Applying Optogenetics in *Synechococcus* sp. PCC 7002

Liam Forbes, BSc, MSc

Thesis submitted for the fulfilment of the requirements for the Degree of Doctor of Philosophy

School of Molecular Biosciences
College of Medical, Veterinary and Life Sciences

University of Glasgow

August 2024

Abstract

Cyanobacteria can be utilised in biotechnology as environmentally sustainable cell machines to convert CO_2 into a diverse range of biochemicals. However, a lack of molecular tools available for precise and dynamic control of gene expression hinders metabolic engineering and often contributes to low product titres. Optogenetic tools enable light-regulated control of gene expression with high tunability and reversibility, though their application in cyanobacteria species with high industrial potential is limited. In this study, two different optogenetic systems were characterised in *Synechococcus* sp. PCC 7002, applications for metabolic engineering were tested, and a transcriptomics study was performed to pinpoint novel optogenetic tools. Firstly, the green/red light-responsive CcaS/CcaR optogenetic system originating from *Synechocystis* sp. PCC 6803 was expressed in *Synechococcus* sp. PCC 7002 and performance was monitored by measuring its regulation of Green Fluorescent Protein (GFP). Characterisation of activation and deactivation was performed under a range of different light wavelengths and intensities using both fluorescence spectrometry and qRT-PCR. The performance of the CcaS/CcaR system was further improved by making targeted genetic modifications to the pCpcG2 output promoter. Secondly, the blue light-regulated YF1/FixJ optogenetic system of non-cyanobacterial origin was expressed in *Synechococcus* sp. PCC 7002. The system demonstrated poor performance with minimal increase in system output in response to blue light or darkness. To evaluate the use of optogenetics for metabolic engineering, the optimised CcaS/CcaR system was tested for light-controlled heterologous production of mannitol and L-alanine. While the former was successful, the latter failed despite careful troubleshooting. Finally, to identify new potential optogenetic tools from *Synechococcus* sp. PCC 7002, RNA sequencing was used to identify genes differentially expressed in blue and orange light. A variety of genes were identified, including transcriptional regulators and kinases which could have roles in wavelength-specific signalling pathways. Overall, this work lays a critical foundation for the use of optogenetics to increase control over metabolic pathways and improve the production of commercially valuable compounds in an industrially promising cyanobacteria species.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Prof Anna Amtmann for the opportunity to work on this project, her guidance, support, constructive criticism, and supervisory approach which allowed me to work independently and effectively. Thank you to my other supervisors Prof John Christie and Dr Douglas McKenzie for their guidance and input when needed.

A huge thank you to the rest of the Amtmann lab past and present who have supported me. Special thanks to Dr Maria Papanatsiou and Dr Jordan Twigg for your constant support, mentoring and troubleshooting. Thank you to Anna Palombo for being an excellent technician and helping me perform several of the experiments in this thesis. Thank you to Dr Mary Ann Madsen and Dr Aleksander Mihnev for your support and the materials provided at the beginning of this project. I would also like to thank all the wider members of the Stevenson Lab for their support over the last 4 years (even George).

Thank you to Dr Graham Hamilton and the rest of the Glasgow Polyomics team for performing the RNA sequencing and bioinformatic analysis presented in this thesis.

Special thanks to the BBSRC and IBioIC for providing the funding for this project as well as an excellent Doctoral Training Programme.

Finally, I would like to thank my family, friends and Abbie for their unwavering support.

AUTHOR'S DECLARATION

I declare that this thesis is the result of my work except where explicit reference is made to the contribution of others. This thesis has not been submitted for any other degree at the University of Glasgow or elsewhere.

Liam Forbes

ABBREVIATIONS

BV - biliverdin

CA – chromatic acclimation

CBCR - cyanobacteriochrome

Ct – cycle threshold

ETC – electron transport chain

FDR – false discovery rate

FMN – flavin mononucleotide

GAF - cGMP-specific phosphodiesterase, adenylyl cyclase and FhlA domain

LOV – Light, Oxygen, Voltage domain

LED – light emitting diode

OCP – Orange Carotenoid Protein

PAS - Per-Arnt-Sim domain

PCA – principal component analysis

PCB – phycocyanobilin

PHB - poly-ß-hydroxybutyrate

PVB - phycoviobilin

PSI – Photosystem I

PSII – Photosystem II

RBS – ribosome binding site

REC – Receiver domain

RNAP – RNA Polymerase

SDM – Site-directed mutagenesis

1 Introduction

1.1 Cyanobacteria

1.1.1 Environmental distribution

Cyanobacteria are photosynthetic prokaryotes which occupy a vast range of terrestrial and aquatic habitats. They are thought to be responsible for the oxygenation of Earth's atmosphere over two billion years ago via photosynthesis (Olson, 2006). Cyanobacteria are also widely accepted as the evolutionary precursors of plant chloroplasts (Turner *et al.*, 1999). Their photosynthetic capabilities and the ability of some cyanobacteria species to fix atmospheric nitrogen have led to this phylum playing important roles in the balance of many ecosystems including oceans and wetland rice fields (Zehr, 2011; Whitton, 2012). Consequently, huge amounts of cyanobacterial biomass are present on Earth, with one research group estimating global wet biomass as being 10^9 tonnes (Garcia-Pichel *et al.*, 2003).

One factor which has contributed to the widespread distribution of cyanobacteria across habitats is their ability to survive under extreme conditions such as desiccation, hypersalinity and high temperatures. For example, one study found 19 different cyanobacterial species, including the filamentous *Nostoc punctiforme,* in the dry soil crust of a North American desert (Flechtner *et al.*, 2008). Other species such as those of the genus *Spirulina* and *Synechococcus* have been found in hypersaline environments including Lake Hamra in Egypt, which contains 4M NaCl and a pH of around 11 (Imhoff *et al.*, 1979). Biofilms containing *Synechococcus* species have been found in hot springs with temperatures reaching 73°C, the upper-temperature limit at which photosynthesis is possible (Castenholz and Garcia-Pichel, 2012). Some cyanobacterial species can switch from autotrophic to heterotrophic growth in the absence of light, using sugars such as glucose and sucrose as a carbon source (Rippka *et al.*, 1979). This can allow some species to survive in areas where light availability is low.

1.1.2 Morphology and cell biology

Various intracellular and extracellular features can influence the growth of different cyanobacterial species in specific environments. Cellular morphologies range from unicellular, spherical cells, such as *Synechococcus* sp. PCC 7002 (Syn7002), to filamentous aggregates (Uyeda *et al.*, 2016).Indeed, cyanobacteria have been grouped into five taxonomic subsections (I-V) based on their morphologies (Table 1-1) (Henson et al., 2004). Subsections IV and V include cyanobacteria which can form a specialised cell type known as a heterocyst which performs nitrogen fixation. As gram-negative bacteria, cyanobacteria usually contain a 1-10 nm peptidoglycan layer between the cytoplasmic and outer cell membranes (Castenholz, 2015). In addition to peptidoglycan, some species such as Syn7002 contain a thick cellulose-based layer between the membranes (Figure 1-1) (Zhao *et al.*, 2015). On the outer membrane, cyanobacteria often contain a polysaccharide capsule or sheath often referred to as the glycocalyx, which in some species contains the UV-absorbing pigment scytonemin, creating a yellow-brown appearance (Castenholz, 2015). However, many cyanobacteria are blue-green in appearance under white light due to the presence of chlorophyll pigments inside the cell. Unlike many other bacteria, cyanobacteria do not contain flagella and some species instead use type-IV pili for surface-dependent motility. These pili also allow some cyanobacteria species to obtain exogenous DNA from their environment, providing the basis for genetic transformation (Brahamsha and Bhaya, 2014). The photosynthetic machinery of the cell is contained within the thylakoid membranes and the carboxysomes (Figure 1-1) (discussed later).

Table 1-1. Phylogenetic classification of cyanobacteria

Subsections	Defining Characteristics
I	Unicellular Cell division via binary fission or budding
II	Unicellular Cell division via multiple fission
III	Filamentous Non-heterocystous Reproduction via trichome breakage
IV	Heterocystous Cell division in one plane
V	Heterocystous Cell division in more than one plane

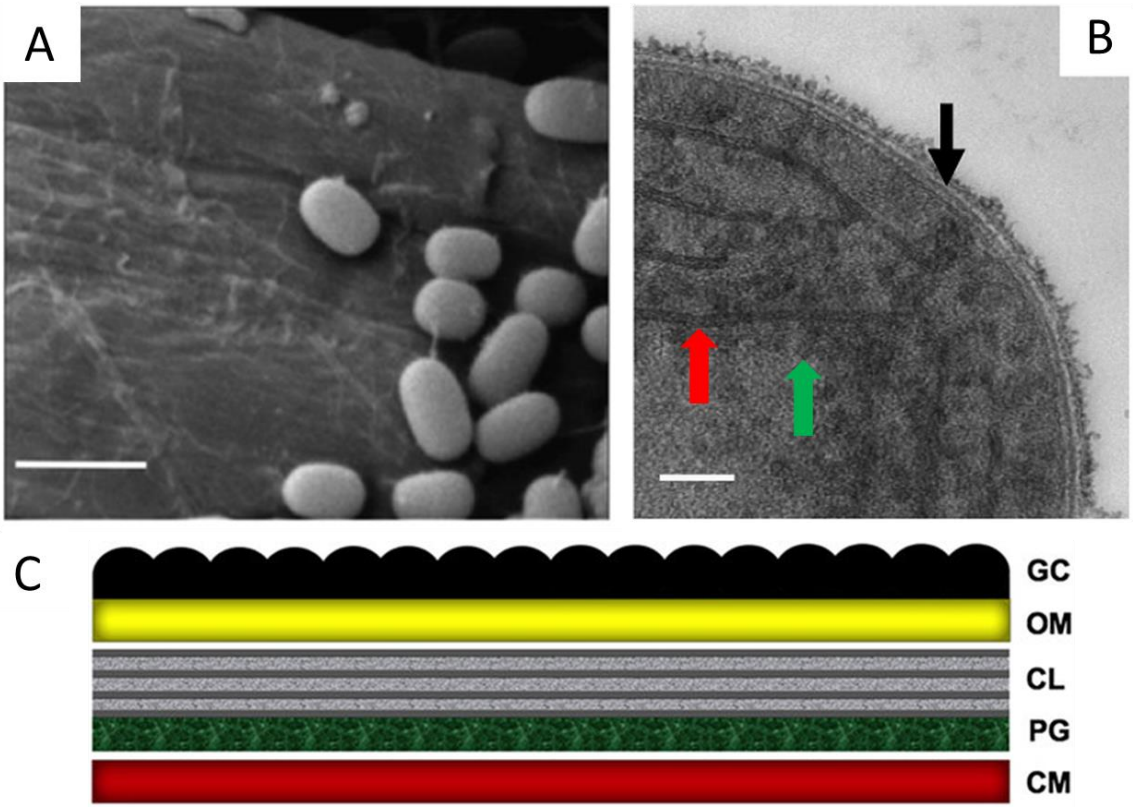


Figure 1-1. Scanning electron microscopy (SEM) and transmission electron microscopy (TEM) images of Syn7002 cells. A) SEM of multiple Syn7002 cells. Scale bar = 2 µm. B) Single Syn7002 cell. Scale bar = 100 nm. The black arrow indicates the cellulose-based layer between the cytoplasmic and outer membranes, the red arrow indicates a thylakoid membrane and the green arrow indicates a carboxysome. C) Schematic showing the layers of the Syn7002 cell wall. CM = cytoplasmic membrane, PG = peptidoglycan layer, CL = cellulose layer, OM = outer membrane, GC = glycocalyx layer. Adapted from Zhao *et al.*, 2015.

Cyanobacteria can adapt their cellular morphology in response to perceived environmental signals. Arguably the most important signal in a photosynthetic organism is light and cyanobacteria can alter their size and shape in response to both different light wavelengths and intensities. In a process termed chromatic acclimation, cyanobacteria alter their pigment composition and cell shape to maximise photosynthesis. *Fremyella diplosiphon* cells grown under green light are often rectangular and form long filaments, whereas in red light they form much shorter spherical cells in a process controlled by the photoreceptor RcaE (Bordowitz and Montgomery, 2008). Similarly, under low light intensity, *F. diplosiphon* cells become elongated in contrast to the shorter and more spherical cells observed under high light intensity. This response has also been observed in the environment by *F. diplosiphon*, where cells near the water surface are round, while cells at lower depths where only certain wavelengths of light penetrate are elongated (Pattanaik *et al.*, 2012). It has been suggested that more rounded cells have increased buoyancy, allowing them to reside closer to the water surface (Pattanaik *et al.*, 2012).

