Comparison of the burden and basal performance of optogenetic expression systems

Thea C T Irvine

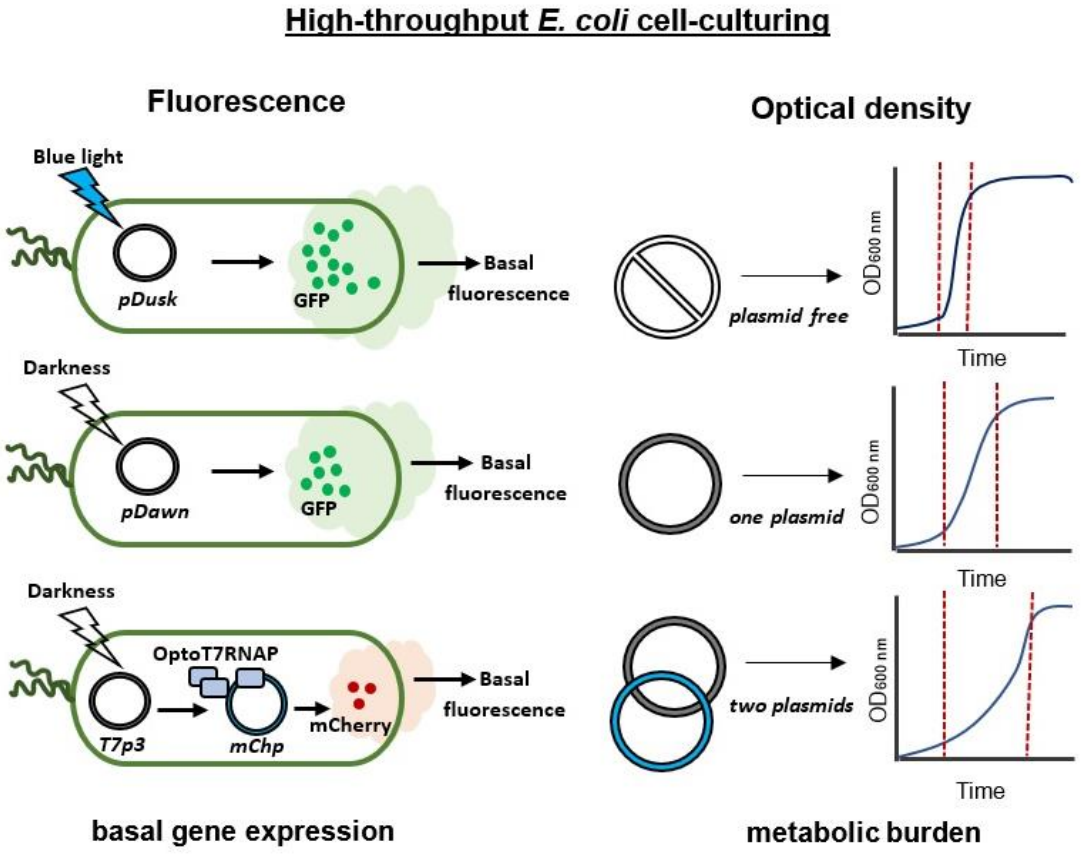
A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Master of Science by Research (MScR) in the Faculty of Life Sciences, School of Biological Sciences.

September 2022

Word count: 26,223

Abstract

Optogenetic circuits offer the unique capabilities of fully reversible and light-tuneable modulation, facilitating precise spatiotemporal control of gene expression. Yet, light-inducible vectors are not broadly implemented in industry and research. Compared to popular chemical induction systems, optogenetic circuits have inherently high basal expression levels. Basal expression is a concern for all synthetic vectors as leaky gene expression during early growth phases can destabilise the bacterial host. The redirection of resources toward recombinant processes impedes host metabolism and slows growth. Comparative research into the basal performance and robustness of optogenetic expression vectors is relatively uncommon. Yet, comparative studies and characterisation of optogenetic circuits in different host strains would expand the usability of optogenetic control. The vectors pDusk and pDawn are constitutively expressed and gene expression is repressed or induced by blue light. Alternatively, the two-plasmid OptoT7RNAP system induces gene transcription via a photo-dimerizing T7 RNA polymerase. In this system, the photoactivable T7 RNA polymerase and gene of interest are transcribed from sperate plasmids. In this work, I compared the key circuit components that improved basal performance, reduced burden, and highlighted the strategies that may aid the broader implementation of optogenetics in the future. To compare the basal performance of pDusk, pDawn and OptoT7RNAP, I used high-throughput cell-culturing of *E. coli* transformants in microplates. Fluorescent reporter proteins were used for fast and efficient analysis of basal gene expression and optical density measurements were used to monitor bacterial proliferation over time. To characterise the metabolic burden associated with plasmid maintenance, replication and basal gene expression, population models were used to estimate bacterial growth properties. I found that photo-dimerizing T7RNAP modulation of gene expression greatly improved basal performance but the expression of a two plasmid system increased the doubling time of the *E. coli* host by 90%.



Acknowledgements

I would like to thank my supervisor, Thomas Gorochowski. Your continued support, guidance, and belief in me has been invaluable and much appreciated. I thoroughly enjoyed my time with the BCL research group.

Thanks to all my friends in life sciences and everyone in lab 140, your positivity and kindness made it a true pleasure to work with you all. Special thanks to Veronica for all your help, I could always rely on your advice when I had a question.

Thank you, Andrew Herman, for your help with flow cytometry and thank you Paulina Dubiel for your help with the plate reader.

Thank you, Shazna, Nat and my family for your love, advice and friendship.

Thank you always to Matt. If you ever wanted a career change, I might recommend motivational speaking. “Don’t let perfectionism be the enemy of progress” was my favourite. Wasn’t such a fan of “This is an intervention…I’m switching the tea and coffee to decaf”.

Author’s declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: Thea C T Irvine DATE: 14/09/2022

Contents

List of Figures

List of Tables

Chapter 1: Background and motivation

1.1 *Escherichia coli* the most popular expression platform

Recombinant DNA technology emerged during the early 1970's when the propagation of eukaryotic DNA in *E. coli* was first demonstrated using recombinant bacterial plasmids (Morrow et al., 1974). Plasmids are naturally occurring small circular DNA strands which replicate independently of the bacterial chromosomes. Transferred from cell-to-cell in a process known as horizontal gene transfer, they facilitate the rapid adaption of a bacterial population to changing environmental conditions (Smalla, Jechalke and Top, 2015). Artificially constructed plasmids, or vectors, are used to clone, amplify, or express heterologous genes of interest in transformed bacteria. Further advancements in molecular biology were made when the polymerase chain reaction (PCR) was developed (Mullis, 1994), using this technique we began to better understand genes and their specific purposes. The PCR technique describes a fast and simple methodology to exponentially amplify DNA fragments using heat-stable DNA polymerase enzymes and an automated heat block capable of rapid changes in temperature, to synthesise copies of a specific DNA fragment (Wages, 2005). PCR and recombinant DNA technology has greatly aided our understanding of genes, their products and protein structures.

Today there are many heterologous gene expression platforms; bacteria, yeast, insect cells, mammalian cells and now plants, are used successfully in both academic research and commercial environments to produce heterologous protein. However, *Escherichia coli* prevails as the most popular gene expression platform. The ability of the gram-negative bacterium to grow rapidly and at a high cell density, as well as its well characterised genetics make the fermentation of *E. coli* relatively easy and cost efficient (Sørensen and Mortensen. 2004). For this reason, many molecular tools, protocols, and cultivation strategies have been developed for heterologous gene expression in *E. coli* and through continued research to improve expression efficiency, a huge catalogue of cloning vectors and engineered strains have become available. Furthermore, the continuous depositing of discovered wild-type or mutated genes into open-source genetic "toolboxes" or gene-banks has enabled the rapid combinatorial assembly of genetic elements into functioning synthetic circuits, with ever growing complexity (Molinari, Tesoriero and Ajo-Franklin., 2021; Cai, Lin and Shi., 2022).

1.2 *E. coli* strains for gene expression

By far one of the most used protein expression strains of *E. coli* is BL21(DE3), which is derived from a succession of strains beginning with strain B, first characterised by Delbruck and Luria in 1942. However, BL21 and some K-12 lineage derivates are still used for protein expression studies (Rosano and Ceccarelli., 2014) particularly when BL21(DE3) is

unsuitable. BL21 and BL21(DE3) are *Lon* protease and *OmpT* protease deficient, as these proteases are problematic for heterologous protein production. The *Lon* protease degrades intracellular foreign protein and the outer membrane protease *OmpT* degrades extracellular protein. After cell lysis and during protein harvest, *OmpT* has been shown to digest heterologous proteins (Rosano and Ceccarelli et al., 2014). The BL21(DE3) expression system contains a chromosomally encoded bacteriophage T7 RNA polymerase (T7RNAP), under the control of the strong promoter *lac*UV5, which can be leakier and more active than the wild type *lac* promoter. A key advantage of the BL21(DE3) expression system is that a heterologous gene, under the control of a T7 promoter, will be transcribed eight times faster by T7RNAP compared to native *E. coli* RNAP (Du et al., 2021). Another key characteristic of expression strains is the *hsd*SB mutation, which inactivates native restriction/methylation activity by eliminating the restriction endonuclease EcoKI. This prevents cleavage of heterologous DNA by an endogenous endonuclease and aids plasmid retention. The *hsd*SB mutation is also present in many common *E. coli* cloning strains, such as DH5α and DH10β. Specific to the K-12 lineage, strains AD494 and OrigamiTM have the *trxB* mutation, which enhances disulphide bond formation in the cytoplasm and the strain HMS174 has a *recA* mutation which enhances plasmid stability by preventing the formation of plasmid multimers. Cloning strains like DH5α and DH10β differ genetically from expression strains, and it is rare to see them used for protein expression. Cloning strains are instead used to generate plasmid libraries. To achieve good plasmid yields the strains are deficient in intrinsic restriction enzymes and produce DNA that is not protected against restriction digestion (i.e., with no DNA methylation). These strains do not need to be optimised for features facilitating protein production and folding, whereas expression strains have been selected/optimised for good quality protein production. The optimal strain for a project will vary depending on the protein of interest and the specific requirements for protein expression, such as harmless or toxic proteins, cytoplasmic or periplasmic expression, whether disulphide bond formation is needed or whether inclusion bodies and in vitro refolding or native expression is required. Strain choice can impact circuit performance. Differences in growth dynamics, ribosome concentration and other features have been shown to create strain dependent variations in circuit performance. Media and growth conditions can also influence circuit performance by altering promoter activity, protein stability and transcription factor dilution. These affects can be so pronounced that transitioning from LB to minimal medium can cause synthetic genetic circuits to fail (Brophy and Voigt., 2014).

1.3 Synthetic genetic circuit design and metabolic burden

Metabolic burden or load can refer to the amount of host cell resources, in the form of amino acids, energy such as ATP or GTP and genetic machinery that is needed to maintain and

express recombinant DNA. Metabolic limitations impact all synthetic circuits used for recombinant protein synthesis. Modifications to the host genome and the genetic components used to generate synthetic expression vectors, introduce tensions between the host metabolism and the synthesis of the recombinant protein. The redirection of host cell resources toward recombinant processes impedes the host metabolism and slows growth. This reduction in growth rate arises from the circuit monopolizing host resources, which slows production of essential proteins and RNAs (Brophy and Voigt., 2014). Recombinant proteins can also cause cellular health declines through varying levels of toxicity, which elicits cellular stress responses that help return the host to a state of functionality and viability, but the additional energy requirement this generates can cause further reductions in biomass growth (Carneiro, Ferreira and Rocha., 2013). Moreover, in some cases recombinant proteins can directly influence host cell metabolism through their innate enzymatic activity (Sørensen and Mortensen., 2004). Most synthetic circuits have mechanism to limit gene expression, in order to time enzymatic activity and protein production to be more active during the later stages of *E. coli* growth (Brophy and Voigt., 2014) because the maximum yield of recombinant protein will be affected if biomass accumulation is stalled. Furthermore, reductions in growth rate can also cause circuits to fail by reducing the dilution rate of circuit components. This can lead to a heavy build-up of proteins or RNAs (Brophy and Voigt., 2014), which reduces protein formation via molecular crowding. As these factors manifest the synthetic circuitry becomes highly disadvantageous to the host and in wild type prokaryotic cells the acquired plasmid would be expelled, which is ultimately why selection markers are required to ensure plasmids are retained. For this reason, recombinant gene expression in *E. coli* usually utilises antibiotic resistance genes. Selection markers also introduce a metabolic burden to the cell by requiring transcription and translation of additional plasmid-encoded genes (Baneyx, 1999). Furthermore, selection markers cannot prevent burden-driven mutations from arising on the synthetic plasmid, meaning that the plasmid is retained, but the precisely engineered genetic circuit could become altered. If these changes sufficiently alleviate the burden on the host, these changes can quickly become propagated in the culture.

1.4 Optogenetics for gene expression

Optogenetics is a tool in cell biology which implements the use of genetically encoded, light-gated proteins (i.e., photoreceptors) to influence and control cellular gene expression and organismal behaviour in a spatiotemporally coordinated manner (Möglich and Moffat., 2010). For many years, the main use of optogenetics was the control of ion channel proteins in mammalian neurobiology (Weitzman and Hahn., 2014) and photopharmocological uses in human medicine (Bamberg, Gartner, and Trauner., 2018). However, recently optogenetic

applications have been expanded to recombinant gene expression. Photoreceptors resemble fluorescent reporter proteins, such as GFP, as they are genetically encoded, non-invasive, and applicable to intact cells and organisms. However, they are explicitly intended to modulate activity, whereas fluorescent proteins generally do not disturb the genetic processes under study. Optogenetic switches are available with excitation wavelengths that span the entire light spectrum. Known photoreceptors fall into the following categories (excitation wavelengths are given): UV receptors (300-400nm), cyanobacteriochromes (400-500nm), BLUF domains (450nm), LOV domains (450nm), cryptochromes (450nm), cobalamin-binding domains (545nm) and finally phytochromes (650-780nm) (Weitzman and Hahn., 2014). This large number of photoreceptors with diverse excitation wavelengths makes it possible to manipulate the expression of multiple genes simultaneously. Optogenetic tools for recombinant gene expression provide precise control, as applied light intensity can be regulated or applied in periodic pulses to tune protein expression. Optogenetic gene expression has the added advantage of full reversibility, as light can be removed instantly (Ohlendorf et al., 2012), unlike the need for complex processes to remove chemical inducers from the media when using small molecule sensors. Optogenetic tools have been applied to metabolic engineering, for instance a study in 2018 fused the transcription-factor-recruitment-protein VP16 to the blue-light sensitive EL222 protein to achieve light-controlled transcription in *Saccharomyces cerevisiae* and enhance the biosynthesis of valuable chemical fermentation products isobutanol and 2-methyl-1-butanol via control of the mitochondrial isobutanol pathway, aerobically (Zhao et al., 2018). Another study used EL222 to create the optogenetic CRISPRi platform for improved chemical production of muconic acid in *E. coli*, achieving increases in production by 130% (Wu et al., 2021). A study in 2016 combined optogenetic gene expression control, using the sensor histidine kinase CcaS, with computational software to design an automated novel experimental feed-back platform containing turbidity, autosampling and light-delivery modules to increase the repeatability and robustness of optical expression control in *E. coli* (Milias-Argeitis., et al. 2016). Optogenetic tools have also be used for light-based regulation of the *lac* operon, using the two-component photoactive protein YF1/FIXJ, to control gene expression from a number of IPTG-inducible promoters using only blue-light (Lalwani et al., 2021). This study illustrated the unique capabilities of reversible and tune-able optogenetic control over chemical induction systems, improving product production by over 25%, compared to IPTG induction, and demonstrating how optogenetics can be scaled to at least two-litre bioreactors. A study using both the Ccas-CcaR and the Cph8-OmpR photoreceptors to create a light-switchable two-component system demonstrated how optogenetic gene expression can be used for localised protein production and to reveal key information on organismal-level processes and protein interactions (Olson and Tabor., 2014). A similar

study used infra-red light and the BphP1-PpsR2 phytochrome, for localised optogenetic gene expression to express target proteins in specific cellular compartments and also to control cell morphology (Kaberniuk, Shemetov and Verkusha., 2016). Optogenetic bacterial gene regulation was recently expanded to the Gram-positive *Bacterium subtilis* for the first time. Using CcaS to study genetic regulatory process in the host helped to make the complex spatial and temporal gene expression signals of *B. subtilis* easier to engineer in the future (Castillo-Hair et al., 2019). This demonstrates how optogenetics can be used to aid the development of new bacterial host strains.

Despite the considerable number of available optogenetic tools and vectors, light-inducible gene expression is yet to be implemented broadly in industry or research applications. Small-molecule-induced (SMI) gene expression prevails as the most popular choice for recombinant protein production, for instance the pET vector series uses IPTG induced T7RNAP gene transcription to activate transcription of a gene of interest (GOI) via the T7 promoter. Even at low concentrations IPTG is a potent inducer. If tuneable protein production is desired, alternative carbon source SMI systems such as ParaBAD, induced by arabinose (Sørensen and Moresnen., 2004) are available. A review investigating the approach taken to create synthetic circuits for GOI production, found over 10 different commonly utilised combinatorial options to create expression circuits which select for strong or weak induction, tuneable or constitutive expression and chemical or thermal induction (Rosano, Morales and Ceccarelli., 2019). However, optogenetic promoters were not considered popular options.

Leaky expression is a common concern for all synthetic gene expression vectors. Leaky expression means there is some basal level of GOI expression seen prior to induction. High levels of leaky expression reduce host growth rate and can lead to plasmid instability or cell death. The commonly used SMI gene expression systems boast low levels of basal expression and a high level of predictability, furthermore their performance has been characterised across several host strains and in a multitude of growth mediums, as well as improvements made in their genetic design (Greco et al., 2022). This affords researchers a good level of confidence when selecting a SMI expression platform for a project. In comparison, there has not been much peer-reviewed comparative research into the basal expression levels and robustness of optogenetic expression vectors. Furthermore, optogenetic vectors are often regarded as having inherently high leaky expression levels due to dark-state dimerization of photoreceptors.

1.5 The optogenetic gene expression vectors pDusk and pDawn

The pDusk and pDawn vectors operate as light repressed and light induced gene expression systems and were designed by Robert Ohlendorf at the Humboldt University, Berlin in 2012 (Ohlendorf et al., 2012), using the two-component system (TCS) YF1/FixJ. Ohlendorf describes the state of induction as the signal polarity of the expression systems: for pDusk, the expression vector is in the active state in the absence of blue-light, for pDawn the signal polarity is inverted, and the expression vector is active in blue-light and inactive in the absence of light. Like most optogenetic expression systems, pDusk and pDawn employ light-gated photoreceptive proteins to influence and control gene expression in a spatiotemporal manner. At the time of engineering pDusk and pDawn the sensor histidine kinase YF1 was the most thoroughly studied fusion protein (Moglich and Moffat., 2010) and was generated via fusion of the *Bacillus subtilis* YtvA LOV photosensor domain to the histidine kinase module of FixL from the *Bradyrizobium japonicum* (Berntsson et al., 2017). In plants, fungi, and bacteria the light, oxygen, or voltage (LOV) family of blue-light photoreceptor proteins are naturally occurring photoreceptors responsible for the activation of light-sensitive signal transduction cascades that allow organisms to sense and respond to light in their extracellular environment (Crosson, Rajagobal and Moffat., 2002). Located at the core of the LOV domain is a conserved cysteine residue that extends from a flavin mononucleotide chromophore. In the dark, YF1 is in a kinase-active state. Excitation of the flavin mononucleotide chromophore by 450 nm blue-light creates an excited singlet state, which rapidly decays into a triplet state and then the signalling state is generated by a thioether bond which forms between the flavin mononucleotide chromophore and the cysteine residue; formation of this bond converts YF1 into a phosphatase-active state (Berntsson et al., 2017). In 2017, five years since the development of pDusk and pDawn, a study using nano- to millisecond time-resolved X-ray scattering identified the structural changes that occur when YF1 is phosphatase-activated by blue light, showing how the LOV domains splay apart resulting in the left-hand rotation of the kinase module (Berntsson et al. 2017). Solving these structural questions is important to continue the advancement of photoreceptor designs and provide a template for the synthesis of new signal transduction pathways which can be utilised in novel synthetic light-regulated vectors. In the absence of blue-light the histidine kinase-active YF1 phosphorylates its cognate response regulator FixJ, which then activates the FixK2 gene, driving robust target gene expression from the FixK2 promoter (Mesa et al., 2008). Upon light absorption, net kinase activity of YF1 and consequently gene expression is greatly reduced. In pDusk and pDawn, YF1 and FixJ are constitutively expressed from the $LacI^{q}$ promoter as a bicistronic operon. Target genes are introduced in the multiple cloning site (MCS) downstream of the FixK2 promoter.

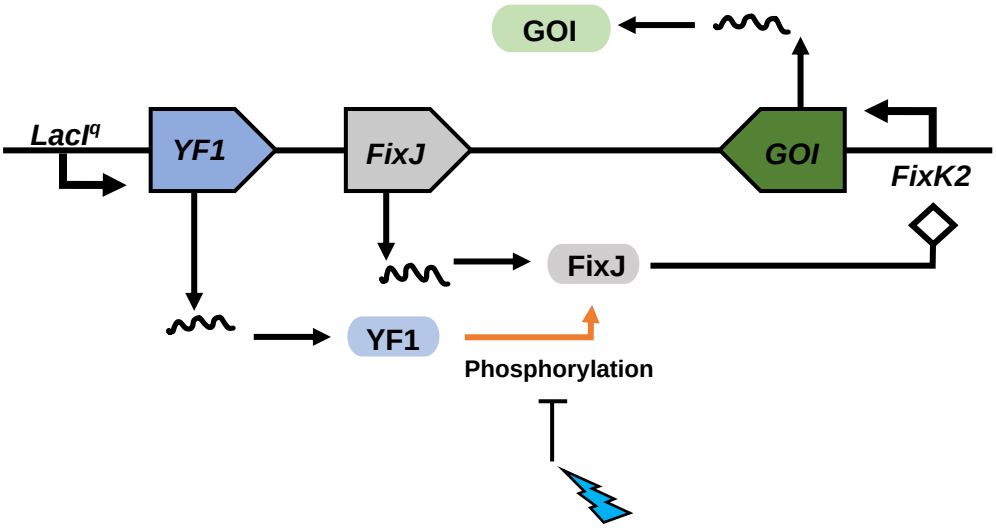


Figure 1.1 Circuit anatomy of pDusk. In the absence of blue light the TCS YF1/FixJ drives GOI expression from the FixK2 promoter. Absorption of 450 nm blue light results in excitation of the YF1 photoreceptor which greatly reduces net kinase activity of YF1 and GOI expression is repressed.

One of the goals Ohlendorf wished to achieve was to construct an optogenetic system that induced gene expression upon blue light absorption rather than repress gene expression (i.e., to reverse the signal polarity of the TCS YF1/FixJ). To do this Ohlendorf used a gene inversion cassette. In pDawn, the TCS YF1/FixJ drives expression of the λ phage repressor cI from the FixK2 promoter which in turn represses expression from the strong λ promoter pR, downstream from which is the MCS and loci of the target gene. In other words, upon blue-light absorption YF1 does not phosphorylate FixJ, thus FixJ does not activate the FixK2 gene and expression of the cI repressor from the FixK2 promoter is not driven. Without cI repressor expression, the promoter pR is not repressed, so target gene expression occurs (Figure 1.2)

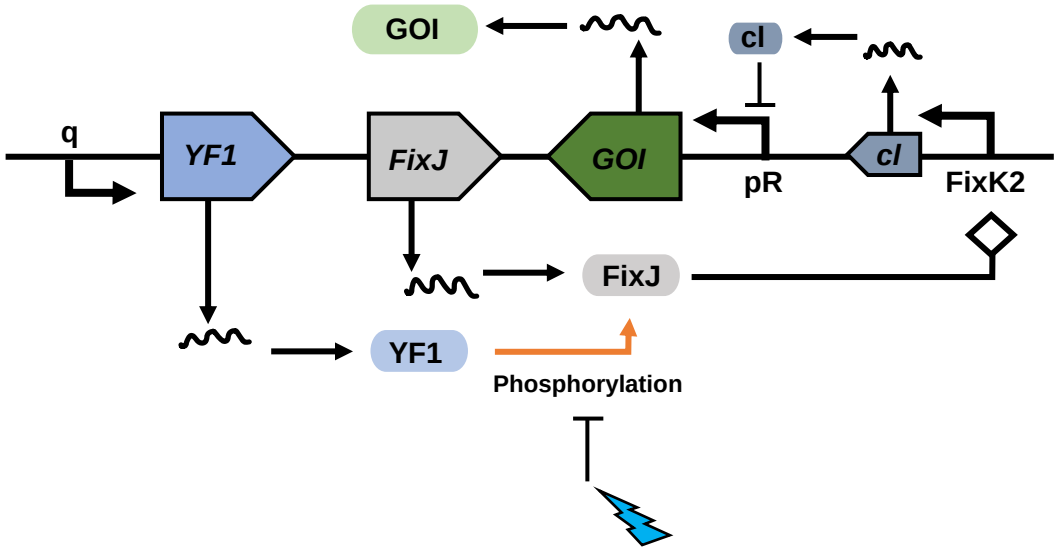


Figure 1.2 Circuit anatomy of pDawn. In the absence of blue light the TCS YF1/FixJ drives cl repressor expression from the FixK2 promoter and cl represses GOI expression downstream of the pR promoter. Absorption of 450 nm blue light results in excitation of the YF1 photoreceptor which greatly reduces net kinase activity of YF1 and cl expression is repressed. Without cl repression, GOI expression occurs downstream of the pR promoter.

1.6 The optogenetic expression system based on OptoT7RNAP

The OptoT7RNAP blue-light inducible expression variants were developed by Armin Baumschlager and co-workers in 2017, at the ETH Zurich institute Switzerland. Unlike pDusk and pDawn, OptoT7RNAP is a two-vector light inducible gene expression system which utilises T7RNAP driven expression. The first vector harbours the gene for T7RNAP, under the control of the *ara*B promoter. To generate a photoactivable T7RNAP, the gene sequence was split into two fragments and a short GGSGG linker fuses each gene fragment to the gene sequence of photoactivable dimerization domains nMag and pMag. The N-terminal fragment was fused to nMag, and the C-terminal fragment was fused to pMag. When arabinose permeates into the cell and interacts with the *araB* promoter transcription of the photoactive T7RNAP fragments is induced. These gene fragments are produced as two separate heterologous proteins. Upon blue light absorption the nMag and pMag domains dimerise, restoring T7RNAP functionality. Following this, on a second plasmid the now functional T7RNAP interacts with its cognate T7 promoter, inducing downstream GOI expression (Figure 1.3). The systematic construction process of the OptoT7RNAP system generated multiple variants. By following a bottom-up approach to engineering and characterisation of each variant generated, the performance of the OptoT7RNAP system was rapidly optimised by Baumschlager. The areas of optimisation focused on were the T7RNAP gene split site, expression levels of each T7RNAP fragment, the reverting time of the photoactivable components and the toxicity of the T7RNAP gene.

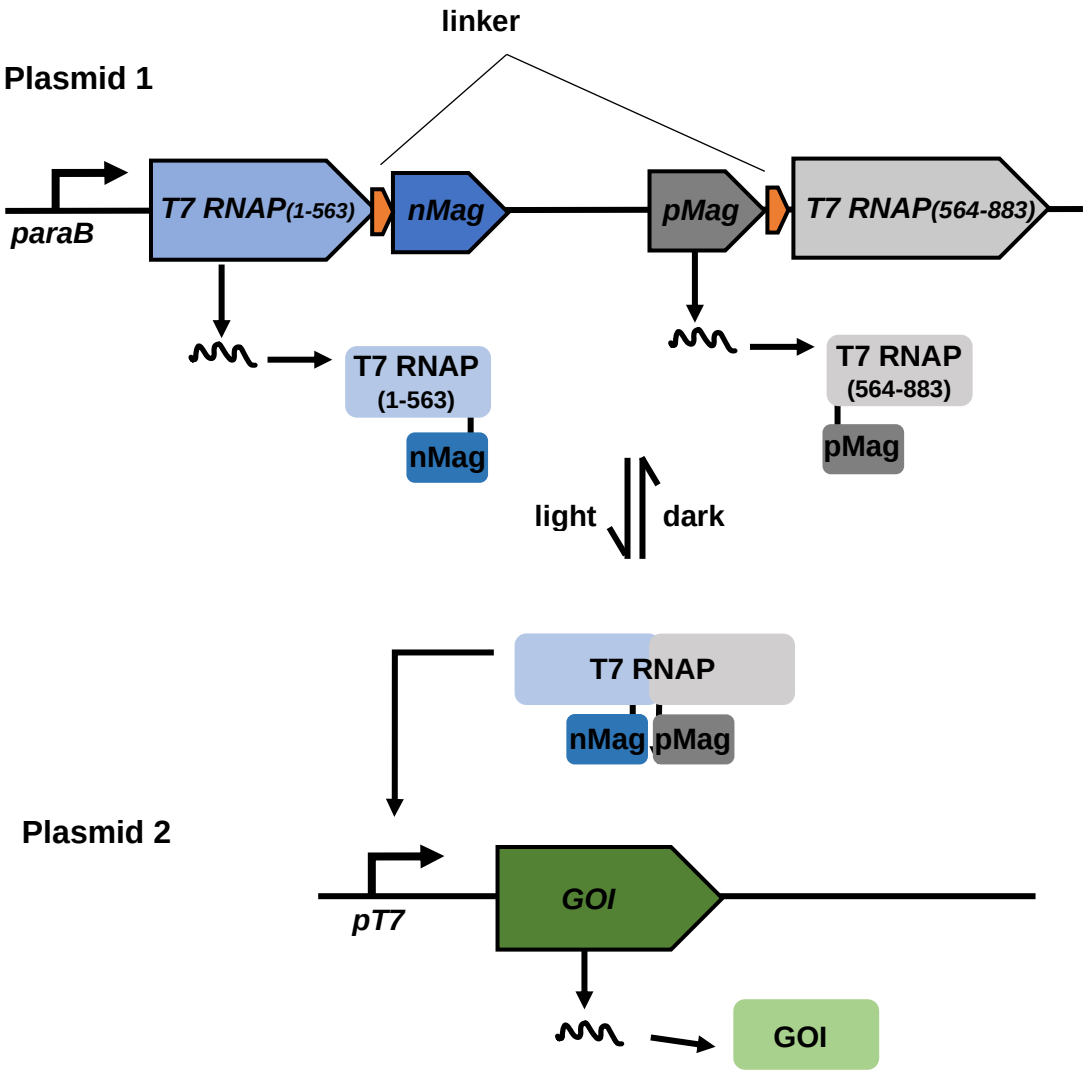


Figure 1.3 Circuit anatomy of the OptoT7RNAP vectors. Upon the absorption of 450 nm blue light the photoactivable domains nMag and pMag dimerise, restoring T7RNAP functionality. The now functional T7RNAP enzyme initiates transcription down-stream of its cognate T7 promoter resulting in GOI expression.

A study by Rui Sousa, in 2004, using X-ray crystallography, resolved the three-dimensional structure of T7RNAP. The enzyme is highly α-helical and is comprised of two domains, the N-terminal domain, amino acids 1-312, and the C-terminal domain, amino acids 313-883. The C-terminal domain is sub-divided into thumb (330-410), palm (411-448, 532-540, 788-838) and finger (541-737, 771-778) subdomains. Nucleic acids bind within a large cleft in this structure (Westover, Bushnell and Kornberg., 2004). A study, in 2014, exploring the structure-function relationships of the T7RNAP found that the protein could be split into four co-expressed fragments that retain protein functionality. By adjusting the concentration of the core fragments the transcriptional capacity of the synthetic system can be regulated

(Segall-Shaprio et al., 2014). Baumschlager tested the functionality of T7RNAP at the following amino acid split sites: 69, 179, 302, 563 and 600. Mean mCherry fluorescence was highest at split site 69, followed by 179, 600, 563 and finally 302. However, the fold-induction in mCherry fluorescence intensities between basal and induced states was highest at split site 563 with a 332-fold-induction and the second best site was 302 with a 146 fold-induction (Baumschlager et al., 2017). Split sites 69 and 179 had the highest induced mean mCherry fluorescence, but also the highest basal mCherry fluorescence. Split site 302 displayed the second lowest basal level of mCherry fluorescence but also had the lowest induced mCherry fluorescence. Split sites 69, 179 and 302 are in the N-terminal domain which is involved in the beginning of RNA binding (Sousa, Mukherjee and Kivie., 2003), thus splitting the T7RNAP gene here may impact sequence specific binding of the enzyme to the promoter and in the separating of the two DNA strands of the promoter during initiation, to expose one DNA template strand for RNA synthesis. At the split site 563, the T7RNAP retained high functionality with very low levels of basal expression. The amino acid site 563 is located in the C-terminal finger subdomains and although the C-terminal domain contains the active site where RNA synthesis occurs (Sousa, Mukherjee and Kivie., 2003), site 563 is located in a surface-exposed flexible loop and is not close to the DNA recognition site (Cheetham and Steitz., 1999).

