pairs long. A 5' exouclease enzyme is used to degrade one strand at each end of the DNA strands (Gibson et al., 2009). The single-stranded overlap regions

then bind together, so that all the sequences bind to one another in a predetermined configuration. Gibson assembly is an ideal method for combining multiple DNA sequences that are hundreds of base pairs long, or more.

For the cloning procedures described in section 3.1, however, Golden Gate assembly was the chosen method. Golden Gate assembly uses Type IIs restriction enzymes and standard DNA ligases to assemble fragments of DNA together (Engler, Kandzia, & Marillonnet, 2008). Type IIs restriction enzymes are those that recognise a specific DNA sequence, and upon recognition, cut a fixed distance away from that recognition sequence. Many type IIs restriction enzymes also leave a 4 base pair overhang at the cut site. An example of a type IIs restriction enzyme is shown in **Figure 1.1**:

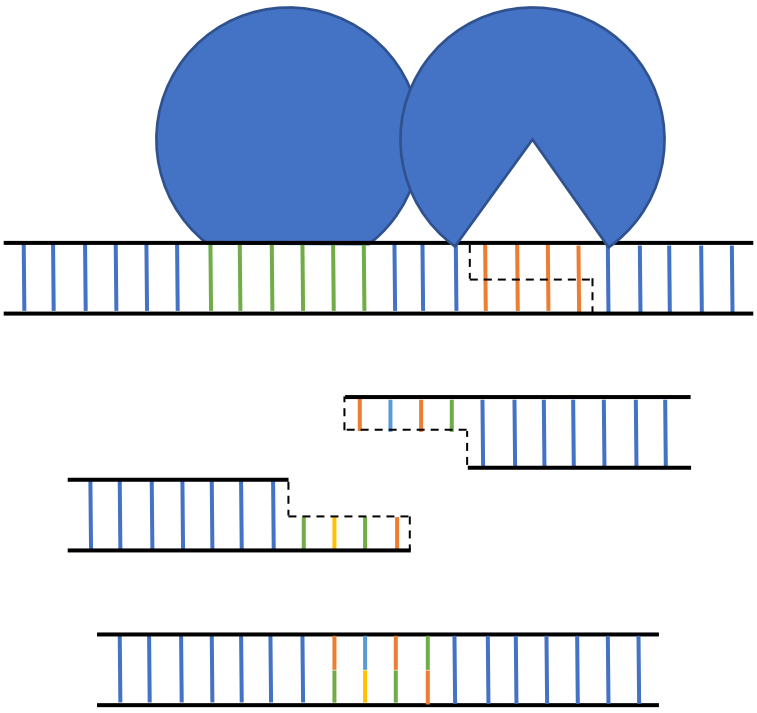

Figure 1.1: *Type IIs restriction enzyme activity.*

Top*: the restriction enzyme binds to the DNA recognition site (green) and cuts at a separate site (orange).* ***Centre****: Each side of the cut site is left with a complementary overhang sequence.* ***Bottom****: Complementary sticky ends can bind together, allowing the DNA strands to be ligated together.*

In Golden Gate assembly, the enzymes' recognition sites are positioned on the edges of the fragments, with the cut sites located inwards of them. The recognition sequences are thus eliminated during the digestion, preventing re-digestion of the ligated product and ensuring no "scar" sequences are left over from the transformation that might complicate the construct's design or function. Complementary overhang sequences ensure that the cut fragments naturally assemble in the desired order and orientation before ligation, resulting in a near-100% theoretical transformation efficiency (Engler et al., 2008; Engler & Marillonnet, 2013; Weber, Engler, Gruetzner, Werner, & Marillonnet, 2011).

By using **consistent** overhang sequences for each type of genetic part (promoters, RBSs, coding sequences and terminators), these parts could be used interchangeably. For example, the constitutive promoters could be exchanged like-for-like during the assembly process to adjust the expression rate of the repressor protein. Likewise, either the GFP or insulin CDS could be used during the assembly of the reporter operon. The overhang sequences chosen were based on the **MoClo** system (Weber et al., 2011; Werner, Engler, Weber, Gruetzner, & Marillonnet, 2012), which presented the opportunity to use genetic parts from other sources that were designed to use the same standard. The overhang sequences are depicted in **Figure 1.2:**

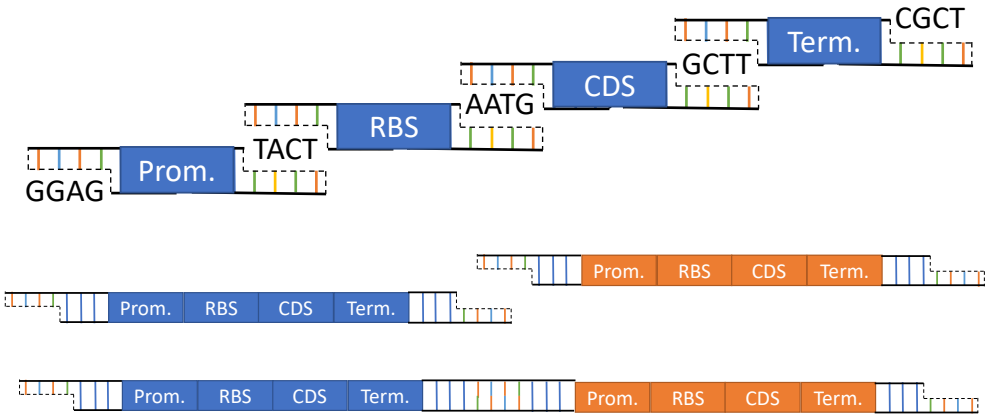

Figure 1.2: *Example of* ***Golden Gate assembly.***

Top: *Individual sequences can be designed with matching overhangs so that they ca be "strung" into a specific sequence in one digestion and ligation reaction.* ***Middle****: Multiple stages can be used to then assemble complex constructs – for example, a two-stage process in which individual operons are first assembled, and then the operons are combined into one construct (****bottom****).*

1.2.2 Recombinant Protein Production

By taking the gene for a particular protein from one organism, and putting it in another organism, it is possible to have the new host produce the protein in question. The gene is typically inserted into a pre-designed expression vector, a plasmid containing a selection marker and a promoter to drive expression of the inserted gene, before transformation into the host (Andersen & Krummen, 2002; Baneyx, 1999; Rosano & Ceccarelli, 2014b).

Heterologous protein production has become an extremely powerful asset in the modern scientific toolbox. Through incorporation into a fast-growing heterologous host with a strong promoter, proteins of interest can be produced in quantities far exceeding that which could be produced in the native organism. Purification tags can be attached to allow those proteins to be purified easily and efficiently through affinity chromatography; linked and co-produced fusion partners can alter their properties, adding a fluorescent or luminescent reporter region, or improving solubility (Rosano & Ceccarelli, 2014a; Sørensen & Mortensen, 2005; Wood, 2014).

A relatively small number of cell types are routinely used for heterologous protein production. The most prevalent is *E. coli*, preferred for its fast growth, ease of engineering, and simple growth conditions. The yeasts *Saccharomyces cerevisiae* and *Pichia pastoris* are also frequently used, particularly for eukaryotic proteins, which the prokaryotic *E. coli* often cannot produce correctly. Mammalian cell lines such as Chinese hamster ovary, human embryonic kidney, and murine hybridoma cells are often used for specific mammalian proteins, such as whole antibodies (Baeshen et al., 2014; Dingermann, 2008).

1.2.3 Applications of Recombinant Proteins

Recombinant proteins are used in a wide variety of pursuits; the ability to produce relatively large amounts of a specific protein is important to many fields. As early as 1979 (Goeddel et al., 1979), human insulin was produced in *E. coli*, and reached the market a few years later, the first of many so-called "biologics", or protein-based drugs (Overton, 2014). Meanwhile, structural and functional studies of proteins frequently use recombinant versions of those proteins.

The advent of recombinant insulin marked a seismic shift in the treatment of diabetes. Previously, therapeutic insulin was obtained from pigs, or from human cadavers. The latter carried the risk of transferring diseases from cadaver to patient, whilst the former yielded porcine insulin, with small differences in amino acid sequence that could lead to dangerous immune responses in some human recipients. Both approaches offered only a very limited supply of insulin with attendant ethical concerns (Walsh, 2005), so the advent of what was essentially vat-grown human insulin was revolutionary. Many other hormone-based treatments followed suit (Overton, 2014).

Other commonly produced proteins include antibodies and enzymes. Antibodies consist of a fixed backbone and a highly variable binding site specific to a particular biological target. Their binding strength and specificity makes them very useful, for

example in immunoassays (such as COVID-19 lateral flow tests) or for purifying other proteins from a homogenous sample.

Antibody fragments can also be produced as fusion proteins, joined by a peptide chain to another protein such as an enzyme, with very little impact on the function of either. This allows specific proteins to be targeted to specific sites in a biological sample (Joosten, Lokman, Van Den Hondel, & Punt, 2003; Nilsson, Ståhl, Lundeberg, Uhlén, & Nygren, 1997).

For example, a wide range of antibody-cytokine fusions are being developed as potential cancer treatments (Helguera & Penichet, 2005; Joosten et al., 2003; Murer & Neri, 2019). The antibody directs the fusion protein to cancer cells expressing a specific antigen, and the cytokine recruits the patient's immune system to attack the tumour. However, these treatments generally result in a range of negative effects on the patient and as a result, only a very few have so far been approved for clinical use (Murer & Neri, 2019).

Enzymes themselves are also often produced through recombinant protein production, and find a variety of uses in industry. Some are present in commercially available products, such as recombinant lipases in biological laundry powder (Borrelli & Trono, 2015). In other cases, the desired product is a product of the enzyme or enzyme pathway, such as an antimicrobial compound or a food additive (Adrio & Demain, 2010; Trono, 2019).

Research into proteins – such as structural imaging, or activity or affinity assaying – usually requires large, purified quantities of the relevant protein. Acquiring these from the wildtype organism can be very difficult, particularly if the organism in question produces very little of that protein under laboratory conditions. As a result, these proteins are commonly obtained through DNA recombination (Goulding & Perry, 2003; Kermani, 2021).

1.2.4 Control of Recombinant Protein Production

Usually, expression in a heterologous host will be regulated so that the gene of interest is only expressed under a specific condition. Constitutive (constant, unregulated) protein production imposes metabolic strain on a cell, diverting resources away from growth and reproduction. This slows the growth of the cell culture, and the reduced cell density severely limits the overall yield, even when the rate of production per cell is high (Sevastsyanovich, Alfasi, & Cole, 2010).

This counterproductive effect is magnified if the product is also toxic to the host cell, with selection pressure often causing the original transformant to be out-competed and rapidly replaced by a mutant that has excised the desired product from its genome, or otherwise rendered it absent or non-functional (Overton, 2014). As a side-effect of slowing growth, protein overproduction creates an evolutionary selection pressure against cells producing the recombinant protein, and in favour of those that mutate so that they produce no or very little of the intended product (Baneyx, 1999; Sørensen & Mortensen, 2005).

Fine control over the production of heterologous proteins is therefore essential for obtaining high yields. Most commercially available plasmids designed for heterologous protein production use the *lac* repressor for this purpose. In this system, the repressor protein LacI is constitutively produced, and binds to the recognition site upstream of the heterologous gene. In doing so, it obstructs the binding of RNA polymerase to the promoter, preventing production of the heterologous product and permitting the host organism the resources necessary to grow. When the desired cell density is reached, isopropyl β-D-1-thiogalactopyranoside (IPTG) is added to the culture medium, binding to LacI and releasing it from the operator sequence, thereby initiating expression of the gene (Overton, 2014; Rosano & Ceccarelli, 2014a; Sørensen & Mortensen, 2005). The *lac* expression system is shown in **Figure 1.3:**

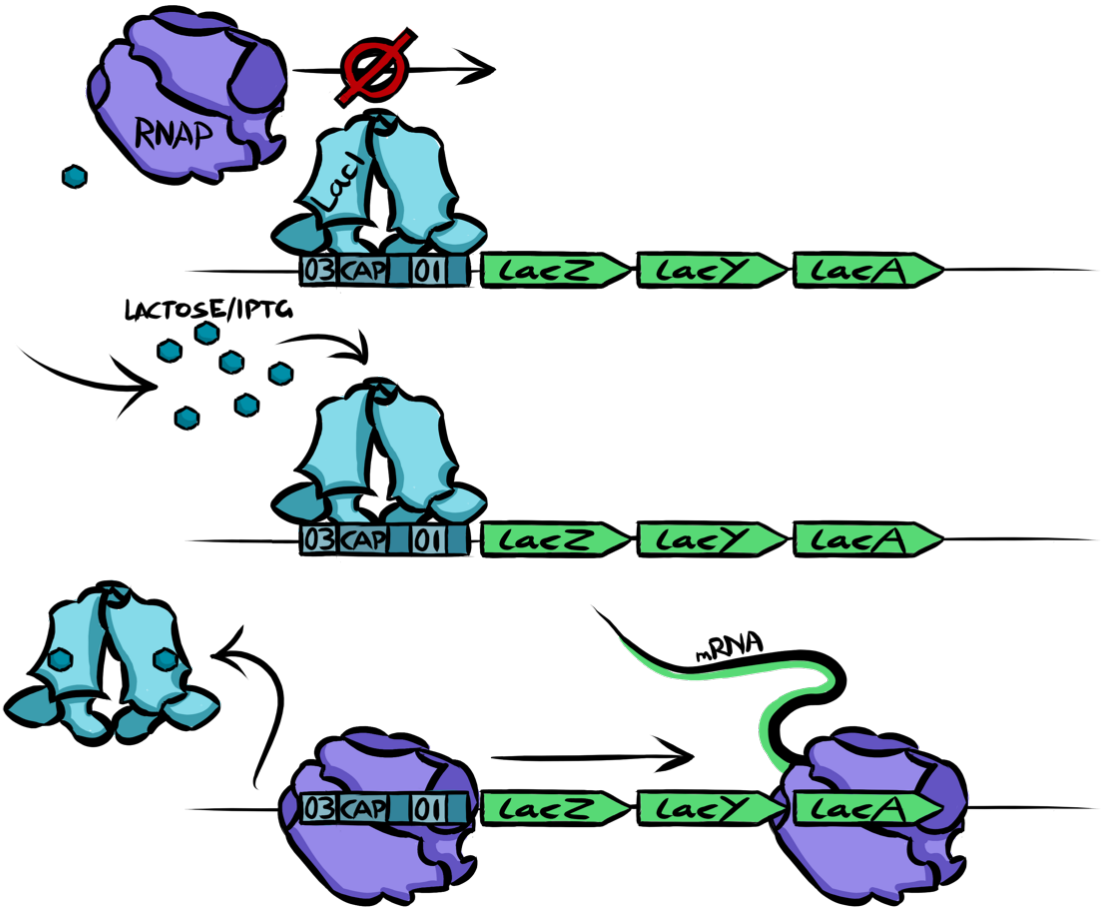

Figure 1.3: *LacI expression system.*

Top*: LacI bound to the O3 and O1 regions of the* lac *operator (the O2 region, downstream from the* lac *operon, is omitted), preventing transcription of the operon.* ***Middle****: LacI inhibitors lactose or IPTG enter the cell.* ***Bottom****: Bound by inhibitor, LacI unbinds from the operator sequence, permitting transcription. Diagram not to scale.*

However, this approach does also carry some limitations. LacI does not bind the operator with perfect affinity, and so expression is not completely inhibited even in the absence of IPTG – a phenomenon known as "leakiness". This causes the expression system to exert some metabolic stress even when repressed, so the Lac repressor is far from a perfect workaround. Furthermore, the production of the repressor itself naturally also exerts some stress on the cell, exacerbating the issue of mutation (Overton, 2014; Rosano & Ceccarelli, 2014a). IPTG is also costly enough to make large-scale production using the Lac repressor viable only for high-value biological products.

This is further complicated in the case of biologics, which generally require IPTG of pharmaceutical-grade purity to be used, adding further expense.

Other methods of inducible expression have been trialled (including metabolite, antibiotic, and thermal induction) by leveraging *E. coli*'s metabolic and heat-shock pathways (Overton, 2014). Nonetheless, while *E. coli* has remained the first choice of heterologous host for biotechnologists, an alternative to the *lac* repressor has yet to be widely adopted (Sørensen & Mortensen, 2005). An alternative expression system that has a reduced cost and reduced leakiness, while maintaining the *lac* system's ease of use, would be highly desirable for heterologous protein production. With the expansive modern-day synthetic biology toolkit, designing and assembling such a system is easier and faster than ever before.

1.3 MmfR & Other ArpA-like TetR-Family Transcriptional Repressors

1.3.1 MmfR, an ArpA-like transcriptional repressor

The *Streptomyces* genus provides a host of natural examples of inducible gene expression. In particular, streptomycetes produce a vast range of antibiotics through biosynthetic gene clusters (BCGs) to help them out-compete other microorganisms. The activation of these BGCs must be strictly regulated – both to reduce the metabolic pressure that would otherwise be incurred through constant biosynthesis, and to ensure that the cell population can co-ordinate production of countermeasures against their own antibiotics (Biarnes-Carrera, Breitling, & Takano, 2015). Examples of biosynthetic gene clusters with complex transcriptional regulation systems in *S. coelicolor* include actinorhodin (Act), (ScbR), and (the most relevant to this thesis) methylenomycin (MM) (Liu, Chater, Chandra, Niu, & Tan, 2013).

Actinorhodin (ACT) is a weak antibiotic produced by *S. coelicolor* that gives the species its characteristic sky-blue colouration. ACT is produced by a cluster of five operons, all of which are regulated by the ActII-ORF4 transcriptional activator. ActII-ORF4 acts as the filter through which physiological information from the cell is

transmitted to the *act* cluster, as at least eight major regulatory proteins are known to interact with its promoter region, and others are strongly suspected to do so. A separate ACT regulatory system exists in the form of the TetR-like repressor ActR, which represses the *actA* operon. *actA* is responsible for the expression of the ACT transport system, and ActR repression is inhibited by the presence of late intermediates in ACT biosynthesis, suggesting a close co-ordination between the biosynthesis and exocytosis of ACT at a transcriptional level (Liu et al., 2013).

Streptomycin is another antibiotic, produced by *Streptomyces griseus*, and its expression is controlled by transcriptional regulator AdpA. AdpA has around 1500 projected genomic binding sites in the *S. griseus* genome (although not all of these are necessarily involved in gene regulation), and is present across multiple *Streptomyces* species. Sometimes regarded as a "master regulator", AdpA's target sites are found upstream of transcriptional regulators situated in a wide range of antibiotic biosynthesis clusters, such as *actII-ORF4* of the ACT pathway. AdpA is known to have multiple binding sites at different operators within some biosynthetic gene clusters, and acts as an activator for some operons whilst repressing others (Liu et al., 2013; Onaka, Nakagawa, & Horinouchi, 1998).

A-factor-specific receptor (ArpA) is a dimeric TetR-like transcriptional repressor (TFTR) that represses expression of the *adpA* gene in *S. griseus*, by binding to a 22-bp target DNA sequence on *adpA*'s operator sequence. As such, it could also be considered (indirectly) a regulator of antibiotic expression across *S. griseus'* genome. ArpA is inhibited by the presence of A-factor, an *S. griseus*-specific gamma-butyrolactone (GBL). GBLs are a class of small, membrane-diffusible chemical messengers that accumulate in the culture over time, and are thus able to signal the density and life cycle stage of the colony as a whole (Liu et al., 2013; Onaka et al., 1998).

Analogues of ArpA are found throughout the *Streptomyces* genus, and often multiple variants are found within one species. ArpA-like repressors typically form part of the regulatory systems for antibiotic-producing biosynthetic gene clusters in various

Streptomyces species (Cuthbertson & Nodwell, 2013; Ramos et al., 2005). *S. coelicolor*, for example, is host to ScbR and MmfR, both of which share a high degree of similarity in sequence and function with ArpA, and which regulate coelimycin and methylenomycin biosynthesis (Tsigkinopoulou, Takano, & Breitling, 2020; Zhou et al., 2021). CprB, meanwhile, is an *S. coelicolor* homologue of ArpA (Natsume, Ohnishi, Senda, & Horinouchi, 2004). The crystal structure of CprB, a typical ArpA-like TFTR, is shown in **Figure 1.4:**

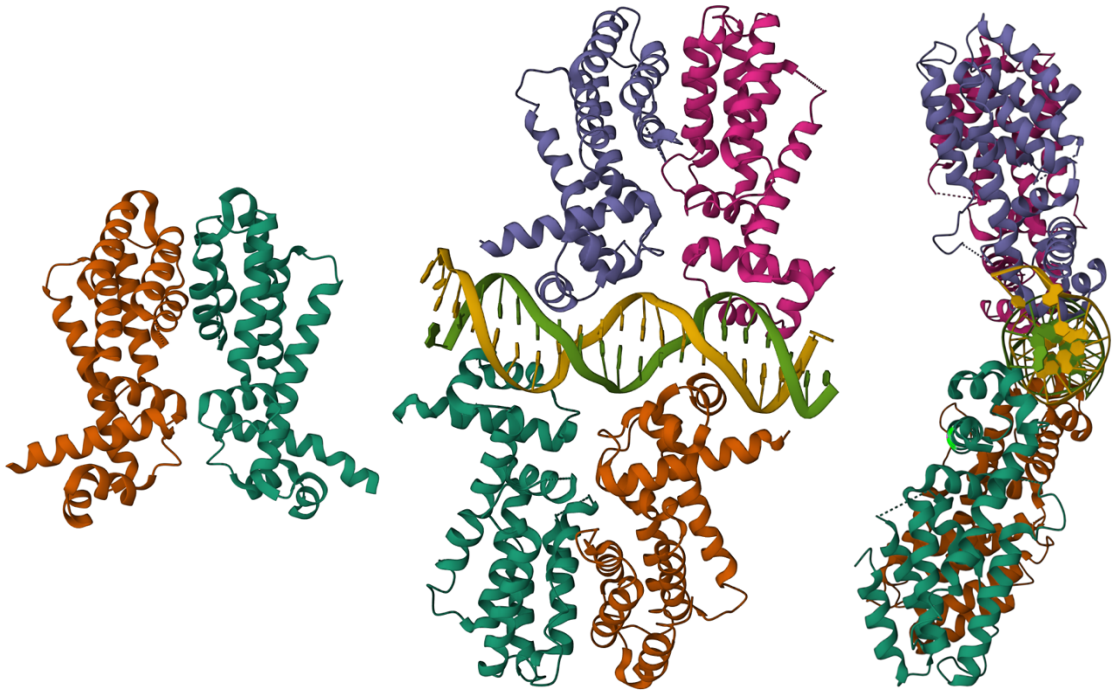

Figure 1.4: *Structure of CprB, an ArpA homologue.* ***Left:*** *Structure of CprB homodimer* (Natsume, R., Senda, T., Horinouchi, 2004; Natsume et al., 2004). ***Middle, right:*** *Two CprB homodimers bound to their DNA target site* (Bhukya, Bhujbalrao, Bitra, & Anand, 2014; Hussain, B., Ruchika, B., Aruna, B., Ruchi, 2014).

CprB forms an "Ω"-shaped homodimer with ten α-helices per monomer. The first three helices form the DNA-binding domain, whilst the remainder form the regulatory ligand-binding domain. The eighth and ninth helices also form the interface for dimerisation. CprB targets a 24-nucleotide consensus sequence in *S. coelicolor*,

ScbR is part of the *cpk* biosynthetic gene cluster of *S. coelicolor*, which produces the polyketide-derived antibiotic coelimycin (CPK; formerly "cryptic polyketide") (Liu et al., 2013). ScbR, like ArpA, is inhibited by an SCB ligand. ScbR is believed to repress its own expression as well as that of ScbA, a GBL synthase that is also present in the *cpk* cluster, by binding to the bidirectional promoter situated between them. ScbR has been shown to bind to over 140 sequences across the *S. coelicolor* genome, making it a far more widespread regulator of gene expression than originally thought (X. Li et al., 2015). ScbR is furthermore believed to form heterodimers with the higly homologous ScbR2 repressor, potentially changing its target sequence specificity as well as the inhibitors that it is able to respond to.

ScbR2, which shares 50% of its sequence with ScbR, binds an even wider range of targets than ScbR – around 500 across the *S. coelicolor* genome. However, despite being closely related to the family of GBL receptors, it does not recognise GBLs – it instead responds to a range of endogenous and exogenous antibiotics (X. Li et al., 2015). Besides having its own target sequences, ScbR2 also seems to bind to ScbR's *scbA* target sequence in the *cpk* cluster, making it an additional regulator for coelimycin expression (Liu et al., 2013).

MmfR is another ArpA-like transcriptional repressor present in *S. coelicolor*, and regulates the biosynthetic pathway for the antibiotic methylenomycin (MM), a gene cluster located on its SCP1 chromosome (Bowyer et al., 2017; Corre, Song, O'Rourke, Chater, & Challis, 2008; Styles, 2016). Like ScbR, MmfR responds to a family of small, non-toxic and diffusible chemical messengers called methylenomycin furans (MMFs). These furans are generated by the MmfLHP enzyme pathway, also part of the MM cluster (Corre et al., 2008). *MmfR* and *mmfLHP* share a bidirectional promoter region which carries an 18-base-pair methylenomycin autoregulatory response element (MARE) – the target sequence for MmfR (O'Rourke et al., 2009). These interactions between MmfR, a MARE and MMF are summarised in **Figure 1.5:**

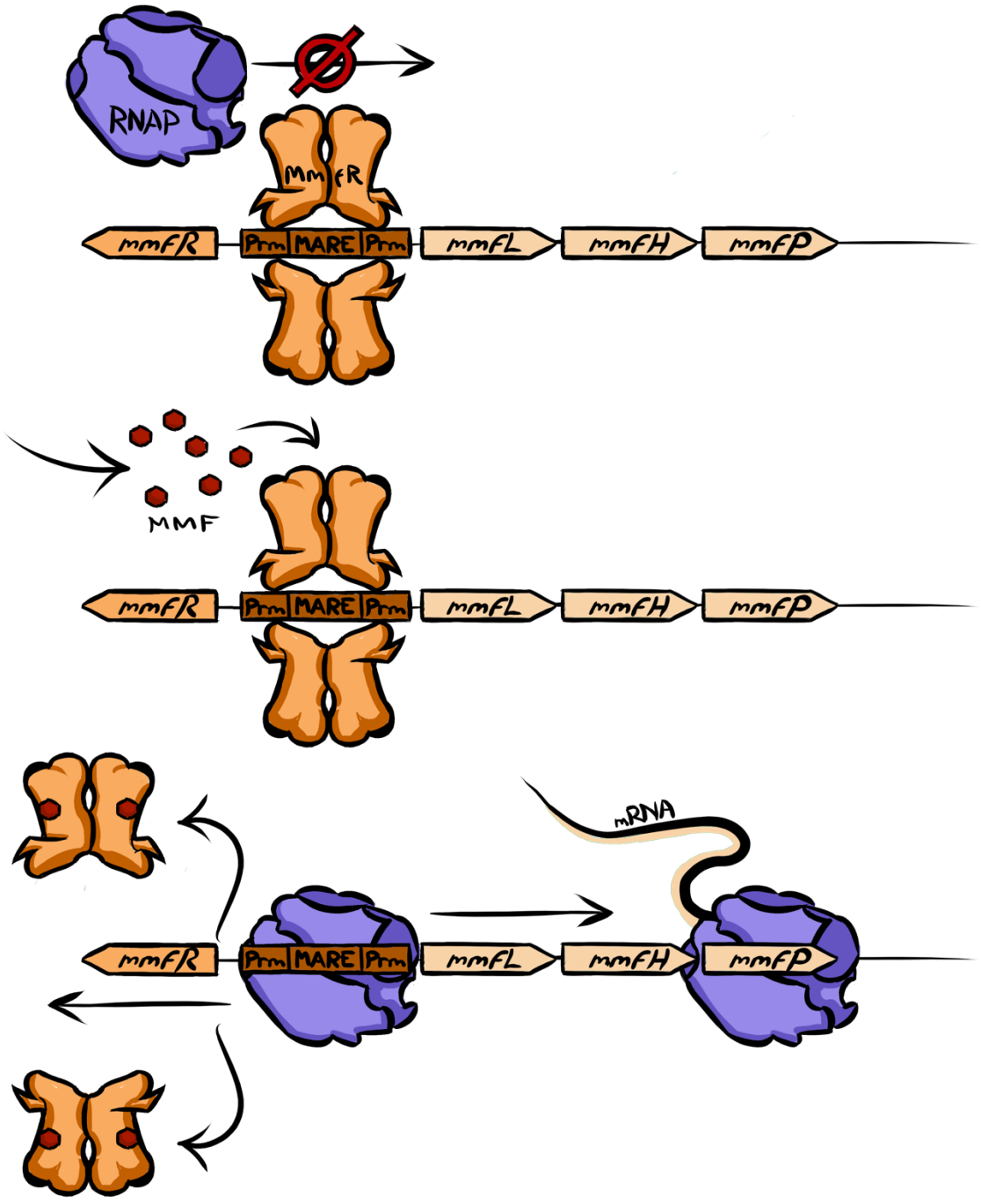

Figure 1.5: *MmfR expression system in* S. coelicolor.

Top: Up to two MmfR dimers can bind to the MARE recognition sequence, preventing transcription of the adjacent operons. **Middle**: MMFs, inhibitors of MmfR, enter the cell. **Bottom**: Bound by MMFs, MmfR unbinds from the MARE sequence and permits transcription of adjacent operons. Diagram not to scale.

In this way, MmfR represses its own expression as well as the biosynthetic pathway for its MMF inhibitor. MARE sequences are most commonly found between the -35 and -10 upstream regions of a promoter, and appear three times in the *S. coelicolor* genome with similar but not identical sequences. One, as previously mentioned, represses expression of *mmfR* itself as well as the MMF biosynthetic pathway. Of the other two MAREs, one regulates MmyR, whose exact function is unclear but whose deletion results in overproduction of methylenomycin (O'Rourke et al., 2009; Zhou & Challis, 2016) and the other regulates production of the methylenomycin biosynthetic pathway (Corre et al., 2008; Zhou et al., 2021; Zhou & Challis, 2016). The layout of the MM cluster is summarised in **Figure 1.6**:

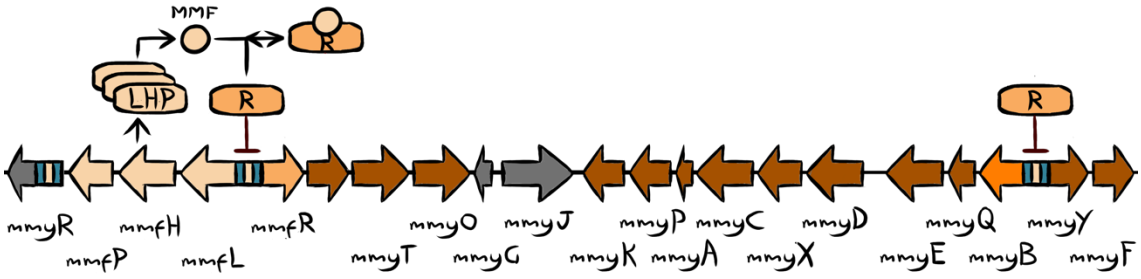

Figure 1.6: *Schematic overview of methylenomycin biosynthetic cluster with focus on MARE sequences.*

Recent work by Zhou et al established the structure of MmfR though X-ray crystallography. MmfR monomers possess 9 α-helices – the first two constitute the DNA-binding domains and the other seven form the furan-binding pocket. The final two α-helices also form the interface for dimer formation; like ScbR and many other ArpA-like TFTRs, MmfR is homodimeric. Ten residues on the latter seven helices interact directly with MMF2, one of several MMFs that MmfR responds to. Electron microscopy data also show that as with many other ArpA-like repressors, two MmfR dimers are able to bind to each MARE sequence. These data are displayed in **Figure 1.7**:

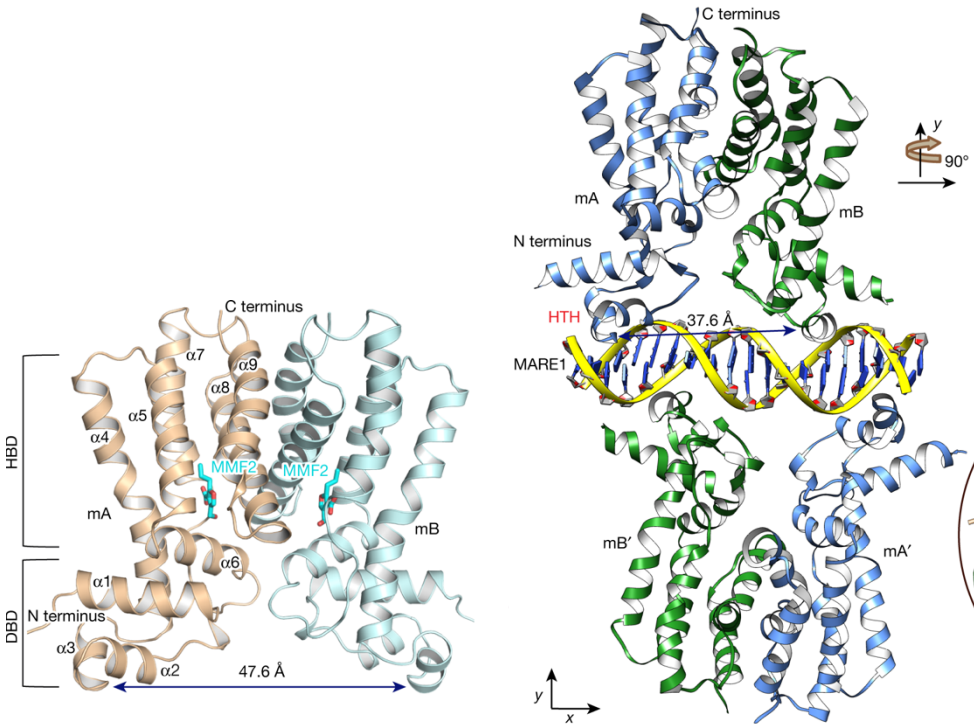

***Figure 1.7: Left**: X-ray crystal structure of dimeric MmfR with MMF2 within the binding pockets. **Right**: Two MmfR dimers bound to MARE DNA sequence. Figure adapted from* (Zhou et al., 2021).