In addition to affecting cell morphology, light also influences cyanobacterial circadian rhythm. This allows coordination of the 24-hour day cycle with perceived changes in light. Circadian rhythm has been studied extensively in the circadian model species *Synechococcus elongatus* PCC 7942 (Syn7942). The master regulators of the circadian clock in Syn7942 are the KaiA, KaiB and KaiC proteins. KaiC is responsible for the timekeeping of the circadian clock across the 24-hour day/night cycle and its phosphorylation status is controlled by KaiA and KaiB. KaiC phosphorylation status is conveyed to the rest of the cell via the action of the histidine kinases CikA and SasA on the master response regulator RpaA (Gutu and O'Shea, 2013). The activity of CikA is also modulated by the oxidation state of quinone molecules, which in turn are controlled by the influence of light and photosynthesis, thus indirectly allowing light input to influence circadian rhythm (Kim *et al.*, 2020). KaiC activates SasA phosphorylation of RpaA, whereas CikA dephosphorylates RpaA. RpaA controls the expression of four sigma factors which are responsible for initiating global patterns of gene expression regulated by circadian rhythm (Fleming and O'Shea, 2018). Circadian rhythm is known to affect a wide range of cellular processes in cyanobacteria including gene expression, cell division and DNA uptake (Liu *et*

al., 1995; Dong *et al.*, 2010; Taton *et al.*, 2020). Indeed, the time of day and onset of light or darkness can influence the transformation efficiency of Syn7942 by two orders of magnitude (Taton *et al.*, 2020).

A notable feature of cyanobacteria is that large differences in chromosome copy number can be observed both between species and within species depending on growth state and the environment. While some species such as *Synechococcus* sp. WH 7805 are monoploid, many others are polyploid (Binder and Chisholm, 1995). *Synechocystis* sp. PCC 6803 (Syn6803) is reported to have a chromosome copy number varying anywhere between 2-218 depending on the sub-strain, growth conditions and technique used to measure copy number (Griese *et al.*, 2011; Pope *et al.*, 2020). Interestingly, while the chromosome copy number of Syn7942 varies between 1-8 per cell, cellular protein concentration remains constant (Zheng and O'Shea, 2017). Increases in genome copy number were also positively correlated with cell volume. Thus, while the molecular basis for this regulation is unknown, the authors propose that increases in cell volume are important in compensating for increases in chromosome copy number. In addition to their chromosomes, many cyanobacteria also contain endogenous plasmids. Syn7002 contains 6 different endogenous plasmids, with copy numbers during the exponential phase ranging between 6-50 (Xu *et al.*, 2011).

Various theories exist as to whether and why polyploidy provides a fitness advantage. Theories include increased repair efficiency of dsDNA breaks and gene redundancy protecting against harmful mutations (Griese *et al.*, 2011). This is convincing, given that cyanobacteria are often found in environments where they are regularly exposed to damaging UV light. Another theory is that polyploidy allows DNA to be used as an emergency phosphate source during starvation. Indeed, it has been demonstrated that both cyanobacteria and haloarchaea reduce genome copy number significantly during phosphate starvation whilst maintaining cell growth (Zerulla *et al.*, 2014; Pope *et al.*, 2020).

1.1.3 Photosynthesis

While purple photosynthetic and green sulphur bacteria can perform anoxygenic photosynthesis, cyanobacteria are the only known bacterial phylum capable of

oxygenic photosynthesis (Nagarajan and Pakrasi, 2016). Consequently, cyanobacteria can use sunlight and water to generate ATP and NADPH. These sources of energy and reducing power are then used to fix atmospheric CO_2 into cellular carbohydrates, using water as an electron donor and producing oxygen as a by-product. Cyanobacteria can also perform respiration, the process of generating energy by breaking down cellular carbohydrates and releasing CO_2 as a by-product to allow for ATP production in the dark (Vermaas, 2001).

All cyanobacteria other than those of the genus *Gloeobacter* perform both photosynthesis and respiration in the thylakoid membranes (Rippka *et al.*, 1974). *Gloeobacter* are thought to have diverged from the cyanobacterial tree early in evolutionary history, which may explain why this genus does not contain thylakoids and instead performs photosynthesis within the plasma membrane (Schirrmeister *et al.*, 2015). Thylakoid membranes often form concentric sheets around the inside layer or throughout the cyanobacteria cell. This differs from algal and plant thylakoids, which often form stacked membrane complexes throughout the cytoplasm known as grana (Figure 1-2) (Mullineaux, 2005).

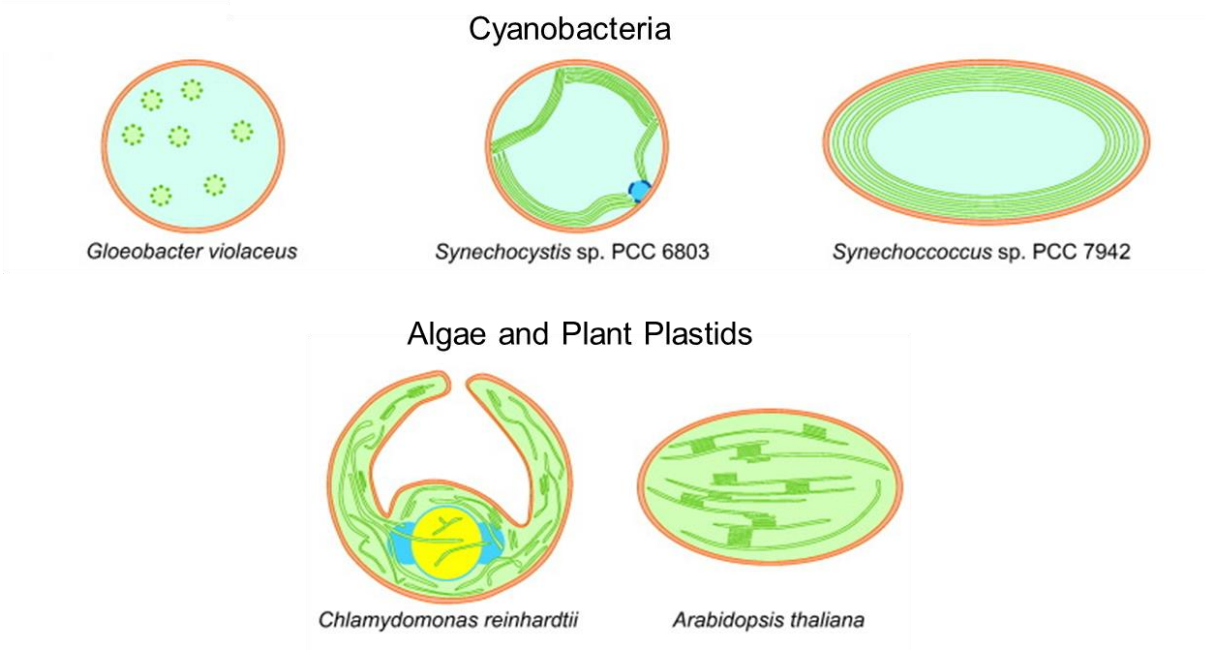


Figure 1-2. Organisation of the photosynthetic machinery in cyanobacteria, plants and algae. In *Gloeobacter*, the photosystems are still located within the plasma membrane, while other cyanobacteria, algae and plants have formed thylakoid membranes which contain the protein complexes for the light reactions and ATP synthesis. Adapted from Rast *et al.*, 2015.

Cyanobacteria generate energy via the so-called light reactions comprising an electron transport chain (ETC) fuelled by two photosystems. The best-characterised ETC process is linear electron transport (Z-scheme), which begins with the dimeric protein complex photosystem II (PSII) comprising 40 protein subunits (Umena *et al.*, 2011) (Figure 1-3). PSII contains a pair of reaction centre chlorophyll *a* molecules termed P680 (λmax = 680 nm), which absorb photons funnelled towards them by various antenna chlorophyll, carotenoids and the phycobilisome (discussed in section 1.1.4). The light energy absorbed by P680 excites an electron in the reaction centre which is rapidly separated from the chlorophyll molecules and passed on to quinones. The electron void is filled by the extraction of electrons from water. The splitting of two water molecules produces molecular oxygen as a by-product and releases four protons into the thylakoid lumen, contributing to the generation of a proton gradient. Additional protons are transported into the thylakoid lumen by the cytochrome b6f complex which acts both as a redox protein catalysing electron transfer from plastoquinone to plastocyanin and as a proton pump. Plastocyanin then passes the electrons to photosystem I (PSI). Like PSII, PSI is a multi-subunit protein complex that contains a pair of reaction centre chlorophyll *a* molecules termed P700 (λmax = 700 nm). Light absorption by energises reduction of ferredoxin. Ferredoxin-NADPH oxidoreductase (FNR) then reduces NADP+, generating NADPH, the molecule which provides the reducing power required for carbohydrate synthesis in the Calvin cycle. The proton gradient which has built up across the thylakoid membrane is used to power ATP synthesis. The multi-protein complex ATP synthase forms a proton channel in the thylakoid membrane which allows protons to flow back from the thylakoid lumen into the cytoplasm and uses the released energy to convert ADP and inorganic phosphate to ATP.

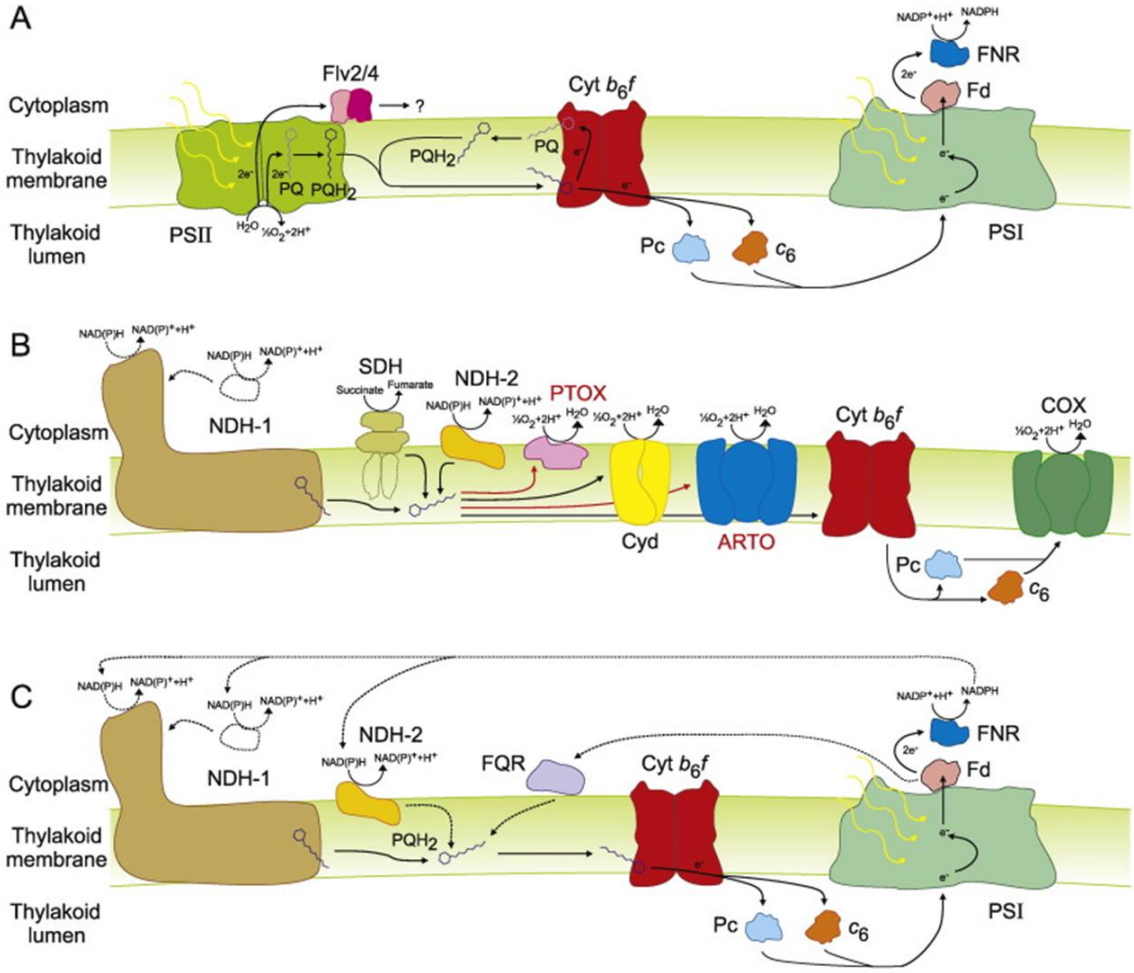


Figure 1-3. Electron transport chains in cyanobacteria. A) Linear electron transport chain. B) Respiratory electron transport. C) Cyclic electron transport chain. Red lines represent proteins and pathways which do not exist in *Synechocystis* sp PCC 6803 whilst broken lines represent pathways and proteins not yet confirmed experimentally. ARTO—alternative respiratory terminal oxidase, COX—cytochrome-c oxidase, Cyd—bd-quinol oxidase, cyt b6f—cytochrome b6f, Cyt c6—cytochrome c6 Fd—ferredoxin, FNR—ferredoxin-NADP+-reductase, Flv2/4—Flavodiiron 2/4, FQR—ferredoxin-plastoquinone reductase, NDH-1—NAD(P)H dehydrogenase 1, NDH-2—NAD(P)H dehydrogenase, Pc—plastocyanin, PSI—Photosystem I, PSII—Photosystem II, PQ—plastoquinone, PQH2—plastoquinol, PTOX—plastid terminal oxidase, SDH—succinate dehydrogenase. Figure from Lea-Smith *et al.*, 2016.