The well-characterised T7RNAP demonstrates stringent promoter specificity and robust transcriptional activity. Perhaps due to its bacteriophage origins (bacteriophages have high replication rates and rapidly overwhelm the infected cell) T7RNAP is a high processivity enzyme which transcribes the GOI at a high level and this results in the overproduction of the target protein (Sousa and Zaher., 2021). The high productivity of systems utilising T7RNAP driven expression is likely the motivating feature behind its frequent use in recombinant gene expression studies. However, in the commonly used T7RNAP based expression systems the gene encoding T7RNAP is stably integrated into the genome of the host *E. coli* cell, under the control of an inducible promoter (Sousa, Mukherjee and Kivie., 2003). For instance, the popular protein expression strain BL21 (DE3) harbours the T7RNAP gene within its own genome and the gene of interest, under the control of a T7 promoter, is introduced to the host cell via a plasmid. High target protein titres are also thought to arise from the independence of the T7RNAP from the biology of the host cell because the expression of the target gene is not limited by the need to share transcription machinery with thousands of host genes (Durbin, 1999). This de-coupling of GOI expression from cellular RNA polymerase concentrations should improve the consistency of target protein titres. Host RNA polymerase concentrations can vary largely between cells and at different growth rates. Furthermore, another benefit of using recombinant transcription machinery is that the activity

and concentration of T7RNAP can be regulated by the synthetic circuit design (Durbin, 1999). In the case of the OptoT7RNAP's, the photoreactive domains are under the control of the inducible promoters *ara*B, and so the maximal expression of the T7RNAP and therefore the GOI can be regulated to desired values by regulating the arabinose concentration in the growth medium (Baumschlager et al., 2017).

In addition to arabinose induced titratable gene expression, gene expression variability was also reduced by using single cell analysis to optimize the levels of the two T7RNAP fragment domains split at amino acid site 563. It was shown that when the C-terminal fragment was more highly expressed (1.9-fold expression level), induced expression of reporter protein was significantly increased and a decrease in protein concentration variability was observed, compared to when the C-terminal and N-terminal fragments were expressed in equal concentrations (1:1 expression level). It is likely that when the fragments are equally expressed the protein titre of functional T7RNAP is dependent on the level of expression of both fragments, however if one fragment is over expressed the variability in protein titre of functional T7RNAP is only dependent on the variability in the level of expression of one fragment (Baumschlager et al., 2017). To create fast reverting T7RNAP fragments, mutations were introduced into the gene sequence of the Magnets which decreased the dissociation time of the T7RNAP fragments. The fast-reverting mutation 185V generating pMagFast1 and the mutations 185V and 174V generating pMagFast2. These mutations were not discovered by Baumschlager but by the research group who used directed evolution to create the optogenetic Magnets from the blue-light fungal photoreceptor Vivid (Kawano et al., 2015). Similarly, the mutation R632S was introduced into the T7RNAP which reduces the toxicity of the enzyme without reducing processivity. This mutation was first characterised by Karsten Temme in 2012 during a study to construct T7RNAP variants with mutations that reduce the toxicity of the polymerase enzyme (Temme et al., 2012). Baumschlager was able to rapidly develop the OptoT7RNAP expression system using a novel idea (i.e., a light inducible T7RNAP) and by incorporating existing and well-characterised mutations. This demonstrates how open-source synthetic parts and research can be used as a genetic toolbox, and with high-throughput techniques, can be quickly assembled in a combinatorial fashion and characterised.

1.7 Comparing the design of pDusk/pDawn and OptoT7RNAP

The pDusk and pDawn expression systems use the light gated TCS YF1/FixJ to activate gene repression cascades and the OptoT7RNAP vectors use heterodimeric "Magnet" photoreceptors to restore T7 RNA polymerase protein functionality which subsequently activates gene transcription. The researchers behind the development of the optogenetic

platforms pDusk/pDawn and OptoT7RNAP had similar aims in mind. Both research groups wanted to expand the applicability of using light as an inducer in research and create an expression platform that could work well during industrial scale-up. Both research groups recognised the advantages of fully reversible optogenetic induction in dynamic gene expression studies over traditional chemical induction systems, which are not easily reversible. Ohlendorf described the full reversibility of light controlled heterologous gene expression as unprecedented. Baumschlager suggested that optogenetic dynamic regulation would be ideal for metabolic engineering studies and characterised dark-state reversal of gene expression by the OptoT7RNAP vectors. Finally, both groups wanted to address the limitations seen frequently in existing light inducible expression systems that had impeded their widespread use. For example: dependence on supplying non-native chromophores exogenously or incorporation of cofactor synthesis genes, a decrease (rather than increase) in gene expression by light absorption, high background activity and a low dynamic range. An obvious difference, however, is the number of vectors in the two expression systems. Ohlendorf, state that a platform with multiple plasmid components has limited portability, specifically they state that other applications of the YF1/FIXJ photoreceptive light gated protein switch had been hindered by its “implementation on two separate and relatively large plasmids”, in addition they state that expression driven from the strong inducible T7-lacO promoter further hindered the portability of the design. The reasoning behind this statement was not expressed but the T7RNAP enzyme is intrinsically toxic and expression systems using this polymerase have been shown to burden the host (Angius et al., 2018). It is not unreasonable to surmise that, coupled with using two large plasmids, this burden could severely impede host growth and reduce portability by limiting the number of host strains which could efficiently cope with this high burden. The pDusk and pDawn vectors are one-plasmid portable expression platforms that do not require T7RNAP to drive gene expression or SMI to induce or suppress gene transcription. This is in stark contrast to the approach taken by Baumschlager who designed and implemented a light inducible T7RNAP and T7 promoter driven gene expression system on two plasmids. By utilising T7RNAP, heterologous gene expression is decoupled from cellular RNA polymerase concentrations. Host RNA polymerase concentrations can vary depending on the growth phase, nutrient conditions, and other extrinsic factors (Bremer and Dennis., 1987), which hampers the consistency of heterologous protein titres and may lead to higher levels of background activity. Furthermore, exclusive heterologous gene expression can be achieved via inhibition of the hosts native transcription mechanisms (i.e., via the addition of the antibiotic rifampicin to the culture medium), which does not affect T7RNAP activity (Tabor, 2001).

1.8 High-throughput *E. coli* growth and gene expression studies

Microplate reader incubation has revolutionised bacterial gene expression studies by using high-throughput growth evaluation to determine maximal growth rate, population density and synthetic vector driven gene expression of up to 96 cultures simultaneously. Analysing bacterial growth in this way provides crucial information on growth-dependent coordination of gene functions and cellular components (Kurokawa and Ying., 2017). Traditionally, colony-forming unit (CFU) assays were used to estimate the number of cells in a culture. This is time and resource intensive compared to optical density evaluation of bacterial growth, as the optical density of bacterial cultures is estimated directly in microplates without the need for laborious sample dilution, plating cells on agar plates and counting of colonies by eye. Furthermore, CFU assays cannot provide temporal records of growth changes (Kurokawa and Ying., 2017). For high-throughput analysis of microbial growth, the most routine method for measuring bacterial proliferation over time is using optical density measurements at 600 nm (OD_{600}) (Stevenson et al., 2016) spectrophotometrically. Optical density measurements have many applications, for example: investigating antibiotic efficacy, growth under different nutritional or stress environments and characterisation of different bacterial strains and recombinant cells transformed with synthetic circuitry (Stevenson et al., 2016). OD measures the degree of light scattering caused by microbial cells within the culture. For small populations microbial cells are well dispersed within the solution and a single scattering regime ensues, but for cultures with a higher population a multiple scattering regime occurs as the likelihood of light being scattered is greater and the probability of incident light being scattered by the cells multiple times increases. So, optical density is only an effective method of estimation for microbial cultures of low cell densities (Stevenson et al., 2016).

Fluorescent reporter proteins

Fluorescent proteins are highly useful for monitoring various aspects of recombinant protein production without the need for invasive, time-intensive and expensive protein purification and quantification techniques. Non-fluorescent protein quantification methods require processing of a crude extract containing all cell proteins, macromolecules, and nutrients, into a purified target protein isolate. Aside from being time and resource intensive, quantification inaccuracies can result from target protein losses during crude extract processing. Furthermore, time-intensive troubleshooting and development is often necessary as host-characteristic, such as the level of oxidisation at the site of protein production, and protein-specific factors such as size, structure and solubility can affect the efficiency of protein purification and quantification (Hui et al., 2018). Therefore, fluorescent reporter proteins are an ideal choice for fast and efficient high-throughput analysis of gene expression. *In vivo* monitoring of fluorescence intensity allows for real-time modelling of dynamic changes in fluorescent protein expression. This is ideal for light-inducible expression systems as a key

advantage of using light as an induction signal is that induction is, in theory, fully reversible and can be dynamically controlled.

Protein databases offer an array of fluorescent proteins with a wide range of spectral properties. Many of the limitations associated with wild-type fluorescent proteins have been circumvented with the development of synthetic variant counterparts (Su, 2005). Thus, a large repository of diverse reporter proteins is available to suit the specific requirements of a study, allowing for rapid monitoring, characterisation, and optimisation of synthetic circuitry. For the pDawn and pDusk expression systems the super-folder green fluorescent protein (sfGFP) was used, which is a mutant of the wild-type green fluorescent protein (GFP) native to *Aequorea victoria*. Compared to the wild-type GFP, the super-folding variant demonstrates faster and more robust folding as well as increased stability at higher temperatures (Frenzel et al., 2018) and in oxidising environments, such as the bacterial periplasm (Aronson, Costantini and Snapp., 2011) where disulphide bond formation can impair protein folding. Furthermore, sfGFP has been designed for *in vivo* high-throughput screening of protein expression levels (Cotlet et al., 2006). For the OptoT7RNAP expression systems the fluorescent reporter protein mCherry was used, which is a red fluorescent protein derived from DsRed and found in *Discosoma* sea anemones. The wild-type protein DsRed has a large tetrametric structure and a slow maturation time. The protein also has a low photostability (Shaner et al., 2004) thus would not be suitable for screening a light-inducible system as the protein could be susceptible to change when exposed to light. DsRed and these issues affecting its practicality are a good example of the types of limitations wild-type proteins can introduce to a synthetic circuit and the motivation driving development of improved fluorescent protein variants. Descendants of DsRed have been developed through directed evolution. For example, the mRFP1 monomeric variant which has a lower molecular weight and faster folding time but compared to DsRed has a decreased photostability. Further development of mRFP1 yielded the monomer mCherry (Shaner et al., 2004) which boasts a long wavelength, high photostability and pH stability, and a fast maturation time. Some fluorescent proteins have a functional requirement for dimer or, in the case of DsRed, tetramer formation which can lead to aggregation of protein in the cell (Hawley, Hawley and Telford., 2017). This tendency to form protein aggregates can limit the application of fluorescent proteins due to cytotoxic affects (Robin et al., 2018), highlighting why the monomeric fluorescent protein family mFruits (of which mCherry is a member) was developed. In contrast the oligomerization state of sfGFP is reported to be a rapidly maturing weak dimer. Both sfGFP and mCherry are widely used reporter proteins and while a more direct comparison of basal expression by the light inducible expression systems may have been achieved using only one fluorescent protein, both proteins were

chosen as they have minimal spectral overlap (Robin et al., 2018). This would have been important for another aim of the project, which was to attempt expression of both the pDusk and OptoT7RNAP expression systems simultaneously, inducing pDusk and supressing OptoT7RNAP in darkness and suppressing pDusk and inducing OptoT7RNAP in blue light. Therefore, basal expression of the pDawn, pDusk and OptoT7RNAP systems was quantified using sfGFP and mCherry, respectively. Furthermore, comparisons can be drawn regarding the ease of fluorescent protein intensity quantification, useful for validating whether high-throughput analysis can be used to screen multiple expression systems with multiple reporter proteins simultaneously and with ease.

1.9 Metabolic burden and bacterial growth rate

Bacteria cells possess a finite quantity of resources which need to be allocated effectively to maintain metabolic equilibrium, heterologous DNA constructs monopolize these resources and disrupt the metabolic balance of resource allocation (Borkowski et al., 2016). Early recombinant DNA studies found that metabolic burden increased almost linearly with increasing foreign DNA length (i.e., plasmid size) (Bentley et al., 1990). In addition, plasmid copy number increases metabolic load as higher copy numbers require a greater amount of metabolic energy to maintain the plasmid DNA (Seo and Bailey., 1985). The consequence of this metabolic burden on recombinant bacterial cells is well-documented and can be observed as: reduced bacterial growth rate, changes in cell morphology, host driven evolution of the recombinant circuity and plasmid expulsion. For example, host cells harbouring plasmids with high copy numbers have decreased relative specific growth rates and reduced concentration of foreign gene expression per plasmid (Glick, 1995). Furthermore, this reduction in specific growth rate is independent of media composition (Bentley et al. 1990). Hence, for maximising the productivity of recombinant protein production, low copy number plasmids can often perform as well as or better than high copy number plasmids in bacterial hosts (Jones, Kim and Keasling., 2000). The cellular changes metabolic burden elicit in a host are complex and hard to model. Stress responses can dramatically alter the physiology and biochemistry of the host (Glick, 1995) and these cellular-level changes cannot be easily identified or characterised, which makes burdensome expression of recombinant proteins unpredictable. Since the 1990's huge advancements have been made in metabolic engineering of bacteria, however the predictability and robustness of gene expression circuits remain major challenges for synthetic biology. Recent studies have identified key transcriptional changes in *E. coli* which happen during expression of foreign protein, this has expanded our understanding of the stress responses which occur to help restore metabolic equilibrium. For example, a study using RNA-seq and *in vivo* assays found that *E. coli* promoters involved in the heat-shock stress response activated

gene expression in response to all of the heterologous DNA expression constructs tested (Ceroni et al., 2018). Characterising metabolic burden at the cellular-physiological-level by identifying the endogenous stress-linked response pathways enables the development of self-regulating feedback loops which tune gene expression via recognition of stress stimuli. This is achieved through integration of recombinant circuitry with stress response systems (Dahl et al., 2013; Ceroni et al., 2018). Other novel *in vivo* tools to reduce the impact of synthetic constructs on host physiology have been developed, for example using highly fragmented T7 RNA polymerase to create a resource allocator for transcription which helps maintain metabolic equilibrium (Segall-Shaprio et al., 2014), ribozyme-based insulators which, when incorporated into the synthetic constructs, help to buffer the gene circuit and reduce context dependence (Lou et al., 2012). Development of antibiotic-free expression systems via plasmid addiction helps to reduce the metabolic burden related to plasmid maintenance (Pasini et al., 2016) and the development of a suppressor tRNA-mediated feedforward loop was shown to eliminate leaky heterologous gene expression in *E. coli* (Ho et al., 2021). The leakiness of a promoter not only introduces burden via resource competition of transcriptional and translational machinery, which can be especially detrimental during the early periods of population growth, but also effects the stability of the synthetic circuit and cell-to-cell heterogeneity. Furthermore, promoter leakiness can be increased by context-dependent interactions (Gyorgy, 2021). Context dependence is an issue in synthetic genetic circuits where the behaviour of a genetic component is influenced by the adjacent or nearby genetic components (Clifton et al., 2018). Thus, for these reasons, leaky gene expression contributes significantly to metabolic burden and the unpredictability of synthetic expression systems.

Time constraints and the high-throughput nature of this study meant that experimental characterisation of metabolic burden introduced by the pDawn and OptoT7RNAP expression platforms via the identification of physiological and biochemical changes at a cellular-level could not be performed. Instead, high-throughput methods to characterise growth properties were utilised. Quantitative study of cell growth provides key insights into the state of metabolic stress in recombinant bacterial populations, for instance a consistent growth rate represents a physiological steady state of an organism (Wang et al., 2010). Non-linear regression analysis can be used to infer various population-based statistics by tracking bacteria populations over time and modelling the sigmoidal curves produced, this can be used to identify changes in the lag, exponential and stationary phases of bacterial growth (Figure 1.4). Hypothetically there is a clear disparity in the metabolic burden introduced to the host via the expression systems pDusk, pDawn and the OptoT7RNAP variants. The pDusk and pDawn plasmids are one plasmid expression platforms with only one selection

marker gene, *kmr*, conveying kanamycin resistance. In contrary the OptoT7RNAP is a two-plasmid expression platform, requiring two selection marker genes conveying chloramphenicol and ampicillin resistance, respectively. The total length of foreign DNA introduced by the pDusk and pDawn systems is 6,220bp and 7,213bp, respectively. For the OptoT7RNAP system the total foreign DNA length of the variants is almost double at ~ 13,156bp (T7RNAP harbouring plasmid: 7,345bp, mCherry harbouring plasmid: 5,811bp). Furthermore, the modified pET vector backbone on which the pDawn/pDusk system is based, containing the reporter protein gene sfGFP, has a medium copy number of ~20-25 per cell. Whereas the OptoT7RNAP reporter plasmid pETM6, containing the gene for mCherry, has a high copy number (Xu et al., 2012). Plus, the pSC101 plasmid for light-sensitive split T7RAP expression, harbours the *RepA* mutation (wild-type pSC101 ~ 5 copies per cell), which increases both plasmid copy number and copy number range of ~31 to ~113 copies per cell (Thompson et al., 2018), which is likely to cause high-levels of cell-to-cell heterogeneity and impede predictability. In addition, the OptoT7RNAPs have the added burden of T7 RNA polymerase driven gene expression and the high processivity of this enzyme has been shown to have toxic effects to the host cell. However, low sfGFP basal expression by pDawn relies on the transcription of four regulatory proteins and the host ability to maintain sufficient stocks of these proteins to repress sfGFP gene expression. In the OptoT7RNAP system low mCherry basal expression is only reliant on sufficient repression of the *araB* promoter.

By using mathematical modelling, the behaviour of host cells expressing optogenetic plasmids can be estimated. However, this does rely on the assumption that the responses of the recombinant cells to environmental factors are reproducible (Pla et al., 2015). A wide range of nonlinear regression models are used in predictive microbiology, focusing on a number of different growth parameters and statistical criteria, thus there is disagreement in the literature over which is the best-fitting model. However, the Gompertz or logistic are commonly used models for analysing bacterial growth (Ratowsky., et al. 1993; Pla et al., 2015).

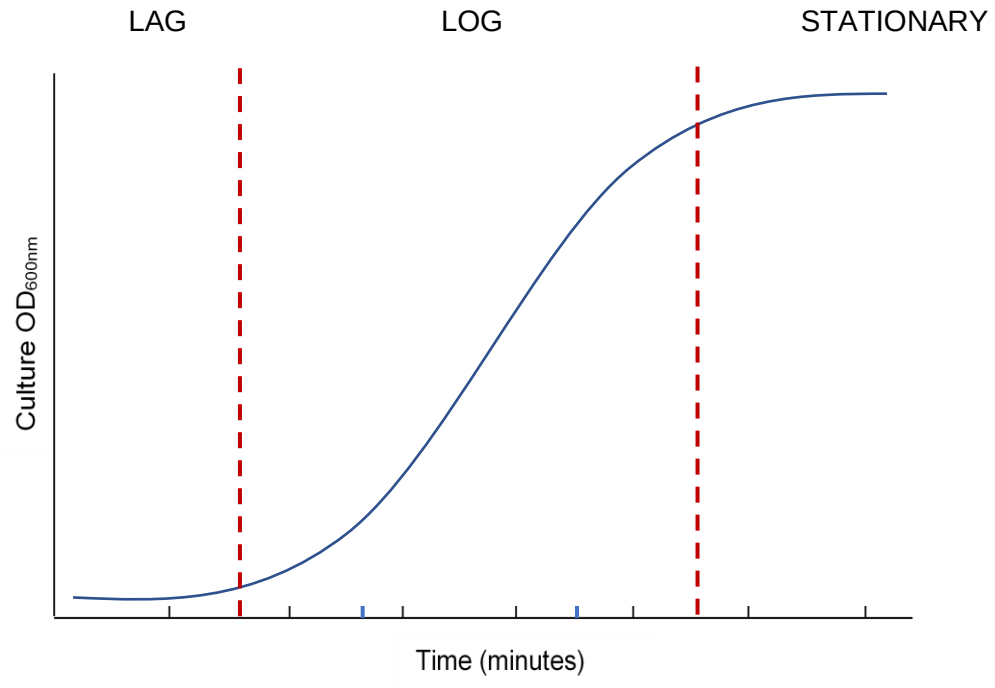


Figure 1.4 Bacterial growth curves have distinct sigmoidal functions.

1.10 Motivation and aims

The optogenetic gene expression platforms pDusk/pDawn and OptoT7RNAP differ in their approach to regulating GOI expression. I set out to compare the basal performance of pDusk/pDawn and OptoT7RNAP, in multiple *E. coli* strains, and identify which genetic circuit had the lowest level of leaky GOI expression. I also wanted to investigate the metabolic burden imposed on the host by the expression of plasmid DNA. I did this using automated non-linear regression analysis of bacterial growth curves obtained from absorbance data at 600 nm. High-throughput screening of multiple synthetic circuits simultaneously was achieved by using a plate reader to measure reporter protein fluorescence and OD_{600}. Thus, I was able to rigorously compare the optogenetic gene expression vectors. Using these comparisons, I aimed to contribute to the understanding of plasmid-host interactions and provide some insight for the development of multichromatic optogenetic vectors for the expression of multiple genes simultaneously.

1.11 Structure of this thesis

Chapter 2 will cover the methods and materials used to conduct the high-throughput bacterial growth and gene expression experiments. Chapter 3 will report the findings of basal performance of the pDusk/pDawn and OptoT7RNAP expression vectors, as well as discuss genetic circuit features that help to reduce basal gene expression. Chapter 4 will report the

result of non-linear regression modelling of bacterial growth properties and discuss how the expression of the OptoT7RNAP plasmid DNA causes metabolic burden. Finally, Chapter 5 will outline the key findings of each Chapter, discuss how these findings contribute to our understanding of optogenetic plasmid-host interactions and using these findings, consider the future directions of this thesis with an aim to develop multichromatic optogenetic gene expression vectors.

1.12 Published work

Related to this project, I contributed to a book Chapter on methods for improving stringent transcription and translation regulation of synthetic gene circuits. This Chapter is published in Methods in Molecular Biology 2518. My contribution to this publication is separate from this thesis and none of my work or the authors work included in this publication is included in my thesis.

Greco, F.V., Irvine, T., Grierson, C.S. and Gorochowski, T.E., 2022. Design and Assembly of Multilevel Transcriptional and Translational Regulators for Stringent Control of Gene Expression. Methods in molecular biology (Clifton, NJ), 2518, pp.99-110.

Chapter 2: Methods

This Chapter will cover the methods and protocols used in this work to create and characterise the OptoT7RNAP and pDawn/pDusk light-inducible expression systems.

2.1 Strains

E. coli DH5α (NEB Inc) was used for all cloning. For characterization, *E. coli* expression strains BL21 (NEB Inc), DH10β (NEB Inc) and AB360 were used. The AB360 strain was gifted by Armin Baumschlager and is derived from *E. coli* BW25113. Arabinose sugar metabolising genes have been knocked out of this strain to achieve titratable arabinose induced expression of the T7 RNA polymerase gene and enable fine-tuned control of OptoT7RNAP expression.

2.2 Antibiotics

Antibiotic stock solutions were made by dissolving 50 mg chloramphenicol powder in 2 mL dd H_2O to produce a stock solution concentration of 25 mg/mL, dissolving 100 mg of kanamycin powder in 2 mL dd H_2O to produce a stock solution concentration of 50 mg/mL and finally, dissolving 200 mg of ampicillin powder in dd H_2O to produce a stock solution of 100 mg/mL. The antibiotic stocks were filter sterilised using a 0.22 µM filter and stored at -20°C.

2.3 Media

Autoclaved 1x Luria-Bertani (Miller) broth (LB broth) was used for strain propagation and supplemented with 1% glucose when used for overnight cultures. Glucose supplementation is a standard practice used to reduce the basal expression of recombinant genes expressed from ParaB promoter. Sterile, syringe-filtered 1x M9 minimal medium salts were supplemented to a final concentration of 0.34 g/L thiamine hydrochloride, 0.4% D- glucose, 0.2% casamino acids, 2mM magnesium sulphate and 0.1mM calcium chloride and used for all gene expression experiments; for some experiments the M9 minimal medium was further supplemented with 0.05%, 0.1% or 0.2% arabinose sugar. For bacterial plates, a 1% LB-agar broth was prepared and autoclaved. All plates were poured and dried in a laminar flow cabinet. For selective growth in liquid broth, minimal media and agar plates, chloramphenicol, kanamycin, and ampicillin stock solutions were diluted to 25 µg/mL, 50 µg/mL and 100 µg/mL, respectively. The *E. coli* strain AB360 was maintained in 50 µg/mL kanamycin during strain propagation but kanamycin was never used during AB360 expression experiments. This decision was made to avoid introducing extra metabolic requirements to the AB360 expression system and keep growth conditions between AB360, BL21 and DH10β the same**.**

2.4 Glycerol stocks

For long-term cryogenic storage of bacterial strains, 1 mL of overnight cell culture was gently mixed in 500 µL of autoclaved 60% glycerol solution and stored at -80°C.

2.5 Preparing chemically competent cells

CCMB80 Buffer is used in bacterial transformation of CCMB80 DH10β, TOP10 and MachI bacterial strains. Sterile, syringe filtered CCMB80 buffer was formulated with 10 mM potassium acetate, 80 mM calcium chloride, 20 mM manganese chloride, 10 mM magnesium chloride, 10% glycerol. The buffer was adjusted to pH 6.4 with 20mM hydrochloric acid. Glycerol stocks of *E. coli* cells were streaked onto LB-agar plates and incubated for 37°C for 16 hours. A single, well-isolated colony was picked to inoculate a 10 mL LB-broth seed culture and incubated for 16 hours at 37°C with 250 rpm shaking. The next day, the seed culture was used to inoculate fresh LB-medium at a 1:50 ratio, this inoculum was grown until an OD_{600} of 0.6 was reached (optical density was measured using a Bioware CO8000 cell density meter, blanked with 1 mL of LB broth). The culture was decanted into chilled 50 mL falcon tubes and the cells were harvested by centrifugation at 4,453 xg for 15 minutes at 4°C. Supernatant discarded, the cell pellet was resuspended in 80 mL of ice cold CCMB80 buffer and incubated on ice for 20 minutes, the suspension was re-centrifuged, and supernatant discarded again. The cell pellets were resuspended in LB-broth, the optical density measured and then 250 µL cell aliquots were flash frozen in liquid nitrogen and stored at -80°C.

2.6 Plasmids, genetic parts and transformations

Table 2.1 The plasmids used in this work

Plasmid/ regulator	Addgene	Vector	Resistance	Reference paper	Reporter protein	Host/s[a]	Referred to as[b]
pDawn	43795	pET	Kanamycin	Ohlendorf et al. 2016	sfGFP	BL21 DH10β	pDawn
pDusk	43795	pET	Kanamycin	Ohlendorf et al. 2016	sfGFP	BL21 DH10β	pDusk
T7 563	101674	pSC101	Chloramphenicol	Baumschlager et al. 2017	N/A	AB360 BL21 DH10β	T7p-1
T7*563	101661	pSC101	Chloramphenicol	Baumschlager et al. 2017	N/A	AB360 BL21 DH10β	T7p-2
T7*563-F1	101666	pSC101	Chloramphenicol	Baumschlager et al. 2017	N/A	AB360 BL21 DH10β	T7p-3
T7*563(B)	101667	pSC101	Chloramphenicol	Baumschlager et al. 2017	N/A	AB360 BL21	T7p(4)[c]
T7*563-F2	101663	pSC101	Chloramphenicol	Baumschlager et al. 2017	N/A	AB360 BL21	T7p(5)[c]
T7 563-F1	101675	pSC101	Chloramphenicol	Baumschlager et al. 2017	N/A	AB360 BL21	T7p(6)[c]
T7*563-F2(B)	101668	pSC101	Chloramphenicol	Baumschlager et al. 2017	N/A	AB360 BL21	T7p(7)[c]
pAB50	101678	pETM6	Ampicillin	Baumschlager et al. 2017	mCherry	AB360 BL21 DH10β	mChp
pAB143	101679	pETM6	Ampicillin	Baumschlager et al. 2017	mCherry	AB360	mChp-tagged[c]

a Hosts for expression experiments. Plasmids were also transformed into *E. coli* DH5α during propagation and cloning experiments.

b Reference names have been given to each plasmid; this name is how the plasmid will be referred to in this work

c These plasmids were not used for the final steady-state growth and expression experiments

All plasmids used in this work are shown in Table 2.1

The optogenetic expression plasmids pDawn and pDusk were from the plasmid repository AddGene (Plasmid Add gene IDs: 43795 and 43795). All other plasmids were kindly gifted by Armin Baumschlager.

2.7 DNA preparation and PCR

The fluorescent protein sfGFP was available in the laboratory plasmid glycerol stock library recombined into an expression vector. The sfGFP-plasmid and the pDawn and pDusk plasmid DNA was prepped using the Monarch™ Plasmid Mini-prep kit from 16 hour bacterial cultures, inoculated from single colonies picked from LB-agar plates. Prepped DNA yields were quantified using Nanodrop DNA analysis. PCR amplification was used to amplify the

sfGFP gene, primers were designed with NdeI and HindIII restriction digestion overhangs to allow for sfGFP digestion and ligation into the pDawn and pDusk expression vectors. Primers were ordered from IDT and resuspended to a final working stock concentration of 10 ng/µL. The forward and reverse amplification primers had an annealing temperature of 71°C, determined by the NEB TM calculator tool specified for Q5 high-fidelity polymerase. A 50 µL PCR reaction was assembled on ice in a 0.2 mL PCR tube using the following components: 25 µL of Q5® High-fidelity polymerase enzyme in a 2x Master Mix (NEB inc), 2.5 µL of forward and reverse primer, 1 µL of template sfGFP at a concentration of 10 ng/µL and nuclease free water up to a total volume of 50 µL. The thermocycler conditions were selected according to the NEB protocol for Q5 high fidelity polymerase: initial denaturation at 98°C for 30 seconds, then 30 cycles of 98°C for 10 seconds, 71°C for 20 seconds, 72°C for 20 seconds (this extension time was determined by the sfGFP gene length, 735 bp at 20-30 seconds/kb). A final extension time at 72°C for 2 minutes concluded the PCR thermocycle.

2.8 Gel electrophoresis

For gel electrophoresis, 1% agarose gels were prepared using 1X TAE buffer and GelGreen dye (diluted from a 50X stock solution) and GelGreen dye. DNA samples were resuspended 5:1 in NEB gel loading dye, purple (6X) and run for 35 minutes at 180 V against the NEB Quick-Load® Purple 1 kb Plus DNA ladder. Gels were imaged using a blue-light transilluminator. 50X TAE buffer is formulated from 2M Tris acetate and 100 mM EDTA.

2.9 DNA digestion, ligation, and transformation

Successful amplification of the sfGFP PCR product was confirmed using gel electrophoresis (figure 2.1). The sfGFP band was extracted and cleaned using the Monarch® DNA Gel Extraction Kit and eluted at a final volume of 6 µL. DNA yield was quantified using Nanodrop DNA analysis. Double digestion of 1000 ng of purified sfGFP and 1000 ng of mini-prepped pDawn and pDusk plasmids was performed using 1 µL each of restriction enzymes NdeI and HindIII and 5 µL of 10x CutSmart buffer (NEB inc.), nuclease free water was added up to a total volume of 50 µL, the reaction was prepared on ice in a 0.2 mL PCR tube and then incubated at 37°C for 1 hour, the reaction was inactivated at 80°C for 20 minutes. Gel electrophoresis confirmed linearisation of the digested plasmid DNA and sfGFP, pDawn and pDusk DNA bands were excised and purified

Fig.2.1. Gel elecrophoresis of double digested pDawn, pDusk and sfGFP digestion screen 1) 1kb plus DNA ladder 2) pDawn digest 3) pDusk digest 4) GFP digest

using the Monarch® DNA Gel Extraction Kit. For T4 DNA ligase ligation protocol a 1:3 vector to insert ratio was required, 90 ng of sfGFP DNA and 30 ng of pDawn and pDusk were combined with 1 µL of T4 DNA ligase, 2 µL of 10x T4 DNA ligase reaction buffer and nuclease free water to 20 µL, the components were mixed gently on ice and then incubated at room temperature for 10 minutes. Then 5 µL of pDawn-sfGFP and pDusk-sfGFP plasmids were incubated on ice for 30 minutes with 50 µL of chemically competent *E. coli* strains BL21 (NEB Inc) and DH10β (NEB Inc.), then heat shocked at 42°C for 45 seconds, after resting on ice for 5 minutes the cells were resuspended in 600 µL of SOC medium and incubated at 37°C for 1 hour. Finally, 200 µL of cells were plated onto pre-warmed LB agar and transformants were selected for by 50 µg/mL kanamycin, then incubated at 37°C for 14 hours. Resistance to kanamycin is present in pDawn and pDusk, and in AB360, thus AB360-pDawn/pDusk transformants were not created as they could not be selected for or maintained. All transformant plates were kept in the dark, and where necessary loosely wrapped in aluminium foil to prevent blue-light induced expression. This was necessary for pDawn and pDusk as these plasmids are constitutively expressed and induced by blue-light, whereas the OptoT7RNAP system is first induced by arabinose, which activates T7 RNA polymerase expression and then blue-light restores T7 RNA polymerase functionality, resulting in T7 RNA polymerase promoter driven mCherry expression.