MmfR responds to a range of MMFs, each of which have slightly different chemical structures, and different MMFs have different affinities for MmfR, as well as for other MmfR-like repressors. MMFs are a family of 2-Alkyl-4-hydroxymethylfuran-3-carboxylic acids, and are synthesised by the MmfLHP enzyme pathway, which is part of the MM cluster (Corre et al., 2008; Styles, 2016). The five MMFs classified by Corre et al. as MmfR inhibitors are shown in **Figure 1.8:**

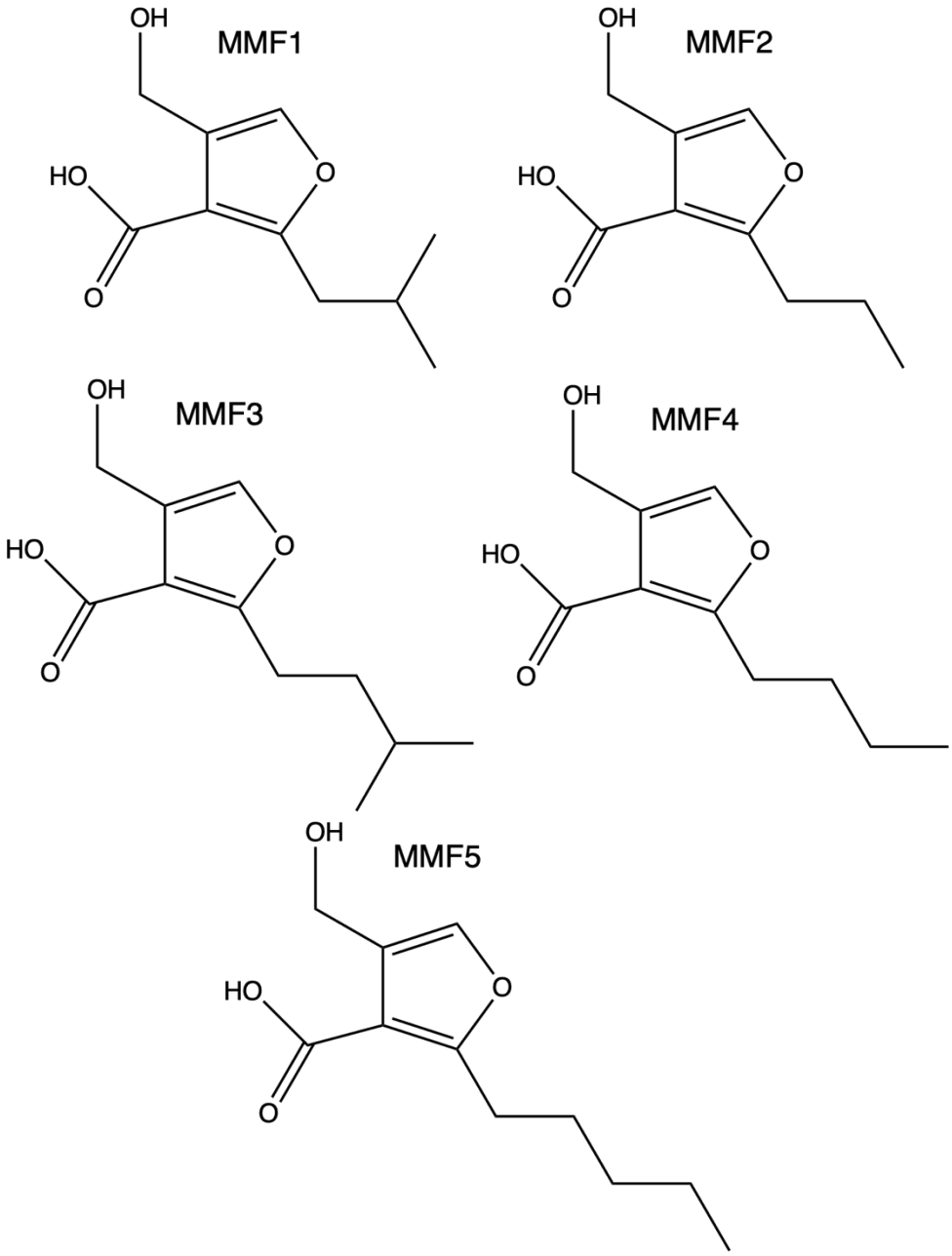

Figure 1.8: *Structures of the five methylenomycin furans (MMFs), inhibitors of MmfR.*

As previously described, *S. coelicolor* has three MARE sequences, target sites for MmfR, located at different locations on the methylenomycin biosynthetic cluster. Each MARE

is a slightly different nucleotide sequence, with different affinity for MmfR. Their nucleotide sequences are shown in **Table 1.1**:

*Table 1.1: MARE sequences in the MM cluster. Non-homologous bases are shown in **bold**.*

Intergenic Region	**MARE Sequence (5' to 3')**
mmfL/mmfR	A**T**ACCTTC**CC**G**CA**GGT**A**T
mmyR	A**T**ACCTTC**CC**G**AG**GGT**A**T
mmyB/mmyY	A**A**ACCTTC**GG**G**AA**GGT**T**T

As a grouping of transcriptional repressors with a very different structure, inducer and target sequence to many repressors commonly used in biotechnological applications, ArpA-like TFTRs are of considerable interest for synthetic biology. SCB-inducible repressors, such as ScbR, have been earmarked for such a purpose (Biarnes-Carrera et al., 2015). MmfR, a comparable repressor that does not respond to the presence of SCB1 (Zhou et al., 2021) (and thus, is apparently orthogonal to the activity of ScbR), holds similar potential.

1.3.2 Characterisation of MmfR in *E. coli*

In unpublished experimental work by Dr Miriam Rodriguez Garcia at the University of Warwick, the MmfR repressor was expressed in *E. coli* and assessed for functionality. Two constructs were transformed into *E. coli* together. The first construct contained the gene *mmfR*, codon-optimised for expression in *E. coli*. The *mmfR* gene was prefaced by the *lac* operator, such that MmfR would only be produced in the presence of IPTG. The second construct contained a *gfp* reporter gene with a MARE sequence included within its promoter, preventing expression of the green fluorescent protein (GFP) reporter whilst MmfR was available to bind to the MARE (Rodriguez-Garcia M; Corre C, 2017). These constructs are summarised visually in **Figure 1.9:**

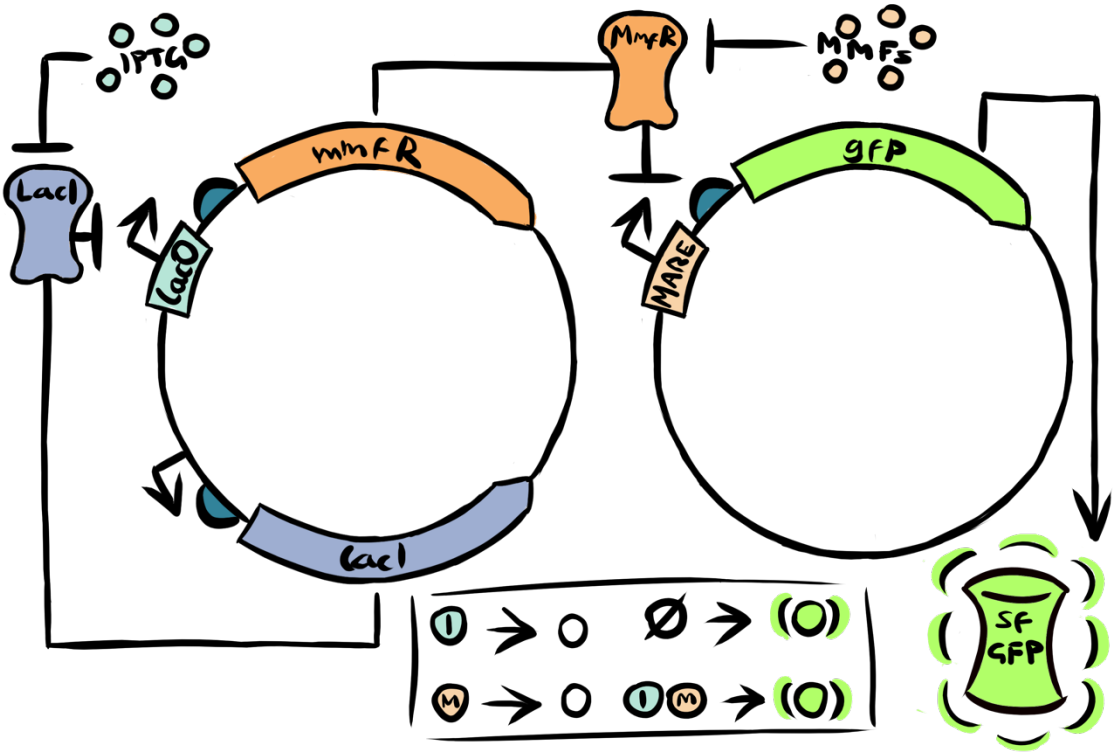

Figure 1.9: *Constructs used by Dr Miriam Rodriguez Garcia to compare MARE sequence affinities.*

The primary purpose of this experiment was to establish whether MmfR would recognise its cognate MARE site when expressed by a heterologous organism, and whether it would also recognise its inhibitor, the MMF family. In both cases, MmfR was found to function, although exact quantifications for its affinity for both target site and inhibitor, as compared to within *S coelicolor*, were not obtained (Rodriguez-Garcia M; Corre C, 2017).

The secondary aim of this work was to establish whether MmfR's affinity for its target site could be improved with a variant MARE sequence. Not all of the wildtype MARE sequences in the *S. coelicolor* genome are identical and so it was expected that at least some affinity would be maintained if one or two of the bases in the consensus sequence were altered. To this end, two of the bases – the second and penultimate of the 18-base pair sequence – were randomised and all of the consequent MARE

sequences (bar one, which by chance was not obtained during the mutagenesis) were trialled experimentally. The MARE sequences can be found in **Appendix 1**. The results of this work showed that some of the variant MARE sequences, particularly MM4, actually displayed stronger affinity for MmfR than the wildtype sequence they were derived from (Rodriguez-Garcia M; Corre C, 2017).

However, some questions remained regarding MmfR activity in *E. coli*. The expression of *mmfR* had been driven by the *lac* operator, and thus had only been induced by the addition of IPTG to the growth media. As a result, MmfR was not present in more than incidental quantities in the cells during early growth, and so its possible effects on the growing cell population were not known. It was thought that constitutively expressing a DNA-binding protein might lead to non-specific binding and toxicity in a growing cell and consequently hinder growth.

Additionally, although a comparative study was made between different MARE sequences with regards to binding affinity for MmfR (and consequently, leakiness), no study had yet been made comparing MmfR-MARE affinity in *E. coli* with other repressor-and-operator systems of biotechnological interest. Such systems include the Lac repressor and Lac operator – an academic and industry standard for recombinant protein expression, as addressed in **Section 1.2.4**. If the MmfR-MARE system compared favourably to LacI-LacO in terms of leakiness in *E. coli*, it would prove to be of immense interest for use as a repressor-operator for heterologous expression platforms.

Besides the prior research that has been conducted regarding the MmfR-MARE regulation system, MmfR was also of particular interest among ArpA-like TFTRs because of the methylenomycin cluster's position on a plasmid, the SCP1 chromosome of *S. coelicol*or (Corre et al., 2008; Yamasaki, Ikuto, Ohira, Chater, & Kinashi, 2003). This means that the *mmfR* gene and each of the MARE targets sites are present in higher copy numbers than many other, genomic ArpA-like TFTRs, suggesting that the MMfR-MARE system might be better adapted for use in a similarly multi-copy expression plasmid.

MmfR's status as a regulator of a biosynthetic gene cluster, rather than a single gene (Corre et al., 2008; Zhou et al., 2021), also suggests its suitability for regulating similar biosynthetic clusters in heterologous expression systems, which may predispose it towards the regulation of BGCs in heterologous expression platforms (although there is no especial reason to believe a particular transcriptional repressor would *not* be suitable for regulating a complex biosynthetic pathway).

Additionally, MmfR was of interest for a broader spectrum of synthetic biology applications rather than strictly for heterologous expression systems. Novel genetic components that are orthogonal to (i.e., do not interfere with) existing components commonly used in synthetic biology also allow experimenters to build constructs with more complex programmed behaviours. For example, the repressilator, a genetic oscillator which is mentioned in **Section 1.4.2**, consists of three transcriptional repressors repressing one another's expression (Elowitz & Leibier, 2000). That the repressors do not interact with each other's cognate target sites is essential to the construct's function. With more repressor-operator combinations to choose from, the potential complexity of genetic constructs increases further, allowing more intricate designs to be trialled.

ScbR has already been used as a transcriptional repressor in *E. coli* for synthetic biology purposes. Built into a construct alongside the biosynthetic pathway encoding its inhibitor ScbR, it was shown to be fully functional, forming a heterologous quorum sensing system in *E. coli*. Furthermore, it was trialled alongside the acetyl homoserine lactonen (AHL) quorum sensing systems of *Aliivibrio fischeri* and *Pseudomonas aeruginosa*, and shown to be entirely orthogonal to both systems (Biarnes-Carrera, Lee, Nihira, Breitling, & Takano, 2018).

MmfR, being a close relative of ScbR, is very likely to be just as orthogonal to pre-existing transcriptional repressors in widespread synthetic biology usage. Furthermore, MmfR has been shown experimentally not to respond to the presence of SCBs (Zhou et

al., 2021), so the two ArpA-like repressors also appear to be orthogonal to one another – allowing both to be employed within the same genetic construct.

In order to design such constructs, as well as to better understand the observed behaviour of complex genetic systems (such as the regulation of biosynthetic gene clusters), mathematical and computational models can be employed, as described in **Section 1.4**. Examples of mathematical models created to explore ArpA-like TFTR regulation of BGCs are discussed in **Section 1.4.5**.

1.4 Mathematical and Computational Modelling

1.4.1 Mathematical modelling

Mathematical modelling is the expression of a real-world process as a formal, discrete set of interactions. As chemical reaction equations describe the interactions of atoms in an understandable, human-readable fashion, so do models of biological systems consist of a set of discrete elements, akin to chemical species, and a set of possible interactions between them. Mathematical models are typically used to identify *emergent behaviour* – properties arising from the interactions within a complex system which would not be obvious from analysing the individual processes in isolation (Segel & Edelstein-Keshet, 2013).

Mathematical models necessarily simplify the systems they describe ("all models are wrong, but some models are useful") (Schrodinger & Penrose, 2012). In biological systems, the individual elements present and their biochemical and biophysical interactions are too numerous, incidental, and complex to replicate in a model. However, by identifying the most pertinent elements and interactions, it is possible to simplify the model down to a finite series of reaction equations that approximates the behaviour of the real system in a specific, limited, but known context.

A model can therefore be used to predict what will happen when conditions of the system are changed – whether specific reaction rate parameters, concentrations of specific species, or otherwise. This can be very useful when, for example, designing gene

circuits to exhibit a particular behaviour. Through predictive modelling, a range of designs can be trialled before physical implementation (J. Murray, 2002; Segel & Edelstein-Keshet, 2013). There are various ways that a mathematical model can be explored, but two commonly-used outputs are steady-states and time-courses.

1.4.2 Mathematical modelling outputs

A steady-state analysis of a system tells the modeller the final resting state. This can be useful knowledge, for example, in the case of a recombinant protein production system, where the experimenter wishes to know how to achieve the highest production rate when induced or the lowest production rate when uninduced.

A steady-state analysis is performed by first generating equations for the rate of change of each of the "species" or "element" present in the system, based on the reactions in which they are generated or consumed, and then by assuming that the system has reached equilibrium, and that the overall rate of change of concentration for each species is zero. By resolving these rate equations, the experimenter can then find the sets of concentrations at which the system is in equilibrium, either as absolute values or relative to the rate parameters of the component reactions of the system (J. Murray, 2002; Segel & Edelstein-Keshet, 2013).

A time-course analysis, meanwhile, tracks the system from predefined starting conditions and over a certain length of time, by breaking the intervening time down into steps and then iteratively determining the condition of the system at each new step from the various rate-determining parameters and the state of the system at the previous step (J. Murray, 2002; Segel & Edelstein-Keshet, 2013).

Time-course analyses can be more useful that steady-state analyses for several reasons. Firstly, they can be used to assess how quickly the reaction reaches its steady state. However, some models can have no fixed steady states, and others can have multiple potential steady states (J. Murray, 2002; Segel & Edelstein-Keshet, 2013).

In the case of a system with no static steady state, it can instead establish if the system settles into a consistent pattern of behaviour. For example, the Brusselator is a model system consisting of four reactions, which has no steady state but instead exhibits a constant cycle like the real-world Briggs-Rauscher iodine clock reaction (Briggs & Rauscher, 1973). The repressilator, meanwhile, is a classic example of an engineered genetic circuit consisting of three repressors that each inhibit one other repressor's production, resulting in an oscillating production rate of each (Elowitz & Leibier, 2000), as shown in **Figure 1.9**:

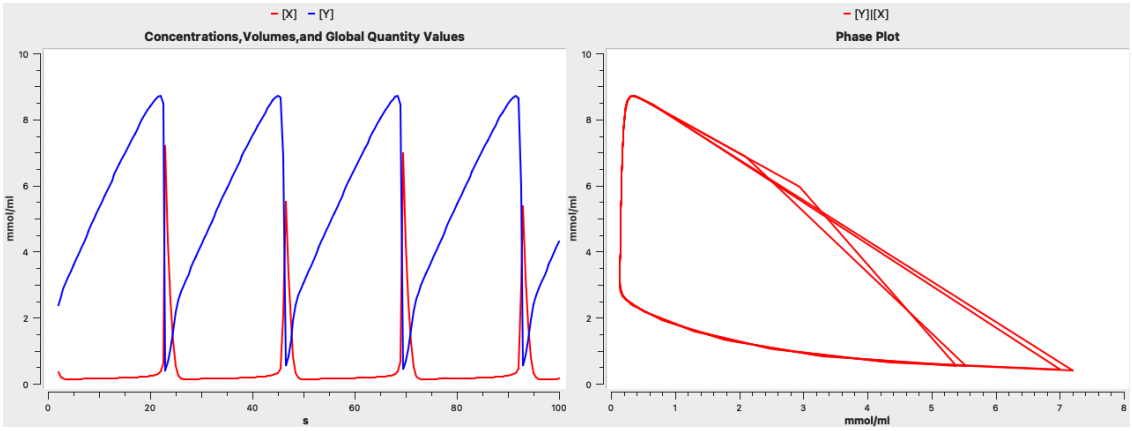

Figure 1.10: *The* ***Brusselator****, a cyclic model reaction system with no equilibrium point, plotted in the Complex Pathway Simulator (COPASI).*

Left*: Time-course of the concentration of two of the six chemical species, X and Y, over time.* ***Right****: Time-course showing the concentrations of X and Y relative to one another at each time-step.*

Meanwhile, systems with multiple potential steady states can be studied using a time-course resolution to determine which state the system is most likely to fall into, given specific starting conditions. The *lux* quorum sensing system of *Aliivibrio fischeri*, for example, consists of an interlinked set of transcriptional regulators that force the expression of the bioluminescent *lux* gene into a binary "on" or "off" state, with both states being possible under some conditions, and the current state determined by the prior set of conditions the cell was exposed to (Williams, Cui, Levchenko, & Stevens, 2008).

However, a time-course analysis requires a new set of calculations to be made at every step, and a time-course of a useful resolution will typically involve hundreds or thousands of steps. Compared to a steady-state analysis, therefore, a time-course is much more computationally intensive (and usually infeasible to do by hand). Relatively simple models with less than 20 reactions and 10^4 steps, like the repressilator model above, can be performed in milliseconds by a modern computer, but more complex models – or simple models run a multitude of times under different conditions – can take considerably more time to resolve.

A model's predictions can only be as accurate as the data used to create it. The parameters that govern reaction rates should be based on experimental data, whether from the modeller's own work or from other publications. Models are inherently imprecise, given that they cannot include every potentially relevant detail. As a result, reasonable accuracy in parameter values is required to obtain useful results, but exact values are generally not necessary.

Mathematical models can be generated and explored in several ways. The oldest method is pencil-and-paper. This approach can be used to identify the steady states of small systems without a computer, but resolving systems with many elements and interactions, or plotting time-courses, becomes extremely laborious.

More commonly, a model is created on a computer. This can be done using a purpose-built program like COPASI, or directly using a scripting language like Python or MATLAB to implement the set of reaction equations that comprise the model and output the results in a human-readable format. Computational models can be vastly more complex than "paper" models, potentially simulating the contents of a whole cell, or running hundreds or thousands of models in sequence to assess the effects of changes in different parameters.

1.4.5 Models of ArpA-like transcriptional regulators

Mathematical and computational models have been applied to the regulatory systems governing the expression of several biosynthetic gene clusters in *Streptomyces* species. Often, these models seek to explore the dynamics of the regulatory system in its native host, for example, to use experimental data already collected to infer the likely behaviour of other elements of the system whose roles are not yet fully understood.

The dynamics of ScbR expression in the coelimycin cluster were explored in a study by Mehra *et al*. ScbR represses its own expression, as well as that of ScbA, which is essential for biosynthesis of SCB1 – the inhibitor of ScbR. In this way, Mehra *et al*. demonstrate a bistable expression pattern in their model under some sets of parameter values: while little SCB1 is present in the system, little SCB1 can be synthesised and thus repression by ScbR is nearly constant. As the ambient concentration of SCB1 increases and ScbR becomes more inhibited, a threshold concentration is reached, above which ScbA expression and thus SCB1 production increase greatly. This shift persists even if SCB1 levels subsequently reduce again, down to a much lower threshold. This phenomenon, in which the state of a system under the same conditions depends on the system's prior state, is known as hysteresis (Mehra, Charaniya, Takano, & Hu, 2008).

Demonstrating that hysteresis occurs in a model system under a specific set of parameter ranges does not prove that hysteresis occurs in the real-life system – whether because the parameter values chosen are unrealistic, or because the purported interactions between model components is simply wrong, or because there are external factors operating upon the system that are unknown or not accounted for by the model. However, the findings of a model can inform experimental design. In the case of SCB1, experimental work did subsequently display a bistable expression pattern in the system, which corresponded closely with the modelling results (Mehra et al., 2008).

Later models by the same group have attempted to explore the apparent interactions between ScbR and ScbA, which are necessary to maintain the bistability of the model system. Earlier modelling efforts suggested that they form ScbR-ScbA

heterodimers (Mehra et al., 2008), while later work posits that bistable behaviour could be due to the precise positioning of the two genes relative to one another, which causes RNA polymerases transcribing the two genes to physically collide and interfere with one another's transcription (Tsigkinopoulou et al., 2020).

Mathematical modelling has also been used to explore the dynamics of MmfR regulation in the MM cluster. Bowyer *et al.* used approximate Bayesian computation (ABC) in an attempt to establish the function of MmyR, an MmfR homologue also present in the MM cluster whose regulatory role is poorly understood. The model incorporates the MARE sequence that control MmfR and MMF production, as well as (in this model) MmyR production, and the MARE sequence that controls MM biosynthesis. Several postulated interactions between the different components were included as biochemical equations, and the model was iteratively run with each possible combination of certain interactions and compared to experimental data to identify the most likely combination of interactions in the real-life system (Bowyer et al., 2017).

The three interactions tested in the model were the ability of MMF to bind to MmfR, and of MM to bind to MmfR and/or MmyR, and the interactions could occur either when the protein was already bound to DNA, when the protein was not bound, both or neither. It was found that, most likely, MMF can bind MmfR under either circumstance and MM does not interact with MmfR or MmyR at all (Bowyer et al., 2017). This example illustrates how, besides being used to direct future experimental work, mathematical modelling can be used in tandem with experimental evidence already obtained to establish the most likely mechanisms behind observed behaviour.

1.5 Research Aims

The project set out to answer the following questions:

- **Is MmfR toxic to *E. coli*? Can it be constitutively produced?**
- **Is MmfR a viable alternative to LacI for regulation recombinant protein production?**

- **How might a MmfR-based expression vector be made a viable alternative to existing LacI-based expression systems?**

1.6 Outline of Thesis Structure

The thesis is organised as follows:

Chapter 2 details the materials and experimental methods used for the work describe in subsequent chapters. The chapter includes details on the consumables, instruments, software, stock solutions, genetic primers and constructs used throughout the project, as well as experimental procedures.

Chapter 3 details the design, assembly, development, and testing of a recombinant protein production system in *E. coli*, using the *S. coelicolor* transcriptional repressor MmfR in place of the research and industry standard *lac* repressor.

Chapter 4 describes mathematical and computational modelling work directed at generating an improved expression construct, using the program Complex Pathway Simulator (COPASI). The modelling uses published biophysical data alongside the experimental results of Chapter 3 to assess the impact of improving specific components of the expression system.

Chapter 5 discusses the results gathered over the previous chapters, examines them in their context within the field of synthetic biology, and ponders the direction of future research into MmfR as a transcriptional repressor for *E. coli* recombinant protein production.

Chapter 2 – Methods

2.1: Materials and Equipment

2.1.1: Consumables

Manufacturer	**Material**
Biotium	GelRed Nucleic Acid Dye
Day-Impex Ltd	Virkon disinfectant
Genewiz	Vectors
Integrated DNA Technologies	Primers
Merck-Millipore	22 µm syringe filter
New England Biolabs	High fidelity restriction enzymes
Thermo Fisher Scientific	Restriction enzymes

2.1.2: Equipment

Manufacturer	**Type**	**Model**
Eppendorf	Microcentrifuge	5452
	Centrifuge	5810 R
	Thermocycler	MasterCycler Nexus
Thermo Scientific	Spectrophotometer	Nanodrop 2000
New Brunswick	Incubator Shaker	C24
StarLab	Heat Block	Mini Dry Bath Incubator
Bio-Rad	Electrophoresis Tray	Mini-Sub Cell GT Cell
BMG Labtech	Microplate Reader	Fluorstar Optima
Jenway	Spectrophotometer	73 Series

2.1.3: Software

Developer	Software
Benchling	Benchling
BMG Labtech	Reader Control
	MARS Data Analysis
COPASI Team	Complex Pathway Simulator (COPASI)

2.1.4: Primers

DT-L0 & DT-M0

Primers for cloning Golden Gate assembly parts; all anneal at **71 °C**:

Primer	Sequence (5'-3')
AmpR 1	
fwd	tctgaagacttagcgccatggggtttcttagacgtcaggtggcac
rev	tctgaagacttcacgctcaccggctcca
AmpR 2	
fwd	tctgaagacttcgtgggtcccgcggtatcatt
rev	tctgaagacttgggacggggtctgacgctcag
LacI 1	
fwd	tctgaagacttaatggtgaaaccagtaacgttatacgatgtcgca
rev	tctgaagacttgtcgcgtaccatcttcatgggag
LacI 2	
fwd	tctgaagacttcgactgggcgtggagcat
rev	tctgaagacttgctgttttcggtatcgtcgtatcccac
LacI 3	
fwd	tctgaagacttcagctcatgttatatcccgccgttaacc
rev	tctgaagactttcactgcccgctttccagtc

psc101	
fwd	aaagaagaccccgctaatattcagcgatttgcccgagcttg
rev	tctgaagacttctccggtctcaggcacattgtcgatctgttcatggtgaacagc
MmfR	
fwd	tctgaagacttaatgaccagcgcacagcagc
rev	atagaagacttaagctctagacgccctttttatgcacgcagtgc
sfGFP	
fwd	tctgaagacttaatgtccaagggcgaggagctg
rev	tctgaagacttaagctcacttgtacagctcgtccatgcc

Single-stranded oligonucleotides to be ligated into double-stranded components for **GGA**:

Oligo	**Sequence (5'-3')**
BBa_J23119	
fwd	tctgaagacttggagttgacagctagctcagtcctaggtataatgctagctactaagtcttcaga
rev	tctgaagacttagtagctagcattatacctaggactgagctagctgtcaactccaagtcttcaga
BBa_J23114	
fwd	tctgaagacttggagtttatggctagctcagtcctaggtacaatgctagctactaagtcttcaga
rev	tctgaagacttagtagctagcattgtacctaggactgagctagccataaactccaagtcttcaga
BBa_J23113	
fwd	tctgaagacttggagctgatggctagctcagtcctagggattatgctagctactaagtcttcaga
rev	tctgaagacttagtagctagcataatccctaggactgagctagccatcagctccaagtcttcaga

DT-M2

For altering constitutive promoters; all anneal at **60 °C**:

Primer	**Sequence**

J23109 (4x)

fwd	ctagggactgtgctagctacttcacacaggaaaagag
rev	gactgagctagctgtaaaatcccaatacgcgtcaattc

J23117 (6x)

fwd	ctagggattgtgctagctacttcacacaggaaaagag
rev	gactgagctagctgtcaaatcccaatacgcgtcaattc

J23114 (10x)

fwd	ctaggtacaatgctagctacttcacacaggaaaagag
rev	gactgagctagccataaaatcccaatacgcgtcaattc

J23105 (24x)

fwd	ctaggtactatgctagctacttcacacaggaaaagag
rev	gactgagctagccgtaaaatcccaatacgcgtcaattc

J23106 (47x)

fwd	ctaggtatagtgctagctacttcacacaggaaaagag
rev	gactgagctagccgtaaaatcccaatacgcgtcaattc

J23104 (72x)

fwd	ctaggtattgtgctagctacttcacacaggaaaagag
rev	gactgagctagctgtcaaatcccaatacgcgtcaattc

J23100 (100x)

fwd	ctaggtacagtgctagctacttcacacaggaaaagag
rev	gactgagctagccgtcaaatcccaatacgcgtcaattc

2.1.5: Vectors

Vector nucleotide sequences are given in **Appendix 1.**

Part sources

Vector	Resistance	Region copied
pC-0221	Apramycin	sfGFP
pC-0032	Ampicillin	MmfR (codon optimised)
pC-0058	Kanamycin	pSC101 origin, MARE (MM4)

5 DT-L/M0 vectors

Vector	Resistance
Repressor	Kanamycin
Reporter (Lac)	Kanamycin
Reporter (MARE + Lac)	Kanamycin
Reporter (MARE)	Kanamycin
Reporter CDS (Insulin)	Ampicillin

Expression Vectors

Vector	Resistance
DT-L1	Kanamycin
DT-M1	Kanamycin

2.1.6: Strains

Transformation

Strain
E. coli TOP10
E. coli BL21

Experimental

Strain	Host	Vector
DT-L0 *E. coli*	*E. coli* TOP10	DT-L0
DT-M0 *E. coli*	*E. coli* TOP10	DT-M0
DT-L1 *E. coli*	*E. coli* BL21	DT-L1
DT-M1 *E. coli*	*E. coli* BL21	DT-M1
DT-M2 *E. coli*	*E. coli* BL21	DT-M2

2.2: Stock Solutions

2.2.1: Growth Media

Lysogeny broth (LB)

Per litre:

10 g tryptone

5 g yeast

10 g sodium chloride

Deionised water to 1 l

Adjusted to pH 7

Autoclaved at 121 ^{o}C, 15 min

Lysogeny broth agar (LBA)

Per litre:

10 g tryptone

5 g yeast

10 g sodium chloride

15 g agar

Deionised water to 1 l

Adjusted to pH 7

Autoclaved at 121 ^{o}C, 15 min

Antibiotics (for selective media):

Ampicillin (100 μg ml^{-1})

Apramycin (50 μg ml^{-1})

Kanamycin (50 μg ml^{-1})

2.2.2: Gel Electrophoresis Media

TAE Buffer, 100x

4 M Tris-acetate

100 mM EDTA

Agarose Gel, 1%, 50ml

500 mg Agar

50 ml 1x tris-acetate-EDTA

Microwaved at full power until dissolved, allowed to cool

5 µl GelRed or SYBR Safe

2.3: Protocols

2.3.1: Bacterial Cultures

E. coli cultures were grown at 37 ^{o}C. Solid cultures were grown on LB agar. Liquid cultures were grown in LB with shaking at 180 RPM. Antibiotics were used where appropriate to the final concentrations listed in **section 2.2.1**.

2.3.2: Vector Assembly

Polymerase Chain Reaction (PCR)

Reactant	**Amount (µl)**
Template (approx. 25 ng µl^{-1})	1
DNA Primer (10 µM)	1 / primer
2x Q5 Hi-Fi PCR Mix	12.5
Nuclease-free water	to 25

Temperature (^{o}C)	**Time (s)**	
98	30	
98	10	**]**
Anneal	30	**] 35 x**
72	30/kbp	**]**
72	120	
4	-	

Oligonucleotide Annealing

Pairs of single-stranded oligonucleotides were annealed in nuclease-free water at a concentration of 10 ng µl^{-1} by heating them up to 94 ^{o}C for 2 minutes using a heat block, and allowing them to cool by turning off the heat block.