In addition to linear electron transport, alternative transport pathways also exist. Cyclic electron flow involves the flow of electrons through PSI only, resulting in ATP but not NADPH production and is important for maintaining cellular redox balance (Figure 1-3c) (Mullineaux, 2014). Another alternative electron pathway is the water-water cycle, in which electrons from PSI are used to reduce oxygen, which is then converted to water. In the process, reactive oxygen species are consumed. This pathway acts as an electron sink, preventing photoinhibition under high light intensity (Raven *et al.*, 2020). Intriguingly, many

of the same proteins involved in the photosynthetic linear and cyclic electron transport chains are also used in respiratory electron transport (Figure 1-3b).

The ATP and NADPH produced from the light reactions are used in a series of reactions that incorporate CO_2 into cellular carbohydrates known as the Calvin Cycle. The Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) enzyme converts CO_2 and ribulose-1,5-bisphosphate (RuBP) into two molecules of 3-phosphoglyceric acid (3-PGA). 3-PGA can then enter sugar metabolism or be converted back into RuBP through various catalytic steps. As Rubisco has an affinity for both O_2 and CO_2, cyanobacteria have evolved a mechanism of increasing local CO_2 concentrations around the enzyme to promote carbon fixation. This carbon-concentrating mechanism occurs within carboxysomes, bacterial microcompartments encapsulating high concentrations of Rubisco and carbonic anhydrase (Kerfeld and Melnicki, 2016). Small pores within the proteinaceous icosahedral carboxysome shell allow inward diffusion of bicarbonate. Here, it is converted to CO_2 by carbonic anhydrase and subsequently used by Rubisco to produce 3-PGA. The carboxysome shell also prevents CO_2 leakage which promotes carboxylation rather than oxygenation of Rubisco (Cai *et al.*, 2009). Interestingly, the structure and number of carboxysomes per cell are influenced by the photoreceptor RcaE in *F. diplosiphon* in response to the absorption of red or green light (Rohnke *et al.*, 2018). In an attempt to enhance carbon fixation in crops, plant chloroplasts have been engineered to produce Rubisco-encapsulated carboxysomes (Long *et al.*, 2018).

1.1.4 Pigments and phycobilisomes

While PSI and PSII contain antenna chlorophyll molecules that funnel photons towards the reaction centre, chlorophyll only absorbs wavelengths in certain parts of the light spectrum. The gaps in the absorbance spectra of chlorophyll are filled by other pigments such as carotenoids, phycocyanin and phycoerythrin (Figure 1-4). Phycocyanin and phycoerythrin are incorporated in phycobilisomes. These large light-harvesting complexes are unique to cyanobacteria and channel wavelengths that are poorly absorbed by chlorophyll towards photosynthetic reaction centres (MacColl, 1998). Phycobilisomes are located on the cytoplasmic side of thylakoid membranes, physically associated with PSI and PSII.

Phycobilisomes consist of phycobiliproteins – proteins which bind bilin chromophores via cysteine residues that allow light absorption (Stadnichuk *et al.*, 2015). Up to 60% of cyanobacterial cellular protein can be comprised of phycobiliproteins, highlighting their importance to cyanobacterial fitness and survival (Singh *et al.*, 2015). During periods of starvation, cyanobacteria can also degrade their phycobilisomes, releasing nutrients such as carbon, nitrogen and phosphate to ensure cell survival (Parmar *et al.*, 2011).

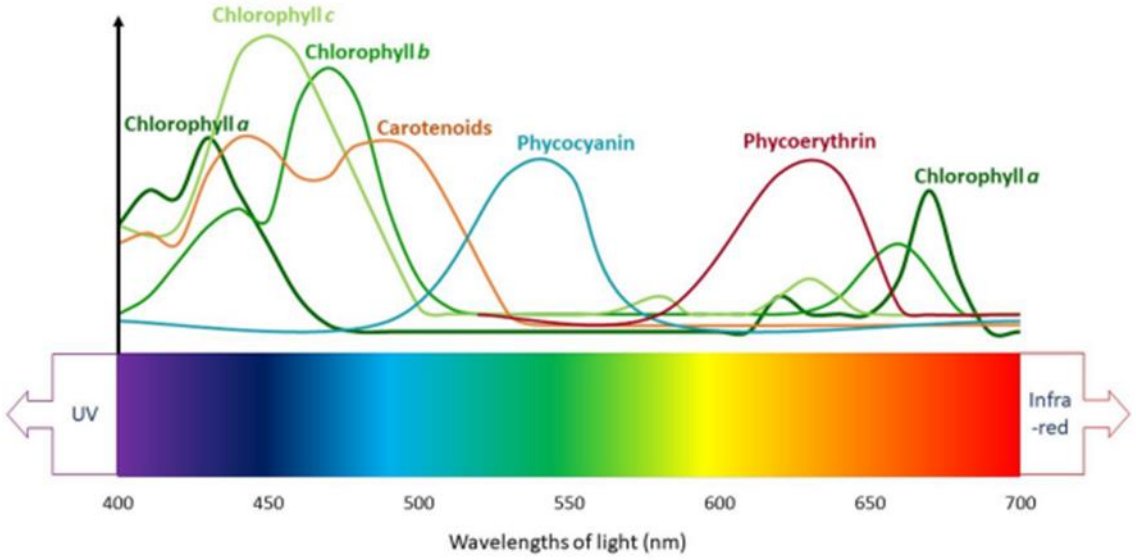


Figure 1-4. Absorbance spectra of typical cyanobacterial pigments. Figure from Roy *et al.*, 2011.

Typically, phycobilisomes contain an allophycocyanin core, with 'rods' projecting out from the core in a semi-circle configuration (Figure 1-5). These rods contain different combinations of phycobiliproteins including phycocyanin, phycoerythrin and phycoerythrocyanin, depending on the species and the prevailing light spectra in their environment (Glazer, 1989). The rods are comprised of phycobiliprotein trimer and hexamer 'discs' which are held together by hydrophobic rod linker proteins. Despite differences in structure between species, all phycobilisomes function by channelling light wavelengths poorly absorbed by chlorophyll to photosynthetic reaction centres via resonance energy transfer (David *et al.*, 2014). Typically, light energy absorbed by either phycoerythrin (λ_{max} = 560 nm), phycoerythrocyanin (λ_{max} = 575 nm) or phycocyanin (λ_{max} = 620 nm) in the outer rods is transferred to the

allophycocyanin (λ_{max} = 652 nm) core and then finally to the reaction centre chlorophyll where it is used for the light reactions.

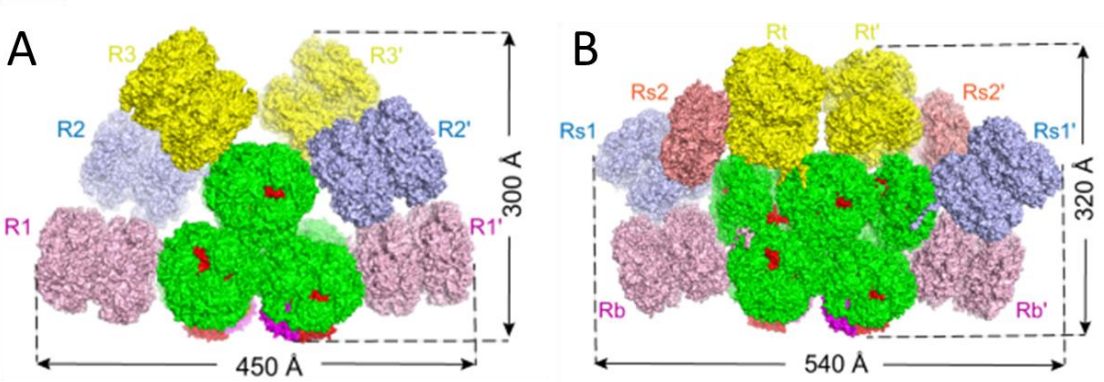


Figure 1-5. Cartoon representation of A) Syn7002 and B) *Anabaena* 7120 phycobilisome cryo-EM structures. Green = allophycocyanin, pink/blue/yellow = phycocyanin, red = core membrane linker protein, magenta = allophycocyanin F. Adapted from Zheng *et al.*, 2021.

Phycobilisomes provide a unique fitness advantage to cyanobacteria by allowing physiological adaptation to perceived changes in light quality and intensity. In a process termed chromatic acclimation, cyanobacteria can reversibly alter the composition of their phycobilisomes to maximise photosynthetic efficiency. This process is controlled by photoreceptors and associated signalling proteins, which link light perception to the physiological response of chromatic acclimation. During chromatic acclimation, pigment, chromophore, rod-core linker protein and chlorophyll composition within PSI and PSII can all be reversibly altered (Sanfilippo *et al.*, 2019).

To date, six different mechanisms of chromatic acclimation (CA1-CA6) have been observed in cyanobacteria, each controlled by a different signalling pathway. CA1-CA3 control green/red light acclimation, CA4 controls blue/green light acclimation, while CA5 and CA6 control red/far-red acclimation. The simplest form of chromatic acclimation, CA1, has been observed in Syn6803 and is controlled by the green/red light-responsive CcaS photoreceptor. Upon absorption of green light by CcaS, the cell increases expression of the phycobilisome rod linker protein CpcG2. Whereas CpcG1 associates with PSII, CpcG2 preferentially associates with PSI and causes a shift in excitation balance between PSI and PSII (Kondo *et al.*, 2007; Hirose *et al.*, 2008). By altering cell pigment composition, chromatic acclimation is responsible for the visible change in colour of cyanobacteria when exposed to different light wavelengths.

While the main function of phycobiliproteins is in photosynthetic light harvesting, other pigments such as carotenoids play important roles in photosynthesis and photoprotection. The two main classes of carotenoids are carotenes and xanthophylls. *β*-carotene is physically associated with the PSI complex in Syn7942 and increases the spectral range for light-harvesting, primarily absorbing light in the 400-500 nm range (Stamatakis *et al.*, 2014). A xanthophyll-deficient mutant of Syn7002 produced significantly more reactive oxygen and nitrogen species under high light intensity than the wild-type, highlighting a photoprotective role for carotenoids (Zhu *et al.*, 2010). Furthermore, a carotenoid-deficient Syn6803 mutant was unable to produce functional PSII complexes and could only grow under light-activated heterotrophic conditions (Sozer *et al.*, 2010). This suggests that carotenoids also play an important role in maintaining PSII structure.

While carotenoids provide photoprotection as individual molecules, they are also found within protein complexes with specific photoprotective functions. The orange carotenoid protein (OCP) binds the carotenoids echinenone, hydroxyechinenone and zeaxanthin and functions by quenching excess energy from excited bilins and chlorophyll. The complex reduces the formation of reactive oxygen species, dissipating excess energy as heat in a process termed non-photochemical quenching (Kirilovsky and Kerfeld, 2016). OCP can be found in two forms: OCP^o and OCP^r. High-intensity blue-green light which is absorbed by the bound keto carotenoid 3′-hydroxyechinenone, triggers a conformational change in OCP^o, resulting in conversion to OCP^r. OCP^r then binds directly to the phycobilisome and induces non-photochemical quenching (Wilson *et al.*, 2008). OCP^o also plays a role in photoprotection by efficiently quenching singlet oxygen molecules under high-intensity orange-red light (Sedoud *et al.*, 2014). Species that lack OCP such as Syn7942 demonstrate significant photoinhibition when exposed to high-intensity light (Boulay *et al.*, 2008).