2.10 Colony PCR

Successful pDawn-sfGFP and pDusk-sfGFP recombinant plasmids were validated by colony PCR and the sfGFP primers. Colony PCR was performed using the following reaction components, prepared on ice: 12.5 µL of Quick-Load Taq 2x Master Mix (NEB inc) and 0.5 µL of 10 µM forward and reverse primers and nuclease free water up to a total volume of 25 µL. Instead of prepped template DNA being added in solution, PCR reactions were inoculated with colonies picked from the transformant LB-agar plate using a 2 µL pipette tip. The annealing temperature for the forward sfGFP primer specified for Quick-Load Taq was 68°C. The thermocycler conditions were set as follows: an initial denaturation at 95°C for 30 seconds, and then 30 cycles of 95°C for 25 seconds, 68°C for 1 minute, 68°C for 7 minutes 12 seconds for pDawn and 6 minutes 12 seconds for pDusk (pDawn 7213 and pDusk 6220bp at 1 minute/kb), a final extension at 68°C for 5 minutes concluded the thermocycler run. Successful transformants were propagated and LB-agar plates and glycerol stocks were prepared. See Supplementary information S.1 for primer sequences.

2.11 OptoT7RNAP's and reporter plasmids

Agar stabs received from Armin Baumschlager arrived containing plasmids transformed into the cloning strain *E. coli* DH5αZI, these were propagated in LB-miller medium before being

plated onto LB-agar and also cryogenically frozen in long term glycerol stocks. In total nine plasmids were gifted by Armin Baumschlager: seven split T7 RNA polymerase expressing plasmids and two T7 RNA polymerase promoter-driven mCherry reporter plasmids. Plasmid DNA was prepped from DH5αZI cultures grown for 16 hours at 37°C in LB-miller medium with 1% glucose supplementation

2.12 mChp, T7p1, T7p2 T7p3 and OptoT7RNAP transformations

All nine plasmids were transformed into *E. coli* strains BL21, DH10β and AB360 using heat shock protocol (as described in section 2.9). Complete OptoT7RNAP expression systems were generated in BL21, DH10β and AB360 strains expressing both a T7 RNA polymerase plasmid (T7p1, T7p2 or T7p3) and an mCherry reporter plasmid (mChp or mChp-tagged) (see Table 2.2). Transformants with mCherry reporter plasmids were selected for by 100 µg/mL ampicillin and T7 RNA polymerase plasmids were selected for by 25 µg/mL chloramphenicol, double transformants were selected for by both 25 µg/mL chloramphenicol and 100 µg/mL ampicillin. Single transformant cells were plated onto pre-warmed LB-agar plates and then incubated at 37°C for 14 hours. Double-transformant colonies were transformed in a one-step protocol and incubated for 24 hours, after this time the colony diameter was approximately equal to the single transformants and the plates were stored in the fridge at 5°C. A 60% glycerol stock library was created, for each transformant 4 glycerol stocks were produced, each derived from a single colony. These glycerol stocks were used to streak fresh LB-agar plates which were used for inoculating medium during expression experiments.

Table.2.2 Combinatorial transformations of E. coli strains to create complete Opto-T7RNAP expression systems

T7 RNA polymerase plasmid	mCherry plasmid	Expression level[a]	Host strains	Referred to as:
T7 563	pAB50	1.9	AB360 BL21 DH10β	Opto-1
T7*563	pAB50	1.9	AB360 BL21 DH10β	Opto-2
T7*563-F1	pAB50	1.9	AB360 BL21 DH10β	Opto-3
T7*563-F1	pAB143	1.9	AB360 BL21 DH10β	Opto-3-tagged
T7*563-F1	pAB50-replacement T7 promoter	1.9	AB360 BL21 DH10β	Opto-3-RepT7
T7*563 (B)	pAB50	1.0	AB360 BL21 DH10β	Opto-4[b]
T7*563-F2	pAB50	1.9	AB360 BL21 DH10β	Opto-5[b]
T7 563-F1	pAB50	1.9	AB360 BL21 DH10β	Opto-6[b]
T7*563-F2 (B)	pAB50	0.5	AB360 BL21 DH10β	Opto-7[b]

[a] The expression level (reported by Baumschlager et al., 2017) describes the level of expression of the C-terminal Opto-T7RNAP fragment compared to the N-terminal fragment.

[b] These plasmid combinations were not used for the final steady-state growth and expression experiments

2.13 Replacing the wild-type T7 RNA polymerase promoter on the reporter plasmid

To replace the wild-type T7 RNA polymerase promoter in the mCherry reporter plasmid, 1000 ng of prepped mChp plasmid was double digested in 5 µL of 10X CutSmart buffer (NEB inc.) using 1 µL of AvrII (NEB inc.) and 1 µL of BamHI (NEB inc.) and nuclease free water up to a total volume of 50 µL, the reaction was prepared on ice in a 0.2 mL PCR tube and then incubated at 37°C for 1 hour. The reaction was inactivated at 65°C for 20 minutes. These restriction enzymes were chosen as they flank the 17-bp wild-type T7 RNA polymerase promoter. Compared to the wild-type promoter sequence the directed evolution T7 RNA polymerase promoter has three point mutations on consecutive nucleotides: at base positions 7, 8 & 9 bases GAC are substituted with bases CGG. The synthetic promoter sequence was synthesised as an oligo (IDT Inc.) with complimentary AvrII and BamHI restriction digestion overhangs (see Supplementary information S.1 for oligo sequence). The oligo was resuspended to a working concentration of 100 ng/µL and then 1000 ng was double digested using the same digestion protocol to create cohesive overhangs. Gel

electrophoresis confirmed plasmid linearisation, the plasmid band was excised and purified using the Monarch® DNA Gel Extraction Kit. DNA clean-up of the digested oligo was performed using the Monarch® PCR & DNA Cleanup Kit. The components were ligated in a 1:3 vector to insert ratio using 10 ng of plasmid DNA and 30 ng of oligo DNA in 2 µL of 10X T4 DNA ligase reaction buffer and 1 µL of T4 DNA ligase, nuclease free water was added up to a total volume of 20 µL, the reaction components were mixed on ice and then incubated at room temperature for 10 minutes. The plasmids were transformed into the *E. coli* cloning strain DH5α (NEB Inc) and plated onto pre-warmed LB-agar.

2.14 Sanger sequencing

Colonies harbouring successfully modified plasmid DNA were identified by DNA sequencing. Plasmid DNA was prepped, and the DNA was quantified using Nanodrop DNA analysis. A sequencing primer was ordered from IDT and resuspended to a working stock concentration of 10 ng/µL (see Supplementary information S.1 for primer sequence). Samples were prepared according to IDT inc. sequencing instructions. Successful colonies were propagated and then the plasmid DNA was prepped and transformed into AB360 and BL21 strains in a double transformation protocol with the plasmid T7p3. The double-transformants were plated on pre-warmed LB-agar and grown for 24 hours at 37°C.

2.15 Steady-state growth and basal expression experiments

For high-throughput analysis of steady state growth and basal expression by the OptoT7RNAP and pDawn/pDusk expression systems, all growth and expression experiments were performed in 96-well conical bottom microplates in biological replicates of twelve for data shown in the main text and supporting information, unless otherwise noted.

2.16 Preparing the microplates

Single colonies were used to inoculate M9 minimal medium overnights for steady-state growth and expression experiments. For overnight cultures, by a Bunsen flame, colonies from LB-agar plates were poked with 2 µL pipette tips. To inoculate, the tips were placed in 200 µL of M9 supplemented medium and gently rotated. The inoculated microplate cultures were covered with a breathable microplate adhesive seal (MicroAmp^{TM})and incubated at 30°C for 16 hours, with shaking at 225 rpm in a Stuart SI500 shaking incubator (Cole-Parmer Ltd.). After 16 hours, 2.5 µL of culture was used to inoculate 200 µL of M9 supplemented medium. An inoculation volume screen was preformed using the following volumes: 1.25 µL, 2.5 µL, 5 µL and 10 µL. A volume of 2.5 µL was decided as a volume of this size produced growth curves that reached stationary phase within approximately 400 minutes, in comparison inoculation volumes of 5 µL and 10 µL caused stationary phase to

be met prematurely and did not capture clearly defined lag and exponential phases. Plasmids were maintained by antibiotics added to the overnight and expression medium in a 1 µL: 1 mL volume/volume ratio, respectively. To investigate how arabinose sugar effects the growth of *E. coli* BL21, DH10β and AB360 strains, arabinose was added to the M9 supplemented medium at a concentration of 0.1%. To investigate the effect of arabinose induction on growth rate and mCherry expression for *E. coli* BL21, DH10β and AB360 hosts transformed with Opto-1, Opto-2 and Opto-3 plasmids, arabinose was added to the M9 medium at concentrations of 0.05%, 0.1% and 0.2%. Arabinose was not added to the overnight plates.

2.17 Plate reader incubation: Monitoring OD_{600} and fluorescence intensity

The 96-well conical bottom microplates were incubated at 37°C, for 10 hours with double orbital shaking continuously at 37°C. Absorbencies at 600 nm (OD_{600}) and fluorescence intensities (absorption and emission wavelengths for mCherry: 579 nm and 616 nm and for sfGFP: 479 nm and 520 nm) were measured every 15 minutes. After 600 minutes the temperature was reduced to 25°C and incubated for a further 30 minutes, this was to allow for any newly translated mCherry or sfGFP proteins to fully mature, fluorescence intensities were measured again at 630 minutes. Data was exported to an excel file.

2.18 Normalising Data

For all fluorescence quantitative analysis, the fluorescence intensity was normalized by dividing the fluorescence intensity at the emission wavelength 520 nm (sfGFP) or 616 nm (mcherry) by the OD_{600} value of the same sample. For culture growth analysis, the optical density of the growth medium at 600 nm was normalised by subtracting the minimum OD_{600} value from all values of the same time course experiment.

2.19 Flow cytometry

To assess, at a single-cell level, the expression of fluorescent proteins in induced and non-induced states, flow cytometry was used for the following transformants: *E. coli* BL21 pDawn-sfGFP, *E. coli* BL21 pDusk-sfGFP and *E. coli* BL21. A single colony was picked from an LB-agar plate and used to inoculate 200 µL of M9 media, in triplicates and grown in either darkness (wrapped in tinfoil) or constant blue light for 16 hours at 250 rpm shaking in a Stuart SI500 shaking incubator (Cole-Parmer Ltd.) (Digimess powerpack supplied blue light emitting LEDs with 0.05A, 2.8V). An X20 analyser (BD LSRFortessa™) analysed the cultures by sorting 10^5 cells based on their sfGFP or mCherry fluorescence. Absorption and emission wavelengths for sfGFP were 479 nm and 520 nm. A full width/half height gating strategy was used to remove any cell aggregations from the data pool.

2.20 Custom-built blue light apparatus

A custom-built light apparatus was constructed to securely hold and illuminate clear-plastic conical bottom microplates with blue-light. A matrix of blue-light emitting diodes spaced equally apart was inserted into the bottom of a plastic container which was light-proofed with aluminium foil and set on non-slip rubber feet. This was inserted into the Stuart SI500 Shaking incubator (Cole-Parmer Ltd.). A powerpack (DC HY300 power supply, Digimess instruments Ltd) supplied a constant current and voltage of 0.05 amp and 2.8 V during blue light experiments.

See Supplementary information S.2 and S.3 for the equipment and reagents used in this work.

Chapter 3: Quantifying basal expression of light inducible expression systems pDawn/pDusk and OptoT7RNAP

3.1 Introduction

Basal gene expression occurs in all synthetic expression circuits. This is because biological parts do not behave as reliably as electrical ones and promoters are never fully repressed (Baumschlager et al., 2017). To compare the levels of "leaky" gene expression of pDusk, pDawn and the different OptoT7RNAP variants, the basal level of GOI expression was quantified using the fluorescent reporter proteins sfGFP and mCherry. In this Chapter, the results of high-throughput fluorescence analysis will be reported and used to characterise the level of basal fluorescent reporter gene expression of pDusk, pDawn and the OptoT7RNAP variants, transformed into *E. coli* strains BL21, DH10β, and AB360. To demonstrate how vector behaviour can vary depending on host genetics, the levels of basal GOI expression will be compared in an expression strain, BL21 and a cloning strain, DH10β. BL21 will also be compared with the *E. coli* expression strain AB360, which was specifically designed for the expression of the OptoT7RNAP optogenetic vectors.

3.2. Flow cytometry of *BL21* pDawn and pDusk

Ohlendorf and co-workers conducted all pDawn expression experiments in the *E. coli* BL21 strain CmpX13, in LB medium with a custom-built blue-light array of 10 x 16 470nm-light-emitting diodes and reported that the pDawn vector had greater levels of induced protein expression and lower levels of basal expression compared to pDusk. A key aim in the Ohlendorf study was to create portable light inducible expression systems that can be easily implemented in the laboratory without any need for specialist equipment. Furthermore, Ohlendorf wanted to create an expression vector that was fully portable and robust to the effects of media composition and host strain on the synthetic circuity. Flow cytometry was used to demonstrate the signal polarity of the pDusk and pDawn vectors and confirm that the optogenetic vectors perform similarly in the *E. coli* strain BL21 and in M9 minimal medium, instead of LB medium.

I used flow cytometry to assess, at a single-cell level, the induced and basal levels of sfGFP gene expression by the pDusk or pDawn plasmid in *E. coli* BL21(Figure 3.1). For both expression vectors, there is little-to-no overlap in the fluorescence distributions between light and dark treated cultures indicating that sfGFP expression is switched on in an all-or-nothing response. Furthermore, after comparing the level of induced sfGFP fluorescence with the level of basal sfGFP fluorescence, it is evident that the fold-induction in GOI expression is much greater for pDawn than for pDusk (Figure 3.B). For BL21 cultures harbouring pDusk,

the mean induced sfGFP fluorescence is approximately 11 times greater than the mean basal sfGFP fluorescence. However, for BL21 cultures harbouring pDawn, the mean induced sfGFP fluorescence was approximately 775 times greater than the mean basal sfGFP fluorescence.

I also find that BL21 cells expressing induced pDusk plasmids have a greater degree of heterogeneity in the level of sfGFP fluorescence compared to BL21 cells expressing induced pDawn plasmids. This is shown by a broader sfGFP fluorescence distribution from induced BL21-pDusk cells compared to the sfGFP fluorescence distribution from induced BL21-pDawn cells. This suggest that after induction, sfGFP gene expression is regulated more stringently by pDawn. Furthermore, the distributions of sfGFP fluorescence from blue-light induced BL21-pDawn cells are narrower, and therefore less varied, than the distributions of basal sfGFP fluorescence in dark-treated BL21-pDawn cells (Figure 3.1B). This suggests that for the BL21-pDawn cells, induced sfGFP expression is more homogenous, compared to the mean, than basal sfGFP expression. This finding is only true for pDawn because both the induced and basal sfGFP fluorescence distributions from BL21-pDusk cells have a similar homogeneity compared to the mean. In addition, BL21 cells harbouring pDawn exhibit lower levels of basal sfGFP fluorescence (i.e., closer to the natural fluorescence of the BL21 cells), compared to pDusk, confirming that pDawn is better at repressing basal sfGFP gene expression.

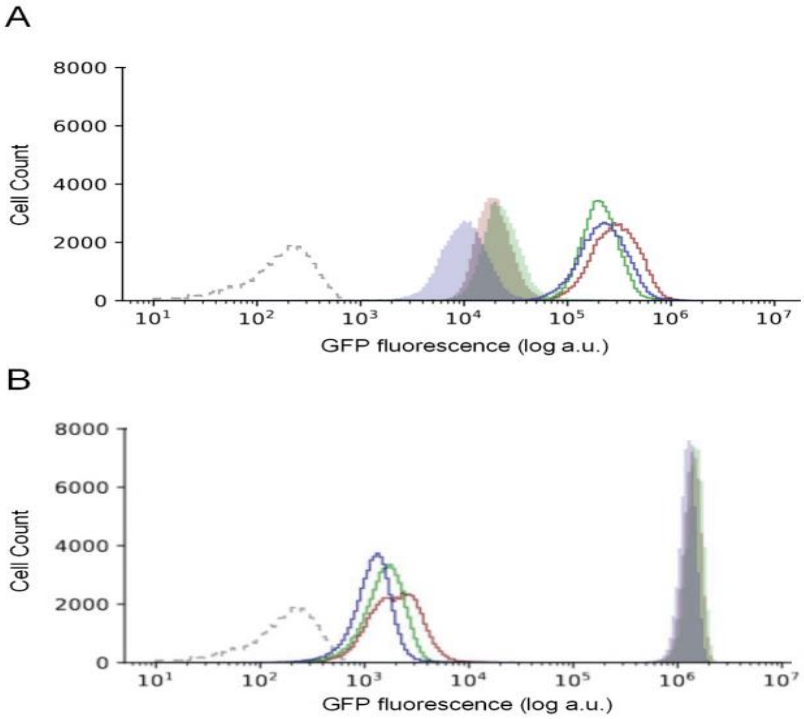


Figure 3.1: Fluorescence distributions of sfGFP produced by E. coli BL21 cells expressing pDawn or pDusk. A) *E. coli* BL21 cells harbouring pDusk-sfGFP **B)** *E. coli* cells harbouring pDawn-sfGFP. **A & B)** a total of 10^5 cells were analysed using flow cytometry for each distribution. Shaded distributions represent cultures grown in constant blue light, non-shaded distributions represent cultures grown in darkness and the grey dashed line represents the mean autofluorescence of plasmid free *E. coli* BL21 control culture. Red, blue and green represent three vector bearing *E. coli* BL21 cell lines and means were calculated from these three biological replicates.

These results confirm the findings reported by Ohlendorf: the pDawn vector exhibits a higher induction ratio and lower levels of basal gene expression compared to pDusk. The flow cytometry results have shown that the inversion cassette of pDawn not only inverted but greatly amplified the effects of blue-light induction. As λ pR is a very strong promoter, expression levels under inducing conditions are higher in pDawn than in pDusk. Conversely, since λ cI is a very strong repressor under non-inducing conditions the level of basal expression is lower in pDawn than in pDusk. A challenge in synthetic biology is the lack of quantitative tools that accurately describe and predict the behaviours of engineered gene circuits (Liao, Blanchard, and Lu., 2017). However, replicating the findings of the Ohlendorf study goes part way to demonstrate the robustness of gene regulation by pDawn compared to pDusk. The use of the TCS YF1/FixJ in pDusk and pDawn is an example of how synthetic photoreceptors and naturally occurring proteins can be combined in a modular fashion to provide allosteric gene regulation (Pathak, Strickland and Vrana., 2014). Furthermore, the

development of the TCS YF1/FixJ inversion cassette, in the pDawn vector, is a great example of how the function of synthetic genetic parts can be manipulated to achieve lower levels of basal gene expression while increasing maximum protein yield.

The incorporation of sfGFP into the MCS of pDusk and pDawn and transformation of *E. coli* BL21 and DH10β with the recombinant vectors was relatively easy. However, difficulties occurred when trying to store the pDusk transformants. On agar plates, pDusk colonies appeared very green even after short amounts of time in the fridge. To conserve the foreign DNA circuit, it is good practice to avoid protein expression when storing and propagating recombinant cells. This helps to prevent mutations being propagated in cell stocks which can spontaneously arise from burden-driven modifications of synthetic circuitry by the host. The pDusk circuitry uses host resources to function (i.e. transcription and translation machinery, DNA replication equipment and metabolites). Depletion of the host resources may be burdensome to the host especially in environmental conditions that are not conducive to the replenishment of essential resources, like the low temperature of a fridge (5°C) or limited nutrient availability of agar plates. This can cause the host to evolve away the burdensome circuit via recombination, point mutations/deletions or copy number reduction (Brophy and Voigt., 2014). Furthermore, basal sfGFP expression in BL21 and DH10β hosts harbouring pDusk could not be quantified using plate reader incubation, as this provided an inducing environment. I have found that although both vectors satisfy Ohlendorf's objective of creating portable expression platforms, the pDawn vector is more easily implemented than the pDusk vector. For these reasons and because of its favourable signal polarity pDawn was focused on for all further basal sfGFP expression and growth studies.

3.3 Basal performance of pDawn in *E. coli* BL21 and DH10β hosts

For the following basal GOI expression experiments the *E. coli* cultures were grown in darkness, this was done to measure GOI expression in the non-induced state. For the plasmid free *E. coli* BL21 and DH10β cultures the mean maximum OD_{600} after 600 minutes was 1.386 and 1.088, respectively. Thus, on average BL21 cultures had a maximum biomass 22% greater than DH10β cultures, assuming that the average cell size and light scattering regime was the same for both strains. However, the mean maximum autofluorescence of plasmid-free BL21 cultures was 13% greater than the autofluorescence of plasmid free DH10β cultures, showing a difference in the colour profile between the strains at a 479 nm absorption and 520 nm emission spectra. This may suggest that there are some differences in the morphology or physiological environment between the two *E. coli* strains. The average maximum OD_{600} after 600 minutes of BL21 cultures expressing pDawn was 1.233 and for DH10β-pDawn cultures it was 1.146. Thus, the mean maximum

OD_{600} for plasmid free BL21 cultures was 11% greater than for BL21-pDawn cultures. For plasmid free DH10β cultures there was a 5.3% decrease in mean maximum OD_{600}, compared to DH10β-pDawn cultures (Table 3.1). Due to the unequal variances between the plasmid-free populations and the populations expressing pDawn seen in both strains, an unequal variances *T*-test (Welch's t-test) was needed to test whether these differences in mean maximum OD_{600} were statistically significant. *P* values < 0.05 were considered significant. The results showed there was no statistically significant difference in maximum OD_{600} between plasmid free BL21 cultures (M = 1.386, SD = 0.053) and BL21 cultures expressing pDawn (M = 1.233, SD = 0.207); t(23) = 1.41, p = 0.188. Neither was there a statistically significant difference in maximum OD_{600} between plasmid free DH10β cultures (M = 1.088, SD = 0.074) and DH10β cultures expressing pDawn (M = 1.146, SD = 0.299); t(23) = 1.47, p= 0.18.

The relative standard deviation (RSD) of the maximum average OD_{600} for BL21-pDawn cultures increased to 16.8% from the RSD of plasmid free BL21 cultures which was 3.8%. For DH10β-pDawn cultures, the RSD increased to 26.1% from 2% RSD of plasmid free DH10β cultures. This increase in variability indicates a decrease in the growth homology of the BL21 and DH10β cells expressing the pDawn vector, compared to the plasmid free host cultures.

After 600 minutes, the percentage difference between the two *E. coli* strains, expressing pDawn, in mean maximum fluorescence was 52% and on average the DH10β cells had a higher concentration of sfGFP protein (Figure 3.2). A Welch's unequal variances *T*-test was also performed to test whether this difference in mean maximum fluorescence was significant. The test showed that there was no significant difference between BL21

Table 3.1: Comparison of optical density and basal sfGFP fluorescence of *E. coli* BL21 and DH10β plasmid free cultures and cultures expressing pDawn

	BL21		DH10β	
	Max OD [a]	%RSD [c]	Max OD [a]	%RSD [c]
no plasmid	1.386	3.8	1.088	6.8
pDawn	1.233	16.8	1.146	26.1
	Max FI [b]	%RSD [c]	Max FI [b]	%RSD [c]
no plasmid	24.7	36.4	21.6	25.0
pDawn	20823.3	70.5	35311.4	88.6

Averages were calculated from 12 biological replicates.

[a] Mean maximum optical density at 600 nm

[b] Mean maximum sfGFP fluorescence (a.u.) normalized to cell densities (OD_{600}).

[c] Relative standard deviation (or coefficient of variation) expressed as a percentage

expressing pDawn (M = 20823.3, SD = 14680.4) and DH10β expressing pDawn (M = 35311.4, SD = 31285.9), t(24) = -0.77, p = 0.46. This is likely due to a large amount of variance in each group and overlapping error bars. In both BL21 and DH10β cultures expressing pDawn there is a large amount of heterogeneity in the basal sfGFP expression levels, however slightly less so for BL21 cultures which has a maximum fluorescence RSD of 70.5% compared to the RSD of DH10β cultures at 88.6%. A larger sample size would be needed to determine whether there is a statistically significant difference in mean maximum sfGFP fluorescence between BL21 and DH10β cultures expressing pDawn.

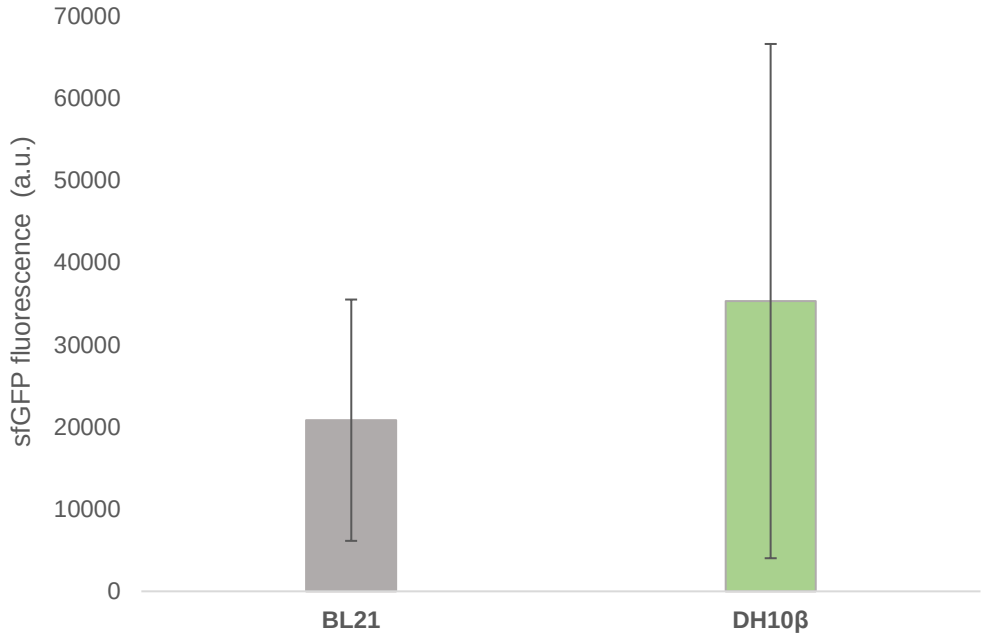


Figure 3.2: Basal sfGFP fluorescence by the pDawn plasmid expressed by BL21 and DH10β *E. coli* host strains. Values are mean maximum sfGFP fluorescence ± SD for at least twelve biological replicates. Fluorescence intensities normalized to cell densities (OD_{600}).

The results suggest that basal sfGFP gene expression from the pDawn vector is more tightly controlled in BL21 host cells, as DH10β cultures have greater mean concentrations of sfGFP protein. The difference in basal sfGFP fluorescence intensities could reflect how the behaviour of an expression vector can be variable depending on host regulation. The plasmids pDusk and pDawn both derive their backbone from the widely used pET-28c vector which has been modified by exchanging the T7-lacO promoter region with the promoter region of the FixK2 protein (Ohlendorf et al., 2012). The pET vector series are compatible with and commonly expressed in the *E. coli* strains BL21 and DH10β (Studier and Moffatt., 1986). The pET backbone carries a PBR322 origin of replication and plasmids with this origin exist in medium copy numbers in *E. coli*. This helps to reduce leaky expression before

induction and is useful in cases where the protein of interest has cytotoxic effects at high concentrations or when expressing proteins with limited solubility (Dubendorff and Studier., 1991). To aid the wider implementation of optogenetic expression platforms in industry and research, modifying existing and commonly utilised vector backbones promotes ease of use and ensures compatibility with existing protocols.

Basal expression of the sfGFP gene occurs in pDawn via incomplete repression of the pR promoter by the cl repressor. The signal transduction cascade effectuating the repression of sfGFP production relies on the transcription and translation of the proteins YF1, FixJ, FixK2 and cl. However, both RNA polymerase and ribosome concentrations are growth-rate dependent. Furthermore, the availability of amino acids, metabolites and other essential resources can vary significantly between host strains and at different growth rates and cell densities (Brophy and Voigt., 2014). Moreover, due to the deletion of multiple metabolic pathways, the range of nutrients DH10β can utilise is limited, making the strain sensitive to nutritional downshifts. In addition, DH10β harbours both the *relA1* and *spoT1* alleles which contribute to lower growth rates and elongated growth lags, compared to its wild-type MG1655 parent strain (Durfee et al., 2008). Cloning strains are optimised for efficient plasmid uptake, minimised recombination and the production of high-quality DNA and so traits associated with good protein production have not been prioritised for optimisation (Durfee et al., 2008). Compared to expression strains such as BL21, cloning strains have less efficient protein folding chaperones and are less able to grow up to a high OD. The expression strain BL21 is a commonly used host for protein expression which boast tight regulation and expression consistency, and greater titres of both native and recombinant proteins (Pan and Malcolm., 2000). These strain dependent differences may explain the difference in mean sfGFP concentration between BL21 and DH10β hosts. As the BL21 strain has a higher maximum specific growth rate, leading to greater RNA polymerase and ribosome titres and more efficient protein folding chaperones it is well suited to the overexpression of regulatory proteins responsible for sfGFP gene repression.

In addition, BL21 has defective *lon* and *ompT* proteases (Paliy and Gunasekera., 2007). Protease deficiency is a key feature of expression strains, which is not commonly seen in cloning strains, such as DH10β. The protease *ompT* is located on the outer membrane and thus only degrades extracellular proteins, however the *lon* protease is mainly associated with the degradation of intrinsic misfolded proteins and protein quality control but *lon* also degrades many regulatory proteins and plays an important role in regulation of physiological processes (Gur, 2013). Protease deficiency may contribute to lower levels of basal sfGFP gene expression from the pDawn vector in the BL21 host because the pDawn vector uses negative gene regulation to repress gene expression. In the DH10β cells, protease

degradation of the regulatory proteins YF1, FixJ, FixK2 and cI, which are responsible for repression of the pR promoter, may have led to a higher mean level of basal sfGFP gene expression compared to that seen in BL21 cells expressing pDawn.

The high degree of variability in sfGFP fluorescence intensity in both strains could be due to differences in RNA polymerase and ribosome concentrations between cells in the culture. RNA polymerase concentrations can vary between 1,800 and 10,200 molecules per cell, depending on the growth phase, nutrient conditions, and other extrinsic factors (Bremmer and Dennis., 1987). Although all efforts were made to mitigate the risk, slight differences in nutrient conditions and extrinsic conditions such as temperature and light exposure between the replicates could have occurred during plate preparation as high throughput experiments were conducted over several days. Ribosome concentrations can also fluctuate significantly between cells in a culture (De Vos et al., 2011). Furthermore, sfGFP fluorescence levels measured by flow cytometry showed that basal sfGFP expression is less homologous compared to induced sfGFP expression. Thus, indicating increased stochasticity in the biological processes responsible for basal GOI expression compared to induced GOI expression. This is because gene expression is rarely completely silenced (Greco et al., 2021) and without the presence of a strong inducer, the levels of leaky gene expression are more susceptible to fluctuations and transcriptional noise.

In summary, these results suggest a difference in gene regulation between the cloning strain DH10β and the expression strain BL21. This difference could have arisen due to BL21 cells containing greater titres of the regulatory proteins responsible for sfGFP gene repression in pDawn. Thus, the basal performance of pDawn is influenced by the choice of *E. coli* strain. To achieve low basal GOI expression levels pDawn should be expressed in strains which produce high amounts of specific protein and the lowest degradation of soluble protein fraction. For example, K-12 host RV308, which has exhibited lower levels of ppGpp (a protein correlated with metabolic stress), and lower degradation of soluble protein compared to *E. coli* hosts in the BL21 lineage (Marisch et al., 2013). In future work, the basal performance of pDawn could be characterised in RV308 to investigate further the robustness of this optogenetic vector in different host strains.

3.4 The OptoT7RNAP variants

Baumschlager observed light inducibility at all the split positions tested. However, split site 563 had the greatest fold change in mCherry expression between light and dark states and the lowest basal expression level. All OptoT7RNAP variants gifted by Baumschlager had a photoreactive T7RNAP split at amino acid 563/564. The leading OptoT7RNAP variant was deemed to be OptoT7RNAP*(563)-F1 due to its low background activity, high-fold induction

ratio, fast reverting time and reduced toxicity to the host (Baumschlager et al., 2017). Thus, for this work I focused on the three variant systems which best demonstrate the systematic improvements made to develop the fast-reverting optogenetic regulator OptoT7RNAP*(563)-F1. Firstly, Opto1: which harbours a T7RNAP fragments split at the 563rd amino acid. Secondly, Opto2: which harbours the T7RNAP fragments split at the 563rd amino acid and the modified C-terminal fragment containing the T7RNAP mutation R632S. Lastly, Opto3 (OptoT7RNAP*(563)-F1) which harbours the T7RNAP fragments split at the 563rd amino acid, the modified C-terminal fragment containing the R632S mutation and the fast-reverting pMag variant pMagFast1. All three of these variants have 1.9-fold expression of the C-terminal fragment compared to the N-terminal fragment. By comparing these three variant systems, the reduction in basal gene expression which is generated by each systematic improvement can be quantified. Basal expression of the Opto1, Opto2 and Opto3 variants was quantified in AB330, BL21 and DH10β hosts at different arabinose concentrations.