Golden Gate Assembly (GGA)

Reactant	**Amount (µl)**
DNA (approx. 100 ng/µl)	0.5 / fragment
10U/ul Restriction Enzyme	1
T7 DNA ligase	0.5

10x T4 DNA ligase buffer	2.5
Deionised Water	to 25

Temperature (°C)	**Time (s)**	
37	10	**]**
16	30	**] 30 x**
60	5	

2.3.3: Transformation by Electroporation

On ice, approximately 100 ng plasmid stock in 1-2 µl buffer was added to 50 ml TOP-10 *E. coli* cells. The mixture was mixed by pipetting, transferred to an electroporation cuvette and incubated for about 1 minute. Electroporation was conducted at **2.5 kV, 25 µF** for **5 ms** before the cuvette was returned to ice, the mixture added to 1 ml sterile LB and incubated for 1 hour at 37 °C and 200 RPM. 10 µl of transformant culture was spread onto selective LB agar plates and incubated overnight at 37 °C.

2.3.4: Plasmid Purification

To create stocks of existing plasmids, colonies from transformant plates were picked and cultured in 5 ml selective LB until turbid. Cells were harvested by centrifugation using the 5180 R at 4 °C, 3220 RCF for 15 min and supernatant discarded. Pellets were processed with a Thermo Scientific GeneJET Plasmid Miniprep Kit according to the manufacturer's instructions.

When extraction of a specific PCR product was necessary, the PCR mixture was separated on a 1 % agarose gel, and the appropriate band was extracted by scalpel and purified using a Thermo Scientific GeneJET Gel Extraction Kit or an NEB Monarch DNA Gel Extraction Kit according to the manufacturer's instructions.

2.3.5: Restriction Digestion Screening

Reaction mixture prepared as follows and incubated at 37 °C for one hour.

Reactant	**Amount (µl)**
DNA (approx. 100 ng/µl)	0.5 per fragment
10U/µl Restriction Enzyme	1

T7 DNA ligase	0.5
T4 DNA ligase buffer	1
Deionised Water	to 25

2.3.6 Platereader Fluorometry

5 ml cell cultures were inoculated from plates and incubated for 1 hour, to an optical density of 0.6 to 0.8 at 600 nm. Cultures were induced with the appropriate concentration of inducer or solvent (**deionised water** or **DMSO**), mixed, and 100 μl aliquots distributed in triplicate in a 96-microwell plate. The microplate was incubated at 37 ^{o}C with shaking during the experiment. The following settings were used:

Basic settings	
Measurement type:	Fluorescence (FI)
Microplate name:	FALCON 96
Plate mode settings	
No. of cycles:	65
Cycle time [s]:	900
No. of flashes per well:	10
Optic settings	
Excitation:	485
Emission:	520
Gain:	1000
Shaking settings	
Shaking width [mm]:	6
Shaking mode:	orbital
Additional shaking time:	300s before each cycle
General settings	
Top optic used	
Settling time [s]:	0.2
Reading direction:	bidirectional, horizontal left to right, top to bottom

Target temperature [°C]: 37

2.4: Construct Design and Development

2.4.1: First-Phase Construct Design

The expression plasmid, designated **DT-M0**, would overall consist of three parts – the **backbone**, making it a functional plasmid, one operon producing the **repressor** protein constitutively, and a second operon producing the **reporter**, with a promoter repressible by the repressor. The plasmids were designed to be assembled from three segments using a **Golden Gate**-based assembly process.

A control plasmid **DT-L0** was also designed. This would be identical to DT-M0 but carry the ***lacI*** gene in place of *mmfR*, and a target site for the *lac* repressor (**LacI**) instead of one for MmfR. This construct would be used to compare the uninduced and induced expression levels of GFP when under control of each of the two repressors. The structure of the construct DT-M0 is shown in **figure 3.1:**

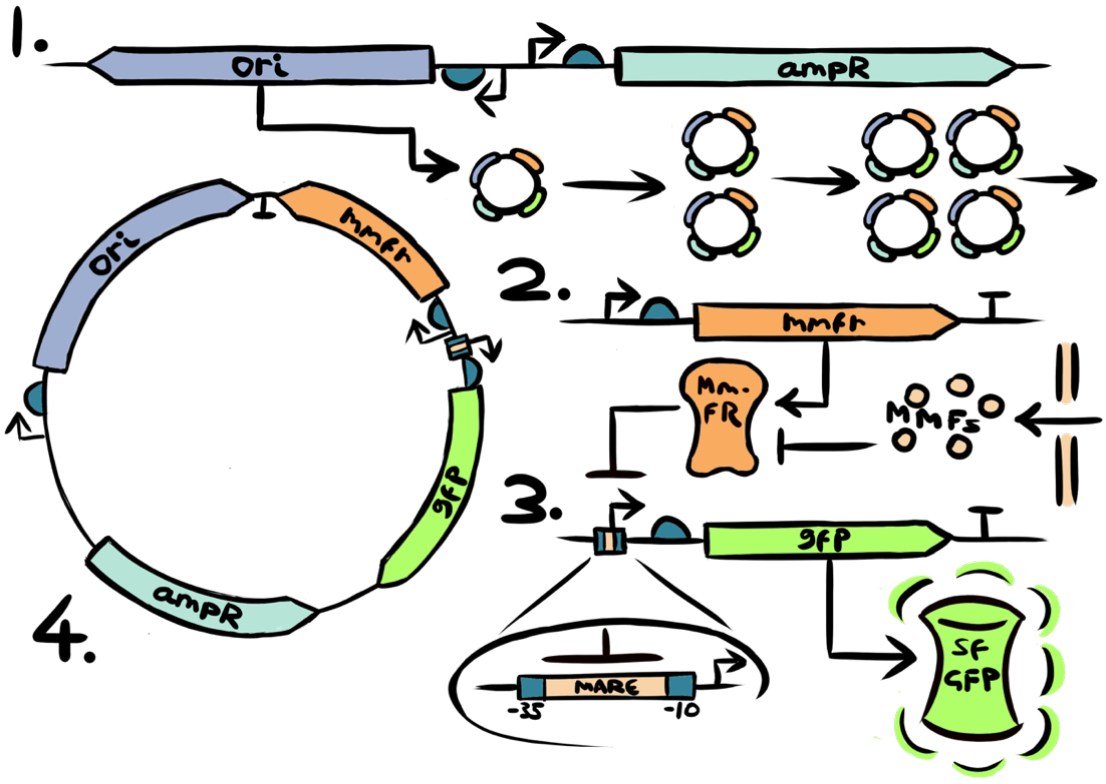

Figure 2.1: *MmfR-based GFP expression plasmid* ***DT-M0****.*

*The fully assembled plasmid (**4**) consists of three parts: the **backbone** (**1**) composed of an **origin** of replication (**blue**) and ampicillin **resistance marker** (**teal**); the **repressor** mmfr (**2**); and the **reporter** gfp (**3**). Each operon includes a promoter and RBS, with the promoter for the reporter gene (bottom right) also incorporating a MARE binding site for MmfR between the -35 and -10 regions.*

The first part, the backbone, would consist of the origin of replication and a resistance marker. A kanamycin resistance marker was chosen for two reasons – firstly, its ubiquity made obtaining and preparing selective media easy, and secondly it lacked any restriction sites (such as **BsaI**) that would complicate the assembly process.

The origin of replication chosen was **pSC101**, an extremely low-copy origin. This would maximise the genetic stability of the construct; by keeping the copy number of the repressor and reporter genes low, the resulting metabolic stress would be minimised.

The second part, the repressor expression cassette, consisted of a transcriptional promoter, ribosome binding sequence (**RBS**), coding sequence and terminator. Other than the coding sequence for the repressor itself, these parts were all sourced from the iGEM registry ("Promoters/Catalog/Anderson - parts.igem.org," n.d.)**.**

The terminator chosen, **BBa_B1006** ("Part:BBa B1006 - parts.igem.org," n.d.), was one of the strongest available. This was done to minimise transcriptional feed-through between the different regions of the construct, which could otherwise cause RNA polymerases to overrun their intended coding sequences and

The RBS **BBa_B0034** ("Part:BBa B0034 - parts.igem.org," n.d.) was likewise chosen for being one of the strongest available, because any necessary adjustment of expression rates could be done by altering the promoter sequences instead.

The promoter sequence was chosen from the range of Anderson promoters, a set of closely related promoter sequences with documented relative strengths ("Promoters/Catalog/Anderson - parts.igem.org," n.d.). From these, the strongest (and original) sequence **BBa_J23119** was chosen. If necessary, it could be replaced with a similar but weaker promoter with a predicable effect.

The repressor would be the primary difference between the constructs. The LacI-based construct would use *E. coli*'s ***lacI.*** The MmfR-based construct would use an already-tested version of the *S. coelicolor* ***mmfr*** gene sequence-optimised for *E. coli*.

The third part, the reporter cassette, consisted of a promoter, RBS, coding sequence and terminator. The terminator and RBS used were the same as for the repressor cassette.

The reporter protein used in both constructs was superfolder green fluorescent protein (**sfGFP**) (Pédelacq, Cabantous, Tran, Terwilliger, & Waldo, 2005), which could be easily quantified through fluorometry. sfGFP is functional when expressed by *E. coli* and

is highly soluble, allowing high concentrations to be present in the cell without generating inclusion bodies (insoluble aggregates of misfolded proteins).

In the **LacI**-based construct **DT-L0** the transcriptional promoter region would be the *lac* operator. The exact sequence used was based on the part **BBa_R0010** from the iGEM registry ("Part:BBa R0010 - parts.igem.org," n.d.). This part is a direct copy of most of *E. coli*'s genomic *lac* operator, including the **O1** and **O3** binding regions for LacI. Unlike the wildtype Lac repressor, the **O2** region is not included, being 401 base pairs downstream of the O1 and part of the *lacZ* gene, but its absence only reduces the repression activity of LacI approximately threefold (Oehler, Eismann, Krämer, & Müller-Hill, 1990).

For the **MmfR**-based construct **DT-M0**, the promoter was a minimally modified version of the ***lac*** operator. Through specific base pair substitutions at the **O1** and **O3** LacI binding sites, the ability of LacI to bind and supress gene expression was effectively eliminated (Oehler et al., 1990). Meanwhile, the addition of a MARE sequence between the -35 and -10 regions of the promoter would allow MmfR to bind. The minimal changes between the two promoter sequences would allow the most direct possible comparison between MmfR and LacI as transcriptional repressors.

2.4.2: Obtaining Components

The components for the assembly process were obtained in several different ways. Some were synthesised; others were modified from existing constructs through PCR. Each of the parts, and their flanking restriction enzyme sites, are shown in **figure 3.2:**

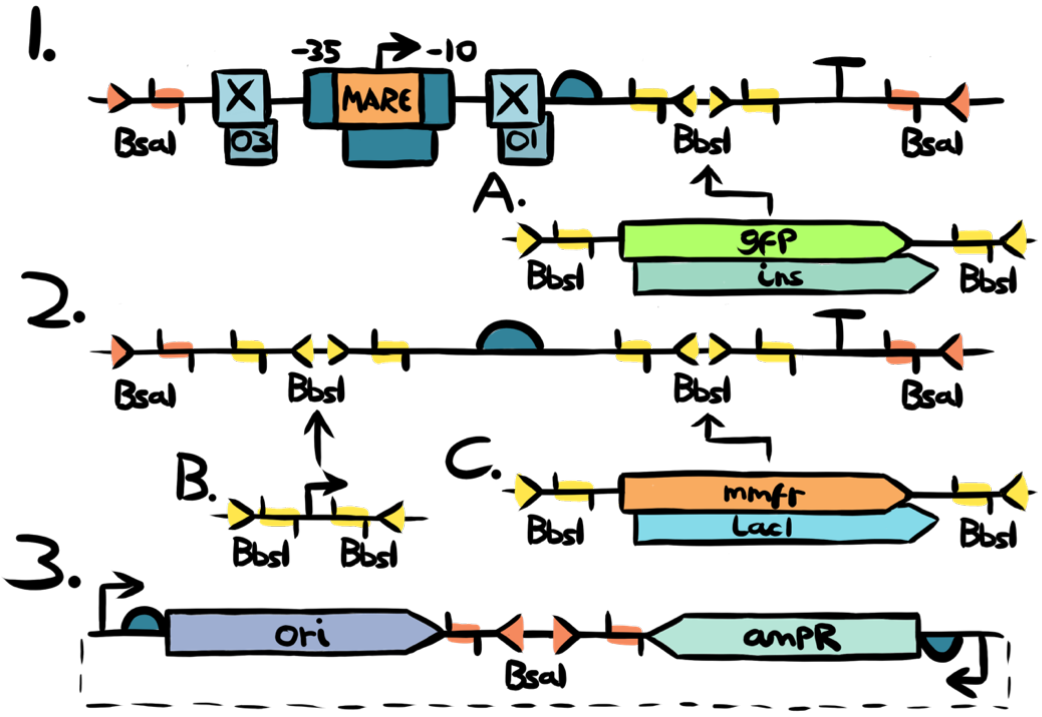

Figure 2.2: *Components and assembly process of the* ***DT-M0*** *MmfR-based GFP expression construct.*

The ***reporter*** *operon (**1**) consists of the lac operator (O3 and O1) and a* ***promoter*** *containing a MARE sequence between the -35 and -10 RNA polymerase binding sites, an* ***RBS*** *and a* ***terminator****. Between the RBS and terminator are a pair of* ***BbsI*** *restriction sites, into which the* ***reporter*** *gene* gfp *or* ins *(**A**) can be inserted. The* ***repressor*** *operon (**2**), similarly, includes an RBS and terminator as well as BbsI sites for the insertion of a* ***promoter*** *(**B**) and a* ***repressor*** *gene* mmfR *or* lacI *(**C**). The* ***backbone*** *(**3**) includes an* ***origin*** *of replication and* ***resistance marker****.*

The vector for the "intermediate" level 0 assembly of the **repressor** operon was synthesised by an external supplier. This vector's backbone consisted of a **pUC** origin of replication and a kanamycin resistance marker, while its functional region included an RBS, a terminator and two **BbsI** sites that would accommodate a promoter and a coding sequence. **BsaI** sites separated the two regions of the vector. This same level 0 vector would be used for both the LacI and MmfR expression cassette. The RBS and terminator were synthesised as part of the vector because any changes to expression rate would be made through assembly with a different promoter sequence.

The two vectors for the reporter level 0 assemblies were also synthesised, one consisting of the unmodified **Lac** operator, and the other the modified operator with a **MARE** sequence. Only the coding sequence for the reporter needed to be inserted into this vector. The same origin of replication and resistance marker were used as the repressor level 0 vector.

The coding sequences for the two reporter genes – insulin and GFP – were obtained differently. The GFP reporter was generated through high-fidelity **PCR** of the

plasmid pC0221, using primers with **BbsI** restriction enzyme sites located on either end, a much cheaper method than full gene synthesis.

The insulin reporter, meanwhile, was synthesised because it could not be obtained from available, pre-existing constructs. The sequence used was the full human *ins* gene, prefaced by a **pelB** periplasmic expression tag and followed by a hexahistidine purification tag. The former tag was included because the periplasmic region of *E. coli* permits disulfide bond formation, which would theoretically save a chemical sulfonation step downstream, and the latter was included to allow easy separation of the insulin from the rest of the cell debris.

The constitutive promoter sequences were synthesised as pairs of single-stranded oligonucleotides, which were then annealed to one another by heating in nuclease-free water. These promoter sequences were synthesised separately so that they could be used interchangeably during assembly, should the need for constructs with different promoter strengths arise.

The coding sequences for **LacI** and **MmfR**, like the sequence for GFP, were generated through PCR of the pre-existing plasmid pC0032, with **BbsI** sites on either side to allow them to be inserted into the level 0 repressor vector.

The parts of the plasmid backbone were also obtained by PCR amplification of pC0058, with **BbsI** sites on the inward-facing primers, where the parts would be joined, and **BsaI** sites on the outward-facing primers for the level 1 assembly. The **ampicillin** resistance marker was made up of two separate parts, to remove an existing BsaI restriction site from the gene as described in Chapter 2.

2.4.3: First-Phase Assembly

As stated in **Section 2.4.1**, the components would be combined in a two-stage process using **Golden Gate assembly.**

In the first stage ("**level 0**") of the assembly process, the backbone and the two operons were assembled separately. The restriction enzyme **BbsI** was used to cleave 4bp overhang sequences, inwards of recognition sites positioned on the edges of each fragment. The overhang sequences of the fragments of the latter two sequences were designed according to the **MoClo** standard assembly system, with specific 4bp sequences used for the joins between the promoters, RBSs, coding sequences (including signal tags) and terminators.

In the second stage ("**level 1")** of the assembly process, the backbone and the two operons were joined together, using the restriction enzyme **BsaI** for the assembly process.

When the plasmids synthesised by external supplier were transformed by electroporation and plated, one did not grow on the kanamycin plate as it should have done, but did grow on ampicillin selective plates. Another transformant showed two different distinct colony morphologies on transformation, suggesting either contamination or a mutation in the plasmid vector immediately following transformation. Repeat transformations of these plasmids from the supplied stocks yielded the same results.

During the assembly process, the assembled level 0 constructs consistently failed to transform correctly into *E. coli* – no growth was observed on selective media following transformation and recovery. This may have been due to a very low rate of survival or growth in cells that received the construct. The metabolic stress of gene expression caused by the strong promoters and RBSs chosen for the coding sequences, alongside the high-copy-number pUC cloning vectors chosen for the level 0 assembly vector, was most likely to blame. Similar problems were encountered – and successfully addressed – in a subsequent phase of construct design.

2.4.4: Second-Phase Construct Design

The assembly process of the construct **DT-M0** failed. Multiple constructs from the supplier were suspected to be incorrectly assembled or contaminated. Additionally, the number of steps in the process – including two levels of Golden Gate assembly following the generation of various components by PCR – would make troubleshooting a long and intensive process. As a result, it was decided to restart the design process. Because of the difficulties in assembling the previous constructs, the new ones would be synthesised and assembled in their entirety by an external supplier. A breakdown of the plasmid is shown in **figure 3.5**:

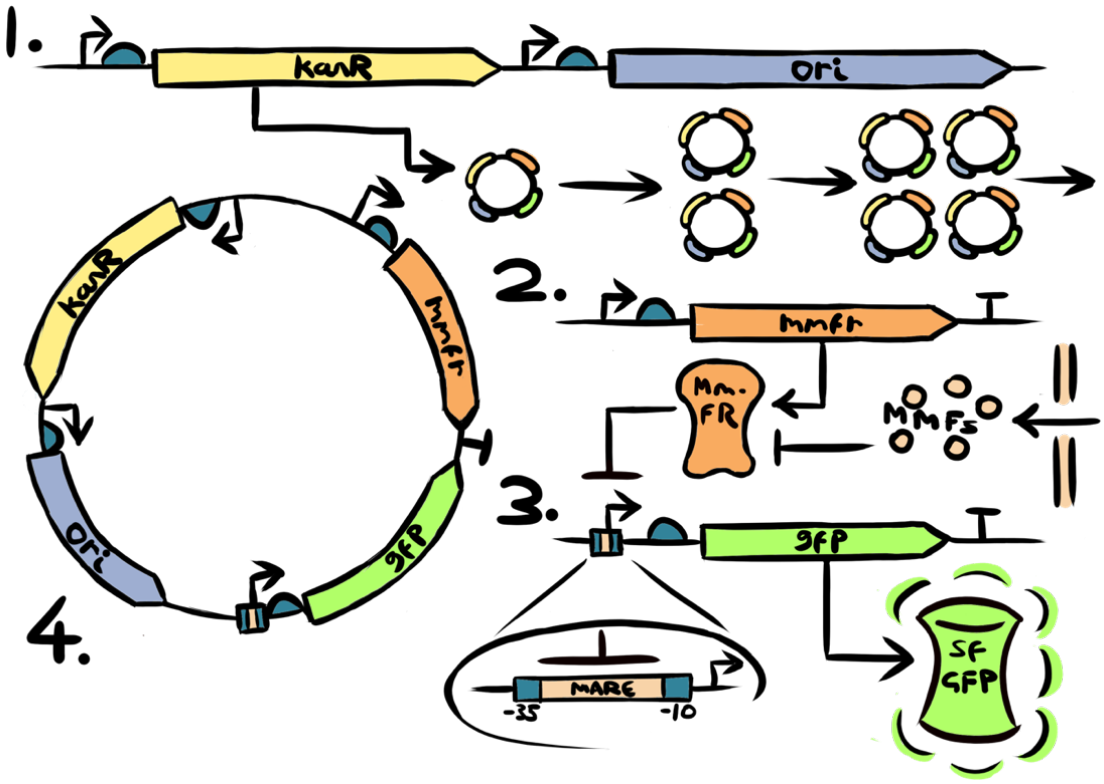

Figure 2.3: *Revised MmfR-based GFP expression plasmid pDT-M1.*

*The fully assembled plasmid (**1**) consists of three parts: the backbone (**2**) composed of an **origin** of replication (**blue**) and kanamycin **resistance marker** (**yellow**); the **repressor** mmfR (**3**); and the **reporter** gfp (**4**). Each operon includes a promoter and RBS, with the promoter for the reporter gene (bottom right) also incorporating a MARE binding site for MmfR between the -35 and -10 regions.*

As before, the two constructs consisted of two operons each, one producing **sfGFP** under the control of a repressor, and one for the constitutive production of the repressor. There were very few changes between the operons of the two constructs.

The repressor operon consisted, again, of a constitutive promoter followed by an **RBS**, the ***lacI*** gene or a codon-optimised ***mmfR*** gene, and finally a terminator sequence. The reporter operon, meanwhile, consisted of the *lac* operator, with (in the MmfR-based construct) a MARE sequence positioned between the -35 and -10 recognition sites of the promoter. The promoter was followed by an **RBS** and the coding sequence for sfGFP.

The two operons would face one another, with a single terminator between them. This was due to concerns about how close together the promoters were in the previous constructs, and whether the binding of a repressor or RNA polymerase in one direction might obstruct transcription in the other direction.

The origin of replication used this time was **pUC**, a high-copy origin of replication typically used for cloning. In contrast with pSC101 origin, it was expected that more

copies of the DNA would result in higher yields of sfGFP. Kanamycin resistance was once again chosen as the selection marker.

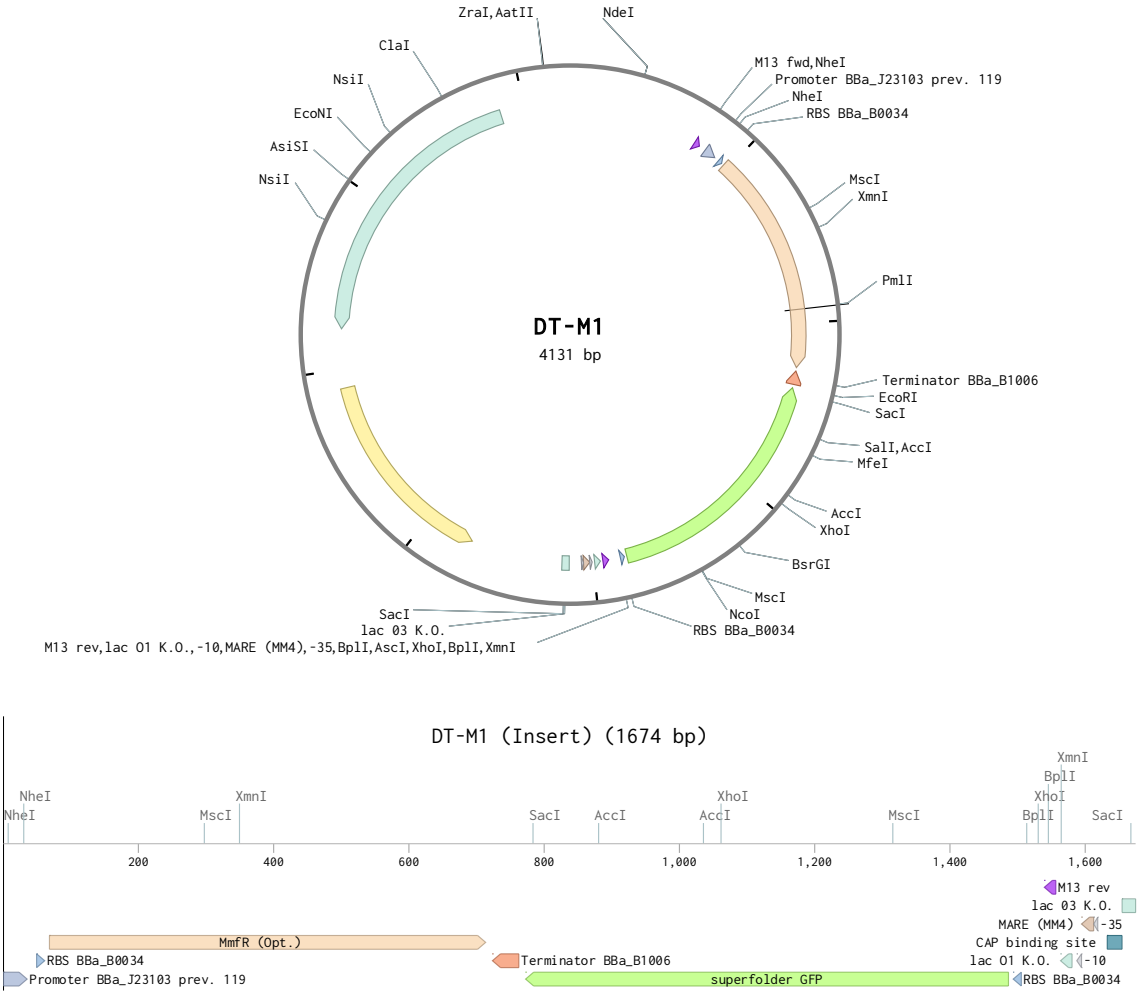

Figure 2.4: *Plasmid map of expression plasmid pDT-M1 (diagram from Benchling).*

Top: *Whole plasmid map of pDT-M1.* ***Bottom:*** *Insert, including repressor operon (left) and reporter operon (right). The repressor operator (**bottom left**) consists of a constitutive promoter, RBS and the* mmfr *gene. The reporter operon (**bottom right**) features a* lac *operator with the key O1 and O3 regions knocked out, and with the MARE target sequence for MmfR positioned between the -10 and -35 regions.*

2.4.5: Second-Phase Assembly

During the transformation process, it was discovered that neither construct was stable in *E. coli*. The LacI-based plasmid pDT-L1 was only stable in *E. coli* with a base insertion at the 10th nucleotide of the LacI gene, with the resulting frame-shift turning the 14th codon of the gene into a stop codon. This would have substantially reduced the metabolic stress of expression, suggesting that the expression rate previously was much too high.

The other, MmfR-operated construct was unable to be assembled at all. It was successfully synthesised as two fragments, separated across the promoter region of the

repressor gene. This prevented the repressor from being expressed until the two halves were assembled (at which point, the cell immediately became non-viable).

The inability of either construct to express its repressor suggested that their rate of expression was too high for their host cells to sustain. The constitutive promoter controlling repressor expression was therefore replaced with a weaker promoter. The original promoter, the Andersen promoter **BBa_J23119**, was replaced with the much weaker promoter **BBa_J23113**. This substitution allowed the assembly of the constructs to be completed without further issues.

The plasmids were transformed into **BL-21** *E. coli* by electroporation, as described in **Section 2.3.3**.

Chapter 3 – Development of an MmfR-Based Protein production System in *E. coli*

3.1 Introduction

The objective of the work described in this chapter is to create a recombinant protein production platform using the *S. coelicolor* transcriptional repressor **MmfR**, the target **MARE** site and the inducer **MMF** to regulate expression in *E. coli*, comparable to a typical LacI/LacO/IPTG expression system.

The work described in this chapter was divided into three stages:

- Firstly, the design of a functioning MmfR-based GFP expression system, with a counterpart LacI-based expression system for comparison. This work is predominantly described in **Section 2.4** and referenced in **Section 3.2**.
- Secondly, a comparison of the two systems, with regards to toxicity, induced expression levels, and leakiness (uninduced expression levels). This work is described in **Section 3.3**.
- Thirdly, the optimisation of the MmfR-based expression system to improve the above characteristics. This work is described in **Section 3.4**.

During the work described in this chapter, the first objective was met, the second was mostly met, and progress towards the final objective was made. Future work to achieve the final objective is described in **Section 3.5.2**.

3.2 Vector Development for MmfR-Regulated Protein production

3.2.1 Introduction

As established in **Chapter 1**, the **aim** of the project described in this thesis is to develop and optimise a recombinant protein production construct using the *S. coelicolor* transcriptional repressor **MmfR** for transcriptional control, following the same general design architecture as the traditional LacI-operated system.

For the purposes of direct comparison, the differences between the two constructs were designed to be minimal – namely the repressor gene, and the target site for that repressor – and to produce a reporter that would be easily detectable, making comparison between the constructs' output as easy to compare as possible. MmfR's performance as a transcriptional repressor would then compared

experimentally with LacI, to establish the extent of further optimisations that would need to be carried out to make it a viable alternative.

3.3 Confirmation of MmfR Functionality in E. coli

3.3.1 Method development

Having designed a construct to express **GFP** under the transcriptional control of **MmfR**, assembled it and transformed it into *E. coli*, the next step was to confirm that the MmfR was functional in its new host, and that it did not interfere with host metabolism in a way that would render the construct useless.

It would also be necessary to narrow down an ideal concentration of MMF to use for induction. For the purposes of initial testing, MMF concentrations were chosen based on previous work with MmfR, which suggested that 1 µM would be sufficient to reach near-maximal induction rates, and a halved and doubled concentration were chosen for comparison.

It was anticipated that the control and pDT-L1 constructs would each show the same levels of fluorescence with or without MMF1, and that the pDT-M1 construct would show roughly the same level of fluorescence as pDT-L1 in the absence of MMF1. This level was likely to be slightly higher than the level displayed by the media control, and potentially increasing slowly over time, allowing for baseline leakiness of GFP regulation.

It was furthermore expected that the fluorescence of pDT-M1 would increase substantially over time following induction with MMF1, due to the MMF binding to MmfR and derepressing the production of sfGFP. It was expected that higher concentrations of MMF would lead to higher levels of fluorescence, although this was not necessarily expected to be a proportional relationship.

A culture of each construct was incubated overnight, divided into microwells, induced and the fluorescence tracked by a platereader for 8 hours. The readings are shown in **figure 3.1**:

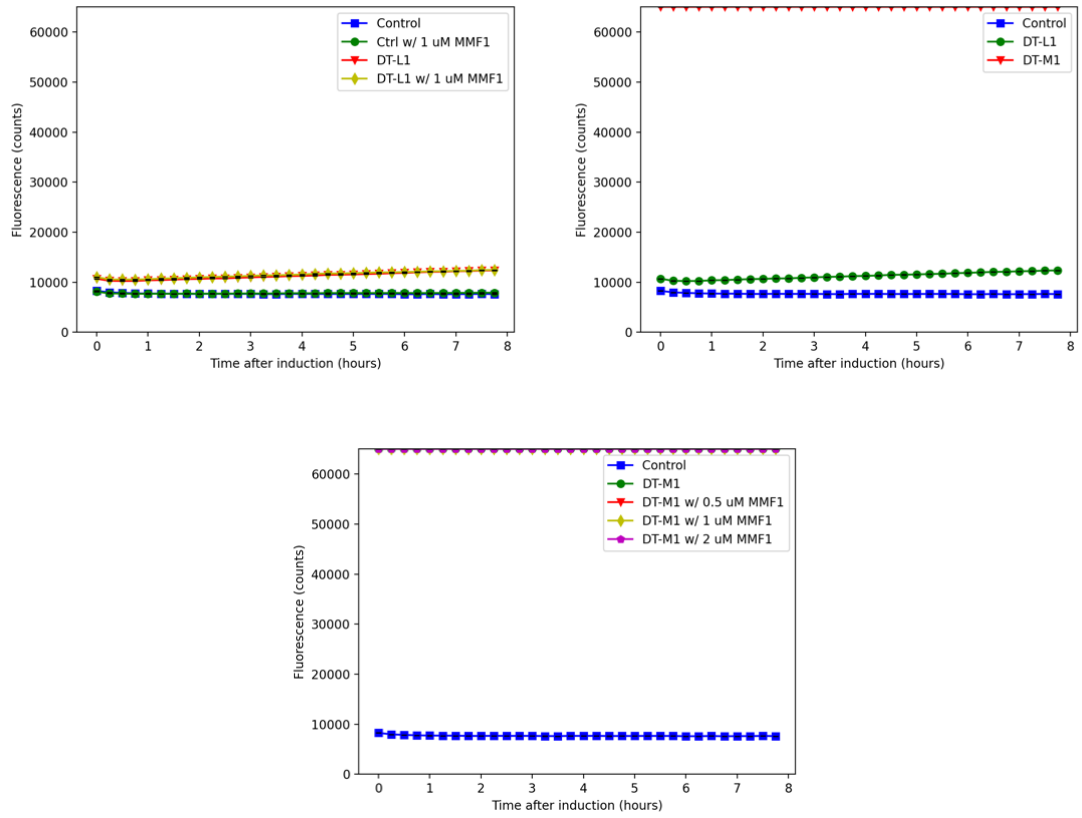

Figure 3.1: *Fluorescence over time of cultures expressing GFP under the repression of LacI (**DT-L1**) or MmfR (**DT-M1**) after **16 hours** of growth, and then the addition of either MmfR's inducer, **MMF1**, or no induction.*

***Top left**: fluorescence of the media control and pDT-L1 cultures when induced with MMF1. **Top right**: the fluorescence of uninduced cultures (media control, pDT-L1 and pDT-M1). **Bottom:** the fluorescence of the pDT-M1 cultures at different induction concentrations.*

The results show that the fluorescence level of the MmfR-regulated cultures already exceeded the detection range of the equipment at the point of induction (and also confirm that that **MMF1** does not itself fluoresce and does not interact with LacI). As a result, it was decided to instead induce the cultures much earlier, to ensure that the MMF was affecting GFP expression levels.