1.2 Cyanobacterial photoreceptors

1.2.1 Structure

As photosynthetic organisms, cyanobacteria have evolved complex molecular mechanisms which allow them to alter their physiology in response to changes in

light. These responses are controlled by photoreceptors, proteins that absorb specific light wavelengths. Photoreceptors can typically be separated into an N-terminal photosensory region and a C-terminal effector region. N-terminal photosensory domains bind chromophores, which upon absorption of a specific light wavelength will elicit a reversible conformational change within the domain. This conformational change is then transmitted to the effector region, allowing either activation or deactivation of effector domains. Effector domains initiate downstream signalling and control physiological responses, often by phosphorylating transcriptional regulators or altering cellular second messenger concentrations (Figure 1-6).

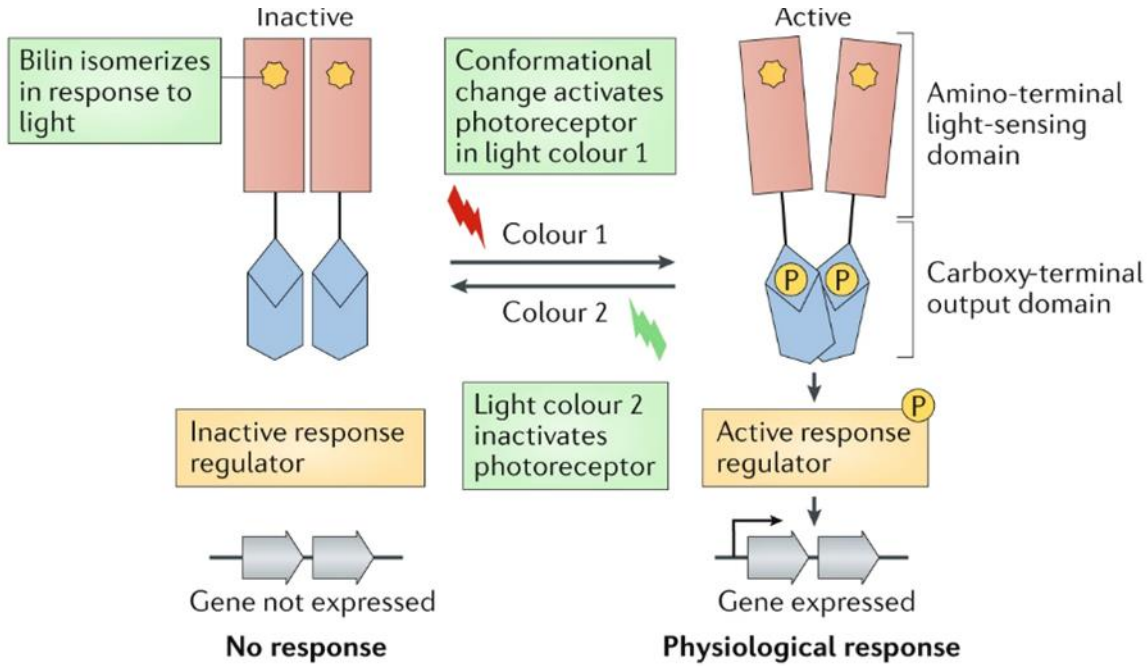


Figure 1-6. Wavelength-specific reversible activation of photoreceptors. Upon absorption of a specific light wavelength (colour 1), the bound chromophore of a photoreceptor (in this case a bilin) will reversibly isomerise, triggering a conformational change in the protein structure. This results in the activation of the C-terminal effector domain, in this case, a histidine kinase domain. This results in the phosphorylation of a downstream active response regulator which can regulate gene expression and alter physiological responses. Irradiation with another wavelength (colour 2) causes reversion to the inactive state. Figure from Wiltbank and Kehoe, 2019.

The vast majority of discovered cyanobacterial photoreceptors belong to the phytochrome superfamily. Phytochromes are present across the plant, algal and bacterial kingdoms and contain a conserved cGMP-phosphodiesterase/adenylate cyclase/FhlA (GAF) domain, which is responsible for light perception (Rockwell and Lagarias, 2017). They also often contain Per-Arnt-Sim (PAS) and

phytochrome-specific (PHY) domains. While plant phytochromes typically contain a PAS-GAF-PHY photosensory module, prokaryotic phytochromes contain more varied structures. These fit into three main categories: PAS-GAF-PHY modules, GAF-GAF or GAF-PHY modules and single GAF domain modules (Figure 1-7) (Wiltbank and Kehoe, 2019). Phytochromes which contain single GAF domain modules are known as cyanobacteriochromes (CBCRs) and are unique to cyanobacteria (Rockwell and Lagarias, 2017). Some cyanobacterial photoreceptors like the CBCR CcaS from Syn6803 contain a single GAF domain, whereas others can contain several. PtxD from *N. punctiforme* contains seven individual GAF domains (Hirose *et al.*, 2008; Campbell *et al.*, 2015). CBCR GAF domains typically bind a linear tetrapyrrole bilin chromophore. In most cases, the binding typically occurs via a thioether linkage between a conserved cysteine residue in the GAF domain ('canonical cysteine') and the C3 atom of the A-ring of the chromophore (Fushimi *et al.*, 2016). However, GAF domains of the Asp-Xaa-Cys-Ile-Pro (DXCIP) lineage such as the CBCR cce_4193g1 from *Cyanothece* sp. ATCC 51472 do not contain the canonical cysteine. Instead, cce_4193g1 binds the bilin chromophore via a second cysteine within the DXCIP motif (Fushimi *et al.*, 2016).

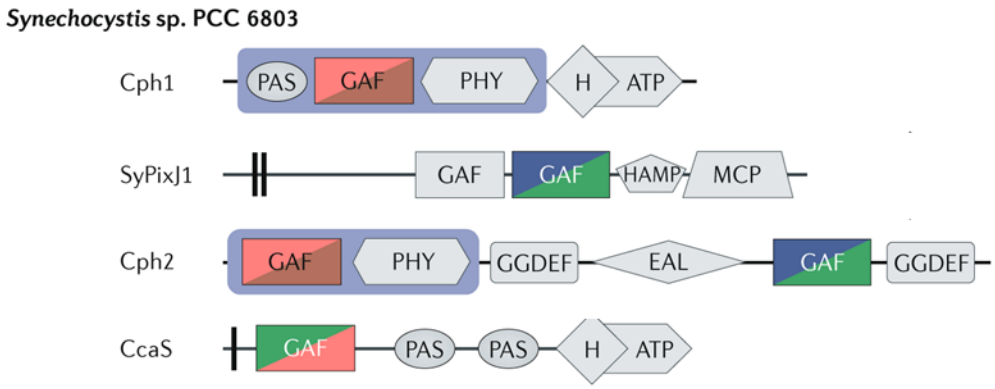


Figure 1-7. Cyanobacterial photoreceptor photosensory domain structure. Examples of *Synechocystis* sp. PCC 6803 phytochrome superfamily members containing PAS-GAF-PHY, GAF-GAF, GAF-PHY and GAF photosensory modules. EAL = glutamate-alanine Leucine, GAF = cGMP-phosphodiesterase/adenylate cyclase/FhlA, GGDEF = glycine-glycine-aspartate-glutamate-phenylalanine, H-ATP = histidine kinase, MCP = methyl-accepting chemotaxis protein, PAS = Per-Arnt-Sim, PHY = phytochrome-specific protein. Figure adapted from Wiltbank and Kehoe, 2019.

Regardless of photosensory module structure, light absorption by photoreceptors triggers Z/E isomerisation of the C15=C16 double bond of the bound chromophore. This reversible isomerisation involves twisting of the D-ring of the

normally planar chromophore, altering the interaction between the chromophore and the photoreceptor protein backbone (Rockwell and Lagarias, 2017). Typically, each isomer will maximally absorb different light wavelengths (e.g., 450 nm and 600 nm) allowing conversion between the active/inactive Z/E photostates. Different isomers of a chromophore can absorb different wavelengths due to changes in the length of the conjugated system. The conjugated system consists of multiple single and double bonds. This allows for delocalised electrons that can move freely within the system and from a low energy molecular orbital to a higher energy orbital upon absorption of light. Z/E chromophore isomerisation is responsible for eliciting the protein conformational change within the sensory region required for effector domain activation/deactivation. Multiple intermediate states will also exist between the fully activated and deactivated photoreceptor states. However, as these are extremely short Lived, they are rarely relevant to the output from the effector domain.

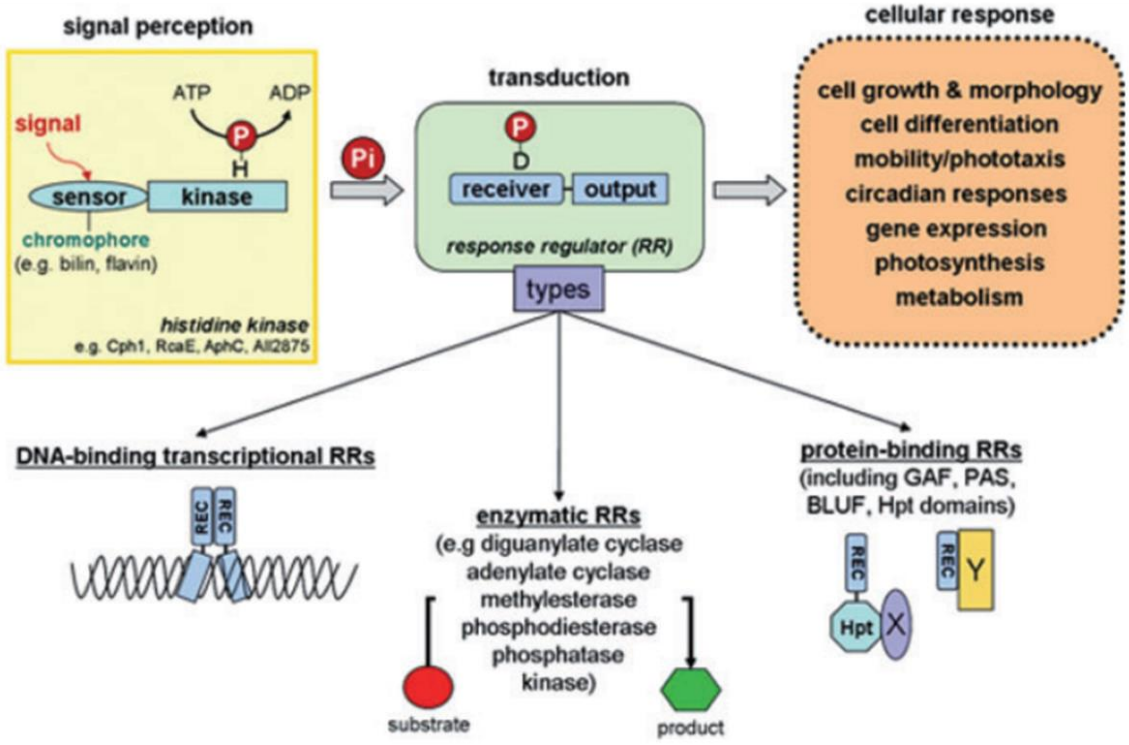
Most CBCR GAF domains are colour sensors, measuring the intensity ratio of a pair of light wavelengths and adjusting the effector response accordingly across a range of outputs. This is only possible as both absorbing forms of these GAF domains are somewhat stable in the dark, usually on the scale of minutes to hours. However, some GAF domains show unidirectional photoconversion which is dependent on light intensity and rapid dark reversion, causing them to instead act as light 'power sensors'. These photoreceptors measure overall light intensity across a broad range of wavelengths rather than the ratio of two wavelengths. An example is NpF2164g7 from *N. punctiforme*, which has a dark reversion half-life of only four seconds. NpF2164g7 absorbs wavelengths between 300-600 nm whilst producing a positive linear output response to increasing light intensity (Rockwell *et al.*, 2012). Similarly, cce_4193g1 of *Cyanothece* sp. ATCC 51472 is thought to act as a green light power sensor, with a dark reversion half-life of 1.1 seconds at 30°C (Fushimi *et al.*, 2016).

Modification of the residues close to the chromophore binding pocket can influence the dark reversion half-life of a photoreceptor. For example, EL222 is a bacterial photoreceptor containing a LOV domain and a DNA-binding domain which originates from *Erythrobacter litoralis* HTCC2594. The dark reversion half-

life of wild-type EL222 is 29 seconds. However, the A79Q mutation increases the half-life to 227 seconds, while the A79R mutation reduces it to around 3 seconds (Zoltowski *et al.*, 2011).