3.5 Basal performance of the OptoT7RNAP variants in different *E. coli* strains and arabinose concentrations

For the AB360 cell cultures expressing Opto1, Opto2 and Opto3 in the absence of arabinose, the basal mCherry fluorescence was 39,914 a.u., 3,485 a.u. and 1,414 a.u., respectively. Thus, the Opto3 expression platform had the lowest level of basal mCherry gene expression in a AB360 host, followed by Opto2 and then Opto1, which had a level of basal mCherry fluorescence greater than Opto3 by 59% and 96%, respectively. For BL21 cell cultures expressing Opto1, Opto2 and Opto3 in the absence of arabinose, the basal mCherry fluorescence was 38,966 a.u., 1,315 a.u. and 833 a.u., respectively. Again, the Opto3 expression platform had the lowest level of basal mCherry gene expression, followed by Opto2 and then Opto1, which had a level of basal mCherry fluorescence greater than Opto3 by 58% and 98%, respectively. Thus, the difference in basal mCherry gene expression between the different OptoT7RNAP variants was similar for AB360 and BL21 host cells, however AB360 cells had an overall higher level of mCherry gene expression compared to BL21 cells (Figure 3.3). For DH10β cell cultures expression Opto1, Opto2 and Opto3 in the absence of arabinose, the basal mCherry fluorescence was 57,082 a.u., 4,269 a.u. and 1,303 a.u., respectively. Thus, the Opto3 expression platform had the lowest level of basal mCherry gene expression, followed by Opto2 and Opto1, which had a level of basal mCherry fluorescence greater than Opto3 by 277% and 4281%, respectively. Therefore, the DH10β host provided the poorest repression of basal mCherry gene expression, compared to AB360 and BL21. However, in all three *E. coli* hosts, the level of basal mCherry gene expression was lowest in cells expressing the more optimised Opto3 expression platform and highest in cells expressing the least optimised Opto1 expression platform.

To determine the relationship between arabinose and basal mCherry gene expression the three OptoT7RNAP variants were cultured in media supplemented with arabinose sugar at concentrations of 0.05%, 0.1% and 0.2%. For all three OptoT7RNAP variants, an arabinose concentration of 0.1% approximately equalled or reduced the level of basal mCherry fluorescence in all three *E. coli* hosts, compared to the basal mCherry fluorescence of the same recombinant cell lines cultured in the absence of arabinose. In summary, for AB360 cells expressing Opto1, Opto2 and Opto3 when cultured with arabinose at a concentration of 0.1%, the basal mCherry fluorescence reduced by 21%, 36% and 27%, respectively. For BL21 cells cultures expressing Opto1, Opto2 and Opto3 cultured with 0.1% arabinose, the basal mCherry fluorescence reduced by 24% for Opto3, slightly increased by 3.8% for Opto2 and reduced by 9.6% for Opto1. For DH10β cells cultured with 0.1% arabinose, basal mCherry fluorescence slightly increased by 0.3% for Opto1 and 5.4% for Opto2 and decreased by 5.7% for Opto3. For all OptoT7RNAP variants in all *E. coli* host strains arabinose concentrations of 0.05% and 0.2% increased basal mCherry fluorescence compared to the corresponding recombinant cells cultured without arabinose. See Supplementary information S.5 for mean maximum optical density of the cultures and relative standard deviations of mean optical density, and Supplementary S.6 information for all mean maximum mCherry fluorescence (a.u.) and relative standard deviations in mCherry fluorescence. Basal mCherry fluorescence values normalized to cell densities (OD_{600}) are shown in Figure 3.3.

QQ plot analysis of basal fluorescence intensity data revealed a non-normal distribution, thus the Kruskal-Wallis test were used to analyse data. *P*-values < 0.05 were considered as significant. To determine whether the three OptoT7RNAP expression variants tested lead to significantly different levels of mCherry fluorescence regardless of *E. coli* host strain and arabinose concentration, the mean maximum mCherry fluorescence for AB360, BL21 and DH10β hosts, at all arabinose concentrations, were pooled into three categories: Opto1, Opto2, Opto3. A Kruskal-Wallis test was performed to determine if mean mCherry fluorescence was the same for Opto1, Opto2 and Opto3 expression systems. A total of 36 measurements were used in this analysis, 12 for each expression systems. The test revealed that the mean FI was not the same ($H = 29.1$, $p < 0.000$) among the three OptoT7RNAP variants [mean mCherry fluorescence: Opto1: 53,776, Opto2: 4,102, Opto3: 1,409]. Thus, there was a statistically significant difference in median average maximum mCherry fluorescence among two or more of the expression systems. A Kruskal-Wallis test was performed to determine if median average maximum mCherry fluorescence of the Opto1, Opto2 and Opto3 expression systems varied significantly in each *E. coli* host regardless of arabinose concentration, for each Kruskal-Wallis test 12 measurements were

used in the analysis, 4 for each OptoT7RNAP variant. The test revealed that for each host the median mCherry fluorescence was not the same (AB360: $H = 9.3$, $p = 0.009$, BL21: $H = 9.8$, $p = 0.007$, DH10β: $H = 9.8$, $p = 0.007$) among the three expression systems. Finally, a test was performed to determine whether arabinose concentration was responsible for the significant difference seen in the average mCherry fluorescence produced by Opto1, Opto2 and Opto3 in all *E. coli* host. Data was pooled into the following categories of arabinose concentration: 0%, 0.05%, 0.1% and 0.2%, a total of 36 measurements were used in the analysis, 9 for each arabinose concentration. The test revealed that there was not a significant difference in median mCherry fluorescence at each arabinose concentration when mCherry fluorescence data of the OptoT7RNAP variants and *E. coli* strains were pooled together ($H = 2.5$, $p = 0.48$). These tests demonstrated that the OptoT7RNAP expression system was the determining factor in maximum fluorescence intensity and not host strain or arabinose concentration. However, at an individual level for each OptoT7RNAP expression system arabinose concentration was shown to affect the level of mCherry fluorescence (Figure 3.3).

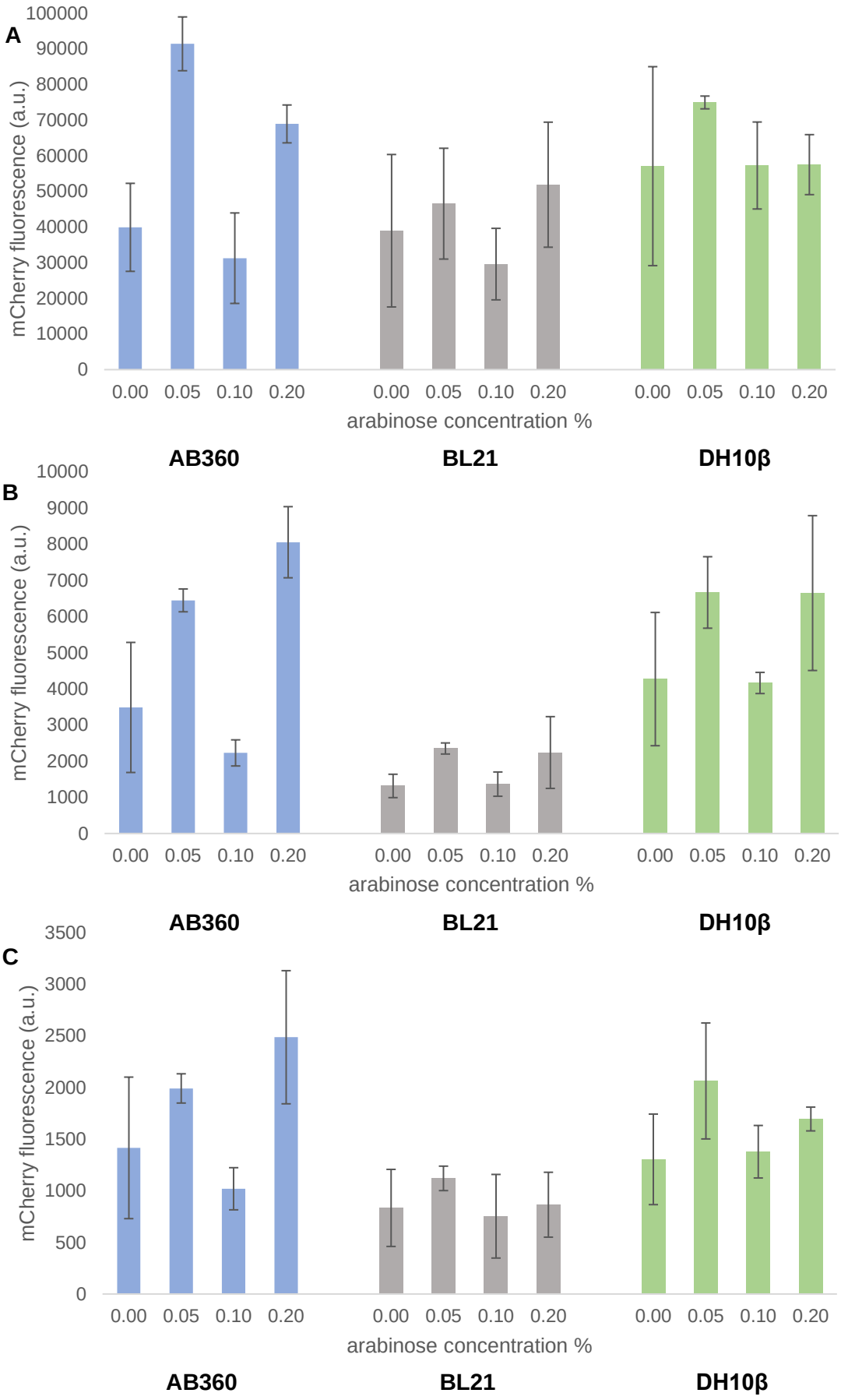

A
mCherry fluorescence (a.u.)
100000
90000
80000
70000
60000
50000
40000
30000
20000
10000
0
0.00 0.05 0.10 0.20 0.00 0.05 0.10 0.20 0.00 0.05 0.10 0.20
arabinose concentration %
AB360
BL21
DH10β
B
mCherry fluorescence (a.u.)
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
0.00 0.05 0.10 0.20 0.00 0.05 0.10 0.20 0.00 0.05 0.10 0.20
arabinose concentration %
AB360
BL21
DH10β
C
mCherry fluorescence (a.u.)
3500
3000
2500
2000
1500
1000
500
0
0.00 0.05 0.10 0.20 0.00 0.05 0.10 0.20 0.00 0.05 0.10 0.20
arabinose concentration %
AB360
BL21
DH10β

Figure 3.3: Quantifying maximum basal GOI expression by the OptoT7RNAP variants using the fluorescent reporter mCherry in *E. coli* strains AB360, BL21 and DH10β cultured in different arabinose concentrations. A) basal fluorescence of mCherry expressed by Opto1 which transcribes the mCherry gene with T7 RNA polymerase that is split at the 563/564 amino acid site. **B)** basal fluorescence of mCherry expressed by Opto2 which transcribes the mCherry gene with T7 RNA polymerase that is split at the 563/564 amino acid site and has a mutated R632S C-terminal fragment. **C)** basal fluorescence of mCherry expressed by Opto3 which transcribes the mCherry gene with T7 RNA polymerase that is split at the 563/564 amino acid site, has a mutated R632S C-terminal fragment and is linked to the pMag variant pMagFast1. **3.3A, 3.3B and 3.3C)** Values are mean maximum mCherry fluorescence ± SD for at least twelve biological replicates. Fluorescence intensities normalized to cell densities (OD_{600}). *Note that the scale of the y-axis varies*.

The OptoT7RNAP variant Opto1 did not contain the T7RNAP mutation R632S or the pMag mutation 185V and had the highest level of basal mCherry expression in AB360, BL21 and DH10β hosts producing an average mCherry fluorescence of 39,914, 38,966, and 57,082, respectively. The introduction of the R632S mutation to reduce T7RNAP toxicity created the Opto2 variant and the average mCherry fluorescence reduced by 91% in the AB360 host strain, 97% in the BL21 host strain and 93% in the DH10β host strain. Basal mCherry expression was further reduced by the mutation 185V in the pMag photoactive heterodimer, which was combined with the R632S T7RNAP, to create the reduced toxicity and fast reverting Magnet variant in Opto3. Using this optimised OptoT7RNAP variant, basal mCherry fluorescence reduced by a total of 97% in the AB360 host strain, 98% in the BL21 host strain and 98% in the DH10β host strain, compared to the basal mCherry fluorescence observed in host cells harbouring the unoptimized Opto1 expression system.

A gene expression study using the BL21(DE3) T7RNAP system, testing several proteins, demonstrated how cell-toxicity arose from GOI overexpression (Miroux and Walker., 1996). Furthermore, genome sequencing revealed that spontaneous burden-driven mutations had arisen in some of the cell-lines. The cell-lines which had overcome the deleterious effects of GOI overexpression had mutations which all resided in the T7RNAP gene (Miroux and Walker., 1996). The toxicity associated with the T7RNAP enzyme has promoted the development of host strains with mutations that overcome the toxic effects of GOI overexpression or T7RNAP variants with mutations that reduce the toxicity associated with high enzyme processivity (Padan, Hunte and Reilander., 2003). Baumschlager reduced this toxicity by introducing the mutation R632S into the T7RNAP gene. Discovered in a study by Temme and co-workers in 2012, the R632S mutation is located in the T7RNAPs active site and was shown to significantly reduce toxicity to the host (Temme et al., 2012).

Characterisation of T7RNAP R632S mutants found no change in polymerase concentration or activity (Temme et al., 2012) compared to wild-type T7RNAP.This was unusual as mutations in this region, the C-terminal 541-737 finger sub-domain (Sousa, Mukherjee and Kivie., 2003), had previously been shown to reduce enzyme processivity (Temme et al., 2012).

Expression of the photo dimerizing T7RNAP gene fragments is under the control of the *araB* promoter, which is induced by arabinose. Thus, basal mCherry gene expression may have arisen from leaky induction of the *ara*B promoter. Studies have shown that the *araB* promoter shows considerable amounts of constitutive leaky expression during log growth (Cao et al., 2017). It is unlikely that the R632S mutation reduced the level of constitutive leaky expression from the *ara*B promoter. However, the introduction of this mutation in the OptoT7RNAP variant Opto2, reduced basal expression of mCherry by over 90% for all host strains, compared to the Opto1 variant (in the absence of arabinose). It seems likely that R632S could have an effect on either T7RNAP enzyme concentration or enzyme activity at the T7 promoter. Especially, since other studies characterising mutations in the C-terminal finger sub-domain showed that in this region mutations reduced enzyme processivity (Westover, Bushnell and Kornberg., 2004). Seeing as the T7RNAP-R632S variant had only been characterised during induced gene expression experiments, the level of reporter protein would be much greater than in this work, which focused on basal gene expression. It is therefore possible that any difference in concentration or activity of the T7RNAP-R632S variant went unobserved due to high reporter protein titres with large variations, typical in synthetic gene networks and other limiting factors to protein titre such as nutrient conditions or biosynthetic precursor availability. Thus, the reduction in basal mCherry fluorescence between Opto1 and Opto2 likely involves the concentration or processivity of the T7RNAP-R632S variant. To further reduce the level of basal mCherry gene expression the *ara*B promoter could be investigated, perhaps using directed evolution techniques, to identify promoter regions which could be modified for tighter gene expression regulation in the absence of arabinose. Thus, reducing basal T7RNAP gene expression.

Kawano and co-workers used directed engineering of the blue-light fungal photoreceptor Vivid to yield the photoactivable Magnet system, which consists of two Vivid protein variants: the negatively charged nMag and positively charged pMag, which bind via electrostatic interactions, thus preventing homodimerization and enhancing light-induced heterodimerization (Kawano et al., 2015). The Magnets dissociate with a half-life of 1.8 hours and cannot be actively deactivated; to improve their practicality in fast dynamic protein expression experiments the mutation 185V was introduced to pMag, resulting in pMagFast1 and the mutations 174V and 185V were introduced to nMag, resulting in pMagFast2.

Reducing the dissociation time from 1.8 hours to 4.2 min and 25 seconds. The pMagFast1 variant, harbouring only the 185V mutant, performed better in dynamic gene expression studies (Baumschlager et al. 2017) so was focused on in this work. Baumschlager incorporated the 185V mutation to Opto2 and generated the OptoT7RNAP variant Opto3. I found that Opto3 further reduced basal mCherry gene expression by over 95% (in all *E. coli* strains), compared to Opto1 (in the absence of arabinose). All expression experiments were conducted in the absence of inducing blue-light (450 nm). So dark-state dimerization of the Magnet domains, leading to restored T7RNAP functionality, was a source of basal mCherry gene expression. Dark-state dimerization of pMag and nMag is likely stochastic and may be due to the heterodimeric Magnets coming into proximity within the bacterial cytoplasm, or perhaps some dimerization was caused by absorption of 587 nm light (used to measure mCherry fluorescence). All efforts were made to prevent blue-light induction, but a small number of blue-light photons could have reached the cell cultures. Unlike with SMI gene expression circuits, it is difficult to prevent inducing wavelengths reaching recombinant cells as most photoreceptors used in optogenetic circuits interact with wavelengths on the visible light spectrum. Therefore, the fast-reverting 185V mutation helped to reduce basal mCherry gene expression by promoting dissociation of the dimers and limiting the amount of time a T7RNAP enzyme remained functional.

Basal activity from the *ara*B promoter and dark-state dimerization of the photoactivable magnet domains leads to basal mCherry expression. ParaB has a slower rate of induction compared to other systems such as T7, lac or tac, due to a smaller rate of IFN-α accumulation on the ParaB promoter (Lim et al., 2000). Thus, the genetic circuitry of the OptoT7RNAP expression platform allows for lower T7RNAP protein titres and higher mCherry (or target protein) titres due to the high processivity of the polymerase enzyme, enabling the host cells to grow continuously despite the metabolic burden introduced by the T7 driven expression of foreign protein.

Baumschlager achieved titratable monosaccharide L-arabinose induction by introducing a mutated lactose permease membrane protein, *lacYA177C,* into the *attB* site of the *E. coli* strain BW25113. Wild-type lactose permease protein exhibits a poor recognition of L-arabinose however several lactose permease mutants have been discovered, due to their ability to grow on L-arabinose plates, which show enhanced recognition of arabinose sugars (Goswitz and Brooker., 1993). The modified BW25113 circumvents all-or-nothing induction of the native arabinose transporter (Baumschlager et al., 2017) because it is deficient in both arabinose uptake and degradation genes (Bowers et al., 2004), allowing for titratable arabinose induction. This modified strain was named ‘AB360’ by Baumschlager and has since been used in other protein expression studies, for instance the study of chemo-

optogenetic anhydrotetracycline and tetracycline transcription (Baumschlager, Rullan, Khammash., 2020). Care needs to be taken when implementing “tunable promoters” which are inducible by sugars such as arabinose, lactose or rhamnose. There is a common misconception that the level of recombinant protein synthesis can be manipulated at will within each cell, but studies have shown that the fold-range in protein expression is dependent on the amount of active sugar that permeases in each cell which is highly variable (Rosano and Ceccarelli., 2014). So, while the final protein yield can be controlled the titre of recombinant protein is not uniform across the population. Meaning that cells in the upper percentiles, with high levels of protein synthesis may be susceptible to increased metabolic strain and, if the expressed protein has innate toxicity (in this instance T7RNAP protein does) cell death can occur. Furthermore, titratable induction suggests that increased arabinose concentration should correlate with increased levels of T7RNAP. Although this thesis focused on basal gene expression (i.e. without blue-light induced dimerization of the Magnets) it stands to reason that with increasing arabinose concentration there should be increased T7RNAP titres which would likely lead to more Magnet dimerization events and result in increased basal mCherry gene expression. However, for all host strains an arabinose concentration of 0.1% lead to the lowest levels of basal mCherry expression, compared to 0.05% and 0.2%. Furthermore, in most cases an arabinose concentration of 0.05% lead to higher levels of basal mCherry expression than an arabinose concentration 0.2% (Figure 3.3). During blue-light expression experiments using the Opto3 expression system, Baumschlager showed that reporter protein expression was maximal with 0.1% arabinose and decreased with 0.2% arabinose and that, interestingly, this was repeated with no light induction. Baumschlager also noted that increasing T7RNAP titres via increasing arabinose concentration had no significant impact on fold change, as dark state and light induced reporter protein expression increased comparably. From this work, it is clear that to achieve truly titratable arabinose induction more work is needed to make the ParaB induction system more robust. Furthermore, very little difference was seen in regulation of basal mCherry gene expression between the AB360 strain and the expression strain BL21 and if differences did occur, in most cases BL21 had lower levels of basal mCherry expression compared to AB360. For example, BL21 cells harbouring the Opto3 expression system had lower basal mCherry fluorescence compared to the AB360 strain harbouring the Opto3 system, at arabinose concentrations of 0.0%, 0.05% and 0.2%. This is perhaps due to arabinose degradation genes being expressed by BL21, resulting in lower arabinose concentrations within BL21 cells.

3.6 Basal mCherry gene expression by the mCherry reporter plasmid

Host AB360, BL21 and DH10β cultures harbouring only the mCherry reporter plasmid showed a significant increase in basal mCherry fluorescence compared to plasmid free cultures lacking the mCherry gene. The average mCherry fluorescence for AB360, BL21 and DH10β cultures harbouring the mCherry reporter plasmid was 285 a.u., 233 a.u. and 181 a.u., respectively. However, the autofluorescence of plasmid free AB360, BL21 and DH10β cultures was 28 a.u., 25 a.u. and 22 a.u., respectively. This shows that the T7 promoter has a basal level of leaky expression in the absence of the OptoT7RNAP enzyme. For AB360, BL21 and DH10β cultures expressing both the modified mCherry reporter plasmid with a replacement T7 “CGG” promoter variant (non-orthogonal) and the OptoT7RNAP plasmid T7p3, the average fluorescence intensity was 281 a.u., 219 a.u. and 212 a.u., respectively. However, for AB360, BL21 and DH10β hosts expressing the unmodified Opto3 expression system with the orthogonal T7 promoter the mean mCherry fluorescence was 1414 a.u., 833.19 a.u. and 1303 a.u., respectively. Thus, for the AB360, BL21 and DH10β host cultures expressing the Opto3 expression system the percentage of basal mCherry expression that results from ‘leaky’ T7 promoter induction independent from dimerised OptoT7RNAP induction is 18%, 26% and 16%, respectively.

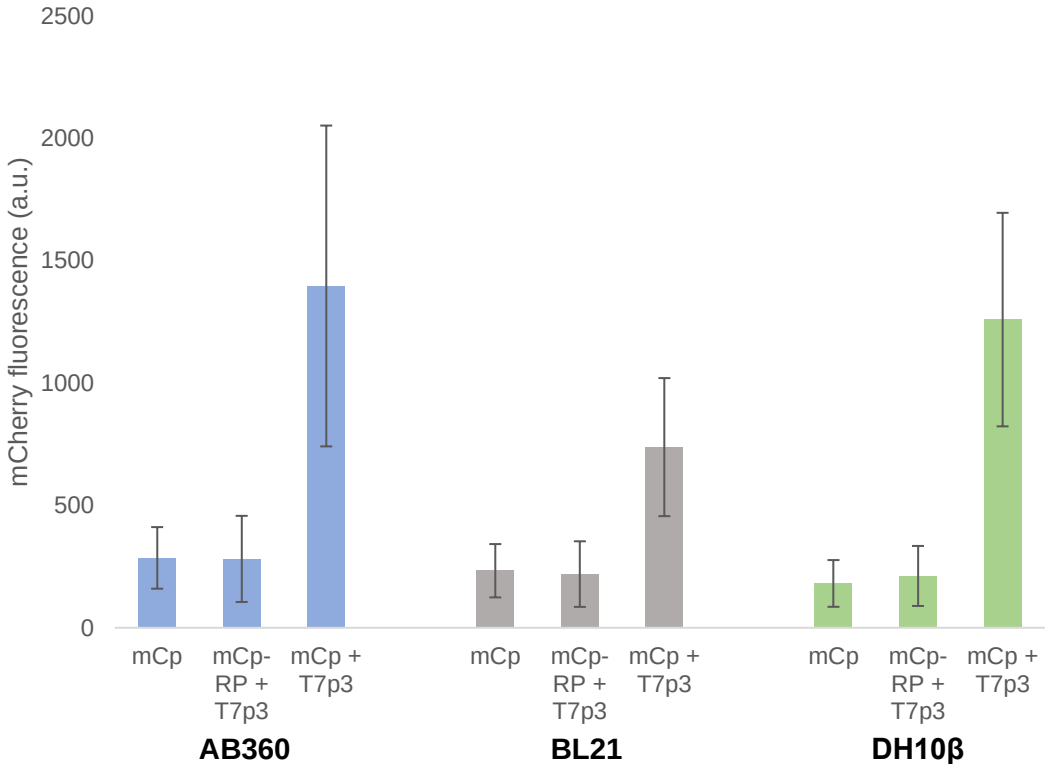


Figure 3.4: Comparison of basal mCherry fluorescence from the plasmid mChp with and without T7 RNAP: T7 promoter interactions, in *E. coli* strains AB360, BL21 and DH10β. Values are mean maximum mCherry fluorescence ± SD. for at least twelve biological replicates. Fluorescence intensities normalized to cell densities (OD_{600}).

Creating AB360, BL21 and DH10β recombinant cell lines expressing only the mCherry reporter plasmid provide an opportunity to investigate the leakiness of the T7 promoter without the presence of the T7RNAP enzyme. In most instances of T7RNAP driven gene expression, the T7RNAP gene is incorporated into the genome of the host. However, the Opto7RNAP optogenetic system incorporates the OptoT7RNAP gene and the T7 promoter on separate plasmids. The mCherry fluorescence data showed that AB360, BL21 and DH10β hosts harbouring the mCherry reporter plasmid expressed the mCherry gene despite the absence of OptoT7RNAP (Figure 3.4). The results also revealed that AB360, BL21 and DH10β hosts expressing the Opto3 expression platform with a mutated T7 “CGG” promoter had approximately equal levels of mCherry fluorescence to cells transformed with only the mCherry reporter plasmid. This suggests that the OptoT7RNAP:T7 promoter interaction is highly specific and the non-cognate T7 promoter was not recognised by the polymerase enzyme. Furthermore, for both the single transformant host cells and the host cells expressing the modified Opto3 system with the T7 “CGG” promoter, average mCherry

fluorescence did not vary significantly between the host strains. This is evident because of a significant overlap in the standard deviation of the mean maximum mCherry fluorescence (Figure 3.4). Therefore, 'leakiness' of the mCherry gene cassette appears consistent between hosts. See Supplementary information S.7 for mean maximum $OD_{600,}$ mean maximum mCherry fluorescence and relative standard deviation data for *E. coli* transformants expressing Opto2- T7 "CGG" promoter variant.

These results suggest that the level of basal mCherry expression from the Opto3 plasmids is mostly due to leaky gene expression of the OptoT7RNAP fragments, which then dimerize in the absence of blue light and induce mCherry gene expression. Thus, to reduce the basal gene expression of the Optogenetic system, the OptoT7RNAP plasmid should be prioritised for optimisation over the mCherry reporter plasmid. This is because leaky gene expression regulators on the mCherry reporter plasmid are only responsible for 18%, 26% and 16% of basal mCherry gene expression in AB360, BL21 and DH10β host cells, respectively.

The low level of basal mCherry expression from the reporter plasmid could be due to the *lac* operator incorporated upstream of the T7 promoter, which has been shown to strongly repress basal T7 promoter transcription (Oehler et al., 1990). However, expression could be further reduced by optimising the distance between the T7 promoter and the *lacI* repressor or the orientation of the *lacI* gene (Dubendorf and Studier., 1990) as the level of control achieved is dependent on the spacing between the second operator and the T7 promoter.

The high throughput fluorescence experiments in this work could be used to rapidly screen *lacI* gene placement and its correlation with T7 promoter repression to improve the basal performance of the optogenetic circuit. In the modified Opto3 system the T7 promoter was excised and replaced with the T7 promoter mutant "CGG". This mutant T7 promoter was first described in 2015, when directed evolution techniques were used to create a panel of orthogonal T7RNAP variants and cognate promoters (Meyer, Ellefson and Ellington et al. 2015). Each of the novel polymerases exhibited highly specific recognition of a synthetic cognate promoter. The OptoT7RNAP study by Baumschlager, articulated the novel idea that by using a T7RNAP variant and exchanging the blue-light sensitive Magnets with other optogenetic dimers, multichromatic control of multiple genes in the same cell could be achieved. To this end, the T7 promoter of the Opto3 system was replaced by the directed evolution mutant with the goal to ensure there was no crosstalk between the wild-type T7RNAP and mutant promoter. Given that the basal mCherry fluorescence for the Opto3-replacement T7 promoter system was comparable to the basal mCherry fluorescence of the host cells expressing only the mCherry reporter plasmid, this goal has been achieved.

3.7 Expression of the T7 RNAP plasmid alters host autofluorescence

Compared to plasmid free *E. coli* cell cultures, autofluorescence (at an absorption and emission spectra of 587 nm and 610 nm) decreased in AB360, BL21 and DH10β host cells expressing only the OptoT7RNAP plasmids T7p1, T7p2 and T7p3. Plasmid free cultures of AB360, BL21 and DH10β had an average autofluorescence of 28 a.u., 25 a.u. and 22 a.u., respectively. Expression of the T7p1, T7p2 and T7p3 variants resulted in a decrease in the mean autofluorescence by an average of 40% in AB360 cultures, by an average of 22% in BL21 cultures and by an average of 35% in DH10β (Figure 3.4).

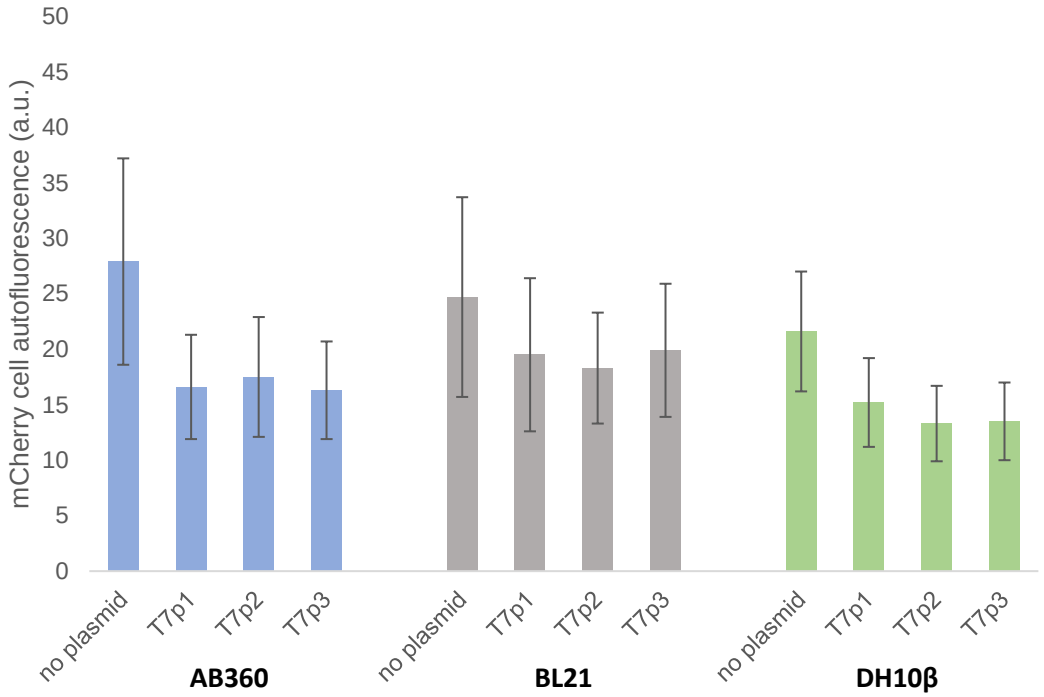


Figure 3.5: Comparison of the cell autofluorescence from plasmid free *E. coli* AB360, BL21 and DH10β with the autofluorescence of *E. coli* AB360, BL21 and DH10β cells expressing the split T7 RNA polymerase plasmids T7p1, T7p2 and T7p3. Fluorescence was measured using the mCherry absorption and emission spectra (587 nm /610 nm) but these plasmids do not have the mCherry gene. Values are mean cell autofluorescence ± SD. calculated from at least twelve biological replicates. Fluorescence intensities normalized to cell densities (OD_{600}).