To demonstrate the ability of the MmfR-regulated construct to respond to MMF, a subsequent experiment was conducted using a similar protocol with some alterations. Firstly, the cultures were incubated for only one hour after inoculation, so that their GFP expression during exponential growth could be observed. Secondly, the incubation period was 16 hours rather than eight, to allow all the cultures to reach steady state. Thirdly, a second MMF, MMF5, was used in addition to MMF1.

DT-M1 was tested under two concentrations of MMF, 1 µM and 0.5 µM. 2 µM was omitted because it was anticipated that 1 µM MMF would already provide a maximal rate of induction (given that in comparable LacI-operated expression systems,

0.1 µM IPTG is generally enough to achieve maximal expression). The results are shown in **Figure 3.2:**

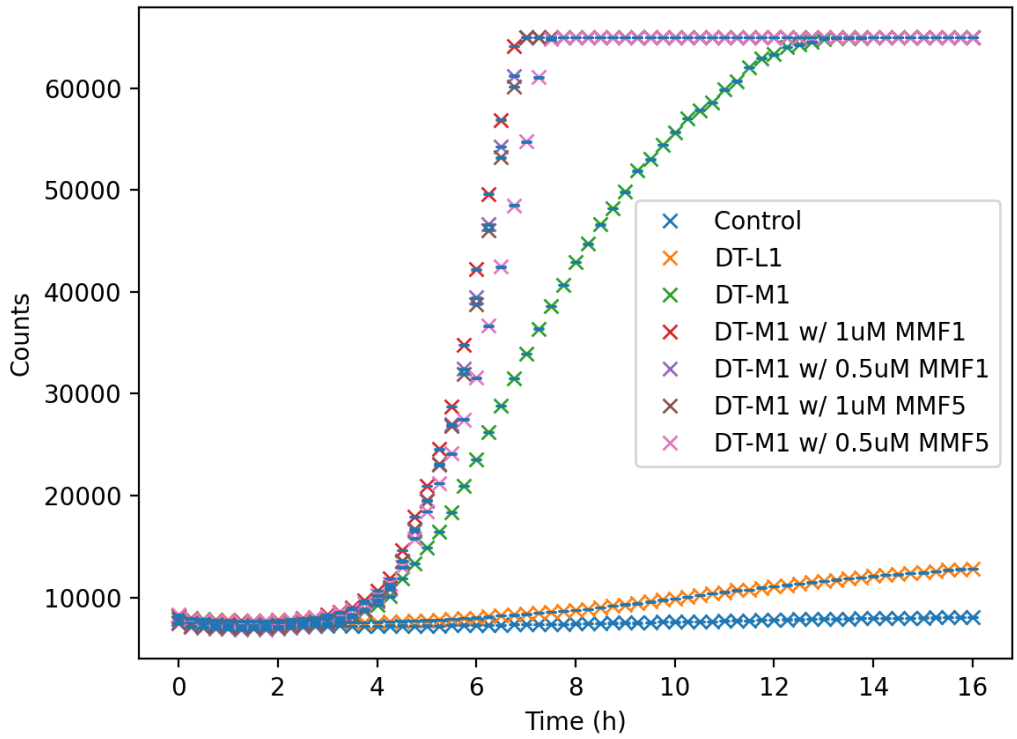

Figure 3.2: *Fluorescence over time of cultures expressing GFP under the repression of LacI (**DT-L1**) or MmfR (**DT-M1**) after 16 hours of growth, and then the addition of either MmfR's inducer, **MMF1**, or no induction.*

The results show fluorescence increasing over time in all samples except the negative control. The increase is sigmoidal in all those cases, with the rate reaching a maximum at about 5-6 hours after induction. The uninduced pDT-L1 shows a small increase in fluorescence over time, peaking in fluorescence at about 1/3rd of the peak of the uninduced pDT-M1, while the control shows none. Meanwhile, pDT-M1 shows a much greater rate increase of fluorescence over time than pDT-L1 without MMF induction, and an even greater rate of increase with either MMF1 or MMF5 at either 0.5 µM or 1 µM, all combinations of which appear to create maximal GFP expression. The fluorescence of all pDT-M1 cultures eventually exceeds the measurement capability of the platereader at around 6 hours, except for the uninduced culture, which does so at about 13 hours. Several important conclusions could be drawn from the data obtained from these two experiments:

Firstly, the MmfR protein does repress GFP expression in the pDT-M1 construct, and this repression is inhibited in the presence of an inducer; in other words, the MmfR functions correctly when expressed by *E. coli*. The addition of either MMF substantially

increases the rate of GFP production in pDT-M1 (and no corresponding increase is seen in pDT-L1, or the control). However, neither the concentration nor type of MMF show much difference, suggesting that both concentrations used were in substantial excess. In future experiments, it is likely that lower concentrations could be used with little impact on expression rate.

Secondly, MMF does not interfere with the binding of LacI to its target site, the *lac* repressor. pDT-L1 did not display a change in GFP output when induced with MMF, with both showing a low level of expression rising slowly over time. This observation also suggests that MMF does not have off-target interactions with any other parts of *E. coli*'s metabolism that would lead to inhibited growth in the culture, confirming its non-toxicity. As the solvent used for storage of the MMF was DMSO, however, it would be necessary to test the effects of DMSO alone to conclusively state this.

Thirdly, the MMF used is not fluorescent, at least not enough to be detectable with this equipment setup. The fluorescence of the control culture did not change with the addition of MMF. However, the addition of DMSO alone was not trialled, so it is technically possible (but unlikely) that MMF *is* fluorescent, but the DMSO solvent somehow counteracts this fluorescence.

Finally, both repressors are leaky, but MmfR is much leakier than LacI. There is a detectable basal expression rate of GFP in both strains, and that the rate is much higher in the MmfR-repressed strain than the LacI-repressed strain. This is most likely the result of the MmfR simply having a lower affinity for the MARE sequence than LacI does for the *lac* operator, either when expressed in *E. coli* or simply in general. It is also possible that MmfR is folding imperfectly after expression, and its binding affinity is thus reduced – because of, for example, the differences in the cell environment of *E. coli* compared to *S. coelicolor*, or because it is being overexpressed. It could, alternatively, reflect a lower overall rate of expression, perhaps due to requiring a balance of amino acids that is better catered for by *S. coelicolor* than by *E. coli*.

3.3.2 Confirmation of activity

The goal of this part of the project (as addressed in section 3.1) was not only to demonstrate that MmfR is effective in *E. coli* and non-toxic to it, but to further develop the MmfR -MMF system as a vehicle for recombinant protein production. The ultimate objective was to provide a viable alternative to LacI-IPTG expression systems. To gain

traction as an alternative expression system, the MmfR-MMF system would need to perform comparably, and ideally better than its competitor.

Specifically, the MmfR system would need to display a comparable leakiness to the LacI system without sacrificing its induced expression rate. Leaky expression systems result in greater metabolic stress on the host, and therefore a higher rate of mutations that render the genetic circuit useless, so a low leakiness is desirable.

Thus far, it had been established that the MmfR protein was functional in *E. coli*, and that it was less effective as a repressor than LacI in the constructs that were tested. However, a quantitative assessment of relative binding effectiveness and inducibility had not yet been conducted.

The protocol used in **section 3.3.1** was repeated. This time, the gain of the fluorometer in the plate reader was adjusted from 1422 to 1000. This was done to avoid the prior issues with cultures rapidly exceeding the maximum detectable fluorescence, which had thus far prevented comparison of the GFP outputs of the two constructs.

Additionally, during preparation a tenfold serial dilution was made of each inducer from the 100 mM stock in the appropriate solvent, and 10 µl of dilution was added to 90 µl of culture to make up each well. Wells that had DMSO present – those induced with an MMF, and the DMSO controls – had a final concentration of 1% DMSO.

This was a refinement on the previous experiments, in which inducer had been added to the different wells directly from a stock solution, despite being added to different final concentrations. As a result, each well had been of a slightly different final volume, and because the MMF was stored in DMSO, different wells also contained different concentrations of DMSO. Although relatively minor differences, it was possible that this had measurably affected the results of prior experiments.

The results of this experiment are shown in **figure 3.3:**

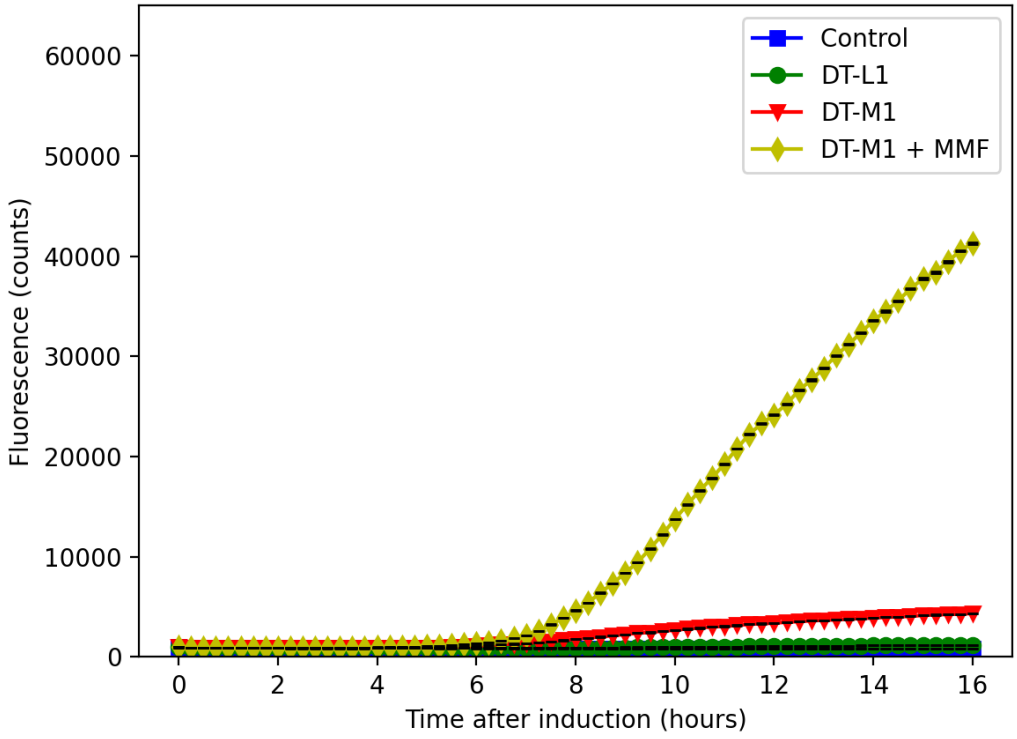

Figure 3.3: *Fluorescence over time of cell cultures carrying one of two GFP expression constructs, pDT-L1 and pDT-M1.*

When not induced, pDT-M1 shows an increase in fluorescence over time, while pDT-L1 shows very little or no increase. After induction with MMF, the culture carrying pDT-M1 shows a much greater increase in fluorescence over time. The difference in baseline expression of GFP between pDT-L1 and pDT-M1 is still obvious under the lower gain setting used, but the fluorescence of all the wells remained comfortably below the maximum detection range of the equipment. However, the increases in fluorescence appear much slower than in previous experiments, taking about 10 hours to reach the maximum rate of increase.

The results show, most importantly, that pDT-M1 displays much higher fluorescence if induced with MMF. This reaffirms that MmfR is repressing GFP expression as expected by binding to the MARE target sequence within its promoter, and that it responds to the presence of MMF by releasing the MARE and enabling GFP expression.

DT-M1 still proved to be much leakier than pDT-L1. This could be due to a lower actual rate of expression (despite their identical promoters and RBSs), or due to MmfR having a lower affinity for its MARE target sequence than LacI does for the *lac* operator.

3.4 Construct Optimisation

3.4.1 Introduction

As a result of the observed leakiness of the pDT-M1 construct, it was decided to experiment with increasing the expression rate of the MmfR. A series of seven pairs of PCR site-directed mutagenesis primers were prepared, targeted at the constitutive promoter driving MmfR expression in pDT-M1. Each pair of primers contained a small number of sequence changes to the promoter sequence.

3.4.2 Mutagenesis

The promoter sequences chosen were Anderson promoters, taken from the iGEM database, with documented expression rates relative to the benchmark promoter **BBa_J23119**. The promoter used thus far, **BBa_J23113**, has a relative strength of **0.01**, and the subsequent promoters were chosen such that there would be an even spread of relative strengths between **0.01** and **0.1**, and between **0.1** and **1** ("Promoters/Catalog/Anderson - parts.igem.org," n.d.). The promoter sequences are listed in the appendix.

The primers were designed using New England Biolabs' NEBasechanger web-based program, which uses the original sequence and the desired changes to algorithmically create the primer sequences for use with NEB's Q5 Site-Directed Mutagenesis Kit. The primer sequences can be found in the appendix.

The PCR mutagenesis of the MmfR promoter sequence in pDT-M1 was conducted as specified in Chapter 2. Amplification was confirmed by gel electrophoresis, and transformation into *E. coli* conducted by electroporation. Following transformation and plating on selective media, it was observed that only one transformant (carrying the promoter **BBa_J23117,** with a strength of 6 times that of BBa_J23113) grew in selective media. All transformations were repeated, and the transformant carrying BBa_J23117 was again the only one capable of growing in selective media.

It is unlikely that the failure of transformation of most of the constructs is due to incomplete or incorrect mutagenesis, as the mutagenesis process was repeated multiple times, and the products were of the expected length according to gel electrophoresis.

The failure of transformation could be due to metabolic stress. Promoter strengths are known to be somewhat sensitive to their genetic surroundings, so it is possible that the other promoters are either strong enough that MmfR is being

massively overproduced in those constructs, or weak enough that GFP production is barely inhibited.

However, ultimately the mutagenesis procedure did yield one construct with a stronger promoter for MmfR than in its predecessor.

3.4.3 Characterisation

The construct pDT-M2, which had a higher rate of MmfR expression than pDT-M1, was successfully created and transformed. It was anticipated that this construct would display lower leakiness in its GFP expression than pDT-M1, and hoped that its leakiness would be comparable to the LacI-operated pDT-L1 construct.

To confirm this, a platereader experiment was set up according to the same protocol used in **section 3.2.2**, and the results are shown in **figure 3.4:**

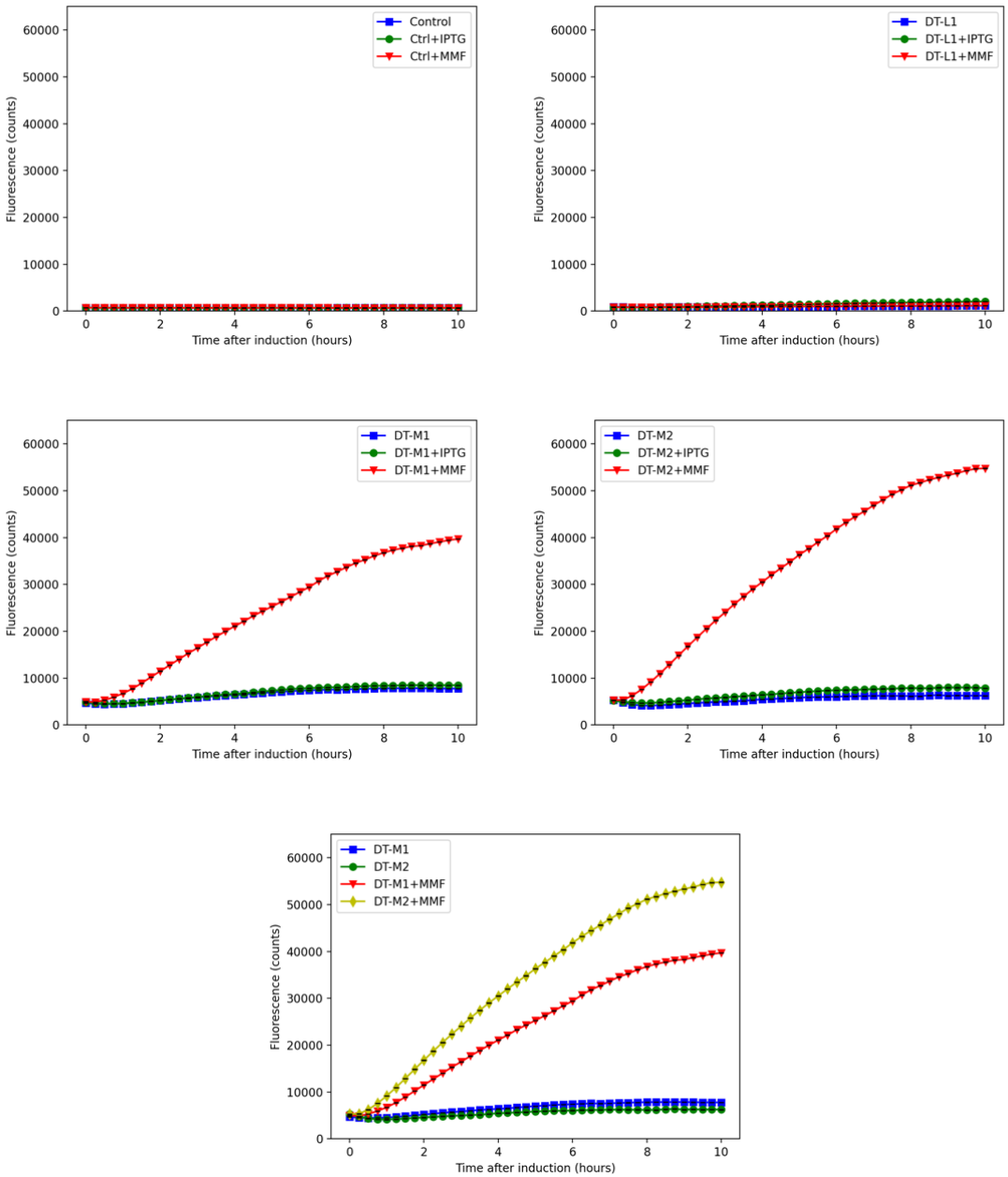

Figure 3.4: *Fluorescence time-course of GFP-expressing constructs pDT-L1, pDT-M1 and pDT-M2 from 0 to 10 hours after induction with* ***water, IPTG*** *or* ***MMF****.*

Top left*: Negative control with IPTG, MMF1, or no inducer added.* ***Top right:*** *pDT-L1 fluorescence with no inducer, IPTG or MMF1.* ***Middle left:*** *pDT-M1 fluorescence with no inducer, IPTG or MMF.* ***Middle right*** *pDT-M2 fluorescence with no inducer, IPTG or MMF.* ***Bottom:*** *Direct comparison between pDT-M1 and pDT-M2.*

DT-M1 and pDT-M2 both show a level of baseline GFP expression higher than that of pDT-L1, although pDT-M2's baseline expression is slightly lower than that of the original. Both reach steady state after about 8 hours.

When induced with MMF1, both constructs show fluorescence increasing rapidly after an hour, and beginning to level off after 7 hours. pDT-M2 shows a much higher rate of increase than the original at all stages.

Notably, pDT-L1 shows only slightly higher fluorescence under induction by IPTG than it does without induction. The extremely low level of GFP production by pDT-L1, even when induced, suggests that an error was made in the setup of the experiment.

Potentially, the initial cell culture was inoculated from a plate colony carrying a mutated construct. This could be ascertained by repeating the experiment using cultures inoculated from one or more other colonies on the plate, or by re-transforming the construct.

It was also possible that pDT-L1's unresponsiveness was due to an issue with the IPTG used to induce it. In this case, a freshly prepared IPTG stock solution would likely solve the problem.

Directly comparing pDT-M1 and pDT-M2 to one another is still possible, however. pDT-M2 shows two favourable properties over the pDT-M1. Firstly, its rate of baseline expression of GFP is lower, showing that the increased production of MmfR is leading to less leakiness in its repression.

Secondly, its level of induced expression is higher than that of pDT-M1. No direct alterations have been made to the rate of GFP expression in the mutant construct, so the difference is most likely due to a higher cell count, owing to the reduced metabolic strain caused by leaky GFP expression.

Because previous experiments had produced inconclusive results, either due to equipment settings, unexpected behaviour from the constructs or other complications, it was decided to comprehensively demonstrate the function of all three constructs developed as part of the previous work.

The experiment conducted in **section 3.3.2** was repeated, with all three constructs (DT-L1, pDT-M1 and pDT-M2) tested using each of the two inducers, IPTG and MMF1. The results are shown in **figure 3.5** and **figure 3.6**:

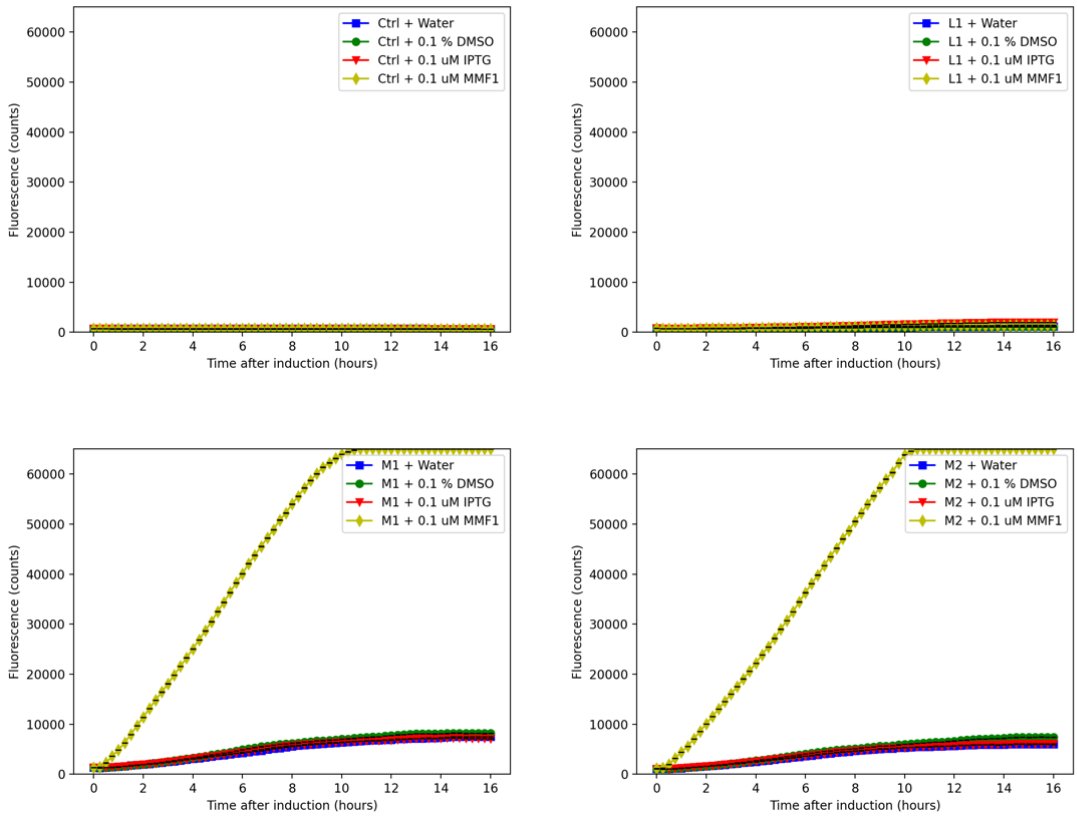

Figure 3.5: *Fluorescence of GFP-expressing constructs from 0 to 16 hours after induction at an OD600 of 0.5 to 0.7 with each of four solutions.*

Top left: *The negative control displaying no fluorescence with or without the addition of inducer.* ***Top right:*** *pDT-L1 displaying little to no fluorescence with or without the addition of inducer.* ***Bottom left:*** *pDT-M1 displaying low fluorescence with water or DMSO only, or with IPTG in water, and high fluorescence with MMF1.* ***Bottom right:*** *pDT-M2 displaying high fluorescence with MMF1 and low fluorescence with all others.*

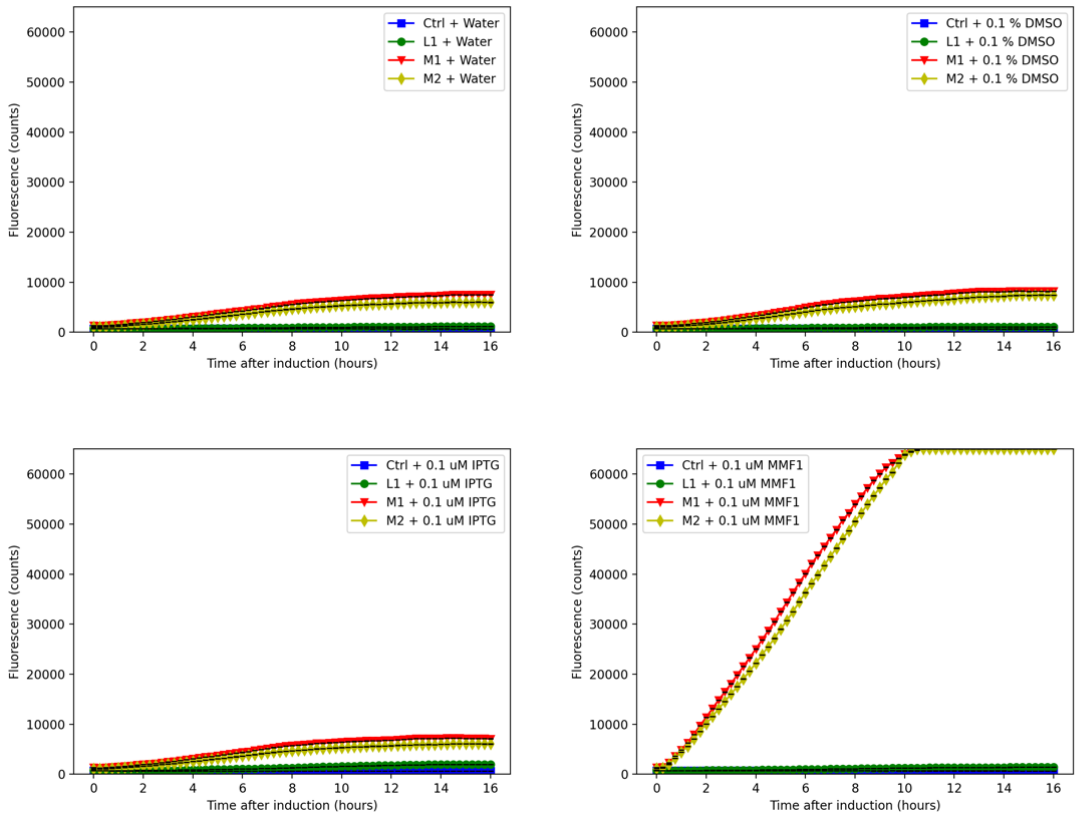

Figure 3.6: *Fluorescence of GFP-expressing constructs from 0 to 16 hours after induction at an OD600 of 0.5 to 0.7 with each of four solutions: water (**top left**), 1 % DMSO (**top right**), IPTG (**bottom left**) and MMF1 in 0.1 % DMSO (**bottom right**). Results are the same as those depicted in **Figure 3.5**.*

The results show little to no difference between those cultures "induced" with deionised water, with DMSO and with IPTG in deionised water. The control culture shows minimal fluorescence under any circumstance, while pDT-L1 shows a very small increase in fluorescence over time regardless of whether IPTG or any other substance is added. Both pDT-M1 and pDT-M2 show a comparatively higher level of fluorescence under all circumstances, but especially when induced with MMF. Induction with MMF causes no change in the control or LacI-based construct, while the responses of both pDT-M1 and pDT-M2 are nearly identical in magnitude.

The near-identicality of the results for cultures that had 0.1 % DMSO added, compared to those with deionised water added, suggests strongly that DMSO at that concentration is low enough not to inhibit cell growth or GFP synthesis. This confirms that any differences in GFP expression level observed in cultures induced with MMF are due to the MMF itself, not the DMSO buffer.

Similarly, and more concerningly, induction with 100 µM IPTG in deionised water has similarly little effect on any of the cultures, including that with the LacI-based construct. Like the results of the previous experiment, this demonstrates that the pDT-

L1 construct is no longer responding to IPTG (which, as it was prepared fresh for the experiment, is unlikely to be degraded). This suggests again that the construct is not very genetically stable, given that multiple colonies from the transformation plate had at this point failed to respond to IPTG.

These results also show very little difference in behaviour between pDT-M1 and pDT-M2, unlike the previous experiment. In fact, the original construct appears to produce GFP slightly faster than the mutant this time. pDT-M2's fluorescence appears to increase more linearly, suggesting it might ultimately peak at a higher level, but this cannot be conclusively shown because both constructs exceed the maximum detection range.

DT-M1 shows slightly higher fluorescence at most timepoints but appears to level off naturally just before reaching the maximum of the detection range. Meanwhile, pDT-M2 appears to be still increasing linearly in fluorescence until just before reaching the detection limit, implying it might have peaked at a higher final fluorescence had the detection limit been higher. However, more data would be required to establish this.

3.5 Conclusions & Discussion

3.5.1 Findings

Several conclusions can be drawn from the experiments described in this chapter:

Firstly, *E. coli* can constitutively express MmfR without suffering a notable decrease in fitness when compared to a strain expressing LacI at the same rate. Overexpression of MmfR may cause the cell to become nonviable, but this is also true of LacI and the effect of MmfR does not appear to be substantially more damaging for the cell.

Secondly, MmfR expressed by *E. coli* is functional, able to bind to its MARE target site and repress downstream genes. This demonstrates that MmfR could be used as a heterologous repressor in E. coli, providing a method of gene expression parallel to the cell's native metabolism and to the *lac* repressor system. This system could have a variety of applications in recombinant protein production, as well as in more complex gene circuits used in synthetic biology.

Thirdly, the attempt to reduce the leakiness of pDT-M1 by increasing the expression rate of MmfR may have worked, generating the pDT-M2 construct, but the difference is minor, and more data are required to confirm this.

However, the pDT-M1/M2 expression system requires further refinement. Despite attempts to increase MmfR expression and thereby reduce the leakiness of GFP expression, the rate of uninduced expression remains well above that of the LacI construct. This elevated metabolic stress reduces the construct's viability for large-volume bioreactors, by encouraging mutation of the expression pathway or product.

3.5.2 Implications for Future Work

The work described in this chapter shows the potential for MmfR's use in recombinant protein production, but is currently limited to the production of GFP. While GFP is a highly visible and non-toxic reporter, it would be beneficial to show that an MmfR-based expression system can also be used to produce a useful protein.

During the first round of construct design, alternative versions of DT-L0 and DT-M0 were designed to produce preproinsulin instead of GFP. However, due to the failure of the GFP-generating DT-L0/M0 to assemble correctly, the assembly of their insulin-producing variants was ultimately never attempted.

The subsequent round of construct design that produced pDT-L1 and pDT-M1 did not include variants producing non-GFP proteins, to save costs (each construct being synthesised as a single part). Once the MmfR-based construct has been further optimised, developing proof-of-concept variants that express other products would be a logical next step.

To further address the original goals of the project, several avenues could be explored:

Firstly, the binding affinity of MmfR for the MARE sequence could be improved. This could be done through mutagenesis of MmfR's DNA binding locations. PCR primers containing randomised bases could be used to create a library of plasmids with altered MmfR binding sites. The library could be screened by transformation into *E. coli* and growing individual strains in the platereader, with fluorescence measured over time. Strains displaying low uninduced GFP production would be identified and re-tested with inducer, to ensure they still produced GFP.

This mutagenesis could yield a more efficient MmfR variant. However, to keep the selection process manageable with conventional laboratory equipment, only a small number of amino acids could be changed. This would require referencing sequence data,

folding predictions, and X-ray crystallography data to identify mutagenesis targets that would achieve maximum effect with minimal changes.

As well as, or instead of, altering the sequence of MmfR, changes could be made to the MARE sequence it binds to. This has already been done (Miriam Rodriguez-Garcia, work not published) targeting the four bases with the greatest effect on MmfR's binding affinity, but the existing combinations could be re-tested with different MmfR mutants, or other bases could be changed.

Secondly, the expression plasmid's origin of replication could be changed. The pUC origin, which was used in pDT-L1, pDT-M1 and pDT-M2, is a very high-copy origin of replication, typically reaching several hundred copies per cell. This makes it ideal for cloning, but is less optimal for protein production. The sheer number of copies of the gene mean that its expression rate is high even with a very weak promoter, and a correspondingly large number of repressor molecules are necessary to inhibit this expression rate.

By reducing the number of copies of the *gfp* gene present and increasing the strength of its promoter to compensate, the leakiness of the expression system could be greatly reduced without altering MmfR or to its binding site. However, pDT-L1 used the same origin of replication, and still displayed much lower leakiness than pDT-M1, so this approach would not make MmfR "better" than LacI but could reduce the disparity between constructs.

Thirdly, the architecture of the genetic circuit could be readjusted to decrease the metabolic strain imposed on the host. In the constructs pDT-M1 and pDT-M2, MmfR does not inhibit its own expression. In *S. coelicolor*, however, the gene *mmfR* shares a bidirectional promoter with the MMF expression pathway *mmfLHP*, and a MARE sequence is positioned on the promoter. MmfR regulating its own expression may reduce overall metabolic stress.

Finally, the use of MMF instead of IPTG as an inducer has hidden advantages. MMF is a quorum sensing messenger, and relatively simple in comparison to IPTG. Crucially, MMF is synthesised by *S. coelicolor*, meaning that its synthesis pathway could be transferred to *E. coli*. If functional, this would allow a user to produce MMF via cell culture rather than chemical synthesis, which might reduce overhead costs.

The MMF-generating strain could be added directly to the culture medium, acting as a living inducer. Alternatively, the MMF expression pathway to be included in

the main expression plasmid, creating a synthetic quorum sensing system. This would allow the expression system to induce itself automatically after reaching a specific concentration. Either approach would circumvent the costs of MMF extraction and purification.