In contrast to sensory regions which always possess a GAF domain, cyanobacterial photoreceptors can contain a range of different effector domains. These include histidine kinase, glutamate-alanine Leucine (EAL); glycine-glycine-aspartate-glutamate-phenylalanine (GGDEF), methyl-accepting and DNA-binding domains (Figure 1-8) (Fushimi and Narikawa, 2019). Histidine kinase domains phosphorylate downstream response regulators such as transcription factors, EAL and GGDEF domains control second messenger concentrations and methyl-accepting domains are involved in the function of chemotaxis-controlling proteins (Galperin and Nikolskaya, 2007). However, some CBCRs contain none of these effector domains, suggesting that signal transduction depends on physical interaction with other photoreceptors or signalling proteins. This is thought to be the case for the red/far-red/green/blue CBCR IflA of *F. diplosiphon* (Bussell and Kehoe, 2013).

A)



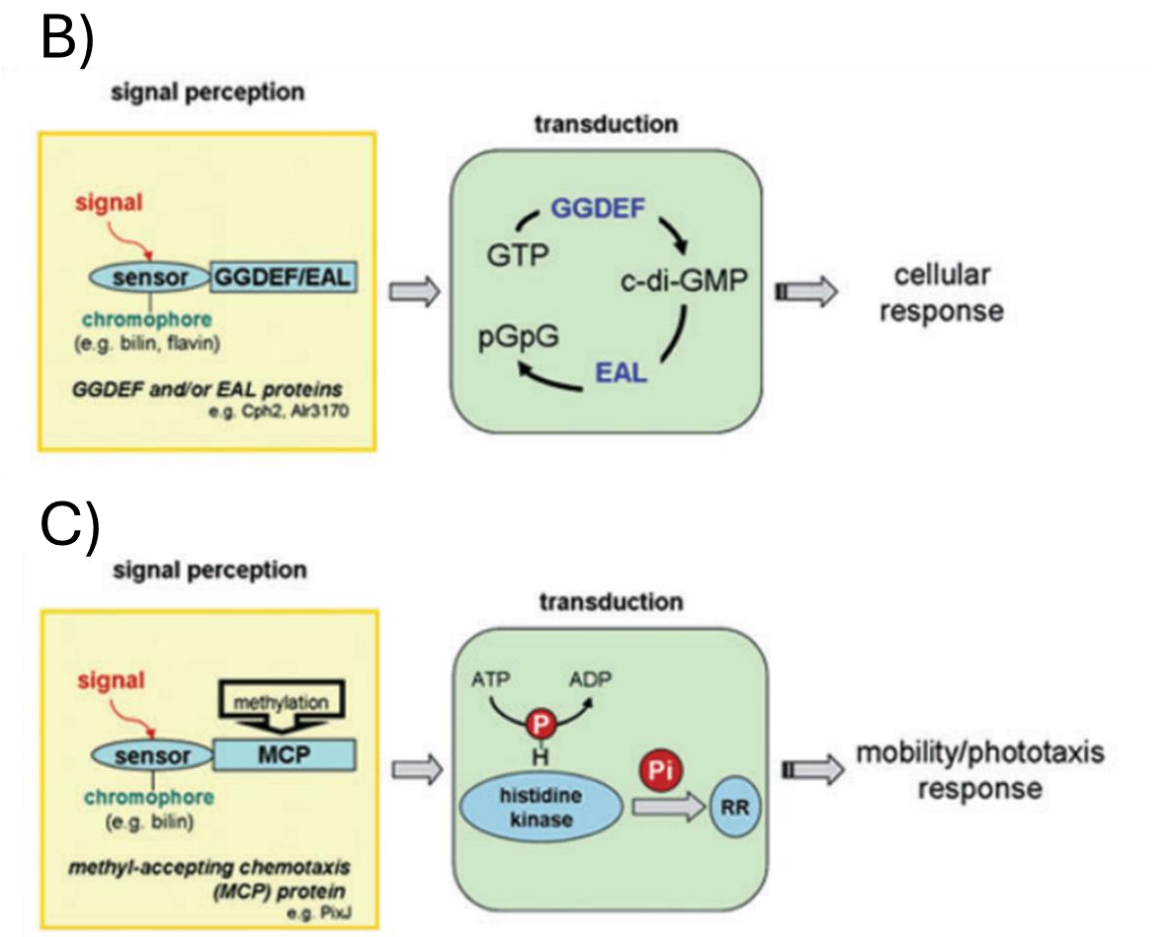


Figure 1-8. Photosensory signalling mechanisms in cyanobacteria. A) Light absorption by a photoreceptor triggers activation/deactivation of its histidine kinase domain. This histidine kinase domain subsequently controls the phosphorylation status and activation/deactivation of a response regulator (RR). The response regulator can be a direct transcriptional regulator, an enzymatic response regulator or a protein-binding response regulator. These regulators can control a range of physiological responses. GAF = cGMP-phosphodiesterase/adenylate cyclase/FhlA, PAS = Per-Arnt-Sim domain, BLUF = blue light sensor using FAD, Hpt = histidine-containing phosphotransfer domain. **B)** Light absorption by a photoreceptor containing a glycine-glycine-aspartate-glutamate-phenylalanine (GGDEF) or glutamate-alanine Leucine (EAL) domain results in activation of GGDEF/EAL catalytic function, modulating the local cellular concentration of second messengers such as bis-(3′, 5′)-cyclic dimeric guanosine monophosphate (c-di-GMP), resulting in a physiological response. **C)** Light absorption by a photoreceptor containing a methyl-accepting chemotaxis protein (MCP) domain triggers reversible MCP domain methylation. The methylation status of the MCP domain controls the activation/deactivation of a histidine kinase, which in turn controls phosphorylation of a response regulator and modulates a physiological response. Figure from Montgomery, 2007.

Photoreceptors can generally be divided into two groups depending on how they transduce signals in response to light through their effector domains. One-component systems are comprised of the photoreceptor only which can control

signalling exclusively through its effector domains. These are often photoreceptors containing EAL, GGDEF, methyl-accepting and DNA-binding domains. Two-component systems, on the other hand, rely on a second protein for transducing the signal. An example is a photoreceptor containing a histidine kinase domain which phosphorylates a transcriptional regulator. It is important to note that the general designation of one and two-component systems can also apply to sensory proteins which respond to signals other than light, such as heavy metals or pH.

1.2.2 Spectral diversity

CBCRs can absorb wavelengths across the full UV and visible light spectrum. Four mechanisms contribute to CBCR spectral diversity, allowing different CBCRs to respond to distinct wavelength signals. These mechanisms are chromophore variation, second cysteine adduct formation, reversible protonation and the trapped twist model (Fushimi and Narikawa, 2019). Firstly, three types of chromophores can be associated with CBCR GAF domains *in vivo*: phycoviolobilin (PVB), phycocyanobilin (PCB) and biliverdin (BV) (Ishizuka *et al.*, 2007; Hirose *et al.*, 2008; Narikawa *et al.*, 2015). Each of these chromophores has a different length of conjugated system, the system of overlapping p-orbitals with delocalised electrons, allowing different wavelengths to be absorbed. BV has the longest conjugated system and thus absorbs longer wavelengths of light than PCB and PVB (Fushimi and Narikawa, 2019).

The second mechanism contributing to CBCR spectral diversity is the formation of a second cysteine adduct, distinct from the canonical cysteine (Figure 1-9). The second adduct forms between the GAF domain and the C10 atom of the chromophore and is only found in certain CBCRs (Fushimi and Narikawa, 2019). The formation of a second cysteine adduct shortens the bilin conjugated system and causes a 'blue shift', resulting in the chromophore absorbing shorter wavelength light. This second cysteine is often contained within the highly conserved Asp-Xaa-Cys-Phe (DXCF) motif found in some CBCR GAF domains (Fushimi and Narikawa, 2019).

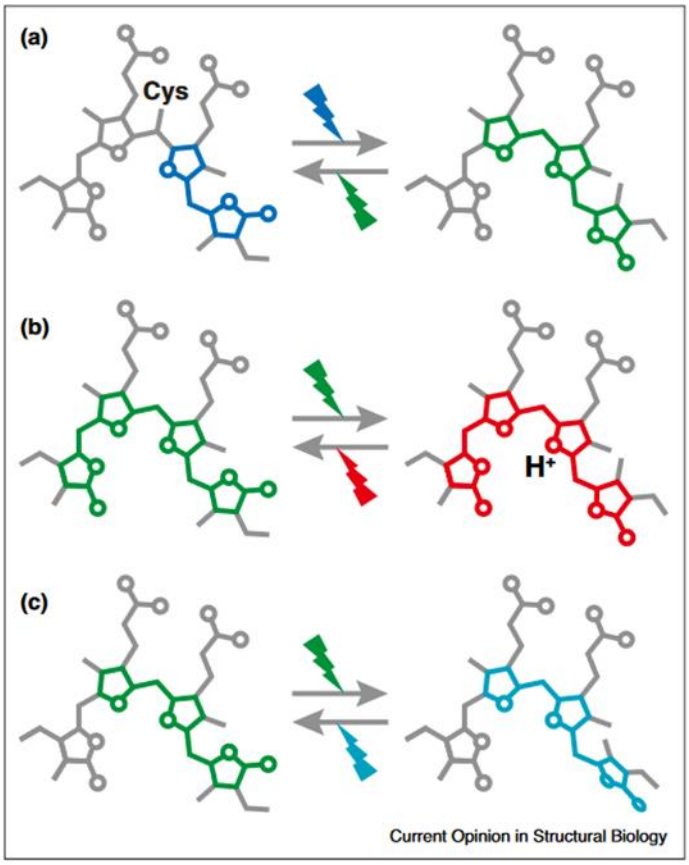


Figure 1-9. Mechanisms of CBCR spectral diversity within a chromophore. a) Second cysteine adduct formation b) reversible protonation c) trapped-twist model of D-ring. The conjugated system is coloured in each chromophore with the remainder of the chromophore in grey. Z-isomers are located on the left and E-isomers on the right. Colours of light used to indicate Z/E isomerisation are only used as examples. Cys = second non-canonical cysteine, H+ = proton. Figure from Fushimi and Narikawa, 2019.

The third mechanism which allows CBCRs to respond to different wavelengths is the less common reversible chromophore protonation, which affects the stability of the conjugated system (Figure 1-5). Chromophore protonation increases the stability of the conjugated system, thus causing a 'red shift' in its absorbance. An example of this is the CcaS photoreceptor, where the red absorbing form of the CBCR CcaS binds protonated PCB, while the green absorbing form binds deprotonated PCB (Hirose *et al.*, 2013).

Finally, the fourth mechanism which contributes to spectral diversity is the trapped twist model. Here, the A or D rings of the chromophore become 'twisted' in relation to the plane of the 'flat' B and C rings (Figure 1-5). This often causes a blue shift by reducing the length of the conjugated system. This model relies on the steric hindrance of the chromophore in one of the two

photochromic states and often involves the action of phenylalanine residues (Fushimi and Narikawa, 2019).

The four mechanisms of CBCR spectral diversity described above have been exploited to engineer CBCRs with alternative photocycles. The CBCR AM1_1499g1 of *Acaryochloris marina* undergoes a green/orange photocycle. After analysing AM1_1499g1 protein structure, researchers targeted mutations at residues computationally predicted to affect which chromophore binds (PCB vs PVB), chromophore isomerisation, cysteine adduct formation and trapping of the chromophore D-ring. This resulted in the identification of seven AM1_1499g1 variants, each responding maximally to a different wavelength pair, consisting of combinations of yellow, teal, green, blue and orange light (Fushimi *et al.*, 2020). This approach could be applied to other CBCRs to alter their colour sensitivity for various biotechnological applications such as optogenetics (discussed in section 1.5.3).

1.2.3 Control of physiological responses to light

Cyanobacterial photoreceptors control physiological responses to light, allowing cells to adapt to changing environmental conditions. For example, chromatic acclimation in Syn6803 is controlled by CcaS photoreceptor-mediated phosphorylation of the transcription factor CcaR as discussed above (Hirose *et al.*, 2008). In contrast to the transcriptional regulation of chromatic acclimation, cell aggregation is controlled by the alteration of local cellular second messenger concentrations. When exposed to violet or blue light under low temperatures, *Thermosynechococcus vulcanus* RKN cells aggregate to induce a cell shading effect, preventing short-wavelength light from damaging photosynthetic machinery (Enomoto *et al.*, 2014). However, under green or teal light, *T. vulcanus* cells do not aggregate. Three CBCRs control cell aggregation in *T. vulcanus*: SesA, SesB and SesC, with SesA playing the most important role (Enomoto *et al.*, 2014; Enomoto *et al.*, 2015). These three CBCRs control cell aggregation via the synthesis and degradation of the second messenger cyclic dimeric guanosine monophosphate (c-di-GMP) using their EAL and GGDEF domains. Absorption of blue/violet light by SesA and SesC triggers increased c-di-GMP production, which activates the enzyme XcsA. XcsA is among several proteins responsible for synthesising cell-aggregating cellulose (Maeda *et al.*,

2018). Under teal light, SesB degrades c-di-GMP through its phosphodiesterase activity, resulting in deactivation of XcsA and a reduction in cell aggregation.