The mCherry fluorescence observed in *E. coli* hosts expressing only the mCherry plasmid, *E. coli* hosts expressing the Opto3 plasmids and *E. coli* hosts expressing the Opto3-replacement T7 promoter plasmids, suggests that the majority of basal mCherry expression occurs due leaky expression of the OptoT7RNAP fragments which dimerize in the absence of blue-light and induce mCherry expression. Therefore, E. coli host cells expressing only the T7p1, T7p2 and T7p3 plasmids will have basal expression of the OptoT7RNAP gene fragments despite the absence of arabinose. Due to time limitations and the high-throughput

nature of this work I could not quantify basal OptoT7RNAP protein titre. To determine protein titres using high-throughput techniques, future experiments could tag the OptoT7RNAP fragments with a fluorescent protein, but this would likely affect structure and functionality (Snapp, 2005). Moreover, tagging the OptoT7RNAP fragments with a reporter protein increases the transcript length which would increase the demand on host resources and may change the T7RNAP protein titres. Despite being unable to quantify OptoT7RNAP concentration, the *E. coli* AB360, BL21 and DH10β cell autofluorescence changed when expressing the OptoT7RNAP plasmid. Thus, basal OptoT7RNAP expression may be effectuating a change in cell autofluorescence. Prokaryotic cells have intrinsic fluorescent properties due to the natural fluorescence of cellular structural components and metabolites. It has been shown that exposure to cellular stressors can change the autofluorescence of *E. coli* cells (Surre et al. 2018), however this stress induced change usually always manifests as an observed increase in autofluorescence, unlike the decrease seen in my results, which is witnessed not only in other bacterial species but also in yeast and human cells. However, a study on the autofluorescence emission spectra of human leukemic cells demonstrated that autofluorescence can decrease as a result of oxidative stress due the conversion of strongly fluorescent NADH to the poorly fluorescent oxidised form NAD+ (Bondza-Kibangou et al 2001). NADH plays a central role in the respiratory metabolism of bacteria (Heikal., et al. 2014) and the ratio of NADH:NAD+ is a key indicator of metabolic state as NAD(H) homeostasis is critical for bacterial survival (Bhat, Iqbal and Kumar., 2016). Thus, it could be plausible that detrimental effects of basal OptoT7RNAP enzyme expression throws the NADH: NAD+ ratio out of kilter, leading to a decrease in host cell autofluorescence. However, OptoT7RNAP activity is highly selective, and toxicity is usually linked to the high processivity of the enzyme. Due to the absence of the mCherry reporter plasmid and T7 promoter it was assumed that specific OptoT7RNAP toxicity would not impede these single transformant cell cultures. There are several biological restraints when producing plasmid DNA in *E. coli* caused by detrimental metabolic changes arising from plasmid maintenance and replication as well as culture conditions (Silva, Queiroz and Domingues., 2012), in addition the elevated energetic demand on the bacterial host cell to produce recombinant proteins can result in oxidative stress (Chevallier, Andersen and Malphettes., 2020). This energetic demand would be further increased by basal OptoT7RNAP expression. This potential source of oxidative stress may have caused the change in cell autofluorescence through impacting the NADH: NAD+ ratio rather than specific OptoT7RNAP toxicity. However, no published literature can be found illustrating a relationship between OptoT7RNAP and NAD(H).

Furthermore, chloramphenicol antibiotic additions for plasmid maintenance can likely be ruled out as a cause for the decrease in autofluorescence. Studies have shown that *E. coli* cells treated with chloramphenicol, ampicillin, kanamycin and norfloxacin antibiotic exhibit increased autofluorescence when analysed by flow cytometry (Renggli, et al., 2013; Surre et al., 2018).

Alternatively, the decrease in cell autofluorescence may be caused by altered cell morphology. Bacteria have evolved complex processes to maintain cell shape and the dimensions of cell diameter and length are strictly controlled for each species (Justice et al., 2008). The cell wall of a bacterium has a primary role in maintaining cell shape and growth of the cell wall is localised to specific regions of peptidoglycan synthesis which can respond dynamically throughout the cell cycle to conserve cell morphology (Cabeen and Jacobs-Wagner., 2005). In nature, certain stressful environments can trigger bacterial cells to alter the highly regulated process used to maintain consistent cell morphologies, for example a transition to filamentation, which is occurs when cell growth continues in the absence of cell division (Justice., et al. 2008). Filamentation promotes survival in the instance of bacteriophage or protist predation or antimicrobial therapies (Rizzo, Plano and Franco., 2020). However, it is unlikely the expression of the OptoT7RNAP plasmid created a large enough metabolic burden to alter cell morphology in such dramatic ways.

A study in 2014 exploring protein expression characteristics, showed that significant temporal shifts occurred during biomass measurements of the recombinant *E. coli* strain MG1655. Time-series experiments revealed sharp drops in culture biomass that were unlikely to be caused by cell lysis. Bacterial cell imaging revealed that over time the bacterial cell shape shifted to be shorter and less variable in cell length. Researchers hypothesised that shifts in cell morphologies, perhaps because of environmental conditions or expression demands, altered the light-scattering dynamics during optical density measurements leading to the observed drops in biomass (Gorochowski et al., 2014). In this work large drops in biomass were not observed, however bacterial cell autofluorescence has been shown to correlate almost linearly with cell density (Bao et al., 2008), thus optical density changes due to shifts in cell morphology may be a factor affecting autofluorescence.

3.8 Summary of findings

This Chapter has characterised the levels of basal reporter protein expression by the pDawn and the OptoT7RNAP expression systems in the *E. coli* strains AB360, BL21 and DH10β. The synthetic regulation of protein expression in these two optogenetic vectors represents two strategies for engineered optogenetic gene expression regulation. The YF1/FixJ protein system in pDawn is an example of allosteric regulation where the photoreceptors undergo a

large structural change upon light absorption to control enzyme activity or binding interactions. The Magnets in OptoT7RNAP are an example of optical dimerizers, containing modular light-interacting domains used to control the interaction of the T7 RNA polymerase fragments. Furthermore, the regulatory proteins in the pDawn system use negative gene regulation to repress GOI expression and the T7RNAP enzyme of the OptoT7RNAP is an example of positive gene regulation to activate GOI expression. The incorporation of the light activated T7RNAP on a second plasmid decoupled gene transcription from host transcription machinery in the OptoT7RNAP vectors. Furthermore, in the pDawn vector the repression of basal reporter gene expression relied on sufficient concentrations of four negative regulatory proteins: YF1, FixJ, FixK2 and cI, whereas avoidance of basal expression by OptoT7RNAP relied on low levels of leakiness from the *araB* promoter to prevent transcription of positive regulatory protein T7RNAP.

I found that the pDawn optogenetic system had a higher mean maximum reporter protein fluorescence, at 20,823 a.u. for BL21 and 35,311 a.u. for DH10β, compared to the highly optimised Optogenetic system Opto3 which had a mean maximum reporter protein fluorescence of 833 a.u. in BL21 and 1,303 a.u. for DH10β, at 0% arabinose. This demonstrates that the regulatory elements of the Opto2 system led to more stringent control of basal gene expression. This is likely because the pDawn plasmid relies on the transcription and translation of four constitutively expressed proteins to supress GOI expression in the non-induced state, fluctuations in these regulatory protein titres interrupt the robustness of the signal transduction cascade. These fluctuations in regulatory protein titres could occur due to fluctuations in the concentration of RNA polymerases and ribosomes, on which the pDawn expression system relies. Furthermore, the Opto2 system uses positive GOI regulation via OptoT7RNAP transcription and the stringent *araB* promoter to control OptoT7RNAP gene expression. Thus, this system is better at controlling basal gene expression due to the requirement of the small-molecule-inducer, arabinose to induce OptoT7RNAP gene expression. Therefore, it could be argued that only pDawn is a true optogenetic circuit and that the OptoT7RNAP vectors are hybrid SMI-light-inducible gene expression systems. As it is difficult to completely prevent incident blue-light from reaching cell cultures, the use of a SMI to induce GOI transcription machinery helps to further reduce basal GOI expression.

The data I obtained also confirmed that the incorporation of the R632S mutation into the OptoT7RNAP gene (perhaps due to a slight and unknown decrease in enzyme processivity and titre) and the 185V mutation to the pMag gene (to decrease dissociation time of the photosensitive dimers) reduced basal reporter protein expression by over 95% in all *E. coli*

strains. This illustrates the huge impact that small changes to protein structure can have on gene expression regulation.

3.9 Future work

To develop this work further, the basal GOI expression levels of Opto3 should be compared to commonly used chemical induction systems such as IPTG/T7RNAP in BL21(DE3). If the basal performance is comparable, this would help to demonstrate the robustness of OptoT7RNAP mediated optogenetic gene expression. The impact of the OptoT7RNAP plasmids was observed due to a decrease in cell autofluorescence for AB360, BL21 and DH10β host cultures, by an average of 40%, 22% and 35%, respectively. Future work could involve imaging the peptidoglycan cell wall of the transformant cells, for example using fluorescence microscopy, to elucidate whether this decrease in autofluorescence is due to a change in cell morphology. If no difference was seen the change in autofluorescence would likely be due to a change in the internal cellular environment. I would then look at NADH: NAD+ ratios to identify if this change is due to oxidative stress. In addition, I would transform cells with the OptoT7RNAP plasmid backbone but with the OptoT7RNAP gene excised or non-functional, to investigate whether this drop in autofluorescence is due to plasmid transcription and maintenance or as a direct result of basal OptoT7RNAP gene expression.

I have also demonstrated that the directed evolution T7 promoter “CGG” was not recognised by the wild-type T7RNAP. Confirming that both the light-inducible wildtype T7RNAP/pT7 and mutant “CGG” T7RNAP/pT7 could be expressed in the same cell for multichromatic gene expression, with minimal to no crosstalk. I also demonstrated that the majority of basal mCherry expression occurs via leaky OptoT7RNAP expression. In the Opto3 system basal expression from the T7 promoter only accounts for 18%, 26% and 16% of basal mCherry fluorescence in AB360, BL21 and DH10β, respectively. Thus, to further reduce basal expression of the Opto3 system the OptoT7RNAP plasmid should be prioritised for development either via increased repression of the *araB* promoter (or exchanging the *ara*B with a more stringent promoter system) or to reduce the likelihood of dark-state dimerization of the Magnets.

Chapter 4: Characterising the growth properties of the recombinant host AB360

4.1 Introduction

In this Chapter, AB360 growth parameters are characterised to investigate the relationship between plasmid replication, or the expression of plasmid encoded genes and metabolic burden. Strategies to reduce metabolic burden are devised and implemented frequently but still improvements to reduce basal protein expression are often at the expense of bacterial growth. For instance, incorporating more regulatory DNA for more stringent control at the risk of increasing metabolic burden. Furthermore, the optimisation of plasmid DNA to improve plasmid stability is still hindered by the lack of information on the host metabolic response to the expression of plasmid encoded genes (Silva, Queiroz and Domingues., 2012). Modelling the bacterial growth rates of the AB360 OptoT7RNAP optogenetic platform could provide a quantifiable benchmark that can be referenced when optimising the system. In this way, improvements made to reduce the level of basal GOI expression and increase predictability of the OptoT7RNAP optogenetic platform, could be linked to the metabolic burden each improvement had introduced.

Optical density measurements at 600 nm (OD_{600}) are used as an indirect measurement of microbial numbers. Absorbance data from OD measurements is fast, non-destructive, inexpensive and easy-to-automate using a plate reader (Beal et al. 2020). This allows for high-throughput screening of multiple synthetic circuits simultaneously. Compared to classic viable count methods, OD measurements are not as accurate at determining maximum specific growth rate (Myers, Curtis and Curtis., 2013). However, non-linear regression analysis using OD has been shown to correlate well with the parameters derived from viable count growth curves (Pla et al., 2015). Some studies convert OD readings into approximate cell numbers using a standard conversion ratio (i.e. 1.0 OD_{600} = log X number of cells/mL) For example, 1.0 $OD_{600} = 8 \times 10^8$ CFU/mL (Bussalleu et al., 2011; Su and Li., 2004; Li et al., 2010; Daly et al., 2002) is often cited. However, the reported conversion value varies in the literature and can change depending on strain, media and cell morphology (Myers, Curtis and Curtis. 2013). Given the evidence of changing cell morphology of recombinant hosts expressing T7RNAP (presented in Chapter 3, Figure 3.4) and because population kinetics of the novel strain AB360 are not well characterised OD_{600} was not converted to approximate cell number. Many nonlinear regression models are available, but the Gompertz non-linear regression model and the re-parameterized Zwietering-Gompertz model were selected to model the change in OD_{600} of AB360 cultures over time. This Chapter will describe how the models were selected and the results of their use.

4.2 Selecting a non-linear regression model

To select a non-linear regression model, the five principles of model selection by Ratkowsky were followed (Ratkowsky, D., 1992). Firstly, the principle of parsimony was considered. Ratkowsky considers parsimony as the preference of a simple model over a complex one to explain the phenomenon under study (i.e. 'entities are not to be multiplied beyond necessity'), because fewer parameters are more likely to obtain a model with a small amount of nonlinearity. Considering this principle, models with more than three coefficients were rejected. For example, the re-parametrized four-parameter logistic, Gompertz (Korkmaz, 2020) and Baranyi models (Baranyi, Pin and Ross., 1992). Models with coefficients that accounted for environmental factors, such as: population-size dependence, temperature, multiple nutrients, or pH (Allen and Waclaw., 2019) were also rejected. The second principle is to select a model or a re-parametrized model with the best estimation properties and ensure that all estimated coefficients are within the range of observed data. In the case of this study, the parametrizations of different nonlinear models were studied and the frequency of their use in biological modelling, specifically for microbiological bacterial growth, was considered (Tjorve and Tjorve., 2017; Pla., et al, 2015; Allen and Waclaw., 2019; Di Crescenzo and Paraggio., 2019). Third, the range of applicability should be considered. This principle is more data specific and advises that the user ensures data describes the full range of applicability of the model. For example, attempting to fit a model to fragmentary data or data that only describes the lower region of the bacterial growth curve, will yield poorly estimated coefficients. The next principle for robust nonlinear modelling is stochastic specification. The stochastic assumptions made by various models are very diverse, however by using a 'ordinary least squares' function equal weight is given to each of the data points and assumes homogeneity of error, which has been described as a valid stochastic assumption (Zwietering et al. 1992). Lastly, the estimations of the parameters should be as meaningful as possible and have real-world interpretability. For example, rate coefficients that denote the rate of increase or rate of decrease of growth rate are found in some models and can be hard to interpret (Ratkowsky, D., 1992).

The Gompertz model and Pearl-Verhulst Logistic model are well-known non-linear regression models frequently used to describe the growth of animals and plants and in predictive microbiology (Tsoularis and Wallace. 2002; Tjorve and Tjorve., 2017; Di Crescenzo and Paraggio., 2019). Both the Gompertz and Pearl-Verhulst models fit the five principles of non-linear regression modelling and work well with my observed OD data. To select the growth model with the best fit, Pearl-Verhulst and Gompertz equations was used to model a diverse selection of 12 average OD_{600} growth curves. This included plasmid-free growth curves and the growth curves of *E. coli* cells transformed with one plasmid and two plasmids. To ensure that the models were robust to different *E. coli* strains average BL21

and DH10β growth curves were also included. The best fitting model was selected using the Akaike information criterion (AIC), residual standard error and residual sum of squares (Table 4.1).

Table 4.1: Selecting a non-linear regression model

	Logistic			Gompertz		
	AIC[a]	RSE[b]	RSS[c]	AIC[a]	RSE[b]	RSS[c]
1	-131.73	0.032	0.019	-163.86	0.020	0.013
2	-144.49	0.027	0.012	-206.51	0.011	0.004
3	-141.71	0.028	0.022	-133.09	0.032	0.031
4	-136.04	0.030	0.009	-200.37	0.012	0.004
5	-146.41	0.026	0.015	-126.24	0.035	0.038
6	-129.92	0.033	0.027	-181.16	0.016	0.007
7	-186.68	0.014	0.005	-140.15	0.029	0.026
8	-109.06	0.045	0.039	-138.73	0.029	0.027
9	-155.11	0.023	0.017	-132.66	0.032	0.032
10	-147.14	0.026	0.026	-168.50	0.019	0.011
11	-136.89	0.030	0.027	-136.50	0.030	0.028
12	-124.11	0.036	0.046	-114.04	0.042	0.055

* For all growth curves the achieved convergence tolerance of the models was 1.49×10^{-8}

* The residual standard error degrees of freedom was 31 for all modelled growth curves

[a] Akaike information criterion

[b] Residual standard error

[c] Residual sum of squares

AIC model selection was employed to help determine the non-linear regression model that best described the relationship between bacterial culture OD_{600} and time. The AIC is calculated from the number of independent variables used to build the non-linear regression model and the maximum likelihood estimate of the model. Maximum likelihood estimates describe how well the model reproduces the data. AIC considers parsimony, i.e. the best fitting model explains the greatest amount of variation using the fewest possible parameters. Lower AIC scores reflect a better fit. Residual standard error (RSE) was also used to compare the fit of the different regression models. Models which make predictions closer to the observed values have smaller RSE. RSE measures the standard deviation of the residuals, i.e. the difference between the observed values and the predicted values. A non-linear regression model that has a small RSE will have data points that are near the regression line. The residual sum of squares is also stated, this is used to measure the amount of variance in the data that is not explained by a regression model itself but instead estimates the variance in the residuals. The smaller the residual sum of squares the better the model fits the data. All models were fitted using the non-linear least-squares Levenberg-Marquardt algorithm. The Gompertz was selected as overall it had the smallest total AIC,

RSE and RSS values. (The twelve sample model predictions are provided in Supplementary information S.10)

4.3 Equations and fitted growth curves

The logistical model for population growth is characterised by a finite carrying capacity, to describe the self-limiting growth of a biological population. The reparametrized Pearl-Verhulst logistic model is commonly used (Ramirez-Cando, Alvarez-Mendoza and Gutierrez-Salazar., 2018) and is given by the equation:

$$\mathrm{N(t)} = \frac{A}{1 + (\frac{A - N0}{N0})\mathrm{e}^{-r\mathrm{t}}} \quad [1]$$

Where N(t) represents the population at the time t. This equation gives estimates for the growth rate (*r*), carrying capacity (A) and initial population size (N0) coefficients. (Spouffske and Wagner., 2016; Di Crescenzo and Paraggio., 2019).

Perhaps only second to the Verhulst model, the Gompertz model is another frequently used non-linear regression model fitted to population growth data. One valuable and commonly found Gompertz model parametrisations is:

$$\mathrm{N(t)} = A \times \exp(-\exp\left(r \times \left(\mathrm{t} - T_{midlog}\right)\right)) \quad [2]$$

In this model N(t) also represents the population at the time t. This equation gives estimates for the time at inflection or mid-log (T_{midlog}), intrinsic growth rate (*r*) and carrying capacity (*A*) coefficients. (Tjorve and Tjorve., 2017; Di Crescenzo and Paraggio., 2019)

There are many forms of the Gompertz model, of varying usefulness (Tjorve and Tjorve., 2017). The reparametrized Zwietering-Gompertz model is often called a “modified Gompertz” (Halmi et al., 2014) and is typically applied to bacterial growth data. Both the Gompertz and the re-parametrised Zwietering Gompertz models had identical AIC, RSE and RSS scores, furthermore the estimation of carrying capacity (a coefficient shared by both models) were identical for all modelled growth curves. The Zwietering-Gompertz model was used to provide lag time estimates and it is given by the equation:

$$\mathrm{N(t)} = A \times \exp(-\exp\left(\frac{\mathrm{e} \times r}{A}\left(T_{lag} - \mathrm{t}\right) + 1\right)) \quad [3]$$

Again N(t) represents the population at time t. This reparametrized model gives estimates for the lag time (T_{lag}), absolute growth rate at lag time (*r*) and carrying capacity (*A*) coefficients. (Tjorve and Tjorve., 2017; Crescenzo and Paraggio., 2019).

The doubling time of a population is the time it takes for the absorbance reading to double and can be calculated from growth rate. Rate coefficients were converted to doubling times (Spouffske and Wagner., 2016) using the equation: [4]

$$\frac{\mathrm{Ln}(2)}{r}$$

Where the natural logarithm of 2 is divided by the intrinsic growth rate coefficient, which is given by the Gompertz model.

To calculate an estimated length of time where the rate of 'biomass accumulation' (given by the rate of increase in absorbance) was theoretically greatest (i.e. there is no restriction on growth and the population is far from the carrying capacity), the lag time coefficient was subtracted from the corresponding mid-log coefficient.

$$T_{midlog} - T_{lag} \quad [5]$$

The fit of the Gompertz model is shown in Figure 4.1.
The coefficient estimates given by the Gompertz and Zwietering-Gompertz model are shown in Figure 4.2.
For non-linear regression modelling R code and example output see Supplementary information S.9

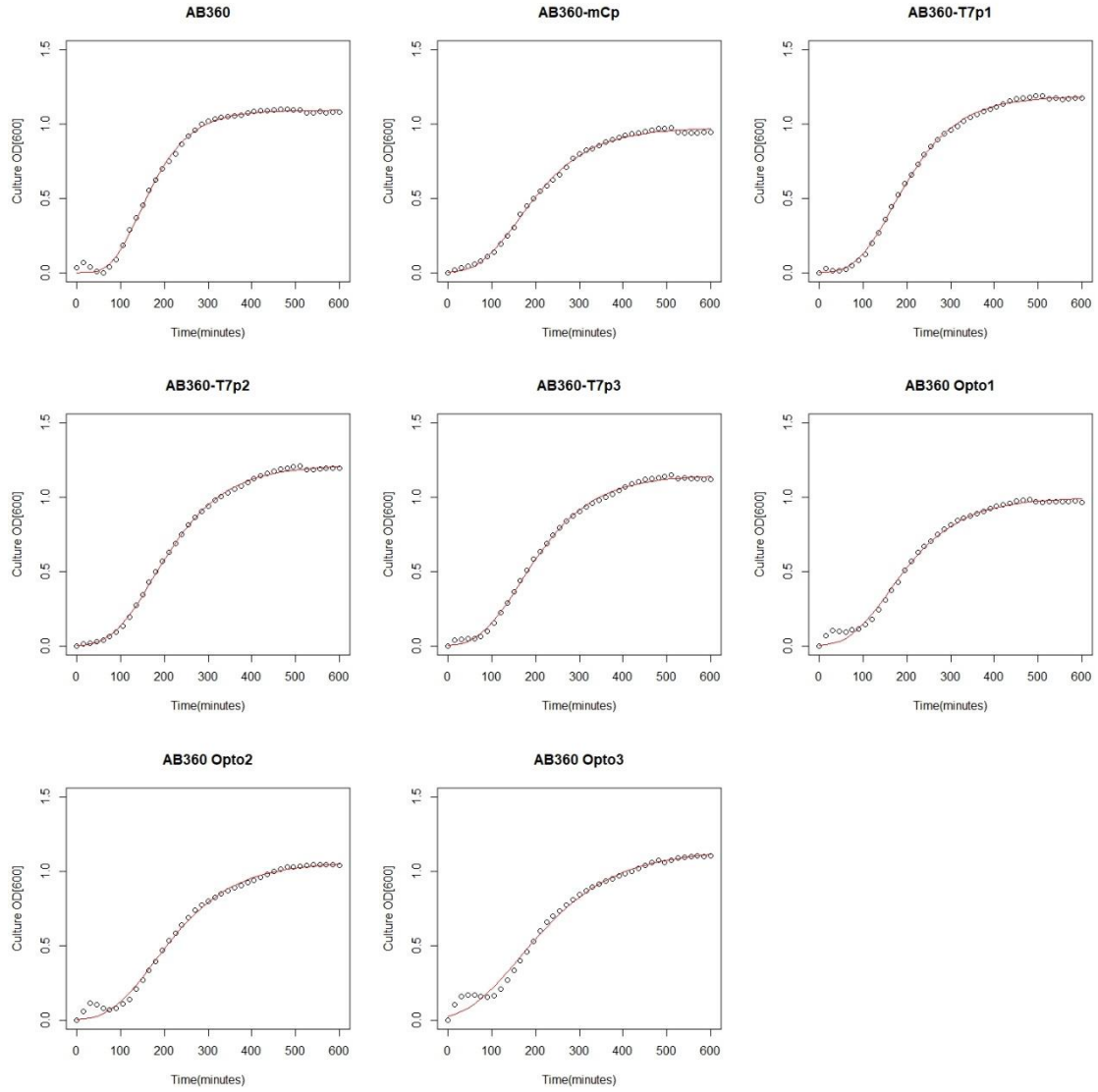
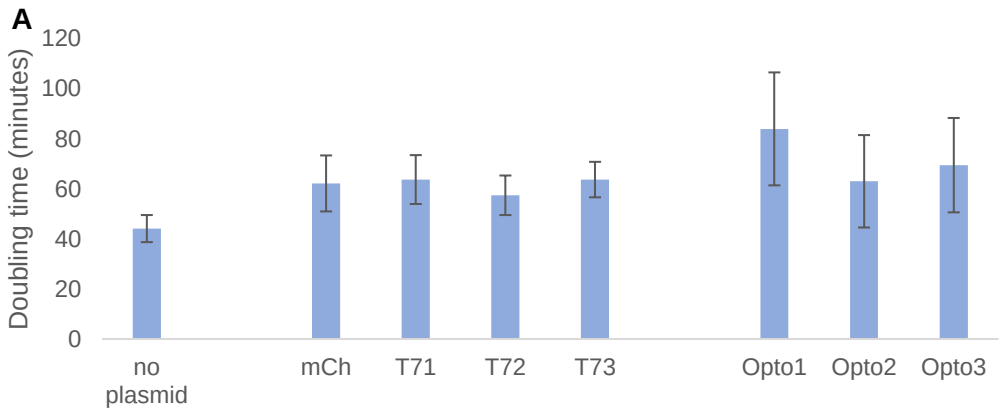
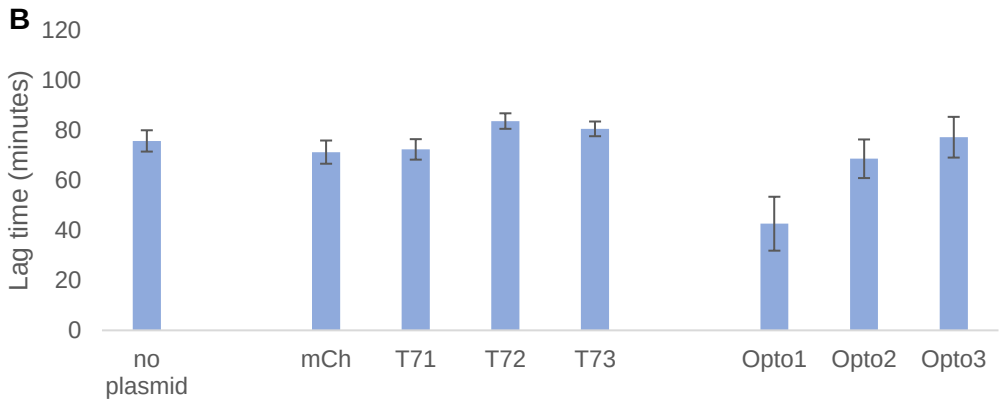
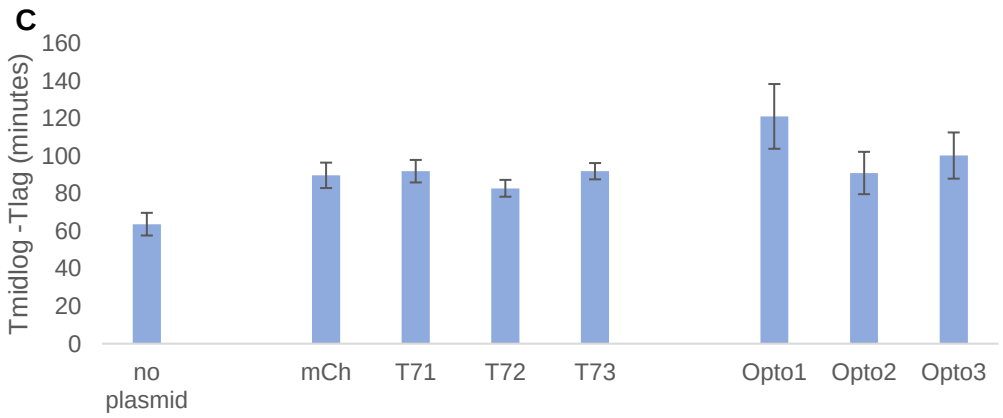


Figure 4.1 Average growth curves of AB360 cultures (OD_{600} over time), represented as circles and the fitted Gompertz models, represented as solid red lines. Average growth curves were calculated from 12 biological replicates. For average growth curves with error bars, see Supplementary S.8


A
Doubling time (minutes)
120
100
80
60
40
20
0
no plasmid
mCh
T71
T72
T73
Opto1
Opto2
Opto3


B
Lag time (minutes)
120
100
80
60
40
20
0
no plasmid
mCh
T71
T72
T73
Opto1
Opto2
Opto3


C
Tmidlog -Tlag (minutes)
160
140
120
100
80
60
40
20
0
no plasmid
mCh
T71
T72
T73
Opto1
Opto2
Opto3

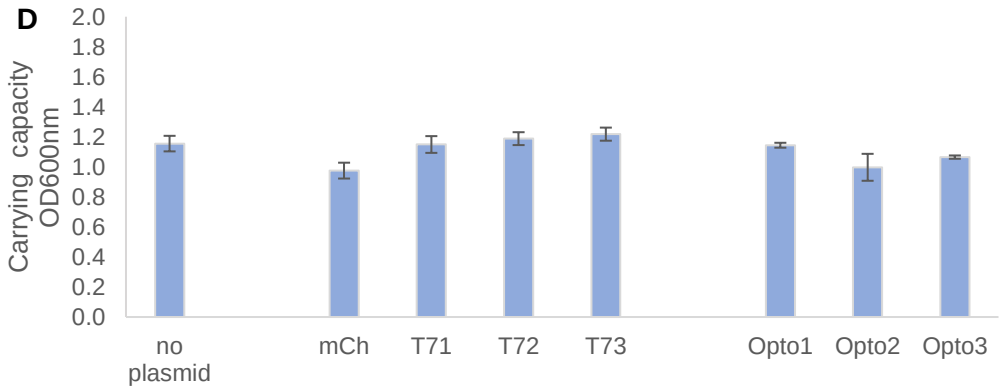

D
Carrying capacity OD600nm
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
no plasmid
mCh
T71
T72
T73
Opto1
Opto2
Opto3

Figure 4.2 Gompertz OD_{600} growth curve coefficients of the AB360 strain expressing recombinant plasmids A) Doubling times were obtained by converting the Gompertz [2] growth rate coefficient into doubling time [4]. **B)** Lag times were obtained from the Zwietering-Gompertz re-parameterised model [3] **C)** T_{midlog} -T_{lag} times [5] were calculated by subtracting the Zwietering-Gompertz lag time coefficient [3] from the inflection time (or time at midlog) given by the Gompertz model [2] **D)** Carrying capacity coefficients were identical for both Gompertz models [2][3] **A-D)** Error bars represent +/- SE. All coefficient estimates had a p value < 0.05. For growth coefficient estimates and errors of fitted Gompertz models and Zwietering-Gompertz models see Supplementary S.11 and S.12, respectively.

4.4 Spearmans Rank correlations of bacterial growth coefficients

I also wanted to investigate to what extent growth rate, or doubling time, correlates with other metrics that summarize growth (Figure 4.3). The perfect correlation between growth rate and doubling time is unsurprising as the definition, and calculation of doubling time, relies on growth rate. The doubling time also correlates with the carrying capacity [Spearmans rank: 0.552 p <2e-16], the time at midlog [Spearmans rank: 0.709 $p < 2 \times 10^{-16}$] and does not correlate strongly with lag time [Spearmans rank: 0.124 $p < 2 \times 10^{-16}$]. Population at mid log and the population at lag time correlate perfectly with each other, and with carrying capacity, as the carrying capacity coefficient is used to calculate these population estimates.

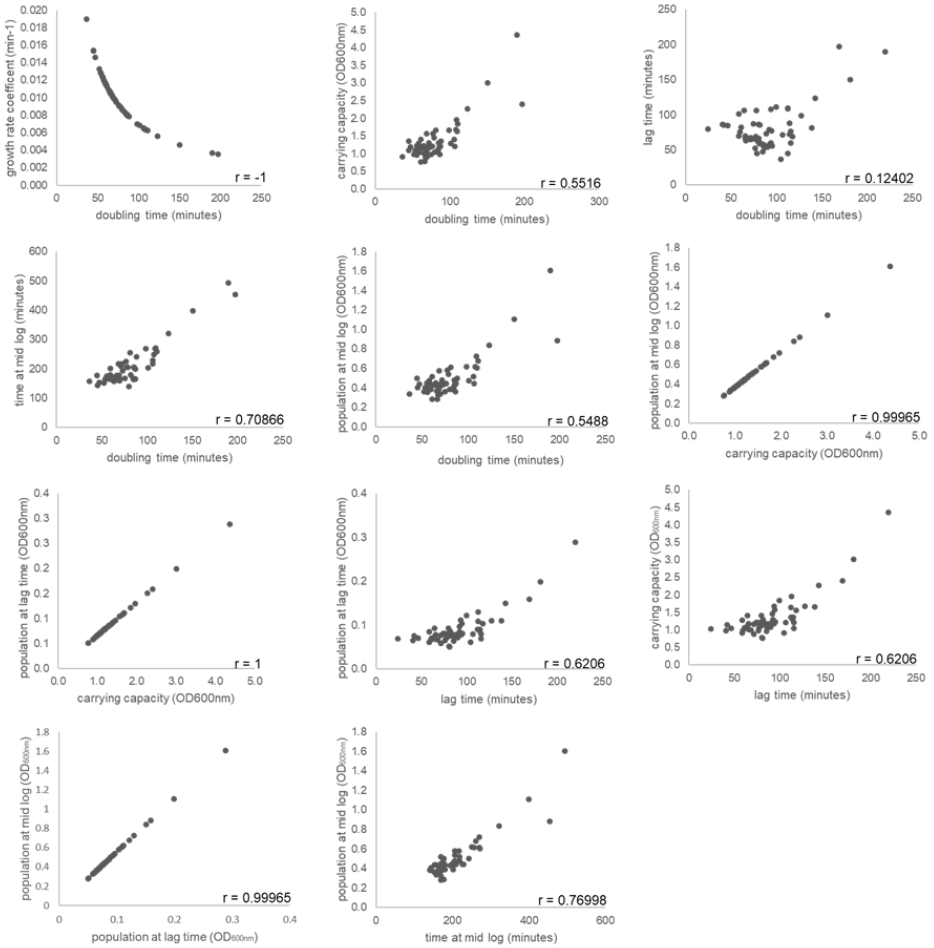


Figure 4.3 Comparisons of Gompertz growth curve metrics for experimental data from 58 individual *E. coli* AB360 populations. The growth curve metrics are plotted in a pairwise fashion to identify correlations between metrics. For example, in the panel comparing growth rate and doubling time, each point is the growth rate and doubling time obtained from a single population replicate. The metrics are listed on the x and y axis (growth rate coefficient, doubling time, carrying capacity, time at mid-log, lag-time, population at lag time, population at mid-log and the carrying capacity). The Spearman correlation of each panel is reported on the panel above the x-axis. All Spearman rank values had a p value < 0.05.