To effectively make changes to the kinetics of the MmfR system, however, it would be necessary to analyse in-depth, requiring mathematical modelling – the focus of **Chapter 4**.

Chapter 4 – Modelling of MmfR Expression System

4.1 Introduction

In **Chapter 3,** the MmfR-operated expression construct pDT-M1 functioned as intended, but its performance left room for improvement. The construct's baseline, uninduced expression of GFP was high enough that the cell cultures became visibly green after several hours' growth; due to this "leakiness", the expression circuit was still putting a sizeable metabolic strain on the cells even before induction. To redress this, several approaches for reducing the leakiness of GFP expression without curtailing induced expression were considered.

To determine ways of decreasing leakiness of the expression system without compromising the induced output, mathematical modelling would be useful. By creating quantitative models of the pDT-M1 expression system, it would be possible to explore the ways in which altering the efficacy of various components – promoters, ribosome binding sites and even terminators – would alter the behaviour of the system. The creation of such a model is described in **section 4.2**. The model was first created on paper, and then implemented and tested in Complex Pathway Simulator (COPASI) (Hoops et al., 2006).

COPASI is a software application designed for the modelling of biochemical reaction networks. Through a simple graphical user interface, it allows the user to quickly set up models by inputting each of the reactions that occur within the system. After assigning starting quantities of biochemical species, units, reaction types and rate parameters, the user can then create run a specific task, such as a steady state analysis, time course, or one of several more complex tasks. COPASI can run models deterministically and stochastically, with several different algorithms available for both (Hoops et al., 2006).

COPASI's primary advantage over other computational modelling approaches is its balance between approachability and flexibility. It provides a range of means of both building and testing mathematical models, without requiring the user to learn scripting or coding languages, and allows data to be collected and displayed in a variety of ways that remain user-friendly to set up (Hoops et al., 2006).

Previous modelling work has aimed to elucidate the dynamics of expression systems regulated by ArpA-like TFTRs, as detailed in **Chapter 1, Section 4.2**. However,

such models generally focus on the repressor in its usual genetic context, such as the regulation of an antibiotic-producing biosynthetic cluster in a *Streptomyces* species. The models presented in this work are instead intended to analyse the findings of experimental work in which MmfR, devoid of its usual genetic context, was used to regulate expression of a reporter in *E. coli*, and extrapolate which avenues of development are likely to yield the greatest improvement.

4.2 MmfR-GFP Models

4.2.1 Model of Chapter 3 Construct

A mathematical model of the pDT-M1 construct was created in **COPASI** (Hoops et al., 2006). **Figure 4.1** offers a visual representation of the model, as well as the stoichiometric equations used in its construction:

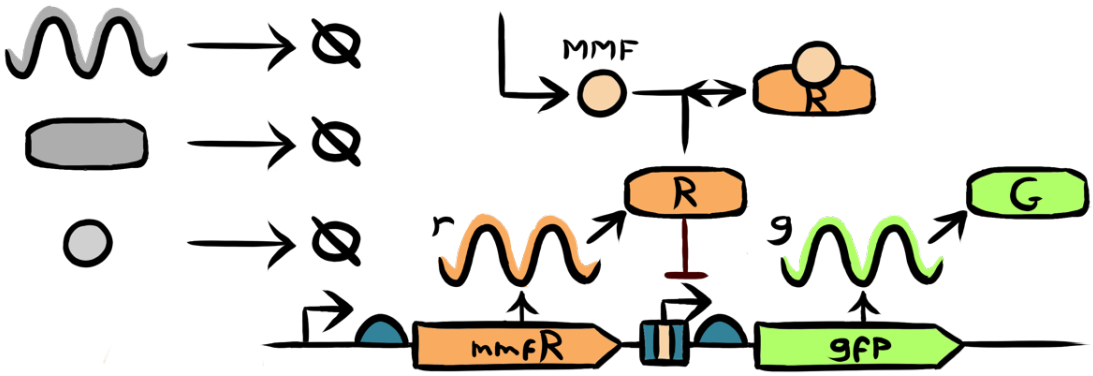

Figure 4.1: *Genetic circuit diagram of Chapter 3 expression construct.*

4.2.2. Model Construction

`The model was constructed by defining a number of biochemical species and the interactions (or reactions) between them. A verbal summary of the model follows:

The gene ***mmfR*** is transcribed constitutively, generating mRNA (**r**), which in turn is translated into the MmfR protein (**R**). When dimerised (**RR**), MmfR binds reversibly to the MARE in the promoter region of ***gfp***, forming a repressor-operator complex (**gfp.RR**). This complex cannot be transcribed, but when *gfp* is not bound by MmfR, its mRNA transcript (**g**) is generated at a constant rate, which is in turn translated into GFP (**G**). **MMF** is not synthesised internally, and must be added from the outside, but when present it binds reversibly to MmfR dimers (**MMF.RR**), which causes the dimer to unbind from *gfp*. The biochemical species are summarised in **Table 4.1:**

***Table 4.1**: Species in MmfR/MARE/MMF expression system model.*

Name	Description
mmfR	*mmfR* operator region
r	*mmfR mRNA*
R	MmfR protein monomer
gfp	*gfp* operator region
g	*gfp* mRNA
G	GFP protein
RR	MmfR protein dimer
gfp.RR	GFP-MmfR dimer complex
MMF	Methylenomycin furan
MMF.RR	MMF-MmfR dimer complex

All the reactions in the model proceed at a rate determined by the law of mass action, although degradation reactions also included a rate of effective depletion due to cell growth rate. Reversible reactions were instead implemented as pairs of non-reversible reactions, because COPASI does not allow stochastic simulations to include reversible reactions, although stochastic simulations were ultimately not explored in this work. The reactions are summarised in **Table 4.2**:

***Table 4.2**: Reaction equations of MmfR/MARE/MMF expression system model.*

#	Name	Reaction	Rate Law
1	*mmfR* transcription	mmfR -> mmfR + r	Mass action
2	*mmfR* translation	r -> r + R	Mass action
3	*gfp* transcription	gfp -> gfp + g	Mass action
4	*gfp* translation	g -> g + G	Mass action
5	*mmfr* RNA degradation	r ->	Degradation
6	MmfR degradation	R ->	Degradation
7	*gfp* RNA degradation	g ->	Degradation
8	GFP degradation	G ->	Degradation
9	MmfR dimerization	2 * R -> RR	Mass action
10	(reverse)	RR -> 2 * R	Mass action
11	*gfp* inhibition	gfp + RR -> gfp.RR	Mass action
12	(reverse)	gfp.RR -> gfp + RR	Mass action
13	Unbound MmfR inhibition	MMF + RR -> MMF.RR	Mass action
14	(reverse)	MMF.RR -> MMF + RR	Mass action
15	Bound MmfR degradation	MMF.RR -> MMF + R	Degradation
16	Bound MmfR inhibition	MMF + gfp.RR -> gfp + MMF.RR	Mass action
17	Dimeric MmfR degradation	RR -> R	Degradation

The reactions are broken down in more detail below:

1: The gene *mmfR* (**mmfr**) is transcribed, generating an mRNA transcript (**r**). This is a first-order reaction – in other words, its rate is dynamically determined by the concentration of *mmfr*, as will be shown in X.Y.Z. The reaction equation is as follows:

(1) *mmfr -> mmfr + r*

2: The transcript r is translated, producing the repressor protein MmfR (**R**), again as a first-order process:

(2) *r -> r + R*

3, 4: The gene *gfp* (**gfp**) is likewise transcribed into mRNA (**g**), which is then translated to produce the reporter protein GFP (**G**), both following the same kinetics as the MmfR counterparts:

(3) *gfp -> gfp + g*

(4) *g -> g + G*

5 – 8: The mRNA transcripts (5, 7) and the proteins degrade through normal cellular processes. In addition, they are functionally diluted as the cells replicate. A doubling in cell count equates to a doubling in cumulative cell volume and, therefore, a halving of the concentration of the contents assuming that none are produced in that time. As a result, the mRNAs and proteins effectively have a half-life equal to the doubling time of the culture, in addition to their normal degradation processes.

(5) *r -> ∅*

(6) *R -> ∅*

(7) *g -> ∅*

(8) *G -> ∅*

9, 10: Two monomers of the repressor MmfR dimerise to form a functioning homodimer $MmfR_2$ (**RR**). This interaction is reversible, with MmfR2 able to dissociate back into the two monomers.

(9, 10) *2R <-> RR*

11: MmfR monomers can also **degrade** while part of a dimer, leaving the other monomer behind.

(11) *RR -> R*

12, 13: The repressor MmfR (**R**) binds reversibly to the **MARE** target site, located on the promoter of *gfp*. The MARE is represented as part of the gene in this model, with the binding of MmfR to *gfp* producing a combined "species" (**gfp.R**) that cannot generate its mRNA transcript (**g**).

MmfR's repressive effect could have been implemented differently, by instead making the rate of *gfp* transcription inversely proportional to $MmfR_2$ concentration. However, like the decision to represent transcription and translation

(12, 13) *gfp + RR <-> gfp.RR*

14: MmfR monomers may degrade whilst their dimer is bound to the MARE, causing the dimer to become a monomer and dissociate from the MARE:

(14) *gfp.RR -> gfp + R*

15, 16: The inducer methylenomycin furan (**MMF**) binds to the repressor MmfR (**R**), creating a combined species that cannot repress *gfp* transcription:

(15, 16) *MMF + RR <-> MMF.RR*

17 MmfR (R) is still able to be degraded while bound by MMF. This degradation does not destroy the MMF, so the MmfR's degradation releases MMF back into the cellular environment.

(17) *MMF.RR -> MMF + R*

No reaction is included for the degradation of MMF. This is for two reasons: firstly, MMF as a small metabolite is known to be very stable, unlikely to be broken down directly by *E. coli*, and degrades naturally only very slowly relative to the timescale of the experiment. Secondly, MMF will not be "diluted" through the act of cell population growth because it is added to the extracellular medium and diffuses into cells, so is evenly distributed throughout the whole cell population *and* the growth medium.

Rate Equations

As stated in the previous section, the rates of all reactions in the system are simply determined by mass action. The rate of flux (change in concentration) of each component is summarised in **Table 4.3**:

Table 4.3: *Rate equations of MmfR/MARE/MMF model expression system as presented in COPASI.*

$$\frac{d([r]\cdot V_{compartment})}{dt} = +V_{compartment}\cdot(k1_{(mmfr\_transcription)}\cdot[mmfR]) - V_{compartment}\cdot\left([r]\cdot\left(v_{(mmfr\_rna\_deg)}+1-0.5^{\frac{1}{division\_time}}\right)\right)$$

$$\frac{d([R]\cdot V_{compartment})}{dt} = +V_{compartment}\cdot(k1_{(mmfr\_translation)}\cdot[r]) + V_{compartment}\cdot\left([MMF{:}RR]\cdot\left(v_{(mmfr\_prot\_bound\_deg)}+1-0.5^{\frac{1}{division\_time}}\right)\right) - 2\cdot V_{compartment}\cdot(k1_{(^{*}mmfr\_dimerisation\ (forward)^{*})}\cdot[R]\cdot[R]) + 2\cdot V_{compartment}\cdot(k1_{(^{*}mmfr\_dimerisation\ (backward)^{*})}\cdot[RR]) + V_{compartment}\cdot\left([RR]\cdot\left(v_{(mmfr\_dimer\_deg)}+1-0.5^{\frac{1}{division\_time}}\right)\right) - V_{compartment}\cdot\left([R]\cdot\left(v_{(mmfr\_prot\_deg)}+1-0.5^{\frac{1}{division\_time}}\right)\right)$$

$$\frac{d([g]\cdot V_{compartment})}{dt} = +V_{compartment}\cdot(k1_{(gfp\_transcription)}\cdot[gfp]) - V_{compartment}\cdot\left([g]\cdot\left(v_{(gfp\_rna\_deg)}+1-0.5^{\frac{1}{division\_time}}\right)\right)$$

$$\frac{d([G]\cdot V_{compartment})}{dt} = +V_{compartment}\cdot(k1_{(gfp\_translation)}\cdot[g]) - V_{compartment}\cdot\left([G]\cdot\left(v_{(gfp\_prot\_deg)}+1-0.5^{\frac{1}{division\_time}}\right)\right)$$

$$\frac{d([gfp]\cdot V_{compartment})}{dt} = +V_{compartment}\cdot(k1_{(mmfr\_bound\_inhibition)}\cdot[MMF]\cdot[gfp.RR]) - V_{compartment}\cdot(k1_{(^{*}gfp\_inhibition\ (forward)^{*})}\cdot[gfp]\cdot[RR]) + V_{compartment}\cdot(k1_{(^{*}gfp\_inhibition\ (backward)^{*})}\cdot[gfp.RR])$$

$$\frac{d([MMF]\cdot V_{compartment})}{dt} = -V_{compartment}\cdot(k1_{(^{*}mmfr\_inhibition\ (forward)^{*})}\cdot[MMF]\cdot[RR]) + V_{compartment}\cdot(k1_{(^{*}mmfr\_inhibition\ (backward)^{*})}\cdot[MMF{:}RR]) + V_{compartment}\cdot\left([MMF{:}RR]\cdot\left(v_{(mmfr\_prot\_bound\_deg)}+1-0.5^{\frac{1}{division\_time}}\right)\right) - V_{compartment}\cdot(k1_{(mmfr\_bound\_inhibition)}\cdot[MMF]\cdot[gfp.RR])$$

$$\frac{d([RR]\cdot V_{compartment})}{dt} = -V_{compartment}\cdot(k1_{(^{*}mmfr\_inhibition\ (forward)^{*})}\cdot[MMF]\cdot[RR]) + V_{compartment}\cdot(k1_{(^{*}mmfr\_inhibition\ (backward)^{*})}\cdot[MMF{:}RR]) + V_{compartment}\cdot(k1_{(^{*}mmfr\_dimerisation\ (forward)^{*})}\cdot[R]\cdot[R]) - V_{compartment}\cdot(k1_{(^{*}mmfr\_dimerisation\ (backward)^{*})}\cdot[RR]) - V_{compartment}\cdot\left([RR]\cdot\left(v_{(mmfr\_dimer\_deg)}+1-0.5^{\frac{1}{division\_time}}\right)\right) - V_{compartment}\cdot(k1_{(^{*}gfp\_inhibition\ (forward)^{*})}\cdot[gfp]\cdot[RR]) + V_{compartment}\cdot(k1_{(^{*}gfp\_inhibition\ (backward)^{*})}\cdot[gfp.RR])$$

$$\frac{d([gfp.RR]\cdot V_{compartment})}{dt} = -V_{compartment}\cdot(k1_{(mmfr\_bound\_inhibition)}\cdot[MMF]\cdot[gfp.RR]) + V_{compartment}\cdot(k1_{(^{*}gfp\_inhibition\ (forward)^{*})}\cdot[gfp]\cdot[RR]) - V_{compartment}\cdot(k1_{(^{*}gfp\_inhibition\ (backward)^{*})}\cdot[gfp.RR])$$

$$\frac{d([MMF{:}RR]\cdot V_{compartment})}{dt} = +V_{compartment}\cdot(k1_{(^{*}mmfr\_inhibition\ (forward)^{*})}\cdot[MMF]\cdot[RR]) - V_{compartment}\cdot(k1_{(^{*}mmfr\_inhibition\ (backward)^{*})}\cdot[MMF{:}RR]) - V_{compartment}\cdot\left([MMF{:}RR]\cdot\left(v_{(mmfr\_prot\_bound\_deg)}+1-0.5^{\frac{1}{division\_time}}\right)\right) + V_{compartment}\cdot(k1_{(mmfr\_bound\_inhibition)}\cdot[MMF]\cdot[gfp.RR])$$

4.2.3. Parameters

The reaction volume was set as 1 μm^3, the approximate volume of an *E. coli* cell (although cell size is somewhat dependent on growth rate). Initial rate parameters for the reactions comprising the model were sourced from literature:

A theoretical maximum rate of transcription is around 20 min^{-1} $gene^{-1}$ copies for extremely strong promoters (Kennell & Riezman, 1977; Pai & You, 2009), but experimental data based on the Anderson promoters – a set of related promoter sequences from which the constructs in Chapter 3 were generated – purports far lower transcription rates (Kelly et al., 2009).

Work by Kelly *et al.* shows that under "certain" conditions, the Anderson promoter J23101 initiates transcription at a rate of 0.03 s^{-1}. The lac promoter, used in the constructs in Chapter 3, is given an RPU of 1.5 (relative to J23101), putting its transcription initiation rate (k_2) at about 0.045 s^{-1} or 2.7 min^{-1} under the same conditions (Kelly et al., 2009). Other research gives a "typical" transcription initiation rate for a gene

as 20 min^{-1} (Pai & You, 2009), so an uncertainty range of 0.27 to 27 min^{-1} (one order of magnitude each way) was chosen for this figure.

The original characterisation of the Anderson promoters, meanwhile, puts J23101 at a strength of 70 relative to the promoter J23119, which was used for MmfR transcription. J23119 can therefore be mathematically inferred to have a transcription initiation rate (k_1) of about 0.00042 s^{-1}, or 0.026 min^{-1}, under similar conditions(Kelly et al., 2009). Other absolute transcription initiation rates for Anderson promoters could not be found, so the uncertainty range was set to an order of magnitude either side of this figure (0.0026 to 0.26 min^{-1}).

The rate of translation, meanwhile, is simpler to calculate. Kelly *et al.* put the rate of translation initiation of the RBS B0032 at 0.4 s^{-1}, or 24 min^{-1} (Kelly et al., 2009). Characterisation work by the 2016 Madras iGEM team puts the strength of the RBS B0034, which was the only RBS used in the constructs, at 0.6 times the strength of B0032, and therefore a translation initiation rate (k_1, k_3) of approximately 0.24 s^{-1} or 14.4 min^{-1}. A theoretical maximum translation rate would be around 60 $min^{-1,}$ based on a typical spacing of 7 nm between the lengths of amino acid chains of sequentially adjacent ribosomes in polysomes (Brandt et al., 2009), corresponding to about 20 amino acids of 0.35 nm each (Kudva et al., 2013), and a maximal rate of translation of about 20 amino acids per second (Bremer & Dennis, 2008).

Cellular mRNA was assumed to have a half-life of approximately one minute, effectively a degradation rate (k_5, k_7) of 0.5 min^{-1} (Kennell & Riezman, 1977; Pai & You, 2009). Some research, however, does suggest a longer half-life for mRNA in *E. coli* – between 3 and 8 minutes (Bernstein, Khodursky, Lin, Lin-Chao, & Cohen, 2002) – and it is reasonable to suggest that similar degradation rates may hold for other bacterial species. Taking 10 minutes as a maximal half-life, this gives a range of 0.05 to 0.5 min^{-1} for mRNA degradation.

Protein half-lives vary greatly, but around 35 % are between 25 h and 70 h, and another 40 % are in excess of 70 h (Maurizi, 1992), a degradation rate on the order of 10^{-2} min^{-1}. A recent model of the comparable ScbR expression system in *S. coelicolor* places the degradation rate at between 5.02 x 10^{-4} and 5.78 x10^{-3} min^{-1} (Tsigkinopoulou et al., 2020). Given that both the GFP and MmfR used in the Chapter 3 constructs lack degradation tags, the latter, slower range of degradation rates was used, and the degradation rate of all proteins set to 10^{-3} min^{-1}.

Rate parameters for the dimerization of MmfR could not be found, but the comparable MmfR-like repressor ScbR is estimated to have a value of approximately 0.307 nM or 0.185 um^{-3} for the dissociation constant (K_d), based on an averaging of reported dissociation constants given in literature for multiple similar proteins. The same work suggests an average dissociation rate k_{diss} of 2.5303 min^{-1}, and the rate of association can be calculated as k_{diss}/K_d = 13.68 μm^3 min^{-1} (Tsigkinopoulou et al., 2020).

The binding and dissociation rates of the interactions between the repressor MmfR and its MARE target site are suggested to be in the order of 10^5 M^{-1} s^{-1} and 10^{-2} s^{-1}, or about 10^{-2} M^{-1} min^{-1} and 0.6 min^{-1} in experimental work by Bowyer *et al* (Bowyer et al., 2017).

Data on the separate binding and dissociation rates of quorum sensing messenger MMFs for MmfR are much harder to find. Zhou et al. claim that the dissociation constant K_d for the binding of MMF1, the primary furan used in Chapter 3, to a Y144F mutant of MmfR is 18.8 nM +/- 1.67 nM (Zhou et al., 2021), or 1.13 * 10^{-2} μm^{-3}. This mutant is noted as having an effectively unchanged affinity for MMFs relative to the wildtype.

However, while the dissociation constant offers a ratio of the binding and dissociation rates, the actual values are not themselves known – so how quickly the MMF-MmfR interaction reaches steady state cannot be derived from the data given. As a result, an assumption was made that equilibrium would be reached more or less immediately, relative to the timescales of the other interactions present in the model. Given the role of MMFs as quorum sensing agents, quick cellular responses to changes in their concentrations seem like a reasonable assumption.

The division time of the cells, which determines the dilution rate (ν) of many of the components, was set initially to an arbitrarily high 2 x 10^6 minutes.

Table 4.4 summarises the reaction rate parameters used in this model. Uncertainty ranges were selected as an order of magnitude centred on the given rate, unless otherwise specified – rate constants with no uncertainty ranges were reverse reactions. All rates assume a reaction volume of 1 μm^3.

Table 4.4: Summary of rate parameters used in this chapter.

Process	**Rate (min^{-1})**	**Range (min^{-1})**
Transcription *(mmfr)*	0.026	0.0026 – 0.26
Transcription *(gfp)*	2.7	0.27 – 27
Translation	14.4	1.44 - 60
mRNA degradation	0.5	0.05 – 0.5
Protein degradation	0.001	0.01 – 0.0001
$MmfR_2$ monomerisation	2.5303	0.25 – 25
MmfR dimerisation	13.68 um^3	-
$MmfR_2$-MARE association	0.01 M^{-1}	-
$MmfR_2$-MARE dissociation	0.6	0.06 – 6
MMF-$MmfR^2$ association	1 um^3	-
MMF-$MmfR^2$ dissociation	0.0113	0.0103 - 0.0123

A summary of the models presented in this chapter is given in **Table 4.5**:

Table 4.5: Summary of models presented in this chapter.

Model	**Description**
00	Constitutive expression of GFP in static cell population.
01	GFP expression repressed by constitutive MmfR expression.
02	MmfR-repressed GFP expression in growing cell population.
03	MmfR-repressed GFP expression induced during growth.
04	As **03**, with various $MmfR_2$-MMF dissociation rates.
05	As **03**, with permanent $MmfR_2$-MMF binding.
06	As **03** with varying transcription rate of MmfR, fast growth.
07	As **06** with slow population growth (100 min).
08	As **06,** with a tenfold lower plasmid concentration.
09	As **08,** with slow population growth (100 min).
10	As **06,** with a hundredfold lower plasmid concentration.
11	As **10,** with slow population growth (100 min).
12	As **08,** with a tenfold increased $MmfR_2$-MMF binding affinity.
13	As **12,** with slow population growth (100 min).
14	As **12**, under a wider range of inducer concentrations.

4.2 Characterisation

4.2.1. Characterisation with static cell population

The system was first characterised as a static cell population**,** with the results shown in **figure 4.2:**

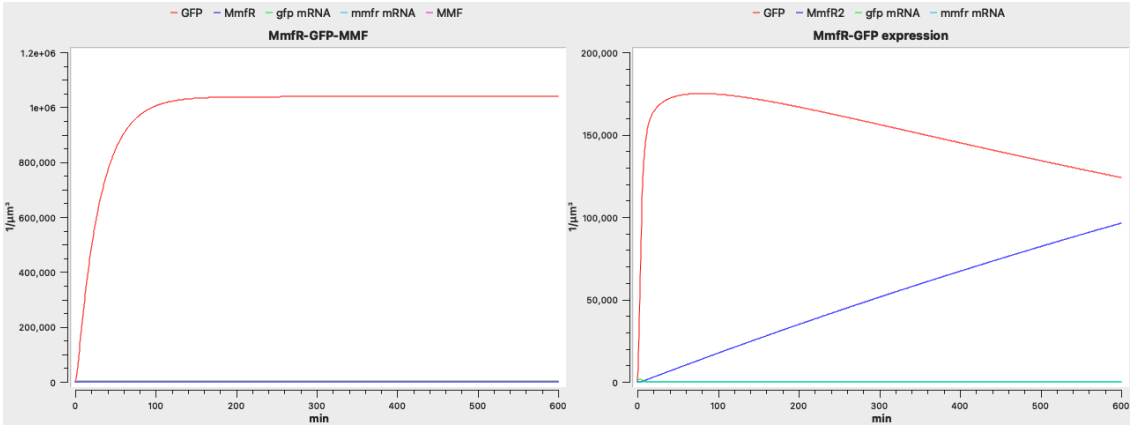

Figure 4.2: *Models 00 (left) and 01 (right) of GFP expression over a period of 10 hours in a* ***static*** *cell population*

Left (00)*: GFP expression with* ***no*** *MmfR expression to repress it. GFP concentration rises rapidly over two hours and peaks at around* ***10^6*** *copies per cell.* ***Right (01):*** *GFP expression* ***controlled*** *by MmfR expression with a promoter* ***1%*** *the strength of that used for GFP expression. GFP concentration peaks at about* ***10^5*** *copies per cell at 20 minutes, before beginning to slowly* ***decrease****. Note the very different scales of the x-axes of the two graphs.*

The characterisation began with the expression of GFP in the absence of any MmfR, as if the gene *mmfR* were not present at all **(00)**. To approximately match the copy number of the pUC origin of replication that was used, the model was initiated with **500** copies of the *gfp* gene.

In this model, GFP expression is essentially constant, and its degradation rate is very low. As a result, the steady state concentration of GFP is very high in this model, peaking at around **10^6** copies per cell.

Given that the total number of proteins in a single *E. coli* cell at any one time is estimated to be between **$2x10^6$** and **$4x10^6$** (Milo, 2013)**,** this illustrates that if GFP expression was not repressed in any way, possessed a reasonably strong promoter and RBS, and possessed no degradation tags, its production could rapidly saturate the host cell's metabolic capabilities.

Introducing **500** copies of the ***mmfr*** gene, as expected, drastically reduces GFP expression **(01)**. An initial burst of GFP production causes it to peak at a little over **10^5** copies per cell, before the concentration of MmfR rises enough to inhibit all *gfp* transcription at about 20 minutes. GFP then slowly degrades, while **$MmfR_2$**

concentration continues to **rise** and eventually overtake GFP concentration. The MmfR concentration does not reach a steady state over the course of the 10 hours, continuing to rise almost linearly.

The metabolic stress imposed on the cell is vastly reduced by the inclusion of **MmfR**. Without it, GFP is produced at a rate of **3.64 x 10^4 min^{-1}** at steady state; repressed by MmfR, it is instead produced at a rate of only **6.2 min^{-1}**, while MmfR monomers are generated at a rate of **374 min^{-1}**, a **100-fold** reduction in total steady-state expression rate of recombinant proteins.

However, this model assumes a static cell population. This might be somewhat believable with unrestrained GFP production, given the metabolic stress incurred. However, with GFP expression repressed by MmfR, the cell population should be able to grow at a rate somewhat approximating a normal rate.

4.2.2 Characterisation with dynamic cell population

The commonly upheld reproduction time of *E. coli* under laboratory conditions is **20 minutes,** so the model was re-run with this division time, as shown in **figure 4.3**:

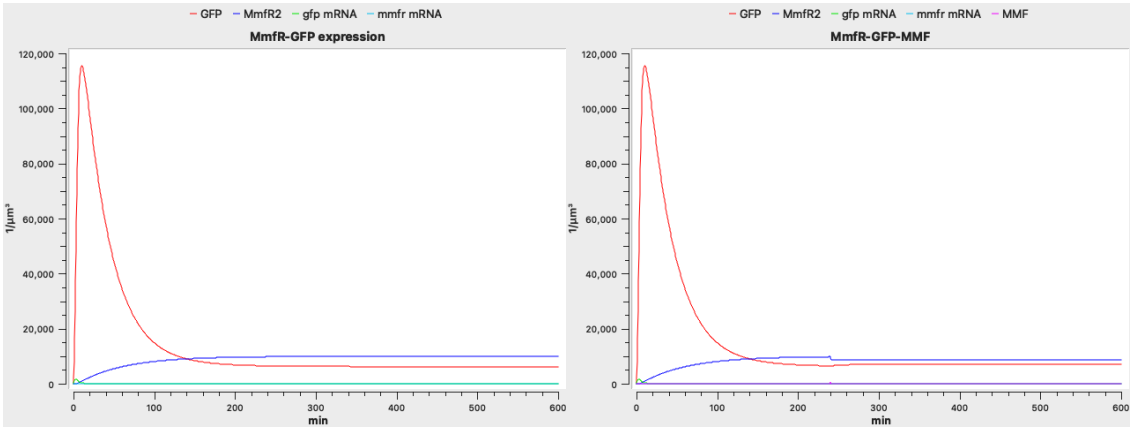

***Figure 4.3: Deterministic** model 02 (left) and 03 (right) of GFP expression over a period of 10 hours and in a **growing** cell population.*

***Left (02)**: GFP expression **repressed** by **1%** relative strength MmfR expression. **Right (03)**: The same system with **1uM MMF** added at 4 hours (240 minutes).*

The concentration of mRNAs and proteins is halved every 20 minutes (in addition to the normal degradation processes) as the cell population doubles in this model **(02)**. As a result, the **GFP** concentration peaks much lower, at about **1.2 x 10^5** copies. The concentration of $MmfR_2$, meanwhile, peaks at **10^4** copies and remains stable there after about four hours.

However, the $MmfR_2$ concentration at equilibrium is insufficient to completely repress GFP concentration. The steady-state rate of GFP production in this growing population model is around **218 min^{-1}**. Given that the rate of MmfR expression is only slightly lower than in the static population at **350 min^{-1}** (the rate being lower due to a reduced mRNA concentration, because of the growth-dilution), the total rate of protein production by the genetic construct is about **1.5 times** that of the static population model.

This is still a very low level of baseline expression compared to unrepressed GFP expression, but with both protein concentrations peaking at around **10^4** copies per cell, they still comprise around **1 %** of the cell's total protein content even before induction. This reflects experimental observations in **Chapter 3**, in which engineered cells were visibly green even when not induced by MMFs.

600 MMF molecules were added to the model (an effective concentration of **1 μM**) after four hours **(03)**. This was done to mimic the experimental protocol of Chapter 3, in which **1 μM MMF1** was observed to drive a substantial increase in GFP production rate and consequent fluorescence.

Conversely, the model predicts that the addition of such a low concentration of MMF will have a minimal effect on the repressive activity of MmfR, and consequently on the rate of **GFP** production, raising it from 218 min^{-1} to 233 min_{-1}. This is understandable given that at steady state there are around **10^5 MmfR** dimers per cell, to only **6 x 10^2** molecules of MMF.

4.2.3 Impact of MMF-MmfR Binding Affinity

The model was re-run with a range of values for the rate of dissociation between MMF and $MmfR^2$. The amount of increased affinity will approach a complete repression of GFP production, as shown in **figure 4.4**:

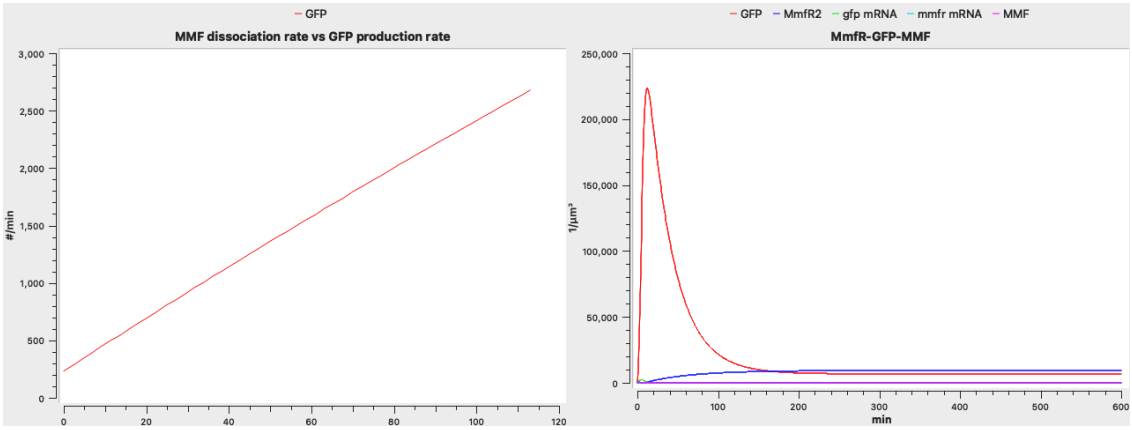

Figure 4.4: *Deterministic model 04 (left) and 05 (right) of GFP expression repressed by MmfR, and derepressed by* ***100 µM MMF****.*

Left (04): *Effect of rate of* ***dissociation of MMF*** *from $MmfR_2$ (x-axis) on* ***GFP expression*** *rate at steady state (y-axis).* ***Right (05):*** *Time-course of GFP expression with dissociation rate of MMF from $MmfR^2$ set to zero.*

A decreased dissociation rate of MMF from MmfR results in a decreased expression rate of GFP, but although the relationship is linear, a dissociation rate of zero does not entirely eliminate GFP expression. Even with MMF permanently binding $MmfR_2$ (until the $MmfR_2$ is degraded), GFP is still transcribed at a rate of **233 min^{-1}**, which in reality would occupy the majority of the cell's protein production capabilities.