Another photoreceptor-mediated physiological response is phototaxis, the process of a cell moving towards (positive phototaxis) or away from (negative phototaxis) a light source. The UV/green light responsive CBCR UirS is thought to control type IV pili-dependent phototaxis in Syn6803. The model proposed for this response is that UV-A absorption by UirS promotes the release of the UirS-bound transcriptional regulator UirR. Released UirR binds to the promoter of the transcriptional regulator lsiR and increases lsiR expression. LsiR then initiates an unknown signalling cascade, resulting in type IV pili-dependent negative phototaxis (Song *et al*. 2011). Negative phototaxis protects cells from potential UV-induced DNA damage, whereas positive phototaxis can increase photosynthetic efficiency under more favourable light conditions.

1.3 Wavelength-specific signalling pathways

1.3.1 Cyanobacterial responses to light

As discussed above, cyanobacterial photoreceptors control various physiological responses to light, often via wavelength-specific signalling pathways. Several wavelength-specific signalling pathways have been characterised in cyanobacteria. However, very few pathways have been characterised compared to the huge number of known and predicted photoreceptors in cyanobacteria. Work from the Amtmann lab predicts there are 16 photoreceptors encoded in the Syn7002 genome alone, none of which have characterised signalling pathways (Mihnev and Amtmann, 2022). Characterising these pathways could lead to a greater understanding of complex cyanobacterial responses to light. This may yield insights which can be applied to improve cyanobacterial growth or fitness under specific light conditions for biotechnological applications. Furthermore, identifying the transcriptional regulators that are linked to photoreceptors in two-component systems is key to discovering new optogenetic tools (discussed later).

The characterisation of wavelength-specific signalling pathways is often a difficult and time-consuming process. First, a photoreceptor which responds to specific light wavelengths must be identified. If analysing a species with a fully

sequenced genome, potential photoreceptors can be identified computationally through the presence of predicted photosensory regions such as GAF domains. Furthermore, the presence of predicted effector domains can also give clues about the potential mechanism through which signalling may occur. For example, the presence of a DNA-binding domain would suggest that the photoreceptor is likely a one-component system that controls the transcription of target gene(s) in response to light absorption. In contrast, the presence of a histidine kinase domain suggests a two-component system. However, the identification of the second component, usually a transcriptional regulator, is much more difficult.

In many model prokaryotes, the sensor and effector proteins are both contained within a gene cluster. For example, only 23% of predicted *Escherichia coli* sensory kinases and 16% of predicted *Bacillus subtilis* sensory kinases are 'orphans', those without a response regulator protein located in the same genomic region. In contrast, 73% of predicted sensory kinases in Syn6803 and Syn7942 are orphans (Kato *et al.*, 2012). This makes the identification of the response regulator associated with the sensory kinase difficult in many cases. Nonetheless, the red/far-red light-responsive two-component systems cphA/rcpA and cphB/rcpB from *Calothrix* sp. PCC 7601 were initially identified as the photoreceptor and response regulator are present within the same operon (Jorissen *et al.*, 2002). Similarly, the CcaS/CcaR system was more easily characterised due to the presence of the two proteins and the targeted pCpcG2 promoter being located within a gene cluster (Hirose *et al.*, 2008).

When both the photoreceptor and assumed response regulator have been identified, their interaction can be confirmed via protein-protein interaction and phosphorylation assays. The gene(s) which are transcriptionally regulated by the two-component system must then be identified. One approach could be to use ChIP-sequencing to identify which regions of the genome the transcriptional regulator binds to. Another approach would be to compare the transcriptome of the wild-type organism under the relevant light wavelengths to that of a strain with the response regulator gene knocked out. This would result in the identification of gene(s) which are transcriptionally controlled by the response regulator.

While the photoreceptor, response regulator and output genes may all have been identified, the physiological response controlled by the system may still be unclear. Clues can be gathered from the annotated or characterised functions of the output gene(s). However, many cyanobacterial genes are annotated as hypothetical proteins or no obvious function can be foreseen from their predicted domains. Furthermore, the annotation of a gene does not confirm all its functions, as it may have roles in several processes. In this case, the output gene(s) can be knocked out and the cells screened for changes in a limited number of known phenotypes (discussed further in section 6.1.2). Therefore, the process of identifying wavelength-specific signalling pathways and their associated components in cyanobacteria is challenging.

1.3.2 Possible approaches to identify genetic components of wavelength-specific signalling pathways

Identification of novel wavelength-specific signalling components enables the development of new optogenetic tools. For a one-component optogenetic system which controls gene expression, a photoreceptor with a DNA-binding domain and an associated promoter region must be identified. For a two-component optogenetic system, a photoreceptor with a histidine kinase or phosphatase domain, an associated transcriptional regulator, and an associated promoter region must be identified. In cyanobacteria, potential photoreceptors can be easily identified by the presence of predicted GAF domains in their sequence. 16 genes containing predicted GAF domains are encoded in the Syn7002 genome and 8 of these genes contain predicted histidine kinase domains (Mihnev *et al.*, 2023). However, linking these photosensory kinases to an associated transcriptional regulator, if one exists, is more difficult.

One approach to identifying components of wavelength-specific signalling pathways is to screen mutant strains for changes in phenotypes such as pigment accumulation, phototaxis and cell aggregation under different lighting conditions. However, reliable quantification of phenotypic changes can be challenging. There are only a limited number of known measurable wavelength-dependent phenotypes and many laboratory strains of cyanobacteria have lost the ability to perform functions such as phototaxis and biofilm formation when compared to the equivalent environmental strain (Yang *et al.*, 2018).

An alternative approach to measuring physiological phenotypes would be to monitor molecular readouts. This would require the identification of a palette of 'reporter genes' that are transcriptionally activated in response to illumination with particular wavelengths. Upstream targets like predicted transcriptional regulators could then be knocked out and reporter gene expression quantified in cultures grown under different light conditions using qRT-PCR. This would help determine whether the transcriptional regulator may be involved in a wavelength-specific signalling pathway (Figure 1-10).

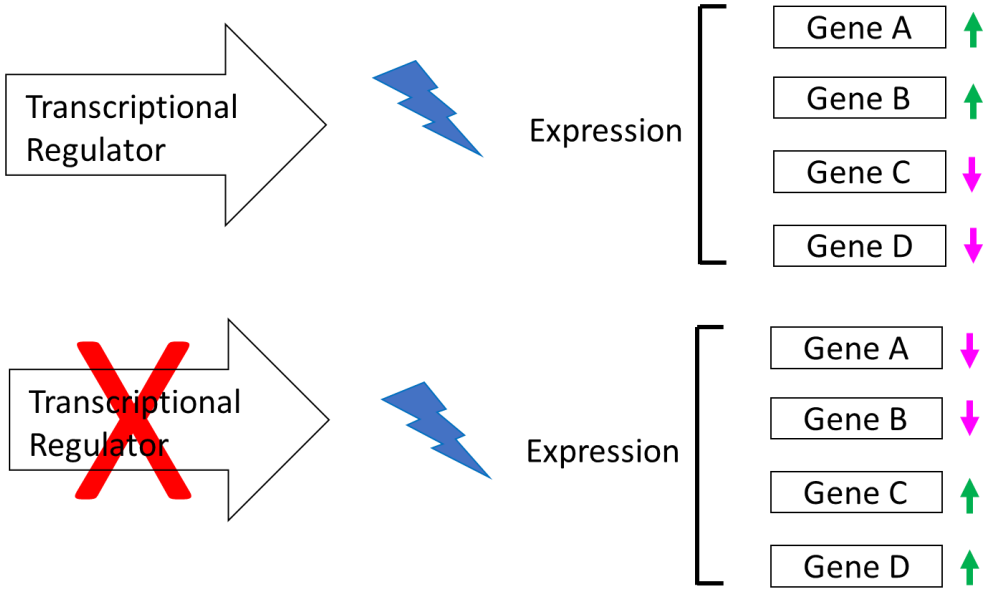


Figure 1-10. Reporter gene-based approach to identifying components of wavelength-specific signalling pathways. qRT-PCR could be used to compare the expression of wavelength-dependent 'reporter genes' in wildtype and mutant strains for a transcriptional regulator suspected to act in a wavelength-specific signalling pathway.

Pairing this reporter gene strategy with CRISPRi could be even more effective than knockout by recombination. Generating constructs expressing sgRNAs that target different transcriptional regulators could avoid the need for re-streaking plates to reach full segregation. Furthermore, some gene knockouts may be lethal, meaning a gene knockdown strategy would be more suitable. Crucially, using CRISPRi would enable many transcriptional regulators to be screened in parallel, reducing the requirement to select a small number of targets for knockout.

To make this approach feasible, first, a palette of reporter genes responsive to a specific light wavelength must be identified. These can be pinpointed by

analysing the transcriptome of cultures grown under different colours of narrowband LED light and comparing them to another colour. Transcriptomics has previously been applied to study gene expression patterns in response to different light wavelengths in cyanobacteria. Using microarrays, it has been demonstrated that Syn6803 displays differential expression of 145 genes in response to blue light, compared with only 5 in red light and 4 in orange light (Luimstra *et al.*, 2020). However, when analysing this list of differentially expressed genes between light colours, the authors mainly focused on their involvement in PSII excitation. The potential involvement of these differentially expressed genes in wavelength-specific signalling pathways was not considered. Furthermore, the reduced sensitivity of microarrays versus RNAseq may have resulted in some differentially expressed genes being overlooked in this study. RNAseq of *N. punctiforme* identified differential expression of 1142 genes in blue light and 1179 genes in red light, a much greater response than that observed in Syn6803 (Han *et al.*, 2019). This work focused on the effect of blue and red light on genes involved in exopolysaccharide production rather than photoreceptors and associated signalling proteins. Differences in growth conditions including light intensity and duration, as well as the statistical methods used to identify differentially expressed genes could explain the large difference between the two experiments,

For Syn7002, RNAseq has been used to study gene expression patterns under other experimental conditions. These include various temperatures, salinities, carbon sources and oxidative stress (Ludwig and Bryant, 2012). RNAseq has also been used to study the effects of high light stress on the Syn7002 transcriptome (Xiong *et al.*, 2015). Over 500 differentially expressed genes were identified as being potentially involved in the ability of Syn7002 to survive under high light stress. The authors produced knockout strains for 12 of these genes and demonstrated their requirement for Syn7002 survival under high light stress. However, the work did not relate to wavelength-specific signalling. Indeed, to our knowledge, there is no literature which discusses the use of transcriptomics to interrogate wavelength-specific signalling pathways in Syn7002.

1.4 Cyanobacterial transcription and translation

1.4.1 Transcription

Transcription is the process by which the genetic information encoded within a gene is transcribed into mRNA. mRNA is transcribed from DNA by a DNA-dependent RNA polymerase (RNAP). Cyanobacterial RNAP is formed of 6 protein subunits (α_2ßß'ω where ß' is split into two subunits) which form the core enzyme (Schneider *et al.*, 1987). The binding of dissociable sigma factors to the core enzyme results in the formation of the holoenzyme complex. Sigma factors enable the binding of RNAP to groups of specific promoter sequences. While two groups of sigma factors are present across bacteria, σ^{70}-type and σ^{54}-type, cyanobacteria contain only σ^{70}-type factors (Gruber and Gross, 2003). There are four groups of σ^{70}-type cyanobacterial sigma factors. Group 1 encodes a single factor which is responsible for transcribing the basic housekeeping genes required for cell growth and maintenance. Group 2, 3 and 4 sigma factors are responsible for transcribing genes under stress conditions such as heat or salt stress, or genes which enable adaptation to specific environmental conditions (Riaz-Bradley, 2019).

Transcription can be described in three distinct stages: initiation, elongation and termination. In E. coli, transcription initiation occurs when the RNAP holoenzyme binds to a promoter region via contact between the sigma factor and the -35 and -10 hexamers of the promoter sequence, which are separated by a spacer region of 16-19 bp (Figure 1-11). The sigma factor also contacts an extended upstream region of the -10 hexamer (EXT) as well as a discriminator region (DISC) of 6-8 bp in length. The C-terminal domains of the RNAP α-subunits also bind to an upstream element (UP element) of the promoter, typically between 40-60 bp upstream of the transcriptional start site (+1) (Ruff *et al.*, 2015).