4.5 Doubling time and T_{midlog} -T_{lag}

Intrinsic growth rate coefficients of the *E. coli* AB360 growth curves were estimated using the Gompertz model Equation [2] and used to calculate the fastest doubling time of the population using Equation [4], which occurred when the population was growing exponentially and far from the carrying capacity (i.e., there are no restrictions on growth) (Rockwood, 2015). The estimated average doubling time for plasmid free AB360 populations was 44 minutes. For single transformant AB360 populations harbouring the mCherry

reporter plasmid the doubling time increased by 41% and for the split T7 plasmids: T7p1, T7p2 and T7p3, the doubling times increased by 44%, 30% and 44%, respectively. For double transformant AB360 populations harbouring Opto1, Opto2 and Opto3 the doubling time increased by 90%, 43% and 57%, respectively (Figure 4.2A).

A one-way analysis of variance test (ANOVA) was used to test whether there was a significant difference in the doubling time estimates of the AB360 populations. *P*-values < 0.05 were considered as significant. The ANOVA test on doubling time estimates did not yield statistically significant variation among the different AB360 populations , $F(7,88) = 0.633$, $p = 0.7274$. A post hoc Tukey's Honestly Significant Difference (HSD) test, at 95% confidence intervals, confirmed that there was no significant difference in doubling times between any of the AB360 populations ($p > 0.05$). This was expected due to the overlapping error bars seen in Figure 4.2A.

The T_{midlog} -T_{lag} times were calculated from the Gompertz model coefficient T_{midlog} and the Zwietering-Gompertz model coefficient T_{lag} [5]. This gave an estimated length of time where the rate of biomass accumulation was at its greatest. All the AB360 transformant populations had a longer period of maximum rate of growth compared to the plasmid free AB360 cultures, which had an estimated average T_{midlog}-T_{lag} time of 64 minutes. Interestingly, the percentage increases in T_{midlog}-T_{lag} for all AB360 populations was also identical to the corresponding percentage increases in doubling time (Figure 4.2C). This is likely because non-linear regression models describe bacterial growth in three distinct phases, log phase, exponential phase, and the stationary phase (or carrying capacity). The doubling time and T_{midlog}-T_{lag} estimates are calculated from parameters in the exponential phase. Thus, if the doubling time increases, the time taken by the exponentially growing population to reach mid-log also increases similarly. Spearman's rank correlation tests were performed to investigate the correlation between doubling times and T_{midlog} and the correlation between doubling times and T_{lag}. A high degree of positive correlation was observed between doubling time and T_{midlog} [Spearman's rank: 0.708] and a weak/no correlation was observed between doubling time and lag time [Spearman's rank: 0.124] (Figure 4.3).

An ANOVA on T_{midlog}-T_{lag} estimates revealed that there was a statistically significant difference in T_{midlog}-T_{lag} between at least two AB360 populations, $F(7,88) = 2.813$, $p = 0.012$. A post hoc Tukey's HSD test was carried out to identify which AB360 populations had significantly different T_{midlog}-T_{lag} estimates from which others and found a significant difference between the plasmid-free AB360 population and the AB360 Opto1 population ($p = 0.0014$, 95% CI = 15.36, 99.48). This could be expected as the plasmid free AB360 population has the smallest T_{midlog}-T_{lag} estimate and the AB360 Opto1 population had the

largest T_{midlog}-T_{lag} estimate (Figure 4.2C). There were no other AB360 populations with T_{midlog}-T_{lag} estimates that differed significantly ($p > 0.05$).

Studies investigating plasmid-host interaction consistently observe an unavoidable metabolic burden associated with plasmid replication and foreign gene expression and that growth rate decreases almost linearly with increasing foreign protein content (Bentley et al., 1990). Factors influencing the severity of this metabolic burden include increased ATP synthesis requirements (Rozkov et al., 2004), properties of the recombinant protein and how the protein interacts with the host (Young, Britton, and Robinson., 2012), the rate of transcription and translation (Hoffmann and Rinas., 2004) and host specific properties (Brophy and Voigt. 2014) such as, the capacity for protein production (i.e., the level of physiological adaption of the *E. coli* strains to the expression of foreign DNA).

There are two categories of plasmid-host interactions I aimed to characterise in this Chapter. Firstly, when only one species of plasmid replicates within the AB360 host cells. This is the case for AB360 cells expressing only one of the T7p1, T7p2 and T7p3 plasmids or only the mCherry reporter plasmid. These cultures were grown in non-inducing conditions (i.e., the *araB* regulated OptoT7RNAP plasmids were cultured in the absence of arabinose and the *T7p* regulated mCherry reporter plasmid was not induced by the presence of T7 RNAP). Therefore, unlike an expression vector during the induction phase these plasmids were not overexpressing any GOI and if any GOI expression occurred this was due to leaky induction of the T7p promoter. Results presented in Chapter 3, Figure 3.4, demonstrated that the T7p promoter is stringent and AB360 cells expressing the mCherry reporter plasmid have very low levels of basal fluorescence. Instead, the majority of foreign genes expressed are those necessary to confer antibiotic immunity, to replicate and to segregate into daughter cells (Million-Weaver and Camps., 2014). In this situation, the factors that most negatively influence host doubling time are plasmid size and copy number (Seo and Bailey., 1985; Smith and Bidochka., 1998). The mCherry reporter plasmid backbone pETM6, with an F1 origin of replication, has a high copy number of ~300-500 copies per chromosome equivalent (Baumschlager et al., 2017) and has a total length of 5811bp. AB360 populations expressing the mCherry reporter plasmid had an estimated average increase in doubling time of 41% compared to the doubling time of plasmid-free AB360 populations. The OptoT7RNAP plasmid backbone pSC101, with a pSC101 origin of replication, has a low copy number ~6-7 copies per chromosome equivalent (Thompson et al., 2018) and a total length of approximately 7345bp. The AB360 populations expressing T7p1 and T7p3 had identical increases in doubling time that was 14% greater than the doubling time of AB360 populations expressing T7p2.

Due to its high copy number, I expected the mCherry reporter plasmid to have the greatest impact on AB360 growth rate, not the low copy number OptoT7RNAP plasmids T7p1 and T7p3. However, some studies have shown that increasing the average copy number or the turnover of RNA improves segregational stability (Paulsson and Ehrenberg et al., 1998; Million-Weaver and Camps., 2014). This is because plasmid copy number stability can be impaired by fluctuating concentrations in RNA and decreases in the rate of plasmid transcription can have severe effects on segregationally stability, which causes excessive metabolic load (Paulsson and Ehrenberg et al., 1998). Perhaps, due to the low copy number of the pSC101 plasmid, the AB360 cells expressing the split T7 plasmids were more susceptible to segregationally instability. This may have led to unequitable plasmid distribution between daughter cells during cell division (Silva, Queiroz and Domigues., 2012; Kramer, 2016). This could have resulted in some daughter cells receiving fewer plasmids and thus, fewer antibiotic resistance gene copies which could slow cell growth in the presence of antibiotic. Or increasing the likelihood of daughter cells in each generation receiving no plasmids. Thus, these cells would perish in the presence of antibiotic and if a significant percentage of daughter cells segregate with no plasmids, the doubling time of the population would increase.

Investigations into cell autofluorescence reported in Chapter 3 (Figure 3.4) found that the autofluorescence of AB360 T7p1 and T7p3 populations reduced by 41% and 42%, respectively. However, the autofluorescence of AB360 T7p2 populations reduced by 37%. I speculated that this reduction in autofluorescence seen for split T7 plasmids in all *E. coli* strains could be related to burden induced shifts in cell metabolism. Therefore, it is interesting that the AB360 T7p2 population had both the smallest reduction in autofluorescence and the smallest increase in doubling time, compared to AB360 populations expressing T7p1 and T7p3. This could suggest that out of the three OptoT7RNAP plasmids T7p2 is the least burdensome.

For the second group of plasmid characterisations, the AB360 host cells harbour more than one species of compatible plasmid, as is the case for AB360 Opto1, Opto2 and Opto3 populations. Despite almost identical plasmid DNA sizes, the change in doubling time for AB360 hosts expressing Opto1 was much greater than AB360 host populations expressing the optimised Opto2 and Opto3; compared to plasmid free AB360 populations. AB360 populations expressing Opto1 had a 90% increase in doubling time. Basal gene expression studies presented in Chapter 3 Figure 3.3, demonstrated that Opto1 had the highest level of basal mCherry expression (over 2723% greater than the basal gene expression seen in Opto3). This was expected as the findings of early recombinant protein production studies, for example da Silva and Bailey (1986), demonstrated that heterologous gene expression

has a significantly greater impact on cell growth rate than plasmid size or copy number. Interestingly, the doubling time for AB360 Opto3 populations increased by 57% and for AB360 Opto2 populations it increased by 43%. Yet, basal mCherry gene expression experiments revealed AB360 populations expressing Opto2 had an average basal mCherry fluorescence 146% greater than AB360 populations expressing Opto3 (Chapter 3 Figure 3.3). As previously mentioned, the percentage increase in doubling time of single-transformant AB360 T7p3 populations is 14% greater than the increase in doubling time of single-transformant AB360 T7p2 populations. Thus, the effect of the single T7 plasmids on doubling time is likely conserved in the Opto2 and Opto3 AB360 populations. When a host harbours two or more plasmids, each plasmid contributes independently to affecting the host growth rate however the combined impact is greater than in a host harbouring a single plasmid (Rhee et al., 1994). This is demonstrated as the increase in doubling-time of double-transformant AB360 populations (i.e., those transformed with Opto1, Opto2 and Opto3) is greater than the increase in doubling time of single-transformant AB360 populations (i.e., those transformed with the mCherry reporter plasmid or T7p1, T7p2 and T7p3).

In addition, the error in the estimates of doubling time increases with increasing number of plasmids (Figure 4.1a). This suggests that the population does not respond homogenously to the metabolic burden introduced by the plasmids. The level of homogeneity of a population has been shown to correlate with the fitness of the population (Thattai and Oudenaarden., 2004). The reducing homogeneity seen in this work may reflect increasing metabolic stress as the smallest error in doubling time is achieved by the plasmid free AB360 population, whereas the greatest error in doubling time is seen in the AB360 Opto1 population.

4.6 Lag time and carrying capacity

The Zwietering-Gompertz model [3] was used to estimate lag time coefficients (Figure 4.2B). The estimated lag time for plasmid free AB360 populations was 76 minutes. In all AB360 transformant populations, except for the AB360 Opto1 population, the lag-time estimates increased or decreased within 10% of 76 minutes. However, the estimated lag time of the AB360 Opto1 population decreased by 44%.

An ANOVA on lag time estimates revealed that there was a statistically significant difference in lag time between at least two AB360 populations, $F(7,88) = 4.057$, $p < 0.000$. A post hoc Tukey's HSD test was carried out to identify which AB360 populations had significantly different lag times from which others and found significant differences between the AB360 Opto1 population and the following populations: the plasmid-free AB360 population ($p = 0.008$, 95% CI = -60.68, -5.48), AB360 mCherry population ($p = 0.037$, 95% CI = -56.20, -

1.00), the AB360 T7p1 population ($p = 0.040$, 95% CI = -57.29, -2.10), the AB360 T7p2 population ($p < 0.000$, 95% CI = -68.61, -13.41), the AB360 T7p3 population ($p = 0.001$, 95% CI = -65.50, -10.30) and the AB360 Opto3 population ($p = 0.005$, 95% CI = 6.96, 62.16). These statistically significant differences between the AB360 Opto1 population and the other AB360 populations is likely due to the above-mentioned large decrease in estimated lag-time which can be seen in Figure 4.2B. As expected, there were no significant differences in lag time estimates between the other AB360 populations ($p > 0.05$).

For most AB360 transformants lag time does not change significantly compared to the plasmid free AB360 populations. This could suggest that the lag time of a population is largely dependent on the uptake of nutrients and the environment the *E. coli* cells are cultured in (Schultz and Kishony., 2013), which was kept consistent between populations. However, there is a large decrease in the estimated lag time of the AB360 Opto1 population. A decrease in lag time would suggest a smaller metabolic burden, but Opto1 is not only a two-plasmid system, but it is shown to have the highest levels of basal mCherry expression (Chapter 3 Figure 3.3). When modelling bacterial growth, lag phase is often more difficult to predict, this is due to the influence of the physiological state of individual cells (Baty and Delignette-Muller., 2004), thus it is not a frequently used metric for studying metabolic burden. Overnight cultures were grown in nutrient rich LB-medium, and a second seed culture was not used as the lag phase needed to be captured for non-linear regression analysis. A second seed culture is normally used when transitioning cells from rich LB medium to minimal M9 medium to prevent metabolic shock. A sudden change in media composition can destabilise bacterial cells and create metabolic noise. If the physiological state of the cells was affected by the change in growth environment, this can trigger stress responses which increase individual cell lag time variability (Swinnen et al., 2004). This may explain the rapid increase and decrease in OD seen in some of the AB360 growth curves (Figure 4.1), due to this phenomenon is it difficult to resolve an accurate lag time for these populations. Another limitation of not using a second seed is increased variability in viable inoculum size. The number of viable cells in the overnight culture used to inoculate the M9 medium could have varied between transformants and between replicate cultures. This may have created variation between lag times. It has been shown that with decreasing inoculum size, lag time increases by a length of time dependent on the variability in lag times between the cells and the maximum specific growth rate of the strain (Baranyi and Pin. 1999). Due to these factors lag time does not correlate well with the metabolic burden imposed by the OptoT7RNAP plasmids on the AB360 populations.

Population carrying capacities were estimated using the Gompertz or Zwietering-Gompertz models [2][3] interchangeably as both models yielded identical carrying capacity coefficients.

For plasmid free AB360 populations the carrying capacity was reached at an OD of 1.16. In all AB360 transformant populations, except for the AB360 mChp single transformant population and the AB360 Opto2 population, the average estimated carrying capacity increased or decreased within 10% of 1.16. For the AB360 population expressing the mCherry reporter plasmid the carrying capacity decreased by 16% and for the AB360 population expressing the Opto2 plasmids the carrying capacity decreased by 14% (Figure 4.2d).

An ANOVA on carrying capacity estimates revealed that there was a statistically significant difference in carrying capacity between at least two AB360 populations, $F(7,88) = 3.050$, $p = 0.006$. A post hoc Tukey's HSD test was carried out to identify which AB360 populations had significantly different mean carrying capacity estimates from which others and found a significant difference between the AB360 mCherry population and the AB360 T7p3 population ($p = 0.024$, 95% CI = 0.019, 0.467), there were no other AB360 populations with mean carrying capacity estimates that differed significantly ($p > 0.05$).

Stable populations have a saturation level characteristic, typically called the carrying capacity, which forms a numerical upper bound on population size (Tsoularis and Wallace., 2002). In *E. coli* the expression of stationary phase-dependent genes is dependent on the growth phase. These genes are tightly regulated and are expressed during post-exponential growth, these genes confer metabolic and cellular resistance to environmental stresses such as oxidative and osmotic stress (Dong et al., 2008). The expression of foreign DNA has been shown to accelerate the depletion of resources and increase the rate of production of harmful secondary metabolites, thus reducing the upper bound the population can reach and thus, reducing yield (Seoane et al. 2009). Plus, increasing plasmid size correlates with decreased maximum cell density and accelerated cell death after the stationary phase (Cheah, Weigand and Stark., 1987). Given this phenomenon I would expect to see the AB360 host populations with the highest level of basal mCherry gene expression reach the saturation phase at a smaller average optical density. This is not observed in the results. Instead for most AB360 transformant populations the carrying capacity did not vary significantly. However, for the single-transformant AB360 population expressing the mCherry plasmid and for the AB360 population expressing the Opto2 plasmids the carrying capacity decreased. Interestingly, these are the AB360 populations that showed the smallest respective increases in doubling time. Spearman's rank tests suggested a moderate correlation between carrying capacity and doubling time [correlation coefficient: 0.5516].

A limitation of using OD_{600} is that absorbance readings cannot distinguish between live and dead cells and so the estimated carrying capacity provides no data on the proportion of cells

that are living. In addition, recombinant proteins have been shown to alter cell morphology and size, which may impact the accuracy of absorbance readings. Furthermore, it is possible that in the stationary phase the absorbance readings at 600 nm could be affected by mCherry fluorescence as the mCherry protein concentration during this phase would be at its greatest. Red fluorescent proteins have been shown to strongly absorb light at 600nm which can artificially inflate cell density readings (Hecht et al. 2016).

4.7 Summary of findings

The consumption of precursors and energy by the synthesis of plasmid-encoded protein induces acute readjustments of metabolic pathways. Changing cellular regulatory protein and enzyme composition, triggering stress responses, and shifting host metabolic activity from growth and cell division to the reorganisation of biomass (Hoffmann and Rinas., 2004). In this Chapter, I used non-linear regression modelling of bacterial growth curves to elucidate the metabolic load placed on the *E. coli* bacterial strain AB360 by the expression of foreign protein from the OptoT7RNAP plasmids. Out of the bacterial growth coefficients investigated, I have found that doubling time seemed to correlate best with metabolic burden in the AB360 host. This is consistent with the literature, as the growth rate coefficient is the most powerful metric to describe differences seen between bacterial populations (Sprouffske and Wagner., 2016). In summary, compared to plasmid free AB360 populations, the doubling times increased ($\geq$ 30%) for all AB360 populations expressing a single plasmid species (i.e., the mCherry reporter plasmid or T7p1, T7p2 and T7p3). This was followed by further increases in the doubling times ($\geq$ 43%) for the corresponding AB360 Opto1, Opto2 and Opto3 populations. Notably, the Opto1 plasmid had the highest level of basal mCherry expression and the largest increase in doubling time. Although not statistically significant these increases in population doubling time do seem to reflect an increasing amount of recombinant DNA and foreign protein. Thus, more work would be needed to identify whether the Null hypothesis can be rejected. The null hypothesis is: increasing recombinant DNA length and foreign protein content has no effect on bacterial growth rate. Conversely, studies have shown that plasmid size and foreign protein production do induce a metabolic burden and thus, do result in growth rate reductions. Furthermore, the homogeneity in population growth rate decreases with increasing number of plasmids and increasing levels of basal mCherry expressions. This may reflect the reduction in population fitness caused by the metabolic burden that plasmid replication and heterologous protein production induces.

Using non-linear regression modelling to estimate doubling time, this Chapter has demonstrated how plasmid maintenance and the expression of plasmid encoded genes can impact bacterial growth dynamics. A key advantage of optogenetic gene expression is

dynamic multichromatic control of multiple genes simultaneously. To achieve this the plasmid-host interaction must be stable, to prevent host modification of the circuit, failure to produce heterologous protein or expulsion of the plasmid by the host (Brophy and Voigt., 2014). Thus, when designing a successful synthetic genetic circuit, considerations should be given to the metabolic burden that will be imposed.

4.8 Future work

In future work, I would improve the accuracy of the non-linear regression analysis by confirming that OD correlates well with AB360 cell numbers. This would be done through plate reader optimisation to improve the accuracy of optical density data and mCherry fluorescence intensity readings. I would produce an average CFU/mL growth curve for each transformant which would serve as a calibration protocol to compare optical density data against. The success of plate reader protocol optimisations would be quantified by how closely the growth rate coefficients obtained from optical density data matched the growth rate coefficients obtained from CFU/mL curves. Then I would resume high-throughput growth rate analysis, with calibrated OD readings that can more easily be related to actual cell count (Stevenson et al., 2016, Beal et al., 2020). Studies have shown that fluorescent proteins can impact accurate absorbance readings (Hecht et al., 2016), thus leaky mCherry protein production from the mCherry plasmid could be artificially inflating absorbance readings at 600 nm and preventing accurate doubling times from being obtained (Meyers, Furtmann and Jose., 2018). Therefore, I would take optical density measurements at 700nm, this would allow for estimations of cell abundance that are unaffected by almost all fluorescent proteins (Hecht et al., 2016). Having done this, I would repeat the non-linear regression analysis and use one-way ANOVA and Tukey HSD post hoc tests to investigate whether statistically significant differences in growth rate parameters are observed.

Chapter 5: Conclusion and future directions

5.1 Conclusion

In this work I have compared the basal performance of the optogenetic gene expression plasmids pDusk, pDawn and the two-plasmid OptoT7RNAP gene expression system. High-throughput techniques were used to characterise basal performance in the *E. coli* strains BL21, DH10β and AB360. The metabolic burden imposed by the two-plasmid OptoT7RNAP platform on AB360 populations was investigated using mathematical modelling of bacterial growth curves. In this final chapter, I outline key features of an optogenetic circuit that I foresee aiding their implementation, on a broader scale, in industry and research.

Ohlendorf coined the term signal polarity for describing the states of gene induction of pDusk and pDawn and the ease of plasmid implementation. My results further illustrated the importance of developing an optogenetic circuit with favourable signal polarity. In the pDusk genetic circuit, GOI expression was induced in the absence of blue-light. This made it difficult or impossible to prevent GOI expression when *E. coli* cells expressing pDusk were stored on agar or propagated for glycerol stocks. In these circumstances GOI expression is best avoided to prevent burden-driven mutations from being propagated in the cell-lines. Ohlendorf inverted the gene induction cassette to create pDawn and GOI expression was induced by blue light. However, this introduced another regulatory protein, cI. The genetic circuits of pDusk and pDawn utilised negative regulation of GOI expression. Thus, stringent suppression of GOI transcription in the non-induced state relied on sufficient and un-fluctuating titres of regulatory protein. As pDusk and pDawn are constitutively expressed the regulatory proteins should be transcribed and translated continuously. However, cellular RNA polymerase and ribosome concentrations are not consistent. Thus, any fluctuations in YF1, FixJ, FixK2 and cI protein titres could introduce bottlenecks in the signal transduction cascade and may have led to increased basal GOI expression.

In the optogenetic system OptoT7RNAP, transcription of the GOI was decoupled from host cell machinery by incorporating a blue-light inducible split T7 RNA polymerase onto a second plasmid under the control of the *araB* promoter. Transcription of the OptoT7RNAP fragments still required host cell RNA polymerase. However, unlike in pDawn, drops in regulatory OptoT7RNAP protein titres did not lead to increased basal GOI expression. This is because the OptoT7RNAP circuit utilised positive gene regulation. Drops in the titres of OptoT7RNAP would only affect GOI transcription in the induced state and because of the high processivity of the T7RNAP enzyme this should not hinder target protein yields. Based on these results positive gene regulation and the decoupling of GOI transcription from host cell machinery seems to be beneficial for improving basal performance.

Transcription of the OptoT7RNAP gene fragments required induction of the *araB* promoter via arabinose. Therefore, the OptoT7RNAP plasmids use both SMI and optogenetic gene expression. As it is difficult to completely prevent blue light reaching cell cultures, I found that the use of chemical induction to drive expression of T7RNAP helped to further reduce basal GOI expression. Thus, OptoT7RNAP demonstrated more stringent control of basal GOI expression compared to pDawn.

Compared to pDusk, the introduction of the bacteriophage promoter pR and bacteriophage repressor cI in the pDawn circuit not only greatly reduced basal GOI expression but allowed for greater GOI expression in the induced phase. Similarly, for the OptoT7RNAP system my results demonstrated that basal activity of the bacteriophage T7 promoter contributed to < 20% of the overall basal mCherry fluorescence and that the majority of basal GOI expression arose from leaky expression of the bacterial *araB* promoter. Bacteriophage repressors and promoters exhibit robust regulation (Lemire, Yehl and Lu., 2018). Thus, I would highlight the advantage of using bacteriophage promoters and repressors for improving the basal performance of genetic circuits.

Another key difference between pDawn and OptoT7RNAP is the mode of optogenetic control. In pDawn, blue light absorption prevents the YF1 protein phosphorylating FixJ initiating a signal cascade that induces GOI expression. In OptoT7RNAP, blue-light absorption of the photo-dimerizing magnets restores T7RNAP functionality driving GOI expression from the T7 promoter. Since the turn of the century, the optogenetic toolkit has been drastically expanded beyond the scope of naturally occurring photoreceptors by the biologically inspired design of synthetic photoreceptors (Moglich and Moffat., 2010; Fraikin et al., 2015). The TCS YF1/FixJ and the Magnets pMag and nMag are both examples of this. The continued development of synthetic photoreceptors is a promising sign that in the future light can be used as a stimulus for biological interrogation on a larger scale in industry and research, as advances in photoreceptor design improve the robustness of optogenetic circuitry. To improve basal performance, photoreceptors that revert quickly to the dark state would be beneficial. My results have shown that the fast-reverting Magnets in Opto3 helped to reduce basal GOI expression by over 90%. However, the cost of using fast-reverting photoreceptors should be considered. For example, since their use by Baumschlager in the OptoT7RNAP system, the fast-reverting Magnets have been further optimised to address a trade-off issue between expression level and binding affinity, as expression levels in the induced state decreased drastically for the fast reverting variants (Kawano et al., 2017). Thus, the new Magnets could be incorporated into the OptoT7RNAP system to continue benefiting from fast dissociation times but increase the maximal yield of target protein in the induced state. In addition, photoreceptors which switch between inactive and active

signalling states in response to different wavelengths could be tested. For example, phytochromes are a family of light-regulated rapidly photoswitchable proteins which switch between active and inactive states in response to red and far-red light (Tabor, Levskaya and Voigt., 2011; Weitzman and Hahn., 2014). Compared to fast-reverting photoreceptor variants, natural or engineered photoswitchable proteins may facilitate faster reversion to the inactive state and improve the basal performance of optogenetic circuits.

Gompertz modelling of AB360 population growth demonstrated that increasing plasmid number increased metabolic burden and therefore, increased population doubling time. Through predicting host behaviour, population growth models can be used to guide the implementation of suitable circuit components (Boada et al., 2018). Thus, mathematical models will aid the design of increasingly complex gene circuits in synthetic biology.

5.2 Limitations

The conclusions drawn by this thesis only describe the basal performance of the optogenetic vectors. My results demonstrate that the basal performance of the Optogenetic system is much better than pDawn and that in this instance, a two-plasmid system allows for more stringent circuit regulation. However, a severe reduction in doubling time was observed for AB360 populations expressing Opto1 which had the highest level of basal mCherry gene expression out of the OptoT7RNAP variants. Demonstrating the impact of plasmid maintenance, plasmid replication and foreign protein production on AB360 metabolism. Considering this, it is possible that pDawn may outperform OptoT7RNAP in the induced phase of bacterial growth; perhaps without the hinderance of a second plasmid, higher target protein titres may be achieved. Further work would be needed to characterise the metabolic burden and level of GOI expression from pDawn and OptoT7RNAP in the induced phase and identify whether a two-plasmid optogenetic system is better overall.

5.3 Future directions for multichromatic gene expression control

Optogenetic regulation is highly advantageous for the design of synthetic circuits with fully reversible modulation capabilities. This allows for more precise temporal control of gene expression (Tabor, Levskaya and Voigt., 2011; Ohlendorf et al., 2016; Baumschlager at al., 2017) and dynamic gene expression studies, which are better suited to investigating host-protein or host-circuit interactions than conventional and often unidirectional chemical induction systems. Multichromatic regulation of gene expression provides more advanced control of synthetic and natural gene regulatory networks. Where light of different wavelengths and corresponding photoreceptors are used to regulate the expression of multiple genes. This would be especially useful, for example in metabolic engineering studies or the study of signal transduction cascades. A key aim of this project was to outline

key circuit components of the pDawn and OptoT7RNAP optogenetic circuits which improve basal performance. Basal performance is an important characteristic of a multichromatic gene regulatory system. A high level of basal gene expression would create noise and reduce the precise temporal control needed to resolve circuit and protein behaviour. Future directions of this project, if time permitted, would utilise the high-throughput techniques discussed to design and screen multichromatic gene expression systems with low basal activity. In an effort to engineer a robust system, the synthetic circuit design would take into consideration the metabolic burden introduced by each genetic component implemented.

The use of bacteriophage genetic components in both the pDawn and OptoT7RNAP vectors, decoupling GOI transcription from native host transcription machinery and inducing T7RNAP expression via arabinose in the OptoT7RNAP system was fundamental to reducing basal gene expression. For these reasons, I would use T7RNAP to control GOI expression. The results of this study revealed minimal to no crosstalk between T7RNAP and the non-orthogonal synthetic T7 promoter “CGG”. Directed evolution T7RNAP variants which show highly specific recognition of their orthogonal synthetic promoters could be used to control the expression of up to six GOI in parallel (Meyer, Ellefson and Ellington., 2015). Each of these T7RNAP variants could be split and linked to photoreactive dimers. This thesis has demonstrated the importance of fast-reverting photoreceptors; thus dimers would be selected based on their reverting-time or photoswitchable capabilities. Importantly, the activation spectra of the photoreactive dimers must not overlap to allow for precise multichromatic control of individual GOI (Schmidl et al., 2014). I would use heterodimeric proteins to prevent dimerization of the same T7RNAP fragment (Baumschlager et al., 2017) and ensure dimerization events lead to reconstituting the split T7RNAPs. For example, phytochrome B variants could be used with different phytochrome-interacting factors e.g. PIF3 or PIF6 (Chia, Lee and Tong. 2022). To reduce the metabolic burden associated with plasmid maintenance and replication, the light inducible T7RNAP gene fragments could be stably integrated into the host genome, eliminating the need for a second plasmid. To initiate transcription of the heterodimerizing split-T7RNAP fragments in a multichromatic optogenetic circuit an inducible promoter would be used. Fluorescence data presented in this thesis demonstrated that the majority of basal mCherry fluorescence occurred due to leaky expression of the OptoT7RNAP fragments from the *araB* promoter. Thus, using constitutive promoters to initiate OptoT7RNAP fragment transcription would likely lead to higher levels of basal GOI expression. There are many SMI promoters and repressors used for GOI expression regulation (Topp and Gallivan. 2007). For a multichromatic optogenetic circuit using T7RNAP driven GOI transcription, I would choose a SMI promoter with a weak affinity for RNA polymerase and the sigma factor (Mikhaylichenko et al., 2018) in the interest of

preventing gross overexpression of the T7RNAP protein fragments. A single weak promoter could be used to induce the expression of all photoactivable T7RNAP's, provided that the promoter had a low level of basal expression and there was minimal interference in the activation wavelengths of the chosen photoreceptors. Using only one SMI promoter would be beneficial because a genetic cicuit employing multiple regulators could be overly burdensome to the host, reducing the reliability of regulatory performance and increasing the risk of cross-reactions or interference (Zhang and Voigt., 2018).

In recent years there have been many advancements in the methods of synthetic circuit design which has enabled the engineering of highly complex genetic circuits. For example, the genetic sequences of the T7RNAP gene fragments could be inverted, along with their regulatory elements, to create alternating directions of transcription. I would do this to help prevent incomplete termination and consequently, reduce the number of non-functional mRNA transcripts. Thus, conserving host resources and helping to alleviate metabolic burden. Furthermore, I would use genetic insulators to reduce context dependence and improve circuit robustness. Genetic insulators, such as the self-cleaving ribozyme RiboJ (Clifton et al., 2018), allow for increased GOI transcript abundance and thus, greater target protein titres. The use of RiboJ should also improve basal performance and predictability of multichromatic optogenetic circuit behaviour.

Multichromatic optogenetic systems will aid our understanding of the spatiotemporal and highly dynamic roles of genes and proteins, in an approach that well exceeds the capabilities of chemical manipulation. Successful engineering of these circuits will undoubtedly be benefited by comparative studies on current optogenetic approaches. In this thesis I have compared the different designs, functions, and basal performance of two optogenetic tools using high-throughput screening and mathematical modelling. The data obtained has provided insight into the optogenetic approaches used to circumvent the challenges of basal gene expression and burden in synthetic biology, and may help to further expand the usability of optogenetic control.