Why such a low concentration of MMF proved to be effective *in vivo* but not *in silico* is a subject for speculation. Given the mismatch between MMF and $MmfR_2$ concentrations, the most probable cause of the discrepancy is that **$MmfR_2$** is much less concentrated in the real system than it is in the model. It is difficult to otherwise envisage how a known concentration of MMF could effectively inhibit a concentration of repressor protein one to two orders of magnitude greater.

The transcription and translation rates used in the model are idealised values, and the actual protein production capabilities of the cells are potentially considerably lower. Based on experiments in other literature, Bremer & Dennis compiled a list of typical cell contents and parameters of *E. coli* at 37 °C at a range of growth rates. At a doubling time of **20** minutes, a cell is suggested to contain around **73** ribosomes, with translation occurring at a maximal rate **of 22** amino acids per second (Bremer & Dennis, 2008). With the **sfGFP** used in Chapter 3 being **237** amino acids long, a ribosome could translate one approximately every **12 seconds**, or **5 min^{-1}**, while **MmfR** subunits, at **214** amino acids each, take a similar length of time. The cell's full complement of ribosomes could therefore reach a theoretical rate of **365** proteins per minute.

The modelling results so far show a steady-state protein production rate of the order of **10^3** proteins per minute. The models, and the calculations above, assume that

the cell population is growing with a doubling time of 20 minutes. This is obviously not possible if the entirety of the cell's ribosomes, or even a significant fraction, are being dedicated to recombinant protein production, so – given that the cells carrying construct were observed to grow and divide in experiments – the protein production rates given so far by the model are far too high to be realistic.

Given that the expression level of **MmfR** alone in this model is over half the theoretical maximum protein output of the cell, and that the concentration of **MMF** used has a much lesser effect in this model than observed experimentally, it seems reasonable to assume that the "real" expression rate of MmfR in the pDT-M1 construct is much lower than the model has so far assumed.

However, the experimental evidence shown in **Chapter 3** shows very little difference between the behaviour of pDT-M1 and pDT-M2, despite pDT-M2 theoretically generating MmfR at seven times the rate. This similarity in behaviour may indicate that the actual expression rate of **MmfR** in either construct is not very different – potentially, it is already being expressed at a maximal rate (as the computational modelling suggests).

Unfortunately, while the experimental data collected for chapter 3 sheds some light on the effective concentration of **MMF**, the experimental results are in the form of fluorescence rather than an empirical measure of GFP concentration, and the two are not necessarily **proportional** to one another. As a result, fitting the model to experimental results is difficult.

However, by undertaking a parameter scan of the model, it is possible to characterise how the inducibility of GFP expression changes as the rate of *mmfr* transcription changes, as shown by **figure 4.5:**

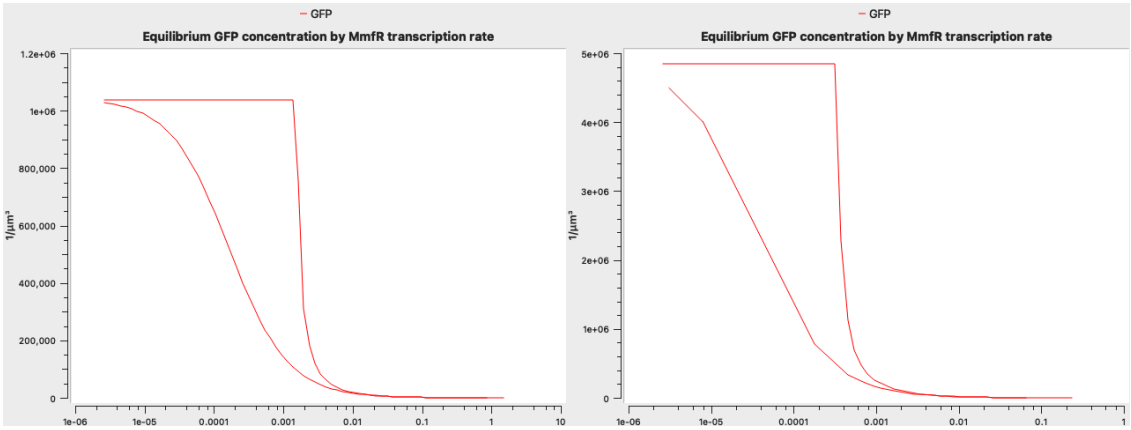

Figure 4.5: *Log/linear plots 06 (left) and 07 (right) of* ***GFP*** *steady state concentration (y axis) by frequency of transcription of* ***mmfr*** *(****min^{-1} gene^{-1}****, x axis).*

Left (06): *Steady states during exponential phase (doubling time* ***20 min****). The upper line represents a post-induction MMF concentration of* ***1 µM*** *(****600 µm^{-3}****), while the lower line represents an absence of MMF.* ***Right (07):*** *Steady states during slow growth (doubling time* ***100 min****) and the same MMF concentrations.*

This parameter scan reveals a specific range of MmfR transcription rates at which GFP expression, when induced with **1 µM MMF**, transitions sharply from maximal expression to no greater than its uninduced expression rate. The transition occurs over about an order of magnitude, with the rate of expression used in the previous models falling outside of the transition range.

Above a transcription frequency in the order of **10^{-2}** min^{-1} gene^{-1}, and below a frequency in the order of **10^{-6}** min^{-1} gene^{-1}, induction with 1 µM MMF has **no effect** on GFP expression rate. At a transcription frequency of **1.3 x 10^{-3}** min^{-1} gene^{-1}, the system displays a threshold; below this rate of transcription, the MMF is sufficient to maximally induce GFP expression, and above this rate, MMF's ability to derepress GFP falls off very rapidly. The rate of *mmfr* transcription used in the model so far is **2.6 x 10^{-2}** min^{-1} gene^{-1}, which falls beyond the upper boundary at which MMF would be expected to have much effect.

Reducing the model's rate of *mmfr* transcription by about twentyfold, to **1.3 x 10^{-3}**, would put the rate of induced GFP expression at the maximum rate that can be achieved, whilst keeping the rate of uninduced expression as low as it could be. This is the threshold; higher values would substantially decrease the rate of induced GFP expression, whilst lower values would gradually increase the rate of uninduced expression.

However, the pattern shifts as growth rate decreases and the culture approaches stationary phase. Due to the lack of dilution through culture growth, GFP concentration peaks at **5 x 10^{6}**, five times that of exponential phase. Additionally, because of the

decreased rate of dilution from cell growth, the steady state concentration of MmfR is much higher than at exponential phase. The threshold between maximal and decreasing GFP expression is at an *mmfr* transcription rate of **3 x 10^{-4}** min^{-1} $gene^{-1}$, substantially lower than that at exponential phase. Consequently, there is a range of MmfR transcription rates at which GFP production **can** be induced with 0.1 µM MMF at exponential phase but **not** at stationary phase, as the MmfR becomes too concentrated in the cell to be fully inactivated by MMF.

Therefore, the model suggests that the threshold expression rate of MmfR for achieving maximal GFP expression in a stationary culture is around **3 x 10^{-4}** min^{-1} $gene^{-1}$. At this expression rate, the exponential phase rate of GFP production would be around **4 x 10^{-5}**, substantially higher than the maximum rate achievable with induction at stationary phase. In other words, depending on the actual rate of transcription of *mmfr*, the construct will either produce GFP faster during its exponential growth phase than after induction at stationary phase, or induction will have little to no effect. This dichotomy renders the repressor system essentially useless as a means of reducing metabolic strain on the cell culture.

The experimental evidence from **chapter 3** demonstrates an obvious increase in GFP expression when induced with MMF, but this increased expression occurs within the first 10 hours after inoculation, while the culture is still growing. Qualitatively, this does correspond to the hypothetical model scenario in which induction is possible during rapid cell growth, but not once culture growth slows down. This occurs when MmfR expression rate is reduced **20-fold** or more from the previous model settings. To reflect the experimental evidence, the rate of MmfR transcription was reduced from **2.7 x 10^{-2}** to **1.3 x 10^{-3}**, for future modelling.

The disparity between the theoretical and actual rate MmfR expression could be explained by one of several factors. The rate of translation may be lower due to limited ribosome availability, although if the host's metabolism were already saturated by MmfR expression, it seems unlikely that the construct would last long before mutating.

A second factor could be that the intracellular concentration of MMF is actually much higher than that of the culture medium. In other words, MMF may be being taken up by the *E. coli* through active transport rather than diffusion through membrane pores. In this case, the effective concentration of MMF within the cells is difficult to

speculate on, but is likely much higher than assumed so far. It would be necessary to model the effects of a range of MMF concentrations to establish the impact of this.

A further factor that has not yet been accounted for is the difference in plasmid concentration between a growing culture and a stationary one. Plasmid concentration is typically directly proportional to cell doubling time, and therefore inversely proportional to growth (Klumpp, 2011). As a result, a fivefold difference would be expected between plasmid concentration at exponential growth (20 min) and at relatively late phase (100 min). Modelling the impact of different plasmid concentrations would be necessary.

4.2.4 Impact of Plasmid Concentration

Typically, different vectors are used for recombinant protein production than those that are used for cloning. Recombinant protein production vectors are generally less prolific than cloning vectors, averaging considerably fewer copies per cell.

This reflects the specific purposes the vectors are chosen for. The pDT-L1, pDT-M1 and pDT-M2 constructs used the pUC origin of replication, which has a copy number of approximately 500 per cell. This many copies of a plasmid make it much easier to extract high concentrations of DNA for cloning and vector engineering. However, for recombinant protein production, more copies of the gene make it very difficult to avoid overexpression and the consequent metabolic stress, even when the expression system is supposed to be inducible.

The high levels of expression of MmfR required for GFP repression in the pDT-M1 and pDT-M2 constructs might be reducible by replacing the origin of expression with one that has a lower copy number.

To explore the consequences of altering the origin of replication on the inducibility of the system, the parameter scan of MmfR expression rate shown in figure 4.6 was repeated with both genes – *mmfr* and *gfp* – having a concentration of 50 or 5 ul^{-1} rather than 500. The results are shown in **figure 4.6:**

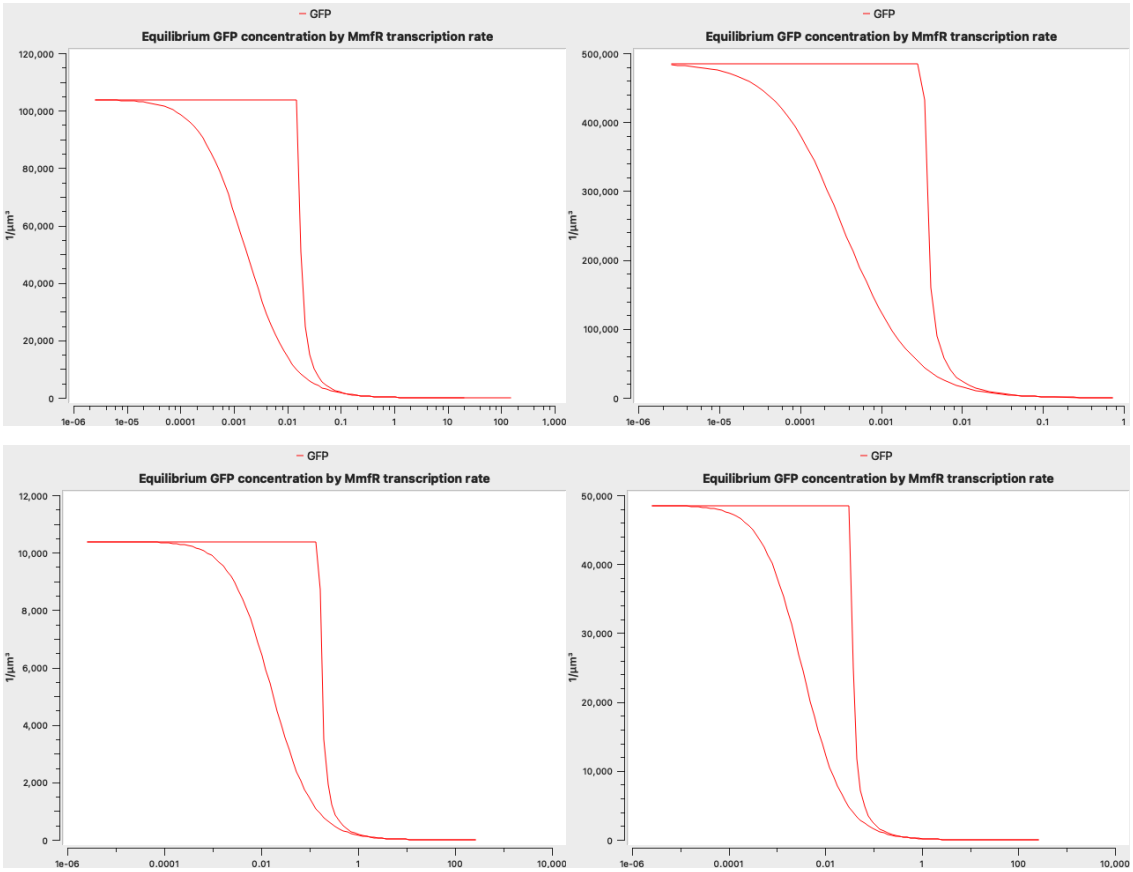

Figure 4.6: *Models 08 (top left), 09 (top right), 10 (bottom left) and 11 (bottom right) of GFP expression with reduced concentration of genes.*

Top left (08): *Parameter scan of GFP expression during exponential growth, assuming a plasmid concentration of* ***50 µm^{-3}****, rather than* ***500****.* ***Top right (09):*** *The same parameter scan during stationary phase growth.* ***Bottom (10, 11):*** *As* ***top****, but with plasmid concentration reduced to* ***5 µm^{3}.***

The results are similar to the system with a concentration of 500 µm^{-1}, but the maximal rate of GFP expression is around 10-fold lower for each 10-fold reduction in plasmid concentration. This makes sense given that there are 10-fold fewer copies of the gene being transcribed. With the cell doubling time at 20 min and plasmid concentration of 50 µm^{-1}, the pattern of expression is shifted an order of magnitude towards higher *mmfr* expression rates (with the transition point being 1.5 x10^{-2} min^{-1} gene^{-1}). In this version of the model, maximal concentration of GFP is around 10^5 µm^{-3} rather than 10^6 µm^{-3}. With the cell doubling time set to 100 min, the model shows a similar shift, with maximal GFP concentration 10-fold lower when the plasmid copy number is also 10-fold lower. When the plasmid concentration is reduced to 5 µm^3, the same pattern continues, with maximal GFP expression rates being reduced tenfold, and the threshold expression rate of MmfR expression increasing tenfold, whether at a doubling time of 20 min (left) or 100 min (right).

The rates of MmfR expression required for induction to be effective are also around 10-fold higher (in other words, the pattern is effectively shifted an order of

magnitude higher on the x-axis). Again, this is logical; with 10-fold fewer copies of MmfR, they would need to be transcribed at ten times the rate to match the concentration of MmfR seen in the previous construct, and thus require the same concentration of MMF to be inhibited.

Interestingly, the number of available target sites for MmfR seems to have very little effect on the behaviour of the system. So long as the concentration of MmfR is high enough that all the MARE sequences are bound constantly, and the concentration of MMF is enough to bind all the MmfR, the actual number of MARE sites appears to make very little difference to the inducibility.

Additionally, the lower boundary of GFP steady state concentration – that which occurs when MmfR is expressed in excess – is functionally 0 $cell^{-1}$ regardless of the actual copy number of the plasmid. This observation is mostly academic, given that at this expression rate, induction is impossible and thus the construct does not serve its intended purpose.

If an effective expression system – one that maintains low production during culture growth and high production after induction and at low growth rates – cannot be achieved merely by up- or downregulating constitutive expression of MmfR, or by altering the number of copies of the expression system per cell, then other improvements will need to be made. Two approaches were considered: optimisation of the repressor or **MARE** affinity, or structural changes to the gene circuit.

4.2.5 Impact of Improved Repressor Affinity

In the conclusion of chapter 3, it was suggested that the leakiness of the pDT-M1/M2 constructs could be mitigated by improving the affinity of the repressor MmfR for the MARE target site. Through direct comparison with the pDT-L1 construct, it was established that pDT-L1 was considerably less leaky, and that the most likely explanation was that LacI simply had a greater affinity for its target than MmfR did. Experimental evidence collected by Kathryn Styles of the University of Warwick supports this conclusion that MmfR is relatively leaky (Styles, 2016).

Some optimisations of the MARE sequence for improved MmfR affinity have already been done by Miriam Rodriguez Garcia, also at the University of Warwick. Through random mutagenesis of two base pairs of the MARE sequence, a total of fourteen sequence variants were generated and characterised, which are shown in

Appendix 1. The sequence with the highest affinity, MM4, was used in the constructs DT-M0, M1 and M2.

Although there could be scope for further optimisation of the MARE sequence, the two base pairs chosen for the mutagenesis were deemed the most likely to yield an improvement in affinity. Further base changes are likely to have a much smaller effect on the MARE sequence's leakiness.

Conversely, modifications to MmfR itself could increase its affinity for the MARE sequence. Mutagenesis of the DNA-binding helix-turn-helix domain, or the surrounding region, might yield a variant MmfR that binds more tightly to the MARE. Randomising a region of multiple amino acids would take a high-throughput screening process to identify high-functioning mutants, but in principle would be quite feasible.

Repositioning the MARE sequence relative to the promoter and gene sequences might also improve its affinity for MmfR. Genetic context is known to substantially affect the functionality of genetic parts, so changing what parts are adjacent to the MARE, and the distance between them, could potentially also improve performance. However, inserting nucleotide spacers between the MARE, promoter and coding sequence might adversely affect the speed of transcription and translation.

Without experimental evidence, it is difficult to estimate the extent to which MmfR-MARE affinity could be improved. The three MARE sequences present in the wildtype *S. coelicolor* alone display a variety of repressor strengths *in vivo*, from a threefold to a tenfold reduction in expression rate when uninduced, according to work by Kathryn Styles (Styles, 2016).

Given the disparity of binding affinities shown by the wildtype MARE sequences alone, it seems reasonable to assume that between modifications to the MARE sequence and MmfR itself, the affinity might be improved by an order of magnitude. A parameter scan as in section 4.4, with a plasmid copy number of 50 and a dissociation rate of MmfR from the MARE sequence of 0.06 min^{-1} (tenfold lower than previously). The results are shown in **figure 4.7:**

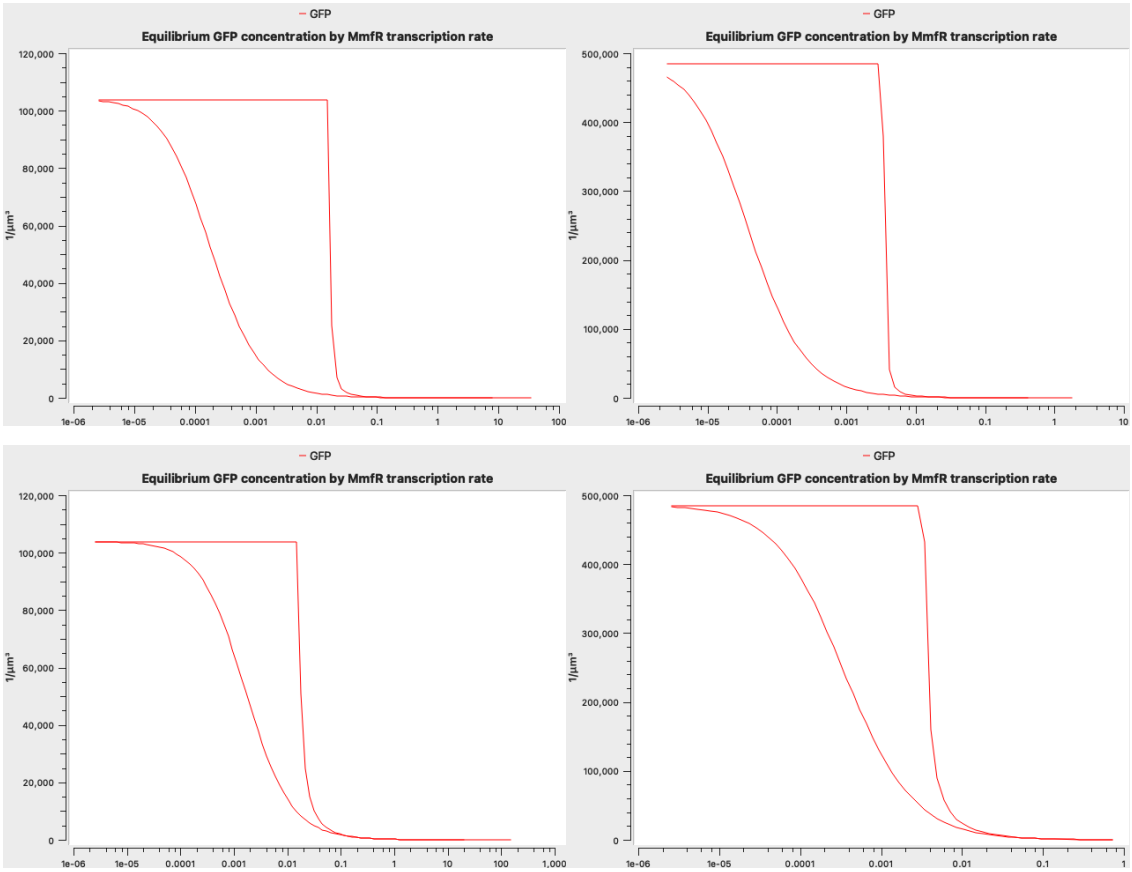

Figure 4.7: *Expression models 12 (top left), 13 (top right), 14 (bottom left) and 15 (bottom right) of GFP under the inducible regulation of MmfR with a binding affinity for the MARE increased by an order of magnitude, and a plasmid concentration of* ***50 µl^{-1}****.*

Top left (12): *GFP steady state concentration during exponential growth (doubling time 20 min).* ***Top right (13):*** *GFP steady state concentration during stationary phase (doubling time 100 min).* ***Bottom left (08):*** *GFP steady state concentration during exponential growth with previous MmfR-MARE binding affinity for comparison.* ***Bottom right (15):*** *GFP steady state concentration during stationary phase with previous MmfR-MARE binding affinity for comparison.*

The maximal level of GFP expression during exponential growth (is unchanged compared to when the MmfR had a reduced affinity (08, bottom left), and the transition point Is the same, but the pattern of repressed expression rates is shifted towards lower MmfR expression rates. As a result, there is a much larger parameter space in which MmfR expression is high enough to mostly repress GFP expression whilst still being fully repressible by 0.1 µM MMF.

A similar relationship is seen during slow growth (doubling time 100 min) between the system at higher MmfR affinity (13, top right) and lower affinity (09, bottom right). At an MmfR transcription rate in the order 10^{-3}, GFP concentration remains relatively low (around 20,000) while fast-growing and uninduced, while achieving maximal concentration (around 500,000) while slow-growing and induced.

The results show a marked improvement in the prospective expression construct. Contrary to previous experimental outputs, the construct shows a range of

MmfR transcription rates at which it would offer a relatively low rate of GFP production whilst not induced and growing quickly, and a maximal rate of production whist growing slowly and induced with 0.1 µM MMF.

Improvements to the affinity of MmfR for the MARE sequence seem unlikely to yield a difference of more than an order of magnitude. Wild-type proteins often have metaphorical room for improvement, but there would be little advantage to be gained from a transcriptional repressor that was orders of magnitude weaker than it could be with a relatively simple set of point mutations.

As a result, further tuning of the expression system might need to take the form of simply expressing more repressor, and using a greater concentration of MMF for induction. This is not an ideal approach, given the metabolic cost of producing additional repressor and the potential financial cost of supplying more repressor, but could still be more cost- and time-efficient than the alternative of a leakier expression system. The effects of adding different concentrations of repressor are explored in **figure 4.8:**

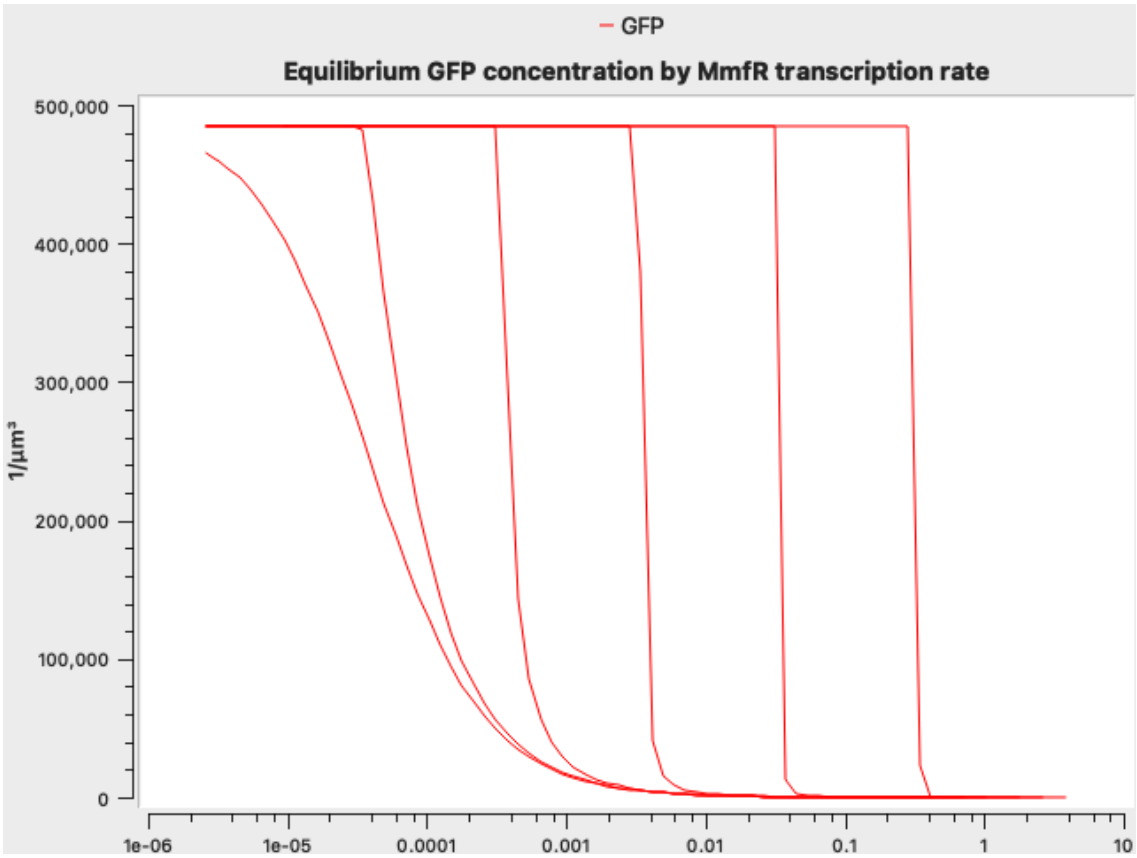

Figure 4.8: *Steady-state concentration of GFP during slow cell growth (**100 min**), with a tenfold increased affinity between MmfR and the MARE and a plasmid concentration of 50 µl^{-1}, under induction by different concentrations of MMF.*

*From **left** to **right**: **0**, **10^{-3}**, **10^{-2}**, **10^{-1}**, **1** and **10 µM**.*

The modelling shows a straightforward linear relationship between the MMF concentration added, and the maximum MmfR transcription rate that can be fully induced by that concentration. A tenfold increase in MMF concentration results in a tenfold increase in the concentration of MmfR that can be fully repressed by it. As the ratio of inducer and repressor per cell to the copy number of the plasmid increases, the threshold between maximal and no induced expression becomes steeper.

As discussed in section 4.3, there is a distinct possibility that the cells concentrate MMF within themselves through active transport. If this is the case, then the effect of a higher MMF concentration is simply to lower the threshold MmfR expression level at which GFP expression transitions sharply from minimum to maximum.

4.3 Robustness Analysis of Modelling Outcomes

4.3.1 Steady-State Cellular GFP Concentrations

Several observations have been drawn so far from the modelling in **Section 4.2**. The first of these is the observation that the theoretical steady-state concentration of GFP in the cell approaches one-quarter to one-half of the logical total cell protein content when GFP expression is not being repressed.

This observation was made based on assumed, fixed values for the reaction rate parameters of the expression system. However, as stated in **Section 4.1**, the reaction parameters given are a best guess based on literature sources. Inevitably, there will be some variation between these values and those actually extant *in vivo*. For this purpose, uncertainty ranges were given for these rate parameters in **Table 4.4**.

The model (**00**) presented in **Section 4.2.1** was repeated using a randomised spread of parameter values, based on the ranges given in **Table x**. The model was run with a range of expression rates of MmfR – from 2.6×10^{-5} min^{-1} to 0.26 min^{-1}. The lower limit was chosen to be as close to an expression rate of zero as possible; the latter is the upper bound of the uncertainty range given for the rate of MmfR transcription presented in **Section 4.1**. The distribution chosen was a logarithmic distribution across the possible range of values – a log-normal distribution might have been more "realistic", but COPASI is not capable of applying a normal distribution to a logarithmic range in this way. The results are shown in **Figure 4.9**:

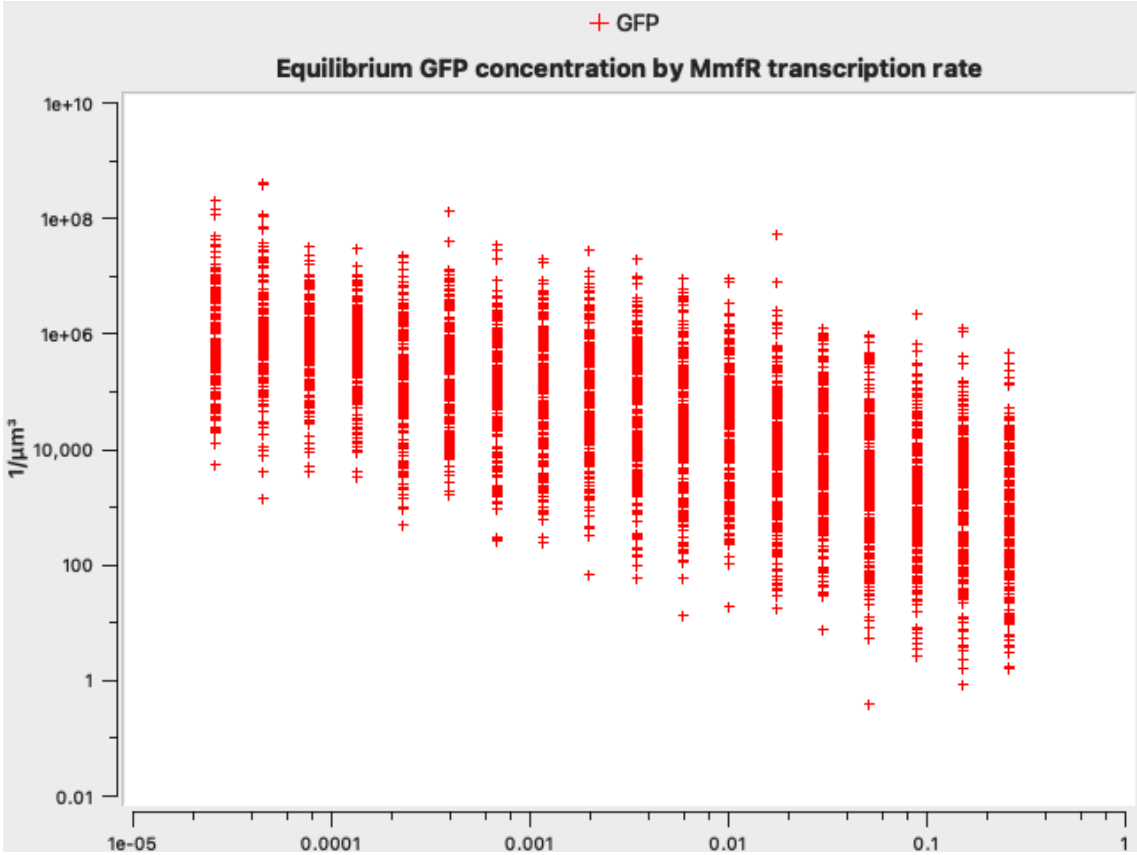

Figure 4.9: *Steady-state concentration of GFP* ***(y axis)*** *against MmfR transcription rate (****x axis****) during slow cell growth (doubling time 100 minutes) with a plasmid concentration of 500 um^{-3}, under 100 random distributions of parameter values within expected uncertainty ranges.*

At the lowest rate of MmfR transcription, which results in the minimum possible repression of GFP transcription, the majority of model outputs display a steady-state GFP concentration between 10^5 and 10^7. This is a wide range, at a full two orders of magnitude, but reflects that multiple parameters are being randomised to within an order of magnitude either side of their default value. Given the aforementioned estimate of 2 to 4 x 10^6 proteins per µm^{-3} cell of *E. coli* given in **Section 4.2**, 10^7 is certainly unfeasible while 10^5 is more within the bounds of physical possibility at 2.5% to 5% of total cell protein. Research suggests that a heterologously expressed protein can reasonably constitute up to 50% of total cell protein (Baneyx, 1999), or around 10^6, so the average is likely to be about right (though the model does not account for metabolic stress imposed by heterologous gene expression, not special constraints).

With MmfR transcription at a maximum rate of 0.26 min^{-1}, meanwhile, the majority of GFP steady states range from around 20 to 2 x 10^5 µm^{-3}. The upper bound constitutes around 5% to 10% total cell protein, and the lower negligible. This again illustrates the sheer variability borne of the degree of uncertainty of the rate constants assumed by the model. Given that even when uninduced, the constructs generated in

Chapter 3 displayed a distinctly green colour after overnight growth, a total GFP concentration of around 5% of total cell protein at steady state is believable, though this would need to be backed up with experimentation.