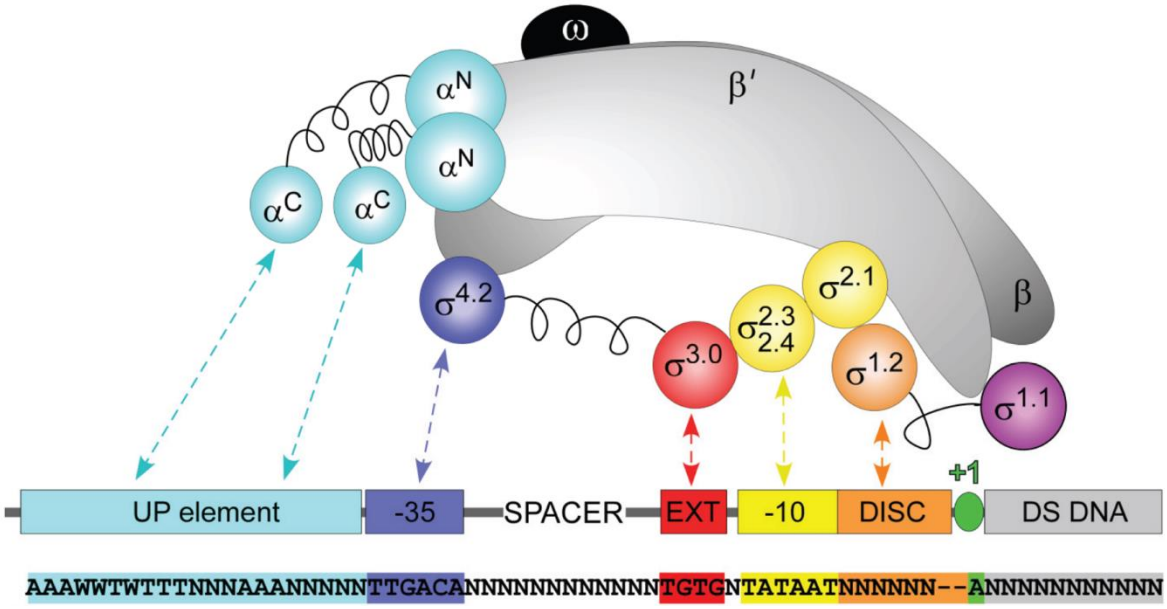


Figure 1-11. RNAP-promoter interactions which affect transcription initiation rate. Schematic of *E. coli* RNAP and the associated consensus promoter sequence. The various subunits (αßß′ω) of RNAP and the promoter regions to which they bind are illustrated with matching colours. All RNAP and sigma factor (σ) subunits are labelled. UP element – upstream -35 region; EXT – extended -10 region; DISC – discriminator region; +1 = transcription start site. Figure from Ruff *et al.*, 2015.

Differences in either the sequence composition or the spacing between any of the discussed promoter elements can influence the transcription initiation rate. For many organisms, a consensus sequence for the -35 and -10 hexamers is known which facilitates the maximum transcription initiation rate. CAGACA and TATAAT are -35 and -10 hexamer sequences respectively which are known to initiate high levels of transcription in Syn7002 (Markley *et al.*, 2015). However, the optimal -35 and -10 sequences for maximising gene expression often vary between species. In *E. coli*, consensus sequences are also known for the UP element and the EXT element (Ruff *et al.*, 2015) (Figure 1-6).

Transcription initiation begins with binding of RNAP to the UP element, the -35 region and the -10 region. This causes the bending of the DNA in the UP element and subsequent wrapping of the upstream sequence around the RNAP C-terminal domains. DNA cannot enter the RNAP active site until the RNAP binds to the promoter which triggers a conformational change. This is due to steric hindrance provided by the $\sigma^{1.1}$ subunit (Murakami, 2013). The DNA downstream of the -10 hexamer is then 'bent' into the active site cleft of RNAP and is melted to form the starting transcription bubble. It is important to note that the exact DNA

region bent into the cleft is variable depending on the specific promoter in question (Ruff *et al.*, 2015). Furthermore, for some promoters, the UP element and the -35 element are bound first, the DNA is unwound, and only then does RNAP bind the single-stranded -10 element (Sclavi *et al.*, 2005). Due to the strong binding between the sigma factor and the template strand, several short, aborted mRNA transcripts are usually produced as RNAP cannot move across the DNA. However, DNA continues to be unwound at this stage, causing it to become 'scrunched' inside the active site cleft. This 'scrunching' results in the energy required to release RNAP and begin productive transcription elongation (Revyakin *et al.*, 2006).

Transcription elongation proceeds, with RNA nucleotides entering the RNAP active site and being successively added to the mRNA chain. Energy derived from the hydrolysis of the nucleotide triphosphate bond is used to power the translocation of the RNAP across the DNA. Each nucleotide diffuses through the RNAP entry channel into the active site, with specifically only ribose-based nucleotides able to enter the channel. The correct RNA nucleotide binds to its complementary base at the active site and is catalytically joined to the growing chain. However, if the incorrect nucleotide enters the active site, the RNAP stalls to allow time for the correct nucleotide to naturally diffuse into the channel and replace it. Alternatively, if the incorrect nucleotide is added to the chain, the RNAP backtracks, releases the 3' end of the mRNA and cleaves it, allowing elongation to proceed again (Riaz-Bradley, 2019).

RNAP-mediated cleavage of the 3' end of the mRNA upon the addition of an incorrect nucleotide is a slow process, thought to be on the order of several minutes. The speed of this process can be greatly enhanced by the Gre proofreading factor. However, cyanobacteria lack a homolog for the GreA proofreading enzyme, which is near-universally present across the bacterial kingdom. It has been suggested that the lack of GreA in cyanobacteria could be compensated for by their polyploid genomes (Riaz-Bradley, 2019). Multiple genome copies allow DNA replication and transcription to proceed on different chromosomes. This prevents damaging collisions from occurring between moving and stalled transcription and replication protein complexes. Indeed, it has been shown that chromosomes are replicated asynchronously in Syn7942 and generally

only one chromosome per cell is being replicated at any point in time (Chen *et al.*, 2012).

Transcriptional termination occurs when the RNAP reaches a terminator sequence, which causes RNAP to dissociate from the DNA template. In prokaryotes, terminators are either Rho-dependent or Rho-independent. Rho is a transcription factor which binds to the terminator sequence to halt transcription. However, only Rho-independent terminators have been identified in cyanobacteria (Vijayan *et al.*, 2011). These terminators are also known as intrinsic terminators, forming hairpin mRNA loops which block the mRNA exit channel of RNAP and 'knock' RNAP off the mRNA molecule, terminating transcription. Some terminators also encode a poly-uracil tract which causes RNAP stalling, providing sufficient time for the hairpin to form (Riaz-Bradley, 2019).

In addition to the sequence composition of a promoter, transcription factors can also influence the transcription initiation rate. The activity of these proteins is often controlled by either the binding of a ligand or by phosphorylation, often via a sensory protein kinase. Transcriptional activators function by enhancing RNAP-promoter binding through three main mechanisms. The first is where the transcriptional activator binds to both a region upstream of the UP element and the C-terminal domain of RNAP. The second is where the activator binds to the region directly upstream of the -35 hexamer and to the $\sigma^{4.2}$ subunit. Finally, the activator can bind to the linker region between the -35 and -10 hexamers without directly binding to RNAP. Binding in the linker region is thought to directly influence the promoter structure to better facilitate sigma factor binding (Browning and Busby, 2016).

Transcriptional repressors also function through several mechanisms which reduce RNAP-promoter binding. The first is where the repressor, an example being OxyR in *E. coli*, binds close to the -35 or -10 hexamer and prevents the binding of RNAP to the promoter through steric hindrance (Browning and Busby, 2016). Alternatively, repressors can bind to both the region upstream of the UP element, as well as within the coding sequence of the gene. Protein-protein interactions between these repressors then result in the looping of the DNA, which reduces the physical space available for RNAP-promoter binding. An

example of this is the GalR repressor of the galactose operon in *E. coli* (Swint-Kruse and Matthews, 2009). Finally, repressors can bind directly to an activator and prevent it from recruiting RNAP to the promoter (Browning and Busby, 2016).

1.4.2 Translation

During translation, the transcribed mRNA is translated into protein. Generally, in bacteria, transcription and translation occur simultaneously through physical interaction between the RNAP and the ribosome (Kohler *et al.*, 2017). However, this is not always the case in cyanobacteria, where certain thylakoid membrane proteins are translated by ribosomes physically associated with the thylakoids (Mahbub and Mullineaux, 2023). The region of the mRNA which is bound by the ribosome is known as the ribosome binding site (RBS). This region is usually located between -20 and +15 base pairs with regard to the start codon. Many, but not all bacterial mRNAs contain a Shine-Dalgarno sequence within the RBS, an 8-10 nucleotide sequence directly upstream of the start codon. The Shine-Dalgarno sequence binds directly to its complementary anti-Shine-Dalgarno sequence contained within the 16S ribosomal RNA (rRNA). Three initiation factors IF1, IF2 and IF3 play various roles in translation including recruiting the initial Methionine-tRNA and preventing the binding of elongation tRNAs (Rodnina, 2018). Many factors are known to affect the translation rate of a given mRNA. These include the Shine-Dalgarno sequence, the spacing between the Shine-Dalgarno sequence and the start codon, the bases present within the start codon, mRNA secondary structure and the presence of AU-rich sequence which binds to the ribosomal protein bS1 (Rodnina, 2018).

Translation precedes through four main stages: initiation, elongation, termination and ribosome recycling. During initiation, the ribosome binds both the mRNA and the initiator Methionine-tRNA. Elongation then proceeds via decoding, where each mRNA codon is bound by the complementary anti-codon of the tRNA, peptide bond formation and subsequent translocation of the mRNA through the ribosome via GTP hydrolysis. The protein forms a compact, partially folded structure during its passage through the ribosomal exit channel before the domain fully folds upon its complete emergence (Holtkamp *et al.*, 2015). Termination occurs when the ribosome encounters an mRNA stop codon. The

ester bond of the termination peptidyl-tRNA is hydrolysed and the peptide is consequently released from the ribosome. Finally, ribosome recycling occurs when the ribosomal subunits split and the mRNA and tRNAs are released, enabling the next cycle of translation to begin (Rodnina, 2018).

1.5 Potential of cyanobacteria for industrial biotechnology

Due to their numerous desirable characteristics, cyanobacteria have great potential in the industrial biotechnology sector as an environmentally sustainable cell chassis to produce biochemicals. They require only light, CO_2 and basic mineral media for growth (Figure 1-12). Many cyanobacterial species can grow in seawater or wastewater, preventing the need for scarce freshwater to be used (Abed *et al.*, 2009). Furthermore, many species can naturally uptake DNA from their environment or are amenable to established transformation techniques such as conjugation, enabling genetic manipulation (Al-Haj *et al.*, 2016). As of early 2024, over 300 cyanobacterial species have fully sequenced genomes contained within the GenBank database, whilst genome-scale models describing most cellular metabolic pathways are available for at least 14 species (Alvarenga *et al.*, 2017; Santos-Merino *et al.*, 2019). Some species also present fast growth rates, with *Synechococcus* sp. PCC 11901 (Syn11901) reportedly reaching a doubling time of around 2 hours and accumulating up to 33 g L^{-1} of dry cell biomass (Włodarczyk *et al.*, 2020).

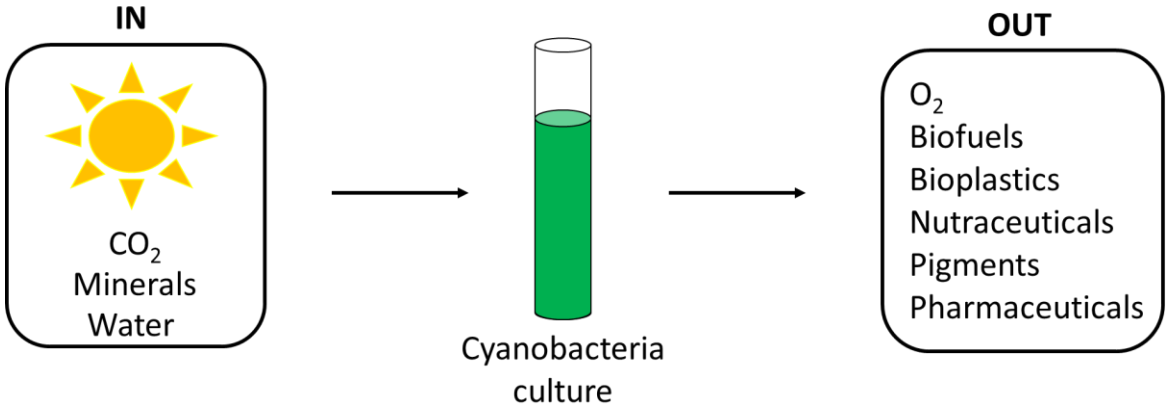


Figure 1-12. Cyanobacteria as a cell chassis for biochemical production. Providing cyanobacteria with simple inputs enables the production of a range of natural and engineered products.