Bibliography

Allen, R.J. and Waclaw, B. (2018). Bacterial growth: a statistical physicist's guide. *Reports on Progress in Physics*, *82*(1), p.016601.

Angius, F., Ilioaia, O., Amrani, A., Suisse, A., Rosset, L., Legrand, A., Abou-Hamdan, A., Uzan, M., Zito, F. and Miroux, B. (2018). A novel regulation mechanism of the T7 RNA polymerase based expression system improves overproduction and folding of membrane proteins. *Scientific Reports*, *8*(1), pp.1-11.

Aronson, D.E., Costantini, L.M. and Snapp, E.L. (2011). Superfolder GFP is fluorescent in oxidizing environments when targeted via the Sec translocon. *Traffic*, *12*(5), pp.543-548.

Bamberg, E., Gärtner, W. and Trauner, D. (2018). Introduction: Optogenetics and photopharmacology. *Chemical Reviews*, *118*(21), pp.10627-10628.

Baneyx, F. (1999). Recombinant protein expression in Escherichia coli. *Current Opinion in Biotechnology*, *10*(5), pp.411-421.

Bao, N., Jagadeesan, B., Bhunia, A.K., Yao, Y. and Lu, C. (2008). Quantification of bacterial cells based on autofluorescence on a microfluidic platform. *Journal of Chromatography A*, *1181*(1-2), pp.153-158.

Baranyi, J. and Pin, C. (1999). Estimating bacterial growth parameters by means of detection times. *Applied and Environmental Microbiology*, *65*(2), pp.732-736.

Baranyi, J., Pin, C. and Ross, T. (1999). Validating and comparing predictive models. *International Journal of Food Microbiology*, *48*(3), pp.159-166.

Baty, F. and Delignette-Muller, M.L. (2004). Estimating the bacterial lag time: which model, which precision?. *International Journal of Food Microbiology*, *91*(3), pp.261-277.

Baumschlager, A., Aoki, S.K. and Khammash, M. (2017). Dynamic blue light-inducible T7 RNA polymerases (Opto-T7RNAPs) for precise spatiotemporal gene expression control. *ACS Synthetic Biology*, *6*(11), pp.2157-2167.

Baumschlager, A., Rullan, M. and Khammash, M. (2020). Exploiting natural chemical photosensitivity of anhydrotetracycline and tetracycline for dynamic and setpoint chemo-optogenetic control. *Nature Communications*, *11*(1), pp.1-10.

Beal, J., Farny, N.G., Haddock-Angelli, T., Selvarajah, V., Baldwin, G.S., Buckley-Taylor, R., Gershater, M., Kiga, D., Marken, J., Sanchania, V. and Sison, A. (2020). Robust estimation of bacterial cell count from optical density. *Communications Biology*, *3*(1), pp.1-29.

Bentley, W.E., Mirjalili, N., Andersen, D.C., Davis, R.H. and Kompala, D.S. (1990). Plasmid-encoded protein: the principal factor in the “metabolic burden” associated with recombinant bacteria. *Biotechnology and Bioengineering*, *35*(7), pp.668-681.

Berntsson, O., Diensthuber, R.P., Panman, M.R., Björling, A., Gustavsson, E., Hoernke, M., Hughes, A.J., Henry, L., Niebling, S., Takala, H. and Ihalainen, J.A. (2017). Sequential conformational transitions and α-helical supercoiling regulate a sensor histidine kinase. *Nature Communications*, *8*(1), pp.1-8.

Bhat, S.A., Iqbal, I.K. and Kumar, A. (2016). Imaging the NADH: NAD+ homeostasis for understanding the metabolic response of Mycobacterium to physiologically relevant stresses. *Frontiers in Cellular and Infection Microbiology*, *6*, p.145.

Boada, Y., Vignoni, A., Oyarzún, D. and Picó, J. (2018). Host-circuit interactions explain unexpected behavior of a gene circuit. *IFAC-PapersOnline*, *51*(19), pp.86-89.

Bondza-Kibangou, P., Millot, C., Dufer, J. and Millot, J.M. (2001). Microspectrofluorometry of autofluorescence emission from human leukemic living cells under oxidative stress. *Biology of the Cell*, *93*(5), pp.273-280.

Borkowski, O., Ceroni, F., Stan, G.B. and Ellis, T. (2016). Overloaded and stressed: whole-cell considerations for bacterial synthetic biology. *Current Opinion in Microbiology*, *33*, pp.123-130.

Bowers, L.M., LaPoint, K., Anthony, L., Pluciennik, A. and Filutowicz, M. (2004). Bacterial expression system with tightly regulated gene expression and plasmid copy number. *Gene*, *340*(1), pp.11-18.

Bremer H., Dennis P.P. (1987). Modulation of chemical composition and other parameters of the cell by growth rate. In Escherichia coli and Salmonella typhimurium: *Cellular and Molecular Biology* pp. 1527–1542

Brophy, J.A. and Voigt, C.A. (2014). Principles of genetic circuit design. *Nature Methods*, 11(5), pp.508-520.

Cabeen, M.T. and Jacobs-Wagner, C. (2005). Bacterial cell shape. *Nature Reviews Microbiology*, *3*(8), pp.601-610.

Cai, G., Lin, Z. and Shi, S. (2022). Development and Expansion of the CRISPR/Cas9 Toolboxes for Powerful Genome Engineering in Yeast. *Enzyme and Microbial Technology*, p.110056.

Cao, Y.Q., Li, Q., Xia, P.F., Wei, L.J., Guo, N., Li, J.W. and Wang, S.G. (2017). AraBAD based toolkit for gene expression and metabolic robustness improvement in Synechococcus elongatus. *Scientific Reports*, *7*(1), pp.1-10.

Carneiro, S., Ferreira, E.C. and Rocha, I. (2013). Metabolic responses to recombinant bioprocesses in Escherichia coli. *Journal of Biotechnology*, 164(3), pp.396-408.

Castillo-Hair, S.M., Baerman, E.A., Fujita, M., Igoshin, O.A. and Tabor, J.J. (2019). Optogenetic control of Bacillus subtilis gene expression. *Nature Communications*, 10(1), pp.1-11.

Ceroni, F., Boo, A., Furini, S., Gorochowski, T.E., Borkowski, O., Ladak, Y.N., Awan, A.R., Gilbert, C., Stan, G.B. and Ellis, T. (2018). Burden-driven feedback control of gene expression. *Nature Methods*, *15*(5), pp.387-393.

Cheah, U.E., Weigand, W.A. and Stark, B.C. (1987). Effects of recombinant plasmid size on cellular processes in Escherichia coli. *Plasmid*, *18*(2), pp.127-134.

Cheetham, G.M. and Steitz, A.T.A., (1999). Structure of a transcribing T7 RNA polymerase initiation complex. *Science*, *286*(5448), pp.2305-2309.

Chevallier, V., Andersen, M.R. and Malphettes, L. (2020). Oxidative stress-alleviating strategies to improve recombinant protein production in CHO cells. *Biotechnology and Bioengineering*, *117*(4), pp.1172-1186.

Clifton, K.P., Jones, E.M., Paudel, S., Marken, J.P., Monette, C.E., Halleran, A.D., Epp, L. and Saha, M.S. (2018). The genetic insulator RiboJ increases expression of insulated genes. *Journal of Biological Engineering*, *12*(1), pp.1-6.

Cotlet, M., Goodwin, P.M., Waldo, G.S. and Werner, J.H. (2006). A comparison of the fluorescence dynamics of single molecules of a green fluorescent protein: one-versus two-photon excitation. *ChemPhysChem*, *7*(1), pp.250-260.

Crosson, S., Rajagopal, S. and Moffat, K. (2003). The LOV domain family: photoresponsive signaling modules coupled to diverse output domains. *Biochemistry*, *42*(1), pp.2-10.

da Silva, N.A. and Bailey, J.E. (1986). Theoretical growth yield estimates for recombinant cells. *Biotechnology and Bioengineering*, *28*(5), pp.741-746.

Dahl, R.H., Zhang, F., Alonso-Gutierrez, J., Baidoo, E., Batth, T.S., Redding-Johanson, A.M., Petzold, C.J., Mukhopadhyay, A., Lee, T.S., Adams, P.D. and Keasling, J.D. (2013). Engineering dynamic pathway regulation using stress-response promoters. *Nature Biotechnology*, *31*(11), pp.1039-1046.

De Vos, D., Bruggeman, F.J., Westerhoff, H.V. and Bakker, B.M. (2011). How molecular competition influences fluxes in gene expression networks. *PloS One*, *6*(12), p.e28494.

Di Crescenzo, A. and Paraggio, P. (2019). Logistic growth described by birth-death and diffusion processes. *Mathematics*, *7*(6), p.489.

Diaz Ricci, J.C. and Hernández, M.E. (2000). Plasmid effects on Escherichia coli metabolism. *Critical Reviews in Biotechnology*, *20*(2), pp.79-108.

Dong, T., Kirchhof, M.G. and Schellhorn, H.E. (2008). RpoS regulation of gene expression during exponential growth of Escherichia coli K12. *Molecular Genetics and Genomics*, *279*(3), pp.267-277.

Du, F., Liu, Y.Q., Xu, Y.S., Li, Z.J., Wang, Y.Z., Zhang, Z.X. and Sun, X.M. (2021). Regulating the T7 RNA polymerase expression in E. coli BL21 (DE3) to provide more host options for recombinant protein production. *Microbial Cell Factories*, *20*(1), pp.1-10.

Dubendorff, J.W. and Studier, F.W. (1991). Creation of a T7 autogene: Cloning and expression of the gene for bacteriophage T7 RNA polymerase under control of its cognate promoter. *Journal of Molecular Biology*, *219*(1), pp.61-68.

Durbin, R., (1999). Gene expression systems based on bacteriophage T7 RNA polymerase. *Gene Expression Systems*, pp.9-44.

Durfee, T., Nelson, R., Baldwin, S., Plunkett III, G., Burland, V., Mau, B., Petrosino, J.F., Qin, X., Muzny, D.M., Ayele, M. and Gibbs, R.A. (2008). The complete genome sequence of Escherichia coli DH10B: insights into the biology of a laboratory workhorse. *Journal of Bacteriology*, *190*(7), pp.2597-2606.

Filipiak, W., Sponring, A., Filipiak, A., Baur, M., Ager, C., Wiesenhofer, H., Margesin, R., Nagl, M., Troppmair, J. and Amann, A. (2013). Volatile organic compounds (VOCs) released by pathogenic microorganisms in vitro: Potential breath biomarkers. *Volatile Biomarkers; Elsevier: Amsterdam, The Netherlands*, p.463.

Fraikin, G.Y., Strakhovskaya, M.G., Belenikina, N.S. and Rubin, A.B. (2015). Bacterial photosensory proteins: Regulatory functions and optogenetic applications. *Microbiology*, *84*(4), pp.461-472.

Frenzel, E., Legebeke, J., Van Stralen, A., Van Kranenburg, R. and Kuipers, O.P. (2018). In vivo selection of sfGFP variants with improved and reliable functionality in industrially important thermophilic bacteria. *Biotechnology for Biofuels*, *11*(1), pp.1-19.

Glick, B.R. (1995). Metabolic load and heterologous gene expression. *Biotechnology Advances*, *13*(2), pp.247-261.

Gorochowski, T.E., Van Den Berg, E., Kerkman, R., Roubos, J.A. and Bovenberg, R.A. (2014). Using synthetic biological parts and microbioreactors to explore the protein expression characteristics of Escherichia coli. *ACS Synthetic Biology*, *3*(3), pp.129-139.

Goswitz, V.C. and Brooker, R.J. (1993). Isolation of lactose permease mutants which recognize arabinose. *Membrane Biochemistry*, *10*(1), pp.61-70.

Greco, F.V., Pandi, A., Erb, T.J., Grierson, C.S. and Gorochowski, T.E. (2021). Harnessing the central dogma for stringent multi-level control of gene expression. *Nature Communications*, 12(1), pp.1-11.

Greco, F.V., Irvine, T., Grierson, C.S. and Gorochowski, T.E. (2022). Design and Assembly of Multilevel Transcriptional and Translational Regulators for Stringent Control of Gene Expression. *Methods in Molecular Biology* (Clifton, NJ), 2518, pp.99-110.

Gur, E. (2013). The Lon AAA+ protease. *Regulated Proteolysis in Microorganisms*, pp.35-51.

Gyorgy, A. (2021). Context-Dependent Stability and Robustness of Genetic Toggle Switches with Leaky Promoters. *Life*, *11*(11), p.1150.

Halmi, M.I.E., Shukor, M.S., Johari, W.L.W. and Shukor, M.Y. (2014). Evaluation of several mathematical models for fitting the growth of the algae Dunaliella tertiolecta. *Asian Journal of Plant Biology*, 2(1), pp.1-6.

Hawley, T.S., Hawley, R.G. and Telford, W.G. (2017). Fluorescent proteins for flow cytometry. *Current Protocols in Cytometry*, *80*(1), pp.9-12.

Hecht, A., Endy, D., Salit, M. and Munson, M.S. (2016). When wavelengths collide: bias in cell abundance measurements due to expressed fluorescent proteins. *ACS Synthetic Biology*, *5*(9), pp.1024-1027.

Heikal, A., Nakatani, Y., Dunn, E., Weimar, M.R., Day, C.L., Baker, E.N., Lott, J.S., Sazanov, L.A. and Cook, G.M. (2014). Structure of the bacterial type II NADH dehydrogenase: a monotopic membrane protein with an essential role in energy generation. *Molecular Microbiology*, *91*(5), pp.950-964.

Ho, J.M., Miller, C.A., Parks, S.E., Mattia, J.R. and Bennett, M.R. (2021). A suppressor tRNA-mediated feedforward loop eliminates leaky gene expression in bacteria. *Nucleic Acids Research*, *49*(5), pp.e25-e25.

Hoffmann, F. and Rinas, U. (2004). Stress induced by recombinant protein production in Escherichia coli. *Physiological Stress Responses in Bioprocesses*, pp.73-92.

Hui, C.Y., Guo, Y., Zhang, W. and Huang, X.Q. (2018). Rapid monitoring of the target protein expression with a fluorescent signal based on a dicistronic construct in Escherichia coli. *AMB Express*, *8*(1), pp.1-10.

Il Rhee, J., Diaz Ricci, J.C., Bode, J. and Schügerl, K. (1994). Metabolic enhancement due to plasmid maintenance. *Biotechnology Letters*, *16*(9), pp.881-884.

Jones, K.L., Kim, S.W. and Keasling, J.D. (2000). Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria. *Metabolic Engineering*, *2*(4), pp.328-338.

Justice, S.S., Hunstad, D.A., Cegelski, L. and Hultgren, S.J. (2008). Morphological plasticity as a bacterial survival strategy. *Nature Reviews Microbiology*, *6*(2), pp.162-168.

Kaberniuk, A.A., Shemetov, A.A. and Verkhusha, V.V. (2016). A bacterial phytochrome-based optogenetic system controllable with near-infrared light. *Nature Methods*, *13*(7), pp.591-597.

Kawano, F., Suzuki, H., Furuya, A. and Sato, M. (2015). Engineered pairs of distinct photoswitches for optogenetic control of cellular proteins. *Nature Communications*, 6(1), pp.1-8.

Korkmaz, M. (2021). A study over with four-parameter Logistic and Gompertz growth models. *Numerical Methods for Partial Differential Equations*, *37*(3), pp.2023-2030.

Kramer, M.G. (2016). Determination of plasmid segregational stability in a growing bacterial population. *Bacterial Therapy of Cancer*, pp. 125-133. Humana Press, New York, NY.

Kurokawa, M. and Ying, B.W. (2017). Precise, high-throughput analysis of bacterial growth. *JoVE (Journal of Visualized Experiments)*, (127), p.e56197.

Lalwani, M.A., Ip, S.S., Carrasco-López, C., Day, C., Zhao, E.M., Kawabe, H. and Avalos, J.L. (2021). Optogenetic control of the lac operon for bacterial chemical and protein production. *Nature Chemical Biology*, *17*(1), pp.71-79.

Lemire, S., Yehl, K.M. and Lu, T.K., 2018. Phage-based applications in synthetic biology. *Annual Review of Virology*, *5*(1), p.453.

Liao, C., Blanchard, A.E. and Lu, T. (2017). An integrative circuit–host modelling framework for predicting synthetic gene network behaviours. *Nature Microbiology*, *2*(12), pp.1658-1666.

Lim, H.K., Jung, K.H., Park, D.H. and Chung, S.I. (2000). Production characteristics of interferon-α using an L-arabinose promoter system in a high-cell-density culture. *Applied Microbiology and Biotechnology*, *53*(2), pp.201-208.

Lou, C., Stanton, B., Chen, Y.J., Munsky, B. and Voigt, C.A. (2012). Ribozyme-based insulator parts buffer synthetic circuits from genetic context. *Nature Biotechnology*, *30*(11), pp.1137-1142.

Marisch, K., Bayer, K., Cserjan-Puschmann, M., Luchner, M. and Striedner, G. (2013). Evaluation of three industrial Escherichia coli strains in fed-batch cultivations during high-level SOD protein production. *Microbial Cell Factories*, 12(1), pp.1-11.

Mesa, S., Hauser, F., Friberg, M., Malaguti, E., Fischer, H.M. and Hennecke, H. (2008). Comprehensive assessment of the regulons controlled by the FixLJ-FixK2-FixK1 cascade in Bradyrhizobium japonicum. *Journal of Bacteriology*, *190*(20), pp.6568-6579.

Meyer, A.J., Ellefson, J.W. and Ellington, A.D. (2015). Directed evolution of a panel of orthogonal T7 RNA polymerase variants for in vivo or in vitro synthetic circuitry. *ACS Synthetic Biology*, *4*(10), pp.1070-1076.

Meyers, A., Furtmann, C. and Jose, J. (2018). Direct optical density determination of bacterial cultures in microplates for high-throughput screening applications. *Enzyme and Microbial Technology*, *118*, pp.1-5.

Mikhaylichenko, O., Bondarenko, V., Harnett, D., Schor, I.E., Males, M., Viales, R.R. and Furlong, E.E., 2018. The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. *Genes & Development*, *32*(1), pp.42-57.

Milias-Argeitis, A., Rullan, M., Aoki, S.K., Buchmann, P. and Khammash, M. (2016). Automated optogenetic feedback control for precise and robust regulation of gene expression and cell growth. *Nature Communications*, *7*(1), pp.1-11.

Million-Weaver, S. and Camps, M. (2014). Mechanisms of plasmid segregation: have multicopy plasmids been overlooked?. *Plasmid*, *75*, pp.27-36.

Miroux, B. and Walker, J.E. (1996). Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of Molecular Biology*, *260*(3), pp.289-298.

Möglich, A. and Moffat, K. (2010). Engineered photoreceptors as novel optogenetic tools. *Photochemical & Photobiological Sciences*, *9*(10), pp.1286-1300.

Molinari, S., Tesoriero Jr, R.F. and Ajo-Franklin, C.M. (2021). Bottom-up approaches to engineered living materials: Challenges and future directions. *Matter*, 4(10), pp.3095-3120.

Morrow, J.F., Cohen, S.N., Chang, A.C., Boyer, H.W., Goodman, H.M. and Helling, R.B. (1974). Replication and Transcription of Eukaryotic DNA in Esherichia coli. *Proceedings of the National Academy of Sciences*, *71*(5), pp.1743-1747.

Mullis, K.B., Ferre, F., Gibbs, R.A., (1994). *The Polymerase Chain Reaction*. 1st edn. Birkhauser Boston, MA.

Myers, J.A., Curtis, B.S. and Curtis, W.R. (2013). Improving accuracy of cell and chromophore concentration measurements using optical density. *BMC Biophysics*, *6*(1), pp.1-16.

Oehler, S., Eismann, E.R., Krämer, H. and Müller-Hill, B. (1990). The three operators of the lac operon cooperate in repression. *The EMBO Journal*, *9*(4), pp.973-979.

Ohlendorf, R., Vidavski, R.R., Eldar, A., Moffat, K. and Möglich, A. (2012). From dusk till dawn: one-plasmid systems for light-regulated gene expression. *Journal of Molecular Biology*, *416*(4), pp.534-542.

Olson, E.J. and Tabor, J.J. (2014). Optogenetic characterization methods overcome key challenges in synthetic and systems biology. *Nature Chemical Biology*, *10*(7), pp.502-511.

Padan, E.T.A.N.A., Hunte, C. and Reilander, H. (2003). Production and purification of recombinant membrane proteins. *Membrane Protein Purification and Crystallization*, pp.55-83.

Paliy, O. and Gunasekera, T.S. (2007). Growth of E. coli BL21 in minimal media with different gluconeogenic carbon sources and salt contents. *Applied Microbiology and Biotechnology*, *73*(5), pp.1169-1172.

Pan, S.H. and Malcolm, B.A. (2000). Reduced background expression and improved plasmid stability with pET vectors in BL21 (DE3). *Biotechniques*, *29*(6), pp.1234-1238.

Pasini, M., Fernández-Castané, A., Jaramillo, A., De Mas, C., Caminal, G. and Ferrer, P. (2016). Using promoter libraries to reduce metabolic burden due to plasmid-encoded proteins in recombinant Escherichia coli. *New Biotechnology*, *33*(1), pp.78-90.

Pathak, G.P., Strickland, D., Vrana, J.D. and Tucker, C.L. (2014). Benchmarking of optical dimerizer systems. *ACS Synthetic Biology*, *3*(11), pp.832-838.

Paulsson, J. and Ehrenberg, M. (1998). Trade-off between segregational stability and metabolic burden: a mathematical model of plasmid ColE1 replication control. *Journal of Molecular Biology*, *279*(1), pp.73-88.

Pla, M.L., Oltra, S., Esteban, M.D., Andreu, S. and Palop, A. (2015). Comparison of primary models to predict microbial growth by the plate count and absorbance methods. *BioMed Research International*. 2015:365025, pp.1-14.

Ramirez-Cando, L.J., Alvarez-Mendoza, C.I. and Gutierrez-Salazar, P., 2018. Verhulst-Pearl growth model versus Malthusian growth model for in vitro evaluation of lead removal in wastewater by Photobacterium sp. *F1000Research*, 7(491), p.491.

Ratkowsky, D.A. (1993). Principles of nonlinear regression modelling. *Journal of Industrial Microbiology*, *12*(3), pp.195-199.

Renggli, S., Keck, W., Jenal, U. and Ritz, D. (2013). Role of autofluorescence in flow cytometric analysis of Escherichia coli treated with bactericidal antibiotics. *Journal of Bacteriology*, *195*(18), pp.4067-4073.

Rizzo, M.G., De Plano, L.M. and Franco, D. (2020). Regulation of filamentation by bacteria and its impact on the productivity of compounds in biotechnological processes. *Applied Microbiology and Biotechnology*, *104*(11), pp.4631-4642.

Robin, G.P., Kleemann, J., Neumann, U., Cabre, L., Dallery, J.F., Lapalu, N. and O'Connell, R.J. (2018). Subcellular localization screening of Colletotrichum higginsianum effector candidates identifies fungal proteins targeted to plant peroxisomes, golgi bodies, and microtubules. *Frontiers in Plant Science*, *9*, p.562.

Rockwood, L.L. (2015). *Introduction to Population Ecology*. John Wiley & Sons.

Rosano, G.L. and Ceccarelli, E.A. (2014). Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in Microbiology*, *5*, p.172.

Rosano, G.L., Morales, E.S. and Ceccarelli, E.A. (2019). New tools for recombinant protein production in Escherichia coli: A 5-year update. *Protein Science*, *28*(8), pp.1412-1422.

Rozkov, A., Avignone-Rossa, C.A., Ertl, P.F., Jones, P., O'Kennedy, R.D., Smith, J.J., Dale, J.W. and Bushell, M.E. (2004). Characterization of the metabolic burden on Escherichia coli DH1 cells imposed by the presence of a plasmid containing a gene therapy sequence. *Biotechnology and Bioengineering*, *88*(7), pp.909-915.

Samuelson, J.C. (2011). Recent developments in difficult protein expression: a guide to E. coli strains, promoters, and relevant host mutations. *Heterologous Gene Expression in E. coli*, pp.195-209.

Schmidl, S.R., Sheth, R.U., Wu, A. and Tabor, J.J. (2014). Refactoring and optimization of light-switchable Escherichia coli two-component systems. *ACS Synthetic Biology*, *3*(11), pp.820-831.

Schultz, D. and Kishony, R., 2013. Optimization and control in bacterial lag phase. *BMC Biology*, *11*(1), pp.1-3.

Segall-Shapiro, T.H., Meyer, A.J., Ellington, A.D., Sontag, E.D. and Voigt, C.A. (2014). A 'resource allocator'for transcription based on a highly fragmented T7 RNA polymerase. *Molecular Systems Biology*, *10*(7), p.742.

Seo, J.H. and Bailey, J.E. (1985). Effects of recombinant plasmid content on growth properties and cloned gene product formation in Escherichia coli. *Biotechnology and Bioengineering*, *27*(12), pp.1668-1674.

Seoane, J., Sin, G., Lardon, L., Gernaey, K.V. and Smets, B.F. (2010). A new extant respirometric assay to estimate intrinsic growth parameters applied to study plasmid metabolic burden. *Biotechnology and Bioengineering*, *105*(1), pp.141-149.

Shaner, N.C., Campbell, R.E., Steinbach, P.A., Giepmans, B.N., Palmer, A.E. and Tsien, R.Y. (2004). Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nature Biotechnology*, *22*(12), pp.1567-1572.

Silva, F., Queiroz, J.A. and Domingues, F.C. (2012). Evaluating metabolic stress and plasmid stability in plasmid DNA production by Escherichia coli. *Biotechnology Advances*, *30*(3), pp.691-708.

Smalla, K., Jechalke, S. and Top, E.M. (2015). Plasmid detection, characterization, and ecology. *Microbiology Spectrum*, *3*(1), pp.3-1.

Smith, M.A. and Bidochka, M.J. (1998). Bacterial fitness and plasmid loss: the importance of culture conditions and plasmid size. *Canadian Journal of Microbiology*, *44*(4), pp.351-355.

Snapp, E. (2005). Design and use of fluorescent fusion proteins in cell biology. *Current Protocols in Cell Biology*, *27*(1), pp.21-4.

Sørensen, H.P. and Mortensen, K.K. (2005). Advanced genetic strategies for recombinant protein expression in Escherichia coli. *Journal of Biotechnology*, *115*(2), pp.113-128.

Sousa, R., Mukherjee, S. and Kivie, M. (2003). T7 RNA polymerase. *Progress in Nucleic Acid Research and Molecular Biology*, 73(73), pp.1-41.

Sousa, R. and Zaher, H. (2021). Transcription: T7 RNA polymerase. *Encyclopedia of Biological Chemistry: Third Edition 3*(4) pp. 352-357.

Sprouffske, K. and Wagner, A. (2016). Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinformatics*, *17*(1), pp.1-4.

Stevenson, K., McVey, A.F., Clark, I.B., Swain, P.S. and Pilizota, T. (2016). General calibration of microbial growth in microplate readers. *Scientific Reports*, *6*(1), pp.1-7.

Studier, F.W. and Moffatt, B.A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology*, *189*(1), pp.113-130.

Su, W.W. (2005). Fluorescent proteins as tools to aid protein production. *Microbial Cell Factories*, *4*(1), pp.1-6.

Surre, J., Saint-Ruf, C., Collin, V., Orenga, S., Ramjeet, M. and Matic, I. (2018). Strong increase in the autofluorescence of cells signals struggle for survival. *Scientific Reports*, *8*(1), pp.1-14.

Swinnen, I.A.M., Bernaerts, K., Dens, E.J., Geeraerd, A.H. and Van Impe, J.F. (2004). Predictive modelling of the microbial lag phase: a review. *International Journal of Food Microbiology*, *94*(2), pp.137-159.

Tabor, S. (2001). Expression Using the T7 RNA Polymerase/Promoter System. *Current Protocols in Molecular Biology*. 16.2.1– 16.2.11

Tabor, J.J., Levskaya, A. and Voigt, C.A. (2011). Multichromatic control of gene expression in Escherichia coli. *Journal of Molecular Biology*, *405*(2), pp.315-324.

Tateno, H., Kuno, A., Itakura, Y. and Hirabayashi, J. (2010). A versatile technology for cellular glycomics using lectin microarray. *Methods in Enzymology* (Vol. 478, pp. 181-195). Academic Press.

Temme, K., Hill, R., Segall-Shapiro, T.H., Moser, F. and Voigt, C.A. (2012). Modular control of multiple pathways using engineered orthogonal T7 polymerases. *Nucleic Acids Research*, *40*(17), pp.8773-8781.

Thattai, M. and Van Oudenaarden, A. (2004). Stochastic gene expression in fluctuating environments. *Genetics*, *167*(1), pp.523-530.

Thompson, M.G., Sedaghatian, N., Barajas, J.F., Wehrs, M., Bailey, C.B., Kaplan, N., Hillson, N.J., Mukhopadhyay, A. and Keasling, J.D. (2018). Isolation and characterization of novel mutations in the pSC101 origin that increase copy number. *Scientific Reports*, *8*(1), pp.1-11.

Tjørve, K.M. and Tjørve, E. (2017). The use of Gompertz models in growth analyses, and new Gompertz-model approach: An addition to the Unified-Richards family. *PloS One*, *12*(6), p.e0178691.

Topp, S. and Gallivan, J.P. (2007). Guiding bacteria with small molecules and RNA. *Journal of the American Chemical Society*, *129*(21), pp.6807-6811.

Tsoularis, A. and Wallace, J. (2002). Analysis of logistic growth models. *Mathematical Biosciences*, *179*(1), pp.21-55.

Wages Jr, J.M. (2005). Polymerase chain reaction. *Encyclopedia of Analytical Science*, p.243.

Wang, R.C. and Levine, B. (2010). Autophagy in cellular growth control. *FEBS Letters*, *584*(7), pp.1417-1426.

Weitzman, M. and Hahn, K.M. (2014). Optogenetic approaches to cell migration and beyond. *Current Opinion in Cell Biology*, *30*, pp.112-120.

Westover, K.D., Bushnell, D.A. and Kornberg, R.D. (2004). Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell,* 119(4), pp.481-489.

Wu, P., Chen, Y., Liu, M., Xiao, G. and Yuan, J. (2020). Engineering an optogenetic CRISPRi platform for improved chemical production. *ACS Synthetic Biology*, *10*(1), pp.125-131.

Xia, Y. (2020). Correlation and association analyses in microbiome study integrating multiomics in health and disease. *Progress in Molecular Biology and Translational Science*, *171*, pp.309-491.

Xu, P., Vansiri, A., Bhan, N. and Koffas, M.A. (2012). ePathBrick: a synthetic biology platform for engineering metabolic pathways in E. coli. *ACS Synthetic Biology*, *1*(7), pp.256-266.

Young, C.L., Britton, Z.T. and Robinson, A.S. (2012). Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications. *Biotechnology Journal*, *7*(5), pp.620-634.

Zhang, S. and Voigt, C.A. (2018). Engineered dCas9 with reduced toxicity in bacteria: implications for genetic circuit design. *Nucleic Acids Research*, *46*(20), pp.11115-11125.

Zhao, E.M., Zhang, Y., Mehl, J., Park, H., Lalwani, M.A., Toettcher, J.E. and Avalos, J.L., (2018). Optogenetic regulation of engineered cellular metabolism for microbial chemical production. *Nature*, *555*(7698), pp.683-687.

Zwietering, M.H., Rombouts, F.M. and Riet, K.V.T. (1992). Comparison of definitions of the lag phase and the exponential phase in bacterial growth. *Journal of Applied Bacteriology*, *72*(2), pp.139-145.

Supplementary information

Supplementary information S.1: Primers and oligos

Name/purpose	Primer sequence	№ bp	GC%	Tm°C *
GFP FWD primer with NdeI overhang	GGTCCATATGCGTAAAGG TGAAGAACTGTTC	31	45.16	70.0
GFP REV primer with HindIII overhang	GGTCAAGCTTATTATTTGT ACAGTTCGTCCATACCG	36	41.67	71.0
T7 replacement promoter oligo (grey) with AvrII and BamHI overhangs	GTAGCCTAGGTAATACCG GTCACTATAGGGAGAGGA TCCTGTC	43	51.16	66.9
Sequencing primer for confirming T7 promoter replacement	GGTGATGTCGGCGATATA GGC	21	57.14	57.1

Primers and oligos were ordered from IDT inc. and resuspended in nuclease free water to a working stock concentration of 10 ng/µL.