4.3.2 Necessary Ligand Concentrations for Induction

Another observation made during prior modelling is that given the parameter values chosen, the addition of 600 uM MMF – as was performed experimentally in **Chapter 3**, to great effect – should have no effect on the expression rate of GFP. This disparity between the observed and expected results of induction suggests that the parameters chosen for the model may not be as accurate as might be hoped. To gauge how great the discrepancy between theoretical and real reaction parameters might be, the model (**03**) was re-run with 50 sets of parameters randomised according to a logarithmic distribution across the uncertainty ranges given in **Table 4.4**, and the steady-state GFP concentration calculated at a range of initial MMF concentrations using each set of parameters. The results are shown in **Figure 4.10**:

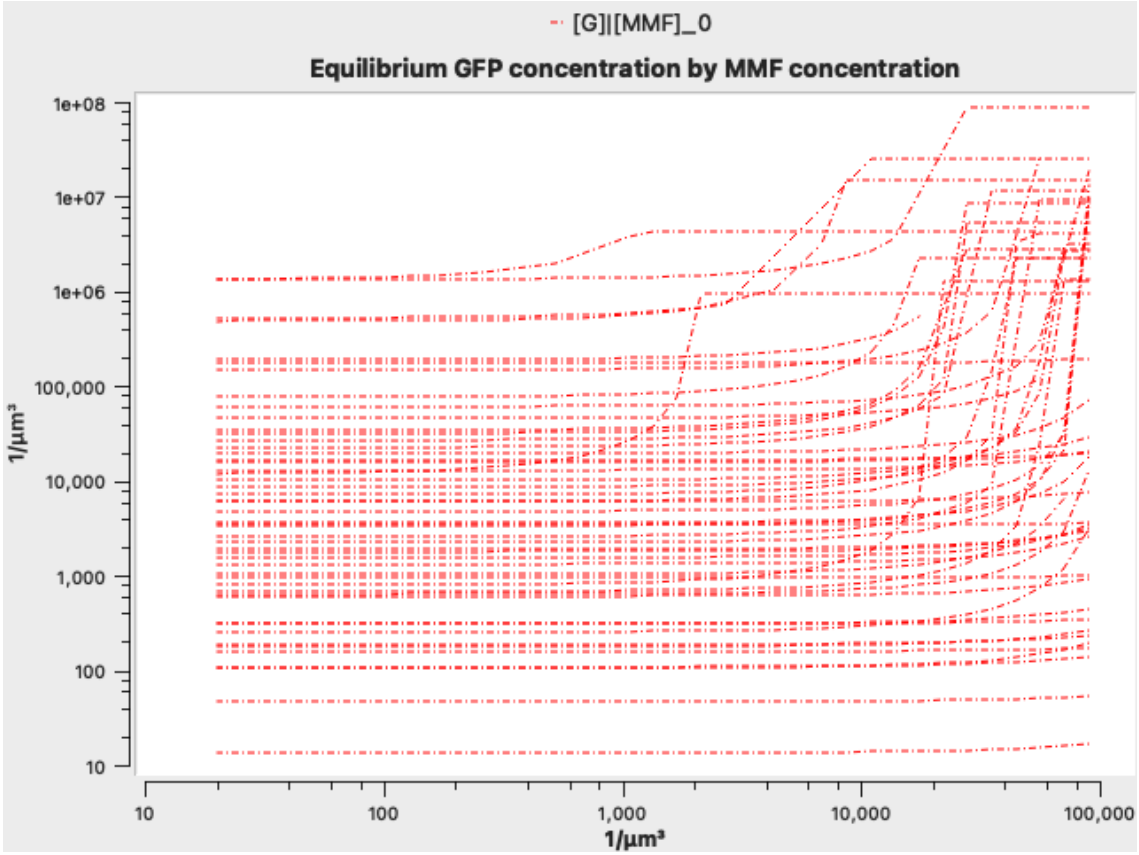

Figure 4.10: *Steady-state concentration of GFP (**y axis**) against MMF concentration (**x axis**), during slow cell growth (doubling time 100 minutes) with a plasmid concentration of 500 um^{-3}, under 50 random distributions of parameter values within expected uncertainty ranges.*

Only a very small proportion of the parameter sets tested show any increase of GFP steady state concentration at 600 µM MMF compared to without MMF, and a steep increase of GFP expression is only seen in any of the constructs at above an approximate MMF concentration of 6 mM. One construct (the uppermost line) does display its greatest increase in GFP expression between 600 µM and 1mM MMF, from about 10^6 μm^{-3} when uninduced to 4 x 10^6 μm^{-3} post-induction (or from one-quarter to one-half of the cell protein to the entirety of the cell protein). Given the experimental evidence discussed earlier in the section – the accumulation of GFP in cells carrying the pDT-M1 construct even when not induced, and the measurable response with an induction of 600 µM MMF – this particular parameter set could be close to the truth.

It is interesting to note the variety in response curves that can be seen between parameter sets. The variation in GFP expression rates pre- and post-induction is a logical outcome of the randomisation of transcription and translation rates for both MmfR and GFP. The different concentrations of MMF at which the expression system responds most likely reflect the steady state concentration of MmfR, as it stands to reason that with fewer copies of the repressor protein present, fewer ligand molecules would be necessary to sequester them. Some parameter sets show over a hundredfold increase of GFP steady state concentration after induction, while others increase by only fourfold. These varied results reinforce the importance of finding appropriate kinetic parameters for modelling, and of the need for the constant interweaving of experimental and modelling work to ensure that the outcomes of each drive the progress of the other.

4.3.3 Effect of Cell Growth Rate on GFP Concentration

It was concluded in **Section 4.2.2** that faster cell growth led to lower steady-state concentrations of GFP, due to the constant dilution of all proteins and mRNA by cell growth. This conclusion, however, was further tested with a wider range of parameters. The model presented there was re-run with parameters randomised as described in **Section 4.3.2**, and with a range of cell division times and MMF concentrations. The results are shown in **Figure 4.11:**

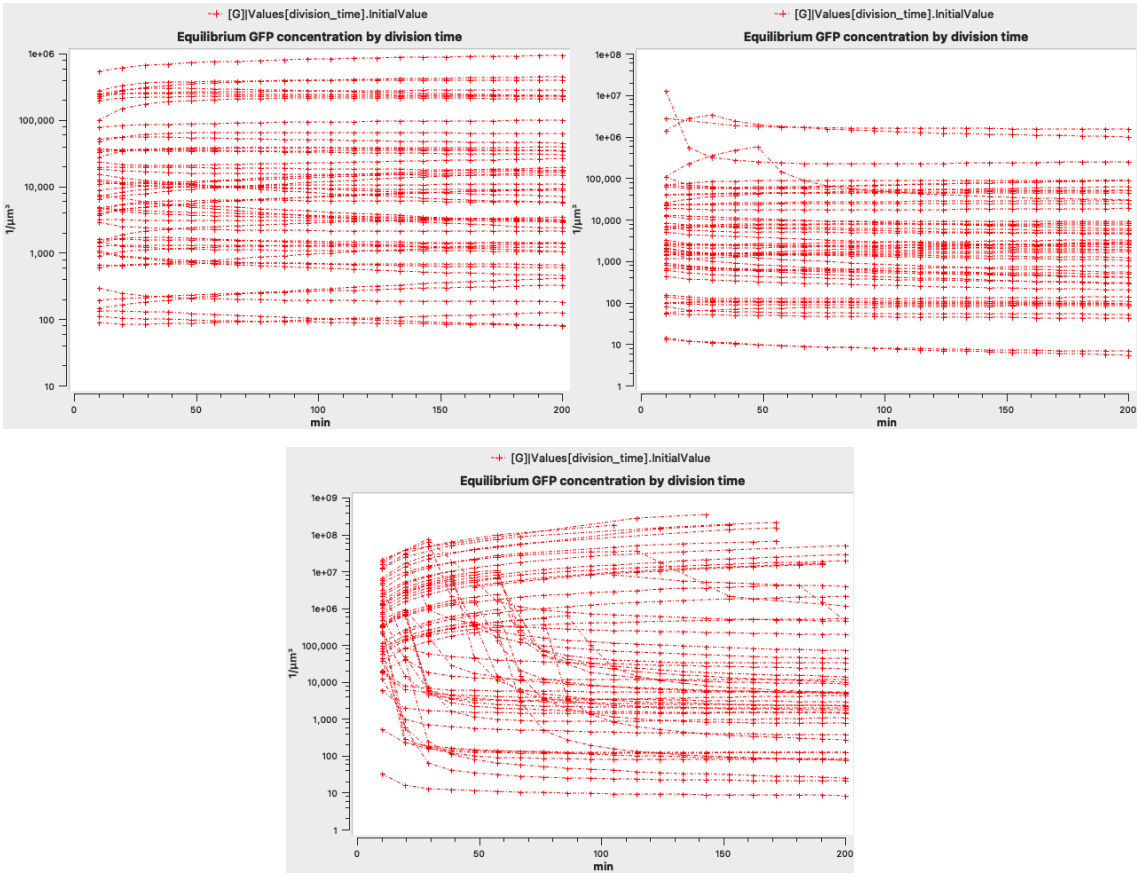

Figure 4.11: *Steady-state concentration of GFP (**y axis**) against cell division time (**x axis**), with a plasmid concentration of 500 um^{-3}, under 50 random distributions of parameter values within expected uncertainty ranges.*

Top left: *MMF concentration of 0.* ***Top right:*** *MMF concentration of 600.* ***Bottom:*** *MMF concentration of 60000.*

The results show that without the addition of MMF, cell growth rate actually has no consistent effect on steady-state GFP concentration. With some parameter sets, an increased doubling time has no or a negligible effect on GFP concentration; in others, it causes an increase *or* a decrease on steady state concentration. In neither case is the change particularly dramatic – less than half an order of magnitude between 20- and 200-miute doubling times, the former being the generally accepted exponential growth rate of *E. coli*. This stands in contrast with the conclusions drawn from Section **4.2.2**.

Further investigations were conducted upon the same system, with MMF added at concentrations of 600 µM and 60 mM. The results do show that, as asserted in the latter part of **Section 4.2**, there is a "threshold" of cell division time for any given parameter set and MMF concentration, above which the addition of the MMF suddenly ceases to take effect. However, where this threshold occurs is wildly variable, and in many parameter sets the chosen concentration of MMF is insufficient to take effect at all, even at the shortest possible division time.

The 600 µM concentration of MMF, which is the concentration that was used experimentally, can be seen to affect GFP concentration in a few cases, mostly in parameter sets that give relatively high steady-state GFP concentrations. In these, the "transition" from derepressed to repressed states also occurs at a cell division time of under 20 minutes, so these transitions would likely not be observable in the lab. The datasets for concentrations of 0 and 600 µM MMF show more or less the same range of GFP steady state concentrations, so it appears that 600 µM MMF has an effect on only a narrow range of models within the chosen parameter space, and possibly then only at physiologically improbable rates of cell division.

More definite transitions between derepression and repression can be seen in the third graph, which uses an MMF concentration of 60 mM. Transitions at this MMF concentration can occur between cell division times of 20 to 100 minutes, and many more datasets without an obvious transition point have a steady state GFP concentration far higher than those observed for lower MMF concentrations.

In conclusion, the modelling data match the expectation that a far higher concentration of MMF would be more substantially more effective at derepression, but also suggest that only under a relatively narrow area of the parameter space afforded by **Table 4.4** does the concentration of MMF that was shown to be effective *in vivo* offer any effect *in silico*. As such, the default parameters used throughout **Section 4.2** likely need to be reviewed and refined with more experimental data to build a more useful and accurate model.

4.4 Discussion

The models developed in this chapter were created to identify the potential benefit of modifications to the expression systems developed in Chapter 3. The existing constructs displayed substantial leakiness in experimental results, a design flaw that leads to slower growth and to reduced genetic stability. Through characterising the system as a set of mathematical interactions, different approaches were be assessed and compared without the need for exhaustive practical experimentation.

A model can only be as accurate as the data used to create it. Therefore, realistic parameters for the chemical and physical interactions present in the sample were sourced from scientific literature. Through qualitative comparisons with experimental data obtained in Chapter 3, the model parameters were able to be refined further.

Several conclusions were drawn from the subsequent exploratory modelling:

Firstly, the existing pDT-M1 and pDT-M2 constructs developed and characterised in Chapter 3 are likely to be saturating their hosts' protein production capabilities, based on a comparison of modelling outputs against experimental data already collected. Replacing the origin of replication of pDT-M1/M2 with one with a lower copy number would reduce the metabolic stress placed on the host during growth, both due to the level of repressor expression necessary to prevent GFP expression, and to the inherent leakiness.

Secondly, the theoretical rate of overall uninduced protein production, maximum induced expression and concentration of inducer required to achieve the latter are all directly proportional to the plasmid count. Finding an appropriate origin of replication for the constructs would be a simple matter of trial and error, as would finding the appropriate MMF concentration for induction.

Thirdly, to optimise the construct for minimal uninduced and maximal induced expression, tuning the expression rate of MmfR is necessary. More MmfR is required to repress product expression whilst the culture is growing rapidly, compared to when it is actively producing, which in turn increases the stress generated by producing MmfR itself as well as the concentration of MMF required to induce the system.

Finally, however, caution should be exercised when interpreting the results in this chapter. While modelling can be used to explore general trends that occur when elements of a system are changed, which may point towards future avenues of practical research, an excessively detailed analysis is generally misleading when using a model built on anything but the most stringently vetted biophysical data.

To optimise an expression system, then, the preferable approach would be to first to trial a variety of expression rates of MmfR, and then trial a range of concentrations of MMF to find the threshold of induction for each construct. Constructs could then be narrowed by comparing their induced and uninduced expression rates, as well as their long-term genetic stability However, with the limitations imposed by the ability to generate MmfR and MARE combinations with greater affinity for one another, further improvements would need to be sought elsewhere.

However, it seems a constant rate of MmfR expression cannot provide both a high level of induced expression and tight control of expression whilst the cell culture is growing, without using very large concentrations of **MMF**, or improving the affinity of

MmfR for the MARE far beyond what is likely to be possible. While inducer oversaturation is a valid option, given MMF's demonstrated non-toxicity, other approaches to improve the expression system would nonetheless be helpful.

The modifications mooted so far are changes to individual components that alter the rates of various interactions of the system, but without changing which interactions occur. For example, as touched upon in the introduction, the *mmfr* gene in *S. coelicolor* is not expressed constitutively, but is under transcriptional control of a MARE sequence and effectively regulates its own expression. Exploring the implications and possible advantages of different system architectures – changes to the overlying structure of the expression system, rather than simply fine-tuning the existing interactions, would be a target for future modelling work.

Chapter 5 – Discussion and Conclusions

5.1 Research Questions

As recalled in **Section 1.5**, the project set out to answer the following questions:

- **Is MmfR toxic to *E. coli*? Can it be constitutively expressed?**
- **Is MmfR a viable alternative to LacI for regulation of recombinant protein production?**
- **How might a MmfR-based expression vector be made a viable alternative to existing LacI-based expression systems?**

These questions were addressed through the design of *de novo* recombinant protein production vectors that used MmfR as a transcriptional repressor for inducible GFP expression in *E. coli*, through time-resolved fluorescence experiments with these constructs and directly comparable LacI-based expression constructs, and through quantitative modelling of the experimentally tested expression systems and of putative improvements. In **section 5.2,** the answers to each individual research question are addressed.

5.2 Findings

5.2.1 Is MmfR toxic to E. coli? Can it be constitutively expressed?

Concerns were had about the potential effects of constitutive expression of MmfR in *E. coli*. As a heterologous transcriptional repressor with a tolerance for some variability in its target sequence, it was not known whether it would engage in off-target binding – potentially causing toxicity if such binding occurred within or around a critical housekeeping gene.

However, the experimental evidence gathered in **Chapter 3** conclusively shows that *E. coli* can grow successfully whilst expressing MmfR constitutively. Genetic constructs were engineered to express MmfR constitutively and transformed into *E. coli*. Compared to a strain expressing the native transcriptional repressor LacI at a comparable rate, MmfR-producing strains showed no inhibition in their ability to grow.

The same constructs were engineered with a *gfp* gene preceded by a transcriptional promoter bearing a MARE operator sequence, the targe site for MmfR. When exposed to an MMF, a specific inhibitor of MmfR, GFP expression increased markedly. As a result, it was concluded that MmfR was being expressed correctly by the

heterologous host of the construct, that the presence of MmfR was not obstructing the growth of the host, and that the MmfR was not only present but functioning as it does in its native host.

For a more quantitative answer to this research question, it would be possible to explore the exact degree to which MmfR expression slows cell growth through a comparative study of culture optical density over time when cells are carrying the MmfR-based expression construct, the LacI-based construct or no construct. This would give a definitive answer to the degree to which constant MmfR expression affects cell fitness, both in comparison to expression of a native (and biotechnological standard) repressor and to no such expression.

5.2.2 Is MmfR a viable alternative to LacI for regulation of recombinant protein production?

MmfR was shown to be a promising alternative to LacI for regulating recombinant protein production. In **Chapter 3**, the genetic construct pDT-M1 was created that constitutively expressed MmfR, and that expressed GFP with a promoter containing a MARE sequence. Experimental data demonstrated that *E. coli* carrying this construct were induced to express GFP when an MMF, the natural inducer for MmfR, was added to the culture medium.

However, MmfR was also shown to be, at present, a leakier repressor than LacI. The construct pDT-L1, identical to pDT-M1 other than that it produced LacI instead of MmfR, and carried the *lac* operator instead of a MARE to regulate GFP production, was produced and tested in parallel. Experimentally, pDT-M1 displayed a higher rate of GFP expression than pDT-L1 when not induced. This leakiness creates a greater selection pressure towards mutations that prevent the expression of GFP altogether. In the industrial context that the construct was intended for, this severely reduces the construct's viability as an inducer.

Furthermore, MmfR was trialled as an expression construct only in *E. coli*. While *E. coli* is very commonly used as an expression host, it is by no means the ideal candidate for all applications. As a prokaryote, it lacks the cellular machinery necessary for the correct post-processing of many eukaryotic proteins – for example, for correct glycosylation, disulfide bridge formation or folding. If not stringently purified, the product can also contain endotoxins produced by the host, which are immunogenic and

thus generally inappropriate for medical applications (Baeshen et al., 2014; Overton, 2014). As a result, it would be desirable to demonstrate that MmfR is not just a viable alternative to LacI-based expression systems, which are specific to *E. coli*, but also to expression systems used in eukaryotic hosts (such as the yeast species *S. cerevisiae* and *P. pastoris*, as well as mammalian cell lines).

5.2.3 How can an MmfR-based expression vector be improved?

Several potential avenues for **improving** an MmfR-based expression system were identified. Mathematical **modelling** of the expression system, based on known interactions and experimental data from **Chapter 3**, was conducted. The modelling explored the consequences of several potential modifications to the expression system.

The effects of reducing the **plasmid concentration** within the cell culture were explored first. By using a plasmid origin of replication with a lower **copy number**, the level of uninduced protein production would be proportionately lower, reducing the metabolic stress caused by the leakiness and the consequent risk of mutation. This approach also reduces the theoretical induced rate of expression, but even with a very low copy number, the **actual** expression rate is likely to still be determined by the metabolic limit of the cell rather than the number of copies of the gene present.

The effects of reducing the **leakiness** of the expression system by increasing MmfR's affinity for the MARE were explored second. Increasing the affinity of MmfR for the MARE (by decreasing the rate of dissociation between the two) has an **inversely proportional** effect on the uninduced expression of product. Such a change would be the ideal way to optimise the expression system, but it is unclear to what degree MmfR-MMF affinity can be improved, and whether it could match LacI's demonstrated affinity for the *lac* operator (based on the relative levels of uninduced GFP expression observed between the pDT-L1 and pDT-M1 constructs).

Experimentally, a set of constructs that would express higher levels of MmfR were created, but only one, pDT-M2**,** was ever able to exist stably in *E. coli*. pDT-M2 carried a promoter for MmfR that was six times as strong as that of pDT-M1, and that construct showed little improvement over pDT-M1 in terms of leakiness. These experimental data suggest that further upregulating MmfR expression is unlikely to reduce the metabolic strain the expression system imposes on its host organism.

5.3 Impact

The MmfR-based expression system pDT-M1 offers a promising alternative to the traditional LacI-based recombinant protein production system for use in *E. coli*. MmfR appears to be a leakier repressor than LacI, but sufficient to maintain a low enough rate of uninduced expression of product to allow for small cultures of non-toxic proteins to be produced. While inappropriate for industrial-scale or toxic protein production, the MmfR expression system pDT-M1 might be sufficient for some laboratory-scale applications without further refinement.

5.4 Outlook

Several avenues for future research and development were identified, with the aim of creating a more stable, more productive, and more versatile recombinant protein production platform:

Firstly, the improvement of MmfR as a repressor. If MmfR and the MARE could be made to have a stronger affinity for one another, the leakiness observed in previous constructs would be mitigated. The stability of the expression construct would thereby improve, by reducing the metabolic strain it places upon the host. There are two methods by which this could be achieved.

The first method would be through site-directed mutagenesis of the MARE sequence. To some extent, this has already been done – the MARE sequence used in the work presented in this thesis was selected from a previous mutagenesis study (Rodriguez-Garcia M; Corre C, 2017). In this study, the mutagenesis was performed on two non-conserved bases on the flanks of the MARE sequence, reasoning that given their variation *in vivo*, they were likely to affect the strength of MmfR-MARE binding without entirely preventing such binding should they be altered. Although it would be exponentially more time-consuming to test every possible sequence for each additional randomised base, further mutagenesis of other bases in the MARE sequence could yet yield a sequence with a stronger affinity than any of the wildtype sequences.

The second method would be through mutagenesis of the MmfR protein itself – specifically, of the first three α-helices of MmfR, which constitute the DNA-binding regions. The a3 helix would be the preferable first candidate for mutagenesis, because it appears to enjoy the most contact with DNA, and because α1 and α2 both have side

chains that appear essential, either directly for ligand binding or for the overall tertiary structure of MmfR (Zhou et al., 2021).

The process for site-directed mutagenesis itself would be straightforward – PCR primers matching the existing pDT-M1 consntruct, with the target bases randomised, would be used to amplify modified variants of the construct. Prospective candidates could then be selected following transformation, and screened for effectiveness with a similar time-resolved fluorescence protocol to that in **Chapter 3**, allowing those constructs generating the lowest baseline fluorescence to be identified, and the addition of MMF used to confirm that they retain their GFP reporter and that their mutant MmfR or MARE still functions as required. A high-throughput liquid handling device capable of performing time-resolved fluorescence readings on multiple microwell plates simultaneously would greatly accelerate this process.

Secondly, mathematical modelling could be used to explore how different gene circuit architectures could yield better results. For example, in the constructs produced and modelled for this project, MmfR is expressed constitutively. However, in *S. coelicolor*, MmfR is under its own transcriptional repression, as is the biosynthetic enzyme pathway MmfLHP, which produces the MMF that inhibits MmfR. The effect that this has on the rate MmfR is expressed both before and after induction could be explored with mathematical modelling, and directly compared with the results of constitutive MmfR expression.

Introducing a MARE sequence to regulate MmfR expression, or introducing the MmfLHP pathway into the construct, could potentially lead to more favourable behaviour (lower uninduced protein production, and higher induced product expression). However, further modelling would be necessary to confirm what effects these changes would be likely to have on the dynamics of the expression system. Given the pre-existing model presented in **Chapter 4**, it would be simple to add more interactions to explore the consequences of, for example, placing MmfR expression under the control of a MARE, placing its expression rate in a negative feedback loop.

Thirdly, the expression of MmfR in hosts other than *E. coli* could be trialled. *S. coelicolor* itself is used as a host for expression of biosynthetic enzymes native to the streptomycetes, which frequently produce antimicrobial compounds, and an MmfR-based expression system has been considered (Styles, 2016).

The use of MmfR as an inducible repressor for yeasts, such as *S. cerevisiae* and *P. pastoris*, is another interesting possibility. Eukaryotic hosts can more reliably produce proteins from other eukaryotic organisms, as they possess post-translational processing capabilities that prokaryotic hosts typically do not. This would make a eukaryotic MmfR-based expression system advantageous for producing proteins that do not express correctly in *E. coli*.

Finally, the use of MmfR as a general-purpose component for synthetic biology applications is also worth consideration. Complex gene circuits, of which the repressilator is a classic example, typically consist of an interlocked set of transcriptional repressors or activators that affect one another's transcription in the desired configuration (Elowitz & Leibier, 2000). For this purpose, it Is desirable to have components that are highly specific to their own operator sequences and ligands. This orthogonality is necessary for the design and construction of gene circuit, and to avoid unintended behaviour arising from unanticipated interactions between components.

MmfR would not be the first ArpA-like TFTR adapted for use in synthetic biology, as mentioned in **Chapter 1** – ScbR has been used successfully as a transcriptional repressor in *E. coli,* in tandem with the biosynthetic pathway for its ligand, SCB. This formed a quorum sensing system entirely orthogonal to both the *A. fischeri* and *P. aeruginosa* quorum sensing systems (Biarnes-Carrera et al., 2018). Given MmfR's sequence and structural homology with ScbR, it is probable that it would be similarly orthogonal to those pre-existing repressor systems, and has been shown not to respond to the presence of SCBs, suggesting that it is also orthogonal to ScbR itself (Zhou et al., 2021) – and that it would therefore make a desirable addition to the library of synthetic biology parts.

5.5 Concluding Statements

MmfR-based expression shows the potential to be a viable alternative to standard LacI-based expression systems. However, further refinement is necessary before this potential can be realised. Potential routes towards improvement have been identified through computational modelling, but need to be experimentally validated.

An MmfR-based expression system could have several advantages over LacI-based alternatives; the inducer MMF can be synthesised in *E. coli* if the biosynthetic enzyme pathway is transferred to it, and so the expression system could generate its

own MMF, activating expression of its intended product through quorum sensing rather than manual induction, or could be induced by cell lysate or a living culture of a separate MMF-producing strain rather than by chemically synthesised MMF. MmfR is also orthogonal to LacI, and to the *E. coli* genome, opening up potential for its use in synthetic biology applications, such as complex gene circuits, that depend on functioning entirely in parallel to the host.

For the present, however, it has been demonstrated that MmfR does not have a deleterious effect on *E. coli* when expressed constitutively, and that an MmfR-based inducible expression system is possible, if not currently quite as efficient as a traditional LacI system.

Appendices

Nucleotide Sequences

Table 0.1: MARE sequences native to S. coelicolor *(non-conserved bases in bold).*

Intergenic Region	MARE Sequence (5' to 3')
mmfL/mmfR	A**T**ACCTTC**CC**G**CA**GGT**A**T
mmyR	A**T**ACCTTC**CC**G**AG**GGT**A**T
mmyB/mmyY	A**A**ACCTTC**GG**G**AA**GGTT**T**

Table 0.2: MARE sequences generated by site-directed mutagenesis (altered bases in bold).

Mutant	MARE Sequence (5' to 3')
Wildtype	A**A**ACCTTCGGGAAGGTT**T**
MM1	A**A**ACCTTCGGGAAGGT**C**T
MM2	A**C**ACCTTCGGGAAGGTT**T**
MM3	A**G**ACCTTCGGGAAGGTT**T**
MM4	A**T**ACCTTCCCGAAGGT**G**T
MM5	A**T**ACCTTCGGGAAGGT**C**T
MM6	A**C**ACCTTCGGGAAGGT**A**T
MM7	A**A**ACCTTCCCGAAGGT**A**T
MM8	A**A**ACCTTCGGGAAGGT**A**T
MM9	A**T**ACCTTCGGGAAGGT**A**T
MM10	A**G**ACCTTCGGGAAGGT**A**T
MM11	A**T**ACCTTCCCGAAGGTT**T**
MM12	A**G**ACCTTCCCGAAGGT**G**T
MM13	A**C**ACCTTCGGGAAGGT**C**T
MM14	A**C**ACCTTCCCGAAGGT**G**T
MM15	A**G**ACCTTCGGGAAGGT**C**T
MM16	A**A**ACCTTCGGGAAGGT**G**T

puc57 - AMPICILLIN

Promoter BBa_J23119 (1x)
gaagacttggagttgacagctagctcagtcctaggtataatgctagctactaa**gtcttc**

Promoter BBa_J23114 (0.1x)
gaagacttggagtttatggctagctcagtcctaggtacaatgctagctactaa**gtcttc**

Promoter BBa_J23113 (0.01x)
gaagacttggagctgatggctagctcagtcctagggattatgctagctactaa**gtcttc**

Promoter - Lac, BBa_C0012
gaagacttggagcaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccga
ctggaaagcgggcagtgagcgcaacgcaattaatgtgagttagctcactcattaggcaccccaggctttacactttatgcttc
cggctcgtatgttgtgtggaattgtgagcggataacaatttactaa**gtcttc**

Promoter - MARE + Lac
gaagacttggagcaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccga
ctggaaagcgggcagtgagcgcaacgcaattaatgtgagttagctcactcattaggcaccccaggctttacaataccttccc
gaaggtgttatgttgtgtggaattgtgagcggataacaatttactaa**gtcttc**

Promoter - MARE Operator
gaagacttggagcaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccga
ctggaaagcgtcgatcgagctcaacgcaattaatgtgagttagctcactcattaggcaccccaggctttacaataccttcccg
aaggtgttatgttgtgtggaattgttagcggagaagaatttactaa**gtcttc**

RBS BBa_B0034
gaagactttacttcacacaggaaaagaggagaaaaacagcaatgaa**gtcttc**

PelB-BCA-His
gaagacttaatgaaatatctgctgcccaccgcggcggccgggctgttgctgctcgccgcgcagcctgcgatggcgtttgtta
accagcacctgtgcggtagccatctggtggaagccctgtacctggtttgcggcgaacgtggcttcttttataccccgaaaacc
cgccgtgaagcggaagatttacaggtgggtcaggttgaactgggcggtggcccgggcgcgggtagcttgcagccgctggc
gctggaaggcagcctgcaaaaacgcggcattgtcgaacagtgctgcacctccatttgctcgctgtatcagcttgaaaactac
tgcaaccaccaccatcatcatcactaagaattcgcttaa**gtcttc**

Terminator BBa_B1006
gaagacttgcttaaaaaaaaaccccgcccctgacagggcggggttttttttcgctaa**gtcttc**

puc57 - KANAMYCIN

Repressor Vector
ggtctcattgctccggaggagaa**gtcttc**gatatc**gaagac**ttcgctgcggccgcggcat**gagacc**

Reporter Vector
ggtctcagcaatgtacaggagaa**gtcttc**gatatc**gaagac**ttcgctcgcgcgcggggat**gagacc**

PCR

AmpR, 1/2

tct**gaagac**ttagcgccatggggtttcttagacgtcaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatt
tttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagt
atgagtattcaacatttccgtgtcgcccttattccctttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtg
aaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgag
agttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtgttgacgcc
gggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatctta
cggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaa
cgatcggaggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccgg
agctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgcagcaatggcaacaacgttgcgcaaactatta
actggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgc
gctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgaa**gtcttc**aga

AmpR, 2/2

tct**gaagac**ttcgtgggtcccgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacg
acggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactg
tcagaccaagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttg
ataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtcccaa**gtcttc**aga

LacI, 1/3

tct**gaagac**ttaatggtgaaaccagtaacgttatacgatgtcgcagagtatgccggtgtctcttatcagaccgtttcccgcgt
ggtgaaccaggccagccacgtttctgcgaaaacgcgggaaaaagtggaagcggcgatggcggagctgaattacattccca
accgcgtggcacaacaactggcgggcaaacagtcgttgctgattggcgttgccacctccagtctggccctgcacgcgccgtc
gcaaattgtcgcggcgattaaatctcgcgccgatcaactgggtgccagcgtggtggtgtcgatggtagaacgaagcggcgt
cgaagcctgtaaagcggcggtgcacaatcttctcgcgcaacgcgtcagtgggctgatcattaactatccgctggatgaccag
gatgccattgctgtggaagctgcctgcactaatgttccggcgttatttcttgatgtctctgaccagacacccatcaacagtatt
attttctcccatgaagatggtacgcgacaa**gtcttc**aga

LacI, 2/3

tct**gaagac**ttcgactgggcgtggagcatctggtcgcattgggtcaccagcaaatcgcgctgttagcgggcccattaagttct
gtctcggcgcgtctgcgtctggctggctggcataaatatctcactcgcaatcaaattcagccgatagcggaacgggaaggcg
actggagtgccatgtccggttttcaacaaaccatgcaaatgctgaatgagggcatcgttcccactgcgatgctggttgccaa
cgatcagatggcgctgggcgcaatgcgcgccattaccgagtccgggctgcgcgttggtgcggatatctcggtagtgggatac
gacgataccgaaaacagcaa**gtcttc**aga

LacI, 3/3

tct**gaagact**tcagctcatgttatatcccgccgttaaccaccatcaaacaggattttcgcctgctggggcaaaccagcgtgg
accgcttgctgcaactctctcagggccaggcggtgaagggcaatcagctgttgcccgtctcactggtgaaaagaaaaacca
ccctggcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgact
ggaaagcgggcagtgaaa**gtcttc**aga

MmfR (codon-optimised)

tga**ggtctc**taatgaccagcgcacagcagccgaccccgtttgcagttcgtagcaatgttcctcgcggtccgcatccgcagca
agaacgtagcattaaaacccgtgcacagattctggaagcagcaagcgaaatttttgcaagtcgcggttatcgcggagcaag
cgttaaagatgttgcagaacgtgttggtatgaccaaaggcgcagtttatttttcattttccgagcaaagaaagcctggccattg
cagttgttgaagaacattatgcacgttggcctgcagcaatggaagaaattcgtattcagggttttactccgctggaaaccgtt
gaagaaatgctgcatcgtgcagcccaggcatttcgtgatgatccggttatgcaggcaggcgcacgtctgcagagcgaacgt

gcctttattgatgcagaactgccgctgccgtatgttgattggacccatctgctggaagttccgctgcaggatgcacgtgaagc
aggtcagctgcgtgcaggcgttgatccggcagcagcagcacgtagcctggttgcagccttttttggtatgcagcatgttagcg
ataatctgcatcagcgtgcagatattatggaacgttggcaagaactgcgtgaactgatgtttttttgcactgcgtgcataaaag
ggctta**gagacc**tat

pSC101 origin

aaa**gaagac**cccgctaatattcagcgatttgcccgagcttgcgagggtgctacttaagcctttagggttttaaggtctgttttg
tagaggagcaaacagcgtttgcgacatccttttgtaatactgcggaactgactaaagtagtgagttatacacagggctggg
atctattctttttatctttttttattctttctttattctataaattataaccacttgaatataaacaaaaaaaacacacaaaggtct
agcggaatttacagagggtctagcagaatttacaagttttccagcaaaggtctagcagaatttacagatacccacaactca
aaggaaaaggacatgtaattatcattgactagcccatctcaattggtatagtgattaaaatcacctagaccaattgagatgt
atgtctgaattagttgttttcaaagcaaatgaactagcgattagtcgctatgacttaacggagcatgaaaccaagctaatttt
atgctgtgtggcactactcaaccccacgattgaaaaccctacaaggaaagaacggacggtatcgttcacttataaccaata
cgctcagatgatgaacatcagtagggaaaatgcttatggtgtattagctaaagcaaccagagagctgatgacgagaactgt
ggaaatcaggaatccttttggttaaaggctttgagattttccagtggacaaactatgccaagttctcaagcgaaaaattagaa
ttagtttttagtgaagagatattgccttatcttttccagttaaaaaaattcataaaatataatctggaacatgttaagtcttttg
aaaacaaatactctatgaggatttatgagtggttattaaaagaactaacacaaaagaaaactcacaaggcaaatatagag
attagccttgatgaatttaagttcatgttaatgcttgaaaataactaccatgagtttaaaaggcttaaccaatgggttttgaaa
ccaataagtaaagatttaaacacttacagcaatatgaaattggtggttgataagcgaggccgcccgactgatacgttgattt
tccaagttgaactagatagacaaatggatctcgtaaccgaacttgagaacaaccagataaaaatgaatggtgacaaaata
ccaacaaccattacatcagattcctacctacgtaacggactaagaaaaacactacacgatgctttaactgcaaaaattcag
ctcaccagttttgaggcaaaatttttgagtgacatgcaaagtaagcatgatctcaatggttcgttctcatggctcacgcaaaa
acaacgaaccacactagagaacatactggctaaatacggaaggatctgaggttcttatggctcttgtatctatcagtgaagc
atcaagactaacaaacaaaagtagaacaactgttcaccgttagatatcaaagggaaaactgtccatatgcacagatgaaa
acggtgtaaaaaagatagatacatcagagcttttacgagtttttggtgcatttaaagctgttcaccatgaacagatcgacaa
tgtgcctgagaccggagaa**gtcttc**aga

sfGFP

tct**gaagact**taatgtccaagggcgaggagctgttcaccggcgtcgtcccgatcctggtcgagctggacggcgacgtgaac
ggccacaagttctccgtccgcggcgagggcgagggcgacgccaccaacggcaagctgaccctgaagttcatctgcaccac
cggcaagctcccggtcccgtggccgaccctggtcaccaccctgacctacggcgtccagtgcttctcccgctacccggaccac
atgaagcgccacgacttcttcaagtccgccatgcccgagggctacgtccaggagcggaccatctccttcaaggacgacggc
acctacaagacccgcgccgaggtcaagttcgagggcgacaccctggtcaaccgcatcgagctgaagggcatcgacttcaa
ggaggacggcaacatcctgggccacaagctcgagtacaacttcaactcccacaacgtctacatcaccgccgacaagcaga
agaacggcatcaaggccaacttcaagatccgccacaacgtcgaggacggcagcgtccagctggccgaccactaccagca
gaacaccccgatcggcgacggcccggtcctgctgccggacaaccactacctgtccacccagtccgtcctgtccaaggaccc
gaacgagaagcgcgaccacatggtcctgctcgagttcgtcaccgccgccggcatcacccacggcatggacgagctgtaca
agtgagcttaa**gtcttc**aga

pDT-L1, M1, M2

Note: These nucleotide sequences include the inserts and lac operator only; the rest of the vector backbone (including resistance marker and origin of replication) is not shown here.

pDT-L1

ctgatggctagctcagtcctagggattatgctagctacttcacacaggaaaagaggagaaaaacagcaatggtgaatgtga
aaccagtaacgttatacgatgtcgcagagtatgccggtgtctcttatcagaccgtttcccgcgtggtgaaccaggccagcca
cgtttctgcgaaaacgcgggaaaaagtggaagcggcgatggcggagctgaattacattcccaaccgcgtggcacaacaac

tggcgggcaaacagtcgttgctgattggcgttgccacctccagtctggccctgcacgcgccgtcgcaaattgtcgcggcgatt
aaatctcgcgccgatcaactgggtgccagcgtggtggtgtcgatggtagaacgaagcggcgtcgaagcctgtaaagcggc
ggtgcacaatcttctcgcgcaacgcgtcagtgggctgatcattaactatccgctggatgaccaggatgccattgctgtggaa
gctgcctgcactaatgttccggcgttatttcttgatgtctctgaccagacacccatcaacagtattattttctcccatgaagatg
gtacgcgactgggcgtggagcatctggtcgcattgggtcaccagcaaatcgcgctgttagcgggcccattaagttctgtctc
ggcgcgtctgcgtctggctggctggcataaatatctcactcgcaatcaaattcagccgatagcggaacgggaaggcgactg
gagtgccatgtccggttttcaacaaaccatgcaaatgctgaatgagggcatcgttcccactgcgatgctggttgccaacgat
cagatggcgctgggcgcaatgcgcgccattaccgagtccgggctgcgcgttggtgcggatatctcggtagtgggatacgac
gataccgaagatagctcatgttatatcccgccgttaaccaccatcaaacaggattttcgcctgctggggcaaaccagcgtgg
accgcttgctgcaactctctcagggccaggcggtgaagggcaatcagctgttgcccgtatcactggtgaaaagaaaaacca
ccctggcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgact
ggaaagcgggcagttataatctagagcttaaaaaaaaccccgccctgtcaggggcggggttttttttttaagcgaattctttgt
agagctcatccatgccatgtgtaatcccagcagcagttacaaactcaagaaggaccatgtggtcacgcttttcgttgggatct
ttcgaaaggacagattgtgtcgacaggtaatggttgtctggtaaaaggacagggccatcgccaattggagtattttgttgat
aatggtctgctagttgaacggaaccatcttcaacgttgtggcgaattttgaagttagctttgattccattcttttgtttgtctgcc
gtgatgtatacattgtgtgagttaaagttgtactcgagtttgtgtccgagaatgtttccatcttctttaaaatcaataccttttaa
ctcgatacgattaacaagggtatcaccttcaaacttgacttcagcacgcgtcttgtaggtcccgtcatctttgaaagatatagt
gcgttcctgtacataaccttcgggcatggcactcttgaaaaagtcatgccgtttcatgtgatccggataacgggaaaagcatt
gaacaccataggtcagagtagtgacaagtgttggccatggaacaggtagttttccagtagtgcaaataaatttaagggtga
gttttccgtttgtagcatcaccttcaccctctccacggacagaaaatttgtgcccattaacatcaccatctaattcaacaagaa
ttgggacaactccagtgaaaagttcttctcctttgctcattgctgtttttctcctctttatcccaatggcgcgccgagcttggctc
gagcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgt
aaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcc

pDT-M1

ctgatggctagctcagtcctagggattatgctagctacttcacacaggaaaagaggagaaaaacagcaatgaccagcgca
cagcagccgaccccgtttgcagttcgtagcaatgttcctcgcggtccgcatccgcagcaagaacgtagcattaaaaccccgtg
cacagattctggaagcagcaagcgaaatttttgcaagtcgcggttatcgcggagcaagcgttaaagatgttgcagaacgtg
ttggtatgaccaaaggcgcagtttattttcattttccgagcaaagaaagcctggccattgcagttgttgaagaacattatgca
cgttggcctgcagcaatggaagaaattcgtattcagggttttactccgctggaaaccgttgaagaaatgctgcatcgtgcag
cccaggcatttcgtgatgatccggttatgcaggcaggcgcacgtctgcagagcgaacgtgcctttattgatgcagaactgcc
gctgccgtatgttgattggacccatctgctggaagttccgctgcaggatgcacgtgaagcaggtcagctgcgtgcaggcgtt
gatccggcagcagcagcacgtagcctggttgcagccttttttggtatgcagcatgttagcgataatctgcatcagcgtgcag
atattatggaacgttggcaagaactgcgtgaactgatgtttttgcactgcgtgcataatctagagcttaaaaaaaaccccgc
cctgtcaggggcggggttttttttttaagcgaattctttgtagagctcatccatgccatgtgtaatcccagcagcagttacaaac
tcaagaaggaccatgtggtcacgcttttcgttgggatctttcgaaaggacagattgtgtcgacaggtaatggttgtctggtaa
aaggacagggccatcgccaattggagtattttgttgataatggtctgctagttgaacggaaccatcttcaacgttgtggcga
attttgaagttagctttgattccattcttttgtttgtctgccgtgatgtatacattgtgtgagttaaagttgtactcgagtttgtgt
ccgagaatgtttccatcttctttaaaatcaataccttttaactcgatacgattaacaagggtatcaccttcaaacttgacttca
gcacgcgtcttgtaggtcccgtcatctttgaaagatatagtgcgttcctgtacataaccttcgggcatggcactcttgaaaaa
gtcatgccgtttcatgtgatccggataacgggaaaagcattgaacaccataggtcagagtagtgacaagtgttggccatgg
aacaggtagttttccagtagtgcaaataaatttaagggtgagttttccgtttgtagcatcaccttcaccctctccacggacag
aaaatttgtgcccattaacatcaccatctaattcaacaagaattgggacaactccagtgaaaagttcttctcctttgctcattg
ctgtttttctcctctttatcccaatggcgcgccgagcttggctcgagcatggtcatagctgtttcctgtgtgaaattcttctccgc
taacaattccacacaacataacaccttcgggaaggtattgtaaagcctggggtgcctaatgagtgagctaactcacattaat
tgcgttgagctcgatcga

pDT-M2

tttacagctagctcagtcctagggactgtgctagctacttcacacaggaaaagaggagaaaaacagcaatgaccagcgca
cagcagccgaccccgtttgcagttcgtagcaatgttcctcgcggtccgcatccgcagcaagaacgtagcattaaaacccgtg
cacagattctggaagcagcaagcgaaatttttgcaagtcgcggttatcgcggagcaagcgttaaagatgttgcagaacgtg
ttggtatgaccaaaggcgcagtttatttttcattttccgagcaaagaaagcctggccattgcagttgttgaagaacattatgca
cgttggcctgcagcaatggaagaaattcgtattcagggttttactccgctggaaaccgttgaagaaatgctgcatcgtgcag
cccaggcatttcgtgatgatccggttatgcaggcaggcgcacgtctgcagagcgaacgtgcctttattgatgcagaactgcc
gctgccgtatgttgattggacccatctgctggaagttccgctgcaggatgcacgtgaagcaggtcagctgcgtgcaggcgtt
gatccggcagcagcagcacgtagcctggttgcagccttttttggtatgcagcatgttagcgataatctgcatcagcgtgcag
atattatggaacgttggcaagaactgcgtgaactgatgtttttgcactgcgtgcataatctagagcttaaaaaaaaccccgc
cctgtcaggggcggggttttttttttaagcgaattctttgtagagctcatccatgccatgtgtaatcccagcagcagttacaaac
tcaagaaggaccatgtggtcacgcttttcgttgggatctttcgaaaggacagattgtgtcgacaggtaatggttgtctggtaa
aaggacagggccatcgccaattggagtattttgttgataatggtctgctagttgaacggaaccatcttcaacgttgtggcga
attttgaagttagctttgattccattcttttgtttgtctgccgtgatgtatacattgtgtgagttaaagttgtactcgagtttgtgt
ccgagaatgtttccatcttctttaaaatcaataccttttaactcgatacgattaacaagggtatcaccttcaaacttgacttca
gcacgcgtcttgtaggtcccgtcatctttgaaagatatagtgcgttcctgtacataaccttcgggcatggcactcttgaaaaa
gtcatgccgtttcatgtgatccggataacgggaaaagcattgaacaccataggtcagagtagtgacaagtgttggccatgg
aacaggtagttttccagtagtgcaaataaatttaagggtgagttttccgtttgtagcatcaccttcaccctctccacggacag
aaaatttgtgcccattaacatcaccatctaattcaacaagaattgggacaactccagtgaaaagttcttctcctttgctcattg
ctgtttttctcctctttatcccaatggcgcgccgagcttggctcgagcatggtcatagctgtttcctgtgtgaaattcttctccgc
taacaattccacacaacataacaccttcgggaaggtattgtaaagcctggggtgcctaatgagtgagctaactcacattaat
tgcgttgagctcgatcga

References

Adrio, J. L., & Demain, A. L. (2010). Recombinant organisms for production of industrial products. *Bioengineered Bugs*, *1*(2), 116–131. https://doi.org/10.4161/bbug.1.2.10484

Andersen, D. C., & Krummen, L. (2002). Recombinant protein expression for therapeutic applications. *Current Opinion in Biotechnology*, *13*(2), 117–123. https://doi.org/10.1016/S0958-1669(02)00300-2

Assenberg, R., Wan, P. T., Geisse, S., & Mayr, L. M. (2013). Advances in recombinant protein expression for use in pharmaceutical research. *Current Opinion in Structural Biology*, *23*(3), 393–402. https://doi.org/10.1016/J.SBI.2013.03.008

Baeshen, N. A., Baeshen, M. N., Sheikh, A., Bora, R. S., Ahmed, M. M. M., Ramadan, H. A. I., … Redwan, E. M. (2014). Cell factories for insulin production. *Microbial Cell Factories*, *13*(1), 1–9. https://doi.org/10.1186/s12934-014-0141-0

Baneyx, F. (1999). Recombinant protein expression in Escherichia coli. *Current Opinion in Biotechnology*, *10*(5), 411–421. https://doi.org/10.1016/S0958-1669(99)00003-8

Bernstein, J. A., Khodursky, A. B., Lin, P.-H., Lin-Chao, S., & Cohen, S. N. (2002). Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(15), 9697–9702. https://doi.org/10.1073/pnas.112318199

Bhukya, H., Bhujbalrao, R., Bitra, A., & Anand, R. (2014). Structural and functional basis of transcriptional regulation by TetR family protein CprB from S. coelicolorA3(2). *Nucleic Acids Research*, *42*(15), 10122–10133. https://doi.org/10.1093/nar/gku587

Biarnes-Carrera, M., Breitling, R., & Takano, E. (2015). Butyrolactone signalling circuits for synthetic biology. *Current Opinion in Chemical Biology*, *28*, 91–98. https://doi.org/10.1016/j.cbpa.2015.06.024

Biarnes-Carrera, M., Lee, C. K., Nihira, T., Breitling, R., & Takano, E. (2018). Orthogonal Regulatory Circuits for Escherichia coli Based on the γ-Butyrolactone System of Streptomyces coelicolor. *ACS Synthetic Biology*, *7*(4), 1043–1055. https://doi.org/10.1021/acssynbio.7b00425

Boo, A., Ellis, T., & Stan, G. B. (2019). Host-aware synthetic biology. *Current Opinion in*

Systems Biology, *14*, 66–72. https://doi.org/10.1016/j.coisb.2019.03.001

Borkowski, O., Ceroni, F., Stan, G. B., & Ellis, T. (2016). Overloaded and stressed: whole-cell considerations for bacterial synthetic biology. *Current Opinion in Microbiology*, *33*, 123–130. https://doi.org/10.1016/J.MIB.2016.07.009

Borrelli, G. M., & Trono, D. (2015). Recombinant lipases and phospholipases and their use as biocatalysts for industrial applications. *International Journal of Molecular Sciences*, *16*(9), 20774–20840. https://doi.org/10.3390/ijms160920774

Bowyer, J. E., de los Santos, E. L. C., Styles, K. M., Fullwood, A., Corre, C., & Bates, D. G. (2017). Modeling the architecture of the regulatory system controlling methylenomycin production in Streptomyces coelicolor. *Journal of Biological Engineering*, *11*(1), 1–12. https://doi.org/10.1186/s13036-017-0071-6

Boyhan, D., & Daniell, H. (2011). Low-cost production of proinsulin in tobacco and lettuce chloroplasts for injectable or oral delivery of functional insulin and C-peptide. *Plant Biotechnology Journal*, *9*(5), 585–598. https://doi.org/10.1111/j.1467-7652.2010.00582.x

Brandt, F., Etchells, S. A., Ortiz, J. O., Elcock, A. H., Hartl, F. U., & Baumeister, W. (2009). The Native 3D Organization of Bacterial Polysomes. *Cell*, *136*(2), 261–271. https://doi.org/10.1016/j.cell.2008.11.016

Bremer, H., & Dennis, P. P. (2008). Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. *EcoSal Plus*, *3*(1). https://doi.org/10.1128/ECOSAL.5.2.3

Briggs, T. S., & Rauscher, W. C. (1973). An oscillating iodine clock. *Journal of Chemical Education*, *50*(7), 496. https://doi.org/10.1021/ED050P496

Carroll, D. (2011). Genome Engineering With Zinc-Finger Nucleases. *Genetics*, *188*(4), 773–782. https://doi.org/10.1534/genetics.111.131433

Carter, P. (1986). Site-directed mutagenesis. *Biochemical Journal*, *237*(1), 1–7. https://doi.org/10.1042/bj2370001

Chen, R. (2012). Bacterial expression systems for recombinant protein production: E. coli and beyond. *Biotechnology Advances*, *30*(5), 1102–1107. https://doi.org/10.1016/j.biotechadv.2011.09.013

Christian, M., Cermak, T., Doyle, E. L., Schmidt, C., Zhang, F., Hummel, A., … Voytas, D. F. (2010). Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. *Genetics*, *186*(2), 757. https://doi.org/10.1534/GENETICS.110.120717

Corre, C., Song, L., O'Rourke, S., Chater, K. F., & Challis, G. L. (2008). 2-Alkyl-4-hydroxymethylfuran-3-carboxylic acids, antibiotic production inducers discovered by Streptomyces coelicolor genome mining. *Proceedings of the National Academy of Sciences*, *105*(45), 17510–17515. https://doi.org/10.1073/pnas.0805530105

Cregg, J. M., Cereghino, J. L., Shi, J., & Higgins, D. R. (2000). Recombinant protein expression in Pichia pastoris. *Molecular Biotechnology 2000 16:1*, *16*(1), 23–52. https://doi.org/10.1385/MB:16:1:23

Cuthbertson, L., & Nodwell, J. R. (2013). The TetR Family of Regulators. *Microbiology and Molecular Biology Reviews*, *77*(3), 440–475. https://doi.org/10.1128/MMBR.00018-13

Di Felice, F., Micheli, G., & Camilloni, G. (2019). Restriction enzymes and their use in molecular biology: An overview. *Journal of Biosciences*, *44*(2), 38. https://doi.org/10.1007/s12038-019-9856-8

Dingermann, T. (2008). Recombinant therapeutic proteins: Production platforms and challenges. *Biotechnology Journal*, *3*(1), 90–97. https://doi.org/10.1002/biot.200700214

Elowitz, M. B., & Leibier, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature 2000 403:6767*, *403*(6767), 335–338. https://doi.org/10.1038/35002125

Engler, C., Kandzia, R., & Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE*, *3*(11). https://doi.org/10.1371/journal.pone.0003647

Engler, C., & Marillonnet, S. (2013). Combinatorial DNA Assembly Using Golden Gate Cloning. In M. Engelhard (Ed.) (Vol. 44, pp. 141–156). Cham: Springer International Publishing. https://doi.org/10.1007/978-1-62703-625-2_12

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, *6*(5), 343–345. https://doi.org/10.1038/nmeth.1318

Goeddel, D. V, Kleid, D. G., Bolivar, F., Heyneker, H. L., Yansura, D. G., Crea, R., … Riggs, A. D. (1979). Expression in Escherichia coli of chemically synthesized genes for human insulin. *Proceedings of the National Academy of Sciences of the United States of America*, *76*(1), 106–110. https://doi.org/10.1073/pnas.76.1.106

Goulding, C. W., & Perry, L. J. (2003). Protein production in Escherichia coli for

structural studies by X-ray crystallography. *Journal of Structural Biology*. https://doi.org/10.1016/S1047-8477(03)00044-3

Helguera, G., & Penichet, M. L. (2005). Antibody-cytokine fusion proteins for the therapy of cancer. *Methods in Molecular Medicine*, *109*, 347–374. https://doi.org/10.1385/1-59259-862-5:347

Hoops, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., … Kummer, U. (2006). COPASI—a COmplex PAthway SImulator. *Bioinformatics*, *22*(24), 3067–3074. https://doi.org/10.1093/BIOINFORMATICS/BTL485

Hussain, B., Ruchika, B., Aruna, B., Ruchi, A. (2014). Elucidation of the Structural and Functional Mechanism of Action of the TetR Family Protein, CprB from S. coelicolor A3(2). https://doi.org/https://doi.org/10.2210/pdb4PXI/pdb

J. Murray. (2002). *Mathematical Biology*. (J. D. Murray, Ed.) (Vol. 17). New York, NY: Springer New York. https://doi.org/10.1007/b98868

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, *337*(6096), 816–821. https://doi.org/10.1126/SCIENCE.1225829/SUPPL_FILE/JINEK.SM.PDF

Joosten, V., Lokman, C., Van Den Hondel, C. A., & Punt, P. J. (2003). The production of antibody fragments and antibody fusion proteins by yeasts and filamentous fungi. *Microbial Cell Factories*, *2*(1), 1. https://doi.org/10.1186/1475-2859-2-1

Joung, J. K., & Sander, J. D. (2013). TALENs: A widely applicable technology for targeted genome editing. *Nature Reviews Molecular Cell Biology*, *14*(1), 49–55. https://doi.org/10.1038/NRM3486

Kelly, J. R., Rubin, A. J., Davis, J. H., Ajo-Franklin, C. M., Cumbers, J., Czar, M. J., … Endy, D. (2009). Measuring the activity of BioBrick promoters using an in vivo reference standard. *Journal of Biological Engineering*, *3*(1), 1–13. https://doi.org/10.1186/1754-1611-3-4/FIGURES/3

Kennell, D., & Riezman, H. (1977). Transcription and translation initiation frequencies of the Escherichia coli lac operon. *Journal of Molecular Biology*, *114*(1), 1–21. https://doi.org/10.1016/0022-2836(77)90279-0

Kermani, A. A. (2021). A guide to membrane protein X-ray crystallography. *FEBS Journal*, *288*(20), 5788–5804. https://doi.org/10.1111/febs.15676

Kim, Y. G., Cha, J., & Chandrasegaran, S. (1996). Hybrid restriction enzymes: Zinc finger

fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(3), 1156–1160. https://doi.org/10.1073/PNAS.93.3.1156

Klumpp, S. (2011). Growth-Rate Dependence Reveals Design Principles of Plasmid Copy Number Control. *PLoS ONE*, *6*(5), 20403. https://doi.org/10.1371/JOURNAL.PONE.0020403

Kudva, R., Denks, K., Kuhn, P., Vogt, A., Müller, M., & Koch, H. G. (2013). Protein translocation across the inner membrane of Gram-negative bacteria: The Sec and Tat dependent protein transport pathways. *Research in Microbiology*, *164*(6), 505–534. https://doi.org/10.1016/j.resmic.2013.03.016

Li, T., Huang, S., Jiang, W. Z., Wright, D., Spalding, M. H., Weeks, D. P., & Yang, B. (2011). TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Research*, *39*(1), 359–372. https://doi.org/10.1093/nar/gkq704

Li, X., Wang, J., Li, S., Ji, J., Wang, W., & Yang, K. (2015). ScbR-and ScbR2-mediated signal transduction networks coordinate complex physiological responses in Streptomyces coelicolor. *Scientific Reports*, *5*(July). https://doi.org/10.1038/srep14831

Liu, G., Chater, K. F., Chandra, G., Niu, G., & Tan, H. (2013). Molecular Regulation of Antibiotic Biosynthesis in Streptomyces. *Microbiology and Molecular Biology Reviews*, *77*(1), 112–143. https://doi.org/10.1128/mmbr.00054-12

Mahfouz, M. M., Li, L., Shamimuzzaman, M., Wibowo, A., Fang, X., & Zhu, J. K. (2011). De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(6), 2623–2628. https://doi.org/10.1073/PNAS.1019533108/-/DCSUPPLEMENTAL

Maurizi, M. R. (1992). Proteases and protein degradation in Escherichia coli. *Experientia*, *48*(2), 178–201. https://doi.org/10.1007/BF01923511

Mehra, S., Charaniya, S., Takano, E., & Hu, W.-S. (2008). A Bistable Gene Switch for Antibiotic Biosynthesis: The Butyrolactone Regulon in Streptomyces coelicolor. *PLoS ONE*, *3*(7), e2724. https://doi.org/10.1371/journal.pone.0002724

Milo, R. (2013). What is the total number of protein molecules per cell volume? A call to rethink some published values. *Bioessays*, *35*(12), 1050.

https://doi.org/10.1002/BIES.201300066

Murer, P., & Neri, D. (2019). Antibody-cytokine fusion proteins: A novel class of biopharmaceuticals for the therapy of cancer and of chronic inflammation. *New Biotechnology*, *52*(April), 42–53. https://doi.org/10.1016/j.nbt.2019.04.002

Natsume, R., Senda, T., Horinouchi, S. (2004). Crystal structure of gamma-butyrolactone receptor (ArpA-like protein). https://doi.org/https://doi.org/10.2210/pdb1UI6/pdb

Natsume, R., Ohnishi, Y., Senda, T., & Horinouchi, S. (2004). Crystal Structure of a γ-Butyrolactone Autoregulator Receptor Protein in Streptomyces coelicolor A3(2). *Journal of Molecular Biology*, *336*(2), 409–419. https://doi.org/10.1016/j.jmb.2003.12.040

Nilsson, J., Ståhl, S., Lundeberg, J., Uhlén, M., & Nygren, P. Å. (1997). Affinity fusion strategies for detection, purification, and immobilization of recombinant proteins. *Protein Expression and Purification*, *11*(1), 1–16. https://doi.org/10.1006/prep.1997.0767

O'Rourke, S., Wietzorrek, A., Fowler, K., Corre, C., Challis, G. L., & Chater, K. F. (2009). Extracellular signalling, translational control, two repressors and an activator all contribute to the regulation of methylenomycin production in Streptomyces coelicolor. *Molecular Microbiology*, *71*(3), 763–778. https://doi.org/10.1111/j.1365-2958.2008.06560.x

Oehler, S., Eismann, E. R., Krämer, H., & Müller-Hill, B. (1990). The three operators of the lac operon cooperate in repression. *The EMBO Journal*, *9*(4), 973–979. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2182324

Onaka, H., Nakagawa, T., & Horinouchi, S. (1998). Involvement of two A-factor receptor homologues in Streptomyces coelicolor A3(2) in the regulation of secondary metabolism and morphogenesis. *Molecular Microbiology*, *28*(4), 743–753. https://doi.org/10.1046/j.1365-2958.1998.00832.x

Overton, T. W. (2014). Recombinant protein production in bacterial hosts. *Drug Discovery Today*, *19*(5), 590–601. https://doi.org/10.1016/j.drudis.2013.11.008

Pai, A., & You, L. (2009). Optimal tuning of bacterial sensing potential. *Molecular Systems Biology*, *5*(286), 1–11. https://doi.org/10.1038/msb.2009.43

Part:BBa B0034 - parts.igem.org. (n.d.). Retrieved April 26, 2022, from http://parts.igem.org/Part:BBa_B0034

Part:BBa B1006 - parts.igem.org. (n.d.). Retrieved April 26, 2022, from http://parts.igem.org/Part:BBa_B1006

Part:BBa R0010 - parts.igem.org. (n.d.). Retrieved April 26, 2022, from http://parts.igem.org/Part:BBa_R0010

Pédelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C., & Waldo, G. S. (2005). Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology 2005 24:1*, *24*(1), 79–88. https://doi.org/10.1038/nbt1172

Pingoud, A. (2001). Structure and function of type II restriction endonucleases. *Nucleic Acids Research*, *29*(18), 3705–3727. https://doi.org/10.1093/nar/29.18.3705

Promoters/Catalog/Anderson - parts.igem.org. (n.d.). Retrieved April 5, 2022, from http://parts.igem.org/Promoters/Catalog/Anderson

Ramos, J. L., Martínez-Bueno, M., Molina-Henares, A. J., Terán, W., Watanabe, K., Zhang, X., … Tobes, R. (2005). The TetR Family of Transcriptional Repressors. *Microbiology and Molecular Biology Reviews*, *69*(2), 326–356. https://doi.org/10.1128/mmbr.69.2.326-356.2005

Rodriguez-Garcia M; Corre C. (2017). [Unpublished Manuscript].

Rosano, G. L., & Ceccarelli, E. A. (2014a). Recombinant protein expression in Escherichia coli: Advances and challenges. *Frontiers in Microbiology*, *5*(APR), 1–17. https://doi.org/10.3389/fmicb.2014.00172

Rosano, G. L., & Ceccarelli, E. A. (2014b). Recombinant protein expression in microbial systems. *Frontiers in Microbiology*, *5*(JULY), 341. https://doi.org/10.3389/FMICB.2014.00341/BIBTEX

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., … Erlich, H. A. (1988). Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase. *Science*, *239*(4839), 487–491. https://doi.org/10.1126/science.2448875

Schrodinger, E., & Penrose, R. (2012). *What is Life?* Cambridge University Press. https://doi.org/10.1017/CBO9781107295629

Segel, L. A., & Edelstein-Keshet, L. (2013). *A Primer on Mathematical Models in Biology*. Philadelphia, PA: Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611972504

Sevastsyanovich, Y. R., Alfasi, S. N., & Cole, J. A. (2010). Sense and nonsense from a systems biology approach to microbial recombinant protein production.

Biotechnology and Applied Biochemistry, *55*(1), 9–28. https://doi.org/10.1042/ba20090174

Sørensen, H. P., & Mortensen, K. K. (2005). Advanced genetic strategies for recombinant protein expression in Escherichia coli. *Journal of Biotechnology*, *115*(2), 113–128. https://doi.org/10.1016/j.jbiotec.2004.08.004

Styles, K. M. (2016). *Investigating Interactions Between Methylenomycin Furan Microbial Hormones And Transcriptional Repressors in Streptomyces coelicolor*.

Tokareva, O., Michalczechen-Lacerda, V. A., Rech, E. L., & Kaplan, D. L. (2013). Recombinant DNA production of spider silk proteins. *Microbial Biotechnology*, *6*(6), 651–663. https://doi.org/10.1111/1751-7915.12081

Trono, D. (2019). *Recombinant Enzymes in the Food and Pharmaceutical Industries*. *Advances in Enzyme Technology, First Edition*. Elsevier B.V. https://doi.org/10.1016/B978-0-444-64114-4.00013-3

Tsigkinopoulou, A., Takano, E., & Breitling, R. (2020). Unravelling the γ-butyrolactone network in Streptomyces coelicolor by computational ensemble modelling. *PLoS Computational Biology*, *16*(7), 1–17. https://doi.org/10.1371/journal.pcbi.1008039

Walsh, G. (2005). Therapeutic insulins and their large-scale manufacture. *Applied Microbiology and Biotechnology*, *67*(2), 151–159. https://doi.org/10.1007/s00253-004-1809-x

Weber, E., Engler, C., Gruetzner, R., Werner, S., & Marillonnet, S. (2011). A modular cloning system for standardized assembly of multigene constructs. *PLoS ONE*, *6*(2). https://doi.org/10.1371/journal.pone.0016765

Werner, S., Engler, C., Weber, E., Gruetzner, R., & Marillonnet, S. (2012). Fast track assembly of multigene constructs using Golden Gate cloning and the MoClo system. *Https://Doi.Org/10.4161/Bbug.3.1.18223*, *3*(1), 38–43. https://doi.org/10.4161/BBUG.3.1.18223

Williams, J. W., Cui, X., Levchenko, A., & Stevens, A. M. (2008). Robust and sensitive control of a quorum-sensing circuit by two interlocked feedback loops. *Molecular Systems Biology*, *4*, 234. https://doi.org/10.1038/MSB.2008.70

Wood, D. W. (2014). New trends and affinity tag designs for recombinant protein purification. *Current Opinion in Structural Biology*, *26*, 54–61. https://doi.org/10.1016/j.sbi.2014.04.006

Yamasaki, M., Ikuto, Y., Ohira, A., Chater, K., & Kinashi, H. (2003). Limited regions of homology between linear and circular plasmids encoding methylenomycin biosynthesis in two independently isolated streptomycetes. *Microbiology*, *149*(5), 1351–1356. https://doi.org/10.1099/mic.0.26102-0

Zhou, S., Bhukya, H., Malet, N., Harrison, P. J., Rea, D., Belousoff, M. J., … Corre, C. (2021). Molecular basis for control of antibiotic production by a bacterial hormone. *Nature*, *590*(7846), 463–467. https://doi.org/10.1038/s41586-021-03195-x

Zhou, S., & Challis, G. L. (2016). Antibiotic Biosynthesis and its Transcriptional Regulation in Streptomyces Bacteria. Retrieved from http://wrap.warwick.ac.uk/89468