Supplementary information S.2: Equipment

Equipment	Model	Manufacturer/Supplier	Use
Stuart SI500 shaking incubator	JI500	Cole-Palmer Ltd.	Culture incubation
Mini centrifuge for Eppendorf tubes	MiniSpin®plus	Eppendorf ®	DNA preparation
Nanodrop™	OneC UV-Vis Spectrophotometer	Thermo Fisher Scientific™	DNA quantification
Thermocycler	C1000	BIO-RAD inc.	PCR amplification
Plate reader	Synergy neo2 multi-mode reader	BioTek©	Bacterial growth experiments and fluorescence measurements

Gen5 microplate reader and imager software	Version 3.04	BioTek©	Bacterial growth experiments and fluorescence measurements
Gel tank	Mini-Sub Cell GT Cell 1704466	BIO-RAD inc.	DNA gel electrophoresis
PowerPac™ Basic Power Supply	1645050	BIO-RAD inc.	DNA gel electrophoresis
UVP MultiDoc-It™ Gel Imaging System	UVP 97019301	Fisher scientific®	DNA gel imaging
LSRFortessa™ X-20 Cell Analyzer	X-20	BD Biosciences®	FACS cell sorting
BD FACSDiva™ Software	Version 9.0	BD Biosciences®	FACS cell sorting
Power pack	HY3003	Digimess®	Blue light apparatus
Autoclave	210047	Prestige Medical BDSI	Sterilisation of media
0.2µm filters	9921-2502	GE healthcare Whatman Uniflow	Filter sterilisation
20mL syringe leur lock	IVL20	Medicina®	Filter sterilisation
Petri dishes	263991	Nunc™	Agar plates
Glass bottles (100mL-500mL)	Duran® original	DWK life sciences	Media and buffers
96 well conical bottom plates	PlateOne® S18339610	Starlabs ltd.	Plate reader growth and fluorescence experiments
24 well plates	PlateOne®	Starlabs ltd.	Blue light expression experiments
Microplate adhesive PCR plate covers	MicroAmp™3345	Fisher scientific ®	To protect overnight 96-

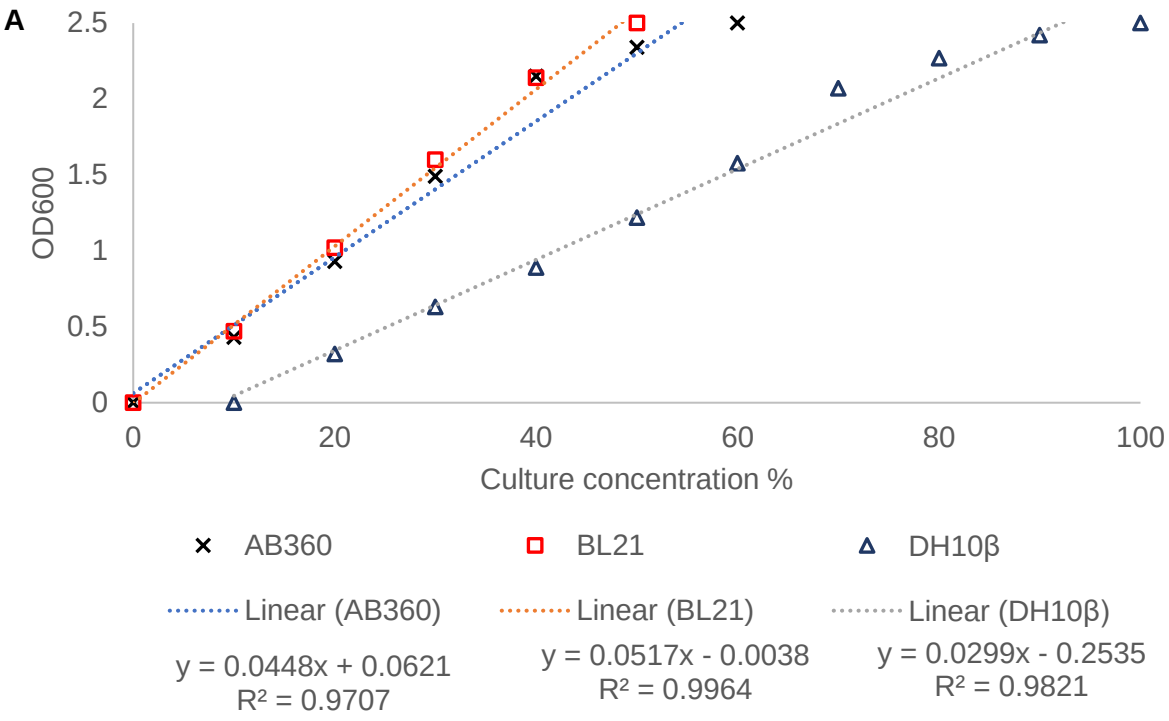
breathable polypropylene			well plates from contamination
Microplate adhesive aluminium PCR plate foils	ABgene AB0626	Thermo Scientific™	Microplate transport to the plate reader
1.5mL microcentrifuge tubes	A0	Fisherbrand™	DNA preparation
Cryogenic vials for long-term storage	AY509X33	Thermo Scientific™	Bacterial glycerol stocks
Centrifuge tubes 50 mL	734-0451	Falcon® Corning life sciences	Media and buffer preparation
PCR tubes 0.2 mL	AI402-3700	Starlabs ltd.	PCR reactions
Sterile culture tubes	1219G07	Cultubes™	Overnight cultures and propagation
Pipette set (0.2-1000 µL)	Finnpipette™ F2 GLP kits	Thermo Scientific™	Liquid handling
Pipette tips	TipOne®	Starlabs ltd.	Liquid handling

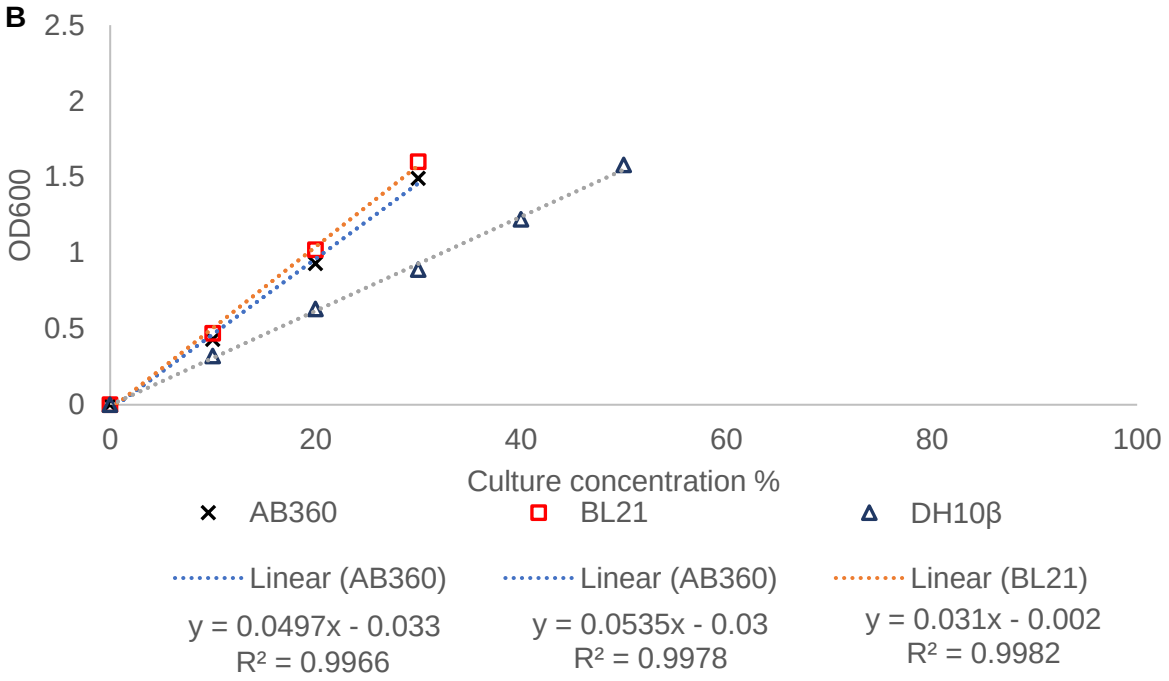
Supplementary information S.3. *E. coli* strains and reagents

Name	Cat №	Manufacturer/Supplier	Use
E. coli DH5α	C2987I	NEB inc.	Bacterial cell cloning strain
E. coli BL21	C2530H	NEB inc.	Bacterial cell expression strain
E. coli DH10β	C3019I	NEB inc.	Bacterial cell cloning strain
LB Broth Miller	1003233066	Thermo Fisher Scientific™	Bacterial growth medium
LB Agar Lennox	22700025	Thermo Fisher Scientific™	Bacterial growth medium
M9 Minimal salts base 5X	1002620708	Sigma-Aldrich®	Bacterial growth medium

Glucose	101783140	Sigma-Aldrich®	LB broth and M9 minimal medium
Thiamine hydrochloride	1002355050	Sigma-Aldrich®	M9 minimal medium
Casamino acid powder	J851-1KG	VWR Life	M9 minimal medium
Magnesium sulphate	101928273	Sigma-Aldrich®	M9 minimal medium
Calcium chloride	1002356226	Sigma-Aldrich®	M9 minimal medium
Arabinose sugar	101717751	Sigma-Aldrich®	Expression experiments
SOC Medium	10129367	NEBBIOLABS	Bacterial cell transformation
Glycerol	327255000	ACROS	Glycerol cryogenic stocks
Kanamycin	KBO286	BIOBASIC	Plasmid maintenance
Ampicillin	A1593-25G	Sigma-Aldrich®	Plasmid and AB360 strain maintenance
Chloramphenicol	1002348528	Sigma-Aldrich®	Plasmid maintenance
Gentamicin	15750060	Thermo Fisher Scientific™	Plasmid maintenance
Potassium acetate, 99%	A16321.36	Thermo Fisher Scientific™	For CCMB80 buffer, making competent cells
Calcium chloride anhydrous 93%	012316.A7	Thermo Fisher Scientific™	For CCMB80 buffer, making competent cells
Manganese (II) chloride, 97%	271412500	Thermo Fisher Scientific™	For CCMB80 buffer, making competent cells
Magnesium chloride hexahydrate, 99%	447155000	Thermo Fisher Scientific™	For CCMB80 buffer, making competent cells
Phosphate buffered saline	BR0014G	Oxoid™	Preserve bacterial cells and arrest growth
Monarch® Plasmid miniprep kit	T1010S	NEB inc.	Plasmid miniprep
Q5®High-Fidelity 2X Master Mix	M0492S	NEB inc.	PCR amplification
10x CutSmart™Buffer	B6004S	NEB inc.	Double restriction digestion buffer

HindIII	R3104	NEB inc.	DNA restriction digestion enzyme
NdeI	R0111S	NEB inc.	DNA restriction digestion enzyme
BamHI	R0136S	NEB inc.	DNA restriction digestion enzyme
EcoRI-HF®	R3101S	NEB inc.	DNA restriction digestion enzyme
SpeI-HF®	R3133	NEB inc.	DNA restriction digestion enzyme
Agarose powder	BP160-100	Thermo Fisher Scientific™	Gel electrophoresis
Tris-acetate-EDTA 10x solution	BP1335500	Fisher scientific®	Gel electrophoresis
1 kb Plus DNA Ladder	N3200L	NEB inc.	DNA gel electrophoresis
6x Gel Loading Dye, Purple	B7024S	NEB inc.	DNA gel electrophoresis
SYBR™ green 10,000x concentrate nucleic acid gel stain	S7585	Thermo Fisher Scientific™	DNA gel electrophoresis
Monarch® DNA gel extraction kit	T1020S	NEB inc.	DNA extraction form gel
T4 DNA ligase	M0202S	NEB inc.	DNA ligation
T4 DNA ligase reaction buffer 10X	B0202S	NEB inc.	DNA ligation
Monarch® PCR and DNA Cleanup kit	T1030S	NEB inc.	PCR product clean up
Quick-load® OneTaq® 2X master mix	M0488S	NEB inc.	Colony PCR
LongAmp® Hot Start Taq 2X Master Mix	M0533S	NEB inc.	PCR amplification





Supplementary information S.4: A) Absorbance calibration curve for *E. coli* strains AB360, BL21 and DH10β up to a culture OD_{600} of 2.5 The spectrophotometer cannot read optical density at a OD_{600} of >2.5. **B) Linear part of the calibration curve for *E. coli* strains AB360, BL21 and DH10β up to an OD_{600} limit of ~1.5.** OD_{600} measurements greater than >1.5 are outside the linear range and may be inaccurate **A, B)** Equations and R^2 values are displayed on the graph. LB-media was inoculated from glycerol stocks and incubated at 37°C, 250rpm shaking, for 14 hours. A culture concentration of 100% represents 1mL of saturated bacterial culture.

Supplementary information S.5: Comparison of optical density of *E. coli* AB360, BL21 and DH10β plasmid free cultures and cultures expressing different plasmid variants of the OptoT7RNAP optogenetic expression system

		AB360		BL21		DH10β	
		Max OD[a]	±RSD[b]	Max OD[a]	±RSD[b]	Max OD[a]	±RSD[b]
no plasmid		1.175	13.7	1.420	3.7	1.197	6.2
mChp		0.977	21.2	1.027	9.4	0.998	19.9
T7p-1		1.148	19.3	1.295	20.9	1.111	16.7
T7p-2		1.190	14.5	1.226	20.5	1.087	9.9
T7p-3		1.208	15.1	1.254	17.5	1.106	11.7
Opto1[c]	0.00%	1.104	10.7	1.179	13.7	1.014	1.9
	0.05%	1.192	10.5	1.464	2.7	0.900	10.8
	0.10%	1.279	10.2	1.472	4.0	1.228	4.6
	0.20%	1.360	1.5	1.301	17.4	1.098	4.9
Opto2[c]	0.00%	0.986	8.0	1.194	11.6	0.966	4.1
	0.05%	1.101	15.0	1.447	0.4	0.772	7.1
	0.10%	1.134	1.2	1.310	9.5	1.260	4.0
	0.20%	1.176	4.7	1.065	5.6	0.780	6.5
Opto3[c]	0.00%	1.045	10.0	1.322	5.4	1.004	2.5
	0.05%	1.202	3.0	1.450	1.1	0.828	14.0
	0.10%	0.963	16.3	1.400	10.8	1.136	2.5
	0.20%	1.206	2.8	1.193	14.5	0.899	12.7

* Averages were calculated from 12 biological replicates.

[a] Mean maximum optical density at 600 nm

[b] Relative standard deviation (or coefficient of variation) expressed as a percentage

[c] *E. coli* strains expressing the Opto1, Opto2 and Opto3 expression systems were cultured at different concentrations of arabinose (%)

Supplementary information S.6: Comparison of basal mCherry fluorescence intensity of *E. coli* AB360, BL21 and DH10β plasmid free cultures and cultures expressing different plasmid variants of the OptoT7RNAP optogenetic expression system

		AB360		BL21		DH10β	
		Max FI[a]	±RSD[b]	Max FI[a]	±RSD[b]	Max FI[a]	±RSD[b]
no plasmid		27.9	33.3	24.7	36.4	21.6	25.0
mChp		278.9	44.1	239.3	46.8	180.7	52.7
T7p-1		17.5	28.3	18.3	35.4	13.3	26.3
T7p-2		16.3	30.9	19.9	27.3	13.5	25.6
T7p-3		16.6	27.0	19.5	30.2	15.2	25.9
Opto1[c]	0.00%	3436.3	28.8	1458.0	55.9	4082.7	50.6
	0.05%	7098.3	7.5	3328.3	23.3	4860.7	2.7
	0.10%	2526.7	41.8	1768.3	23.4	5230.0	17.4
	0.20%	9322.7	5.8	2083.0	28.9	4903.7	13.9
Opto2[c]	0.00%	1459.2	52.3	975.7	22.2	1264.1	45.1
	0.05%	2368.3	4.4	1590.3	4.6	1707.8	20.3
	0.10%	981.0	14.3	1028.7	19.0	1564.3	5.6
	0.20%	2986.0	10.6	978.0	47.6	1447.0	43.7
Opto3[c]	0.00%	42871.3	46.9	38254.1	38.2	55157.0	34.6
	0.05%	99993.7	6.0	66820.3	7.4	66820.3	32.8
	0.10%	30393.3	20.8	42803.0	39.3	70341.3	16.2
	0.20%	92161.3	21.6	60632.3	32.1	60632.8	7.9

* Averages were calculated from 12 biological replicates.

[a] Mean maximum mCherry fluorescence (a.u.)

[b] Relative standard deviation (or coefficient of variation) expressed as a percentage.

[c] *E. coli* strains expressing the Opto1, Opto2 and Opto3 expression systems were cultured at different concentrations of arabinose (%)

Supplementary information S.7: Comparison of optical density and basal mCherry fluorescence of *E. coli* AB360, BL21 and DH10β cultures expressing the Opto3 plasmids with the "CGG" T7 promoter variant

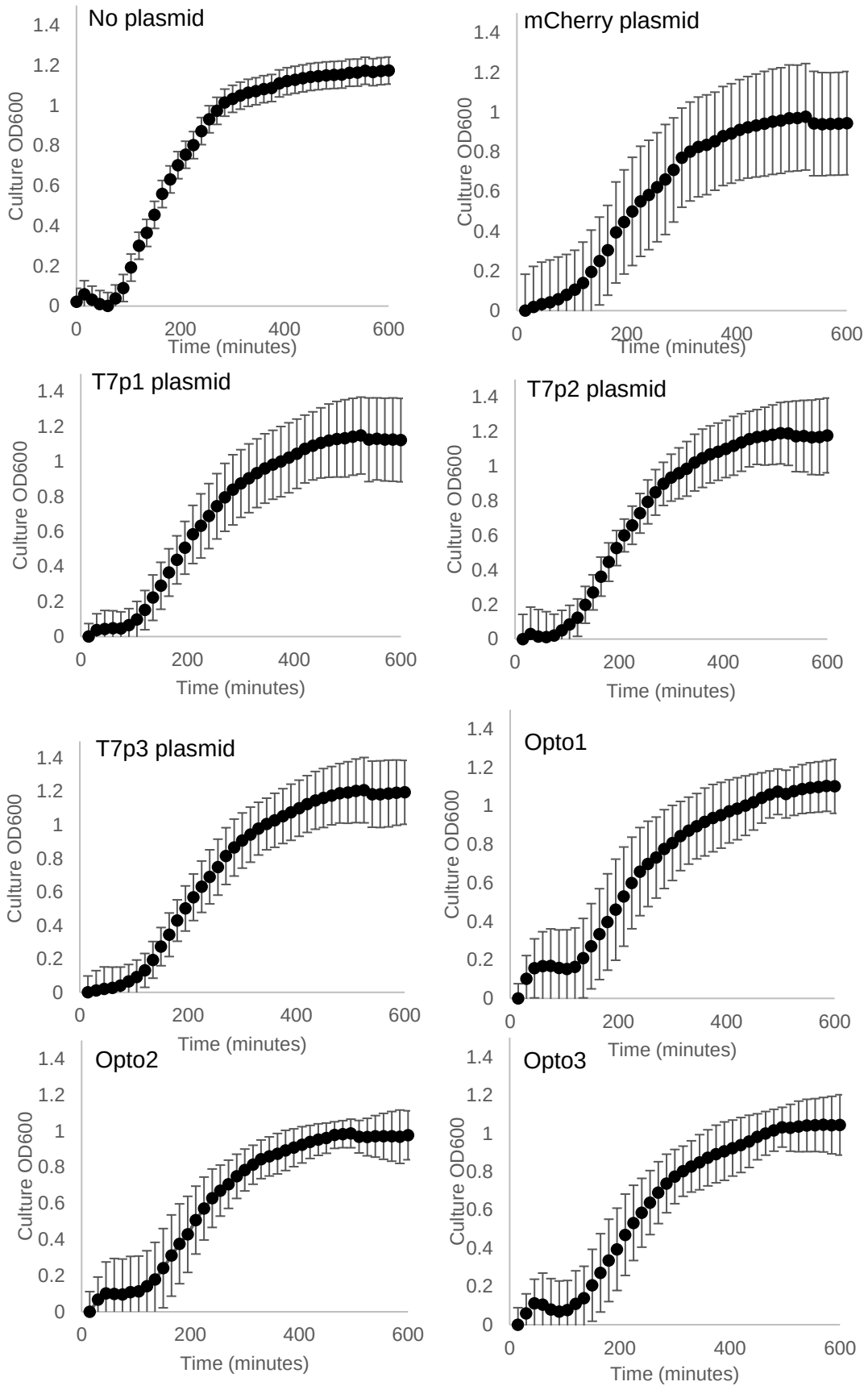
	AB360		BL21		DH10β	
	Max OD[a]	±RSD[c]	Max OD[a]	±RSD[c]	Max OD[a]	±RSD[c]
Opto-3-RepT7	0.687	18.6	0.651	15.1	0.596	19.1
	Max FI[b]	±RSD[c]	Max FI[b]	±RSD[c]	Max FI[b]	±RSD[c]
	251.3	48.1	219.3	61.3	211.7	57.9

* Averages were calculated from 12 biological replicates.

[a] Mean maximum optical density at 600 nm

[b] Mean maximum mCherry fluorescence (a.u.).

[c] Relative standard deviation (or coefficient of variation) expressed as a percentage.



Supplementary information S.8: *E. coli* strain AB360 and AB360 transformant average growth curves. Averages were calculated from a minimum of 12 individual growth curves and error bars represent ± SD.

Supplementary information S.9 Non-linear regression modelling of bacterial growth data using R

##Packages used

```
install.packages("nnls")

install.packages("minpack.lm")
```

##Read in data

```
file_name <- "C:/Users/ti17744/OneDrive - University of Bristol/Documents/Rgrowthcurver/All_GC.csv"

All_GC <- read.csv(file_name, header = TRUE, sep = ",", stringsAsFactors = FALSE)

View(All_GC)
```

Normalise OD600 readings

```
d_loop <- All_GC[, c("Time","A1")]

min_value <- min(d_loop[, "A1"])

d_loop[, "A51"] <- d_loop[,"A1"] - min_value

d_loop[,"A1"]
```

Pearl-Verhulst equation

```
Pearl_Verhulst <- nlsLM(d_loop[,"A1"]~K/(1+((K-No)/No)*exp(-r*All_GC$Time)), data=All_GC, start=list(K=1.479,No=0.216, r=0.016), control=list(maxiter=500))

coef(Pearl_Verhulst)

summary(Pearl_Verhulst)

AIC(Pearl_Verhulst)
```

Gompertz equation

```
Model_Tinf <- nlsLM(d_loop[,"A1"]~ A*exp(-exp(-KG*(All_GC$Time-Tinf))), data=All_GC, start=list(KG=0.006, A=1.049, Tinf=395))

Model_Tinf

summary(Model_Tinf)

AIC(Model_Tinf)
```

Example output

```
> summary(Model_Tinf)

Formula: GC$A1 ~ A * exp(-exp(-KG * (GC$Time - Tinf)))

Parameters:
      Estimate Std. Error t value Pr(>|t|)
KG   1.396e-02  3.351e-04   41.65   <2e-16 ***
A    1.160e+00  5.476e-03  211.84   <2e-16 ***
Tinf 1.469e+02  1.236e+00  118.84   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01962 on 38 degrees of freedom

Number of iterations to convergence: 10
Achieved convergence tolerance: 1.49e-08

> AIC(Model_Tinf)
[1] -201.1181
```

where KG is the rate coefficient at midlog, A is the carrying capacity and Tinf us the time at inflection (i.e., the time at midlog).

Zwietering Gompertz equation

```
Zweitering_Gompertz <- nlsLM(d_loop[,"A1"]~A*exp(-exp(exp(1)*r/A*(Tlag-All_GC$Time)+1)), data=All_GC, start=list(A=1.049, r=0.019, Tlag=0.018), control=list(maxiter=500))

Zweitering_Gompertz

summary(Zweitering_Gompertz)

AIC(Zweitering_Gompertz)
```

to plot a growth curve

```
plot(All_GC$Time , d_loop[,"A1"] , ylim=c(0,1.5), ylab= "Culture OD[600]", xlab= "Time(minutes)", title("e.g. AB360 no plasmid"))

lines(All_GC$Time, predict(Four_parameter_Gompertz), col = "Grey")

lines(All_GC$Time, predict(Model_Tinf), col = "red"
```

##Packages used to compare metrics obtained from my non-linear regression analysis

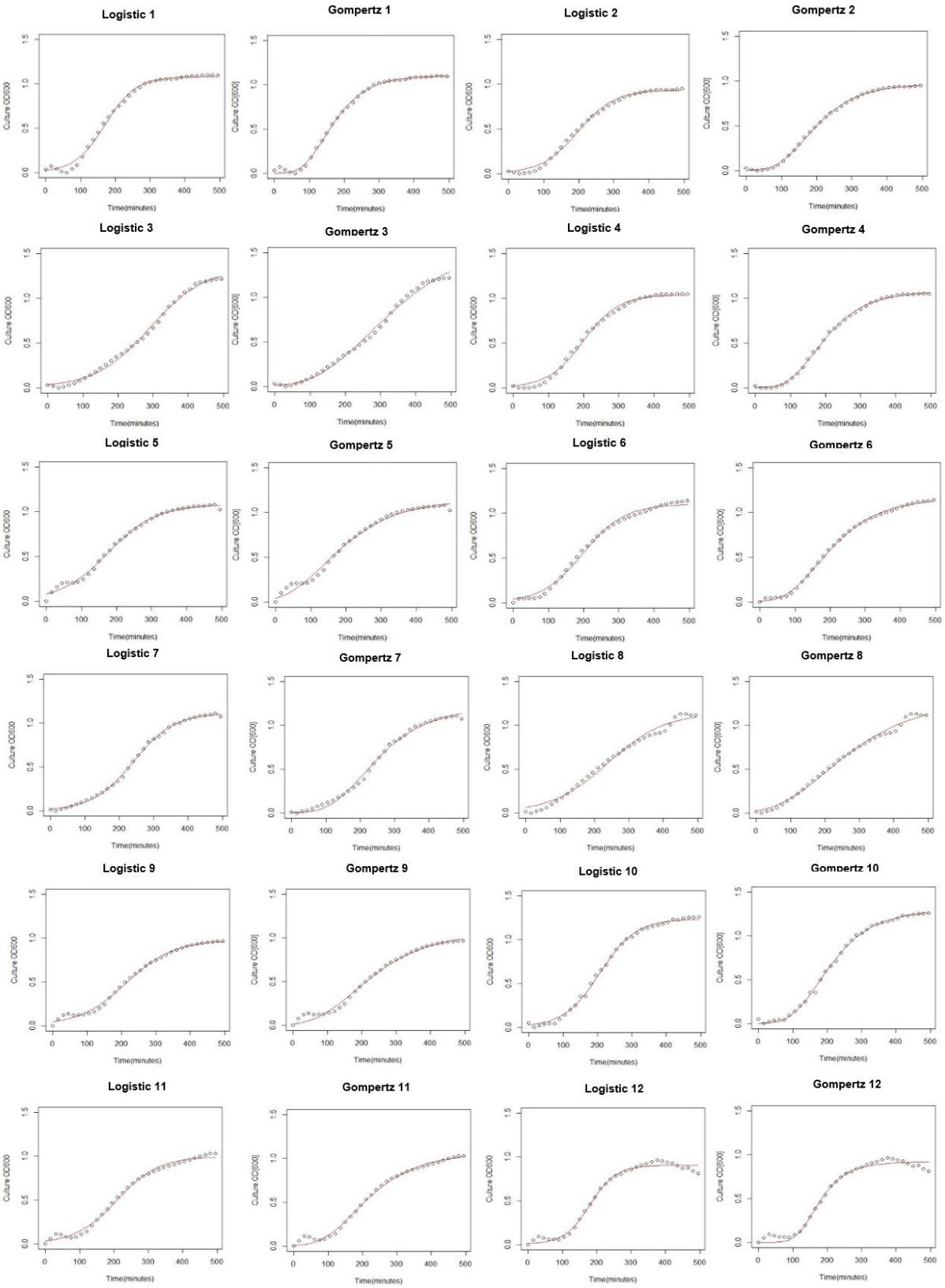
```
Install.packages(“growthcurver”)
```

An R package for logistical modelling of microbial data absorbance data, obtained from a plate reader in a high throughput fashion (Sprouffske and Wagner., 2016)

```
install.packages("growthrates")
```

R package used to estimate growth rates from experimental data, including absorbance data by fitting a linear regression to a subset of data with the steepest log-linear increase (Petzoldt., 2022)

```
#Analyse all growth data using growthcurver

library(growthcurver)

file_name <- "C:/Users/ti17744/OneDrive - University of Bristol/Documents/Rgrowthcurver/ALL_GC.csv"

All_Curves <- read.csv(file_name, header = TRUE, sep = ",", stringsAsFactors = FALSE)

View(All_Curves)

gc_out <- SummarizeGrowthByPlate(All_Curves, plot_fit = TRUE, plot_file = ("ALL_CURVE_data.pdf"))

gc_out
```



Supplementary information S.10: **Growth curves of 12 sample cultures (OD_{600} over time) and the fitted logistic and Gompertz models.** Average growth curves were calculated from 12 biological replicates.

Supplementary information S.11: Gompertz coefficient estimates for the carrying capacity, time at midlog and maximum growth rate of the *E. coli* strain AB360 and AB360 transformant average growth curves

	Carrying capacity				Time at midlog				Rate coefficient				Description of model fit	
	Estimate	std. error[a]	t-value[b]	p value[c]	Estimate	std. error[a]	t-value[b]	p value[c]	Estimate	std. error[a]	t-value[b]	p value[c]	AIC Value[d]	RSE 38 d.f. [e]
no plasmid	1.155	0.0518	222.84	<0.05	146.9	16.04	118.28	<0.05	1.6 x 10 -2	1.3 x 10 -1	38.91	<0.05	-201.9	0.0194
mCh	0.975	0.0528	184.8	<0.05	160.9	13.540	118.89	<0.05	1.1 x 10 -2	6.2 x 10 -2	44.67	<0.05	-221.0	0.0154
T71	1.149	0.0558	206.09	<0.05	164.2	12.010	136.75	<0.05	1.1 x 10 -2	7.1 x 10 -2	51.18	<0.05	-219.1	0.0158
T72	1.188	0.0429	277.21	<0.05	166.4	8.970	185.55	<0.05	1.2 x 10 -2	8.8 x 10 -2	63.31	<0.05	-233.9	0.0132
T73	1.218	0.0436	279.25	<0.05	172.4	8.681	198.6	<0.05	1.1 x 10 -2	9.8 x 10 -2	70.5	<0.05	-241.42	0.0120
Opto1	1.144	0.0164	69.962	<0.05	163.7	17.25	47.46	<0.05	8.3 x 10 -3	3.1 x 10 -2	22.16	<0.05	154.9	0.0345
Opto2	0.997	0.0898	111.07	<0.05	159.5	11.285	70.67	<0.05	1.1 x 10 -2	3.8 x 10 -2	27.08	<0.05	-178.0	0.0260
Opto3	1.065	0.0111	96.32	<0.05	177.4	12.31	72.05	<0.05	9.9 x 10 -3	3.7 x 10 -2	26.52	<0.05	-173.5	0.0275

Supplementary information S.12: Zwietering Gompertz coefficient estimates for the carrying capacity, lag time and rate at lag time of the *E. coli* strain AB360 and AB360 transformant average growth curves

	Carrying capacity				Lag time				rate at lag time				Description of model fit	
	Estimate	std. error[a]	t-value[b]	p value[c]	Estimate	std. error[a]	t-value[b]	p value[c]	Estimate	std. error[a]	t-value[b]	p value[c]	AIC Value[d]	RSE 38 d.f.[e]
no plasmid	1.155	0.0518	222.8	<0.05	75.69	4.3	37.22	<0.05	6.0 x 10-3	1.5 x 10 -4	42.6	<0.05	-201.9	0.0194
mCh	0.975	0.0528	184.8	<0.05	71.21	4.6	30.72	<0.05	4.0 x 10 -3	7.6 x 10 -5	52.6	<0.05	-221.0	0.0154
T71	1.149	0.0558	206.1	<0.05	72.30	4.1	35.36	<0.05	4.6 x 10 -3	7.6 x 10 -5	60.7	<0.05	-219.1	0.0158
T72	1.188	0.0429	277.2	<0.05	83.62	3.1	53.92	<0.05	5.3 x 10 -3	7.2 x 10 -5	73.3	<0.05	-233.9	0.0132
T73	1.218	0.0436	279.3	<0.05	80.51	2.9	54.69	<0.05	4.9 x 10 -3	5.8 x 10 -5	84.0	<0.05	-241.42	0.0120
Opto1	1.144	0.0164	70.0	<0.05	42.61	10.8	28.781	<0.05	3.5 x 10 -3	1.2 x 10 -4	28.8	<0.05	154.9	0.0345
Opto2	0.997	0.0898	111.1	<0.05	68.56	7.7	31.94	<0.05	4.0 x 10 -3	1.3 x 10 -4	31.9	<0.05	-178.0	0.0260
Opto3	1.065	0.0111	96.3	<0.05	77.17	8.1	32.55	<0.05	3.9 x 10 -3	1.2 x 10 -4	32.6	<0.05	-173.5	0.0275

…continued (notes for Supplementary Tables S.11 and S.12)

*All fitted models, both Gompertz and Zwietering-Gompertz, had an achieved convergence tolerance of 1.49×10^{-8}

* The Levenberg-Marquardt algorithm (LM) was used to solve non-linear least squares problems for all models, which interpolates between the Guass-Newton algorithm (nls) and the method of gradient descent

a Standard error of the estimate describes how much variation there is around the estimates of the non-linear regression coefficients

b The t value is the test statistic from a two-sided t-test. The larger the test statistic the less likely the results occurred by chance

c The *p* – value describes how likely the corresponding t-value occurred by chance if the null hypothesis of no effect of the parameter were true

d The Akaike information criterion for each model

e The residual standard error describes how well the model fits the data set and degrees of freedom, calculated from the total number of observations.