In addition to their attractive growth characteristics, cyanobacteria contain a key environmental advantage over heterotrophic chassis typically used in

industrial biotechnology such as *E. coli* and *Saccharomyces cerevisiae*. These heterotrophs require a carbon source such as glucose or other simple sugars for growth. As these sugars are often derived from food crops, arable land is required to supply heterotrophic organisms with energy (Chan *et al.*, 2012; Singh *et al.*, 2013). In contrast, cyanobacteria use CO_2 as a carbon source during photosynthesis. Therefore, using cyanobacteria for biochemical production can help reduce global CO_2 emissions and alleviate pressure on valuable land resources. Potential opportunities also exist for applying cyanobacteria with carbon capture, storage and utilisation technologies. CO_2 produced by carbon-intensive industries could be directly captured by cyanobacteria and converted to valuable products.

Owing to their beneficial properties, cyanobacteria have been used to produce various biochemicals. Cyanobacteria naturally produce several high-value pigments including phycocyanin, which is used as a blue dye in the food and pharmaceutical industries (Dasgupta, 2016). Many species naturally produce extracellular polymeric substances which can be used in the formation of bioplastics, or even possess anti-viral, anti-microbial and anti-cancer properties (Pereira *et al.*, 2019). Other products naturally produced by cyanobacteria include vitamins, amino acids, fatty acids and anti-toxins (Lau *et al.*, 2015). Indeed, over 2000 different secondary metabolites have been identified in cyanobacteria (Jones *et al.*, 2021). Many more natural cyanobacterial products with useful functions will likely be discovered in the future.

In addition to natural products, cyanobacteria have been genetically engineered to produce various non-native products including biofuels like isobutanol and ethanol (Atsumi *et al.*, 2009; Luan *et al.*, 2015). They have also been engineered to produce ethylene, a precursor molecule used extensively in the chemical synthesis industry (Guerrero *et al.*, 2012). Furthermore, cyanobacteria have been engineered to produce poly-β-hydroxybutyrate (PHB), a polymer used in biodegradable plastics (Carpine *et al.*, 2017).

Cyanobacteria can also be used in industrial applications other than biochemical production. Syn6803 and Syn7002 have been engineered for the biodesalination of water via the expression of light-activated chloride pumps, causing cells to accumulate chloride ions (Minas *et al.*, 2015). They have also been used for

bioremediation, due to their natural or engineered abilities to accumulate and metabolise toxic waste products. A range of waste products have been remediated by cyanobacteria including petroleum effluent and heavy metals (Joseph and Joseph, 2001; Zinicovscaia *et al.*, 2018).

One cyanobacterial species of interest for biotechnological applications is the unicellular euryhaline Syn7002. Syn7002 is a fast-growing species, with a reported doubling time of 2.6 hours when grown under optimal conditions and a reduced nitrogen source (Ludwig and Bryant, 2012). Syn7002 is naturally transformable and has a fully sequenced genome of 3.3 Mbp, allowing ease of genetic manipulation. Syn7002 can grow in the brackish water found in some seas and lakes as opposed to freshwater and can also withstand extremely high light intensities of up to 4500 µmol m^{-2} s^{-1} (Nomura *et al.*, 2006). Syn7002 can also grow at various temperatures, with exponential growth observed between 22°C and 38°C, albeit 38°C is the optimal temperature for growth (Ludwig and Bryant, 2012). This allows the cultivation of Syn7002 in open pond systems where cells must withstand high temperatures, high light intensity and fluctuating salt concentrations due to evaporation (Ruffing *et al.*, 2016). Cultivation in open pond systems is often less expensive at industrial scale than photobioreactors and may facilitate the economics of using cyanobacteria for producing medium-value, high-volume products. The recently discovered Syn11901 demonstrates faster growth rates and reaches higher cell densities than Syn7002 under optimal growth conditions. However, the physiology of Syn7002 is better understood and many more genetic tools have been characterised in this species.

Despite their clear potential, using cyanobacteria for biochemical production has several drawbacks. Although they can achieve very high final culture densities, the doubling times of even the fastest-growing cyanobacterial species, such as the 2 hours presented by Syn11901 are significantly longer than the 20-minute doubling time of *E. coli* (Włodarczyk *et al.*, 2020). In addition, although many synthetic biology tools have been developed for cyanobacteria including promoters, ribosome binding sites and plasmids, the toolbox lags those available for *E. coli* and *S. cerevisiae*. Furthermore, several tools which are well-characterised in *E. coli* perform differently in cyanobacteria. For example, the

commonly used *pTrc* inducible promoter is tightly regulated in *E. coli* but enables constitutive expression in cyanobacteria (Guerrero *et al.*, 2012). These differences could be due to the structural variation between *E. coli* and cyanobacterial RNA polymerases, as well as the absence of σ^{56}-family sigma factors in cyanobacteria (Schneider and Hasekorn, 1988; Gruber and Gross, 2003). As a consequence of these drawbacks, the achievable titres of most biochemicals produced in cyanobacteria are not high enough to be economically competitive with those produced by heterotrophic bacteria or petrochemical-based synthesis.

1.6 Tools for controlling cyanobacterial gene expression

1.6.1 Plasmids

Successful genetic engineering of cyanobacteria for biotechnology requires a suite of tools to control gene expression. Plasmids are a common vector used for expressing heterologous genes and two types are used in cyanobacterial genetic engineering: integrative and self-replicating plasmids. Integrative or 'suicide' plasmids allow the integration of DNA into the genome by placing the sequence to be inserted between two 'flank' sites on the plasmid. The flank sites contain sequences homologous to the sequences flanking the target region of the genome. The natural ability of cyanobacteria for homologous recombination allows the sequence of interest to be inserted between the genomic flanks. Thus, integrative plasmids allow DNA insertion in potentially any area of the genome, including both chromosomes and endogenous plasmids. Various "neutral" sites have been identified across cyanobacterial genomes, in which insertions are reported to have minimal effect on cell growth or physiology (Ruffing *et al.*, 2016; Nagy *et al.*, 2021).

The integrative approach to cyanobacterial genetic engineering is complicated by their polyploid genomes. To obtain a homozygous strain, several rounds of streaking successful transformants on antibiotic selection plates and screening by colony PCR are often required. This can take several weeks depending on both the copy number of the chromosome or plasmid which is being targeted and the growth rate of the species on agar plates. An approach has been developed to shorten this process in Syn6803, where the culture is grown in reduced

phosphate media to reduce genome copy number before transformation (Pope *et al.*, 2020). However, this also resulted in a 200-fold decrease in transformation efficiency when compared to growth in standard media, so may be more useful during the colony re-streaking process required to achieve complete segregation.

An alternative approach to integrative plasmids is the expression of genes on a self-replicating plasmid. This shortens the transformation process as unlike integrative plasmids, several rounds of segregation and screening are not required. Transformation of cyanobacteria with these plasmids requires either conjugation or electroporation, which are often more time-consuming procedures than natural transformation. However, these can be used for the transformation of theoretically any bacterial species, whereas natural transformation is limited to a handful of naturally competent species.

The use of both integrative and self-replicating plasmids is complicated by the fact that chromosome, endogenous and even heterologous plasmid copy numbers can be highly variable across different growth conditions (Zerulla *et al.*, 2016). The expression of target genes can therefore be subject to a gene dosage effect, where increased gene copies result in increased gene expression. This can be an issue when the balance of the expression of genes in a metabolic pathway is important to maximising the production of a target molecule. However, this can also be exploited by targeting high copy number endogenous plasmids for overexpression of target genes, such as targeting the small pAQ1 plasmid of Syn7002 which has around 50 copies per cell during the early exponential phase (Xu *et al.*, 2011).

1.6.2 Promoters

Genetic engineering applications often use heterologous promoters to control the expression of target genes. This prevents the potential interference of native regulatory mechanisms which can occur when using endogenous promoters. In addition, using heterologous promoters reduces the probability of genetic instability arising from homologous recombination with native sequences (Gordon and Pfleger, 2018). Various heterologous promoters, often from other cyanobacterial species, have been applied in cyanobacteria. These include the strong constitutive promoter *pCpc560* originating from Syn6803, which can

enable target protein accumulation of up to 15% of total cellular protein content, albeit this is dependent on the specific protein (Zhou *et al.*, 2014).

While strong constitutive promoters are useful when maximal gene expression is required, they are not ideal for the expression of toxic proteins. Protein over-production can also induce a significant metabolic burden on the cell, reducing growth rates. Over time, this may lead to the selection of cells with mutations which reduce promoter activity, causing culture heterogeneity (Pouzet *et al.*, 2020). Indeed, computational analysis of cyanobacterial metabolism has highlighted trade-offs between biomass accumulation and target protein production in terms of ATP/NADPH demand and the re-integration of metabolic by-products (Knoop and Steuer, 2015).

The drawbacks of strong constitutive promoters have led to the use of inducible promoters to maximise product titres during bioproduction. These promoters allow the separation of the production process into a biomass accumulation phase and a product synthesis phase, thus preventing product-mediated growth rate inhibition. Inducible promoters used in cyanobacteria include the nickel-inducible *PnrsB*, L-arabinose-inducible *ParaBAD* and rhamnose-inducible *PrhaBAD* promoters (Englund *et al.*, 2016; Cao *et al.*, 2017; Kelly *et al.*, 2018). While inducible promoters allow greater control over gene expression than constitutive promoters, they suffer from inherent drawbacks. These include potential inducer toxicity, the high cost of inducers at industrial scale and the need to remove the inducer from the culture during downstream processing (Badary *et al.*, 2015). However, their main drawback is the irreversible activation of gene expression, preventing more dynamic metabolic pathway control strategies from being used.

An alternative to inducible promoters has been the use of promoters which become activated during the stationary phase of culture growth. This includes the *PndbA600* promoter originating from Syn6803. When used to control gene expression in Syn7002, *PndbA600* allows distinct separation of the biomass accumulation and product synthesis stages without requiring the addition of an inducer molecule (Madsen *et al.*, 2021). However, stationary phase promoters still do not allow for dynamic control over gene expression i.e. reversible activation and deactivation.

1.6.3 Ribosome binding sites

The strength of a range of RBS have been characterised in cyanobacteria, seemingly allowing them to be used to control protein translation rates. However, RBS strengths vary widely depending on both the species and the genetic context. Both the complementarity of the RBS to the 16S rRNA and the mRNA secondary structure proximal to the RBS influence the overall translation rate (Gordon and Pfleger, 2018). Therefore, a RBS which enables a high translation initiation rate for one gene can result in a low translation initiation rate for another gene. Software tools such as the Salis RBS calculator are available which can predict the translation initiation rate for a given RBS, gene and species, as well as design RBS to user-defined translation initiation rates (Reis and Salis, 2020). The tool uses a thermodynamic model to predict the strength of interaction between the mRNA region directly upstream of the start codon and the ribosome, as well as the mRNA secondary structure. Previous work using the Salis RBS calculator to design RBS for Syn7002 demonstrated that the predicted RBS strength does not always correlate well with the *in vivo* strength in cyanobacteria (Markley *et al.*, 2015). However, this work used version 1.1 of the calculator, so it remains to be seen whether version 2.1 is more effective (Reis and Salis, 2020).

1.6.4 Optogenetics

Optogenetics describes the use of specific light wavelengths to control biological function within a cell. This approach relies upon photoreceptors which respond to irradiation with specific wavelengths often within milliseconds (Deisseroth, 2011). The first application of optogenetics was in the field of neuroscience to control neuron activity using Channelrhodopsin-2 (ChR2), a light-sensitive algal cation channel (Boyden *et al.*, 2005). In rat neuronal cells expressing ChR2, irradiation with a brief pulse of blue light caused rapid depolarisation of the neuron membrane potential. Since then, optogenetic tools have been used to control various molecular functions including protein localisation, protein degradation, gene expression, epigenetic modifications and enzyme activity (Niopek *et al.*, 2014; Usherenko *et al.*, 2014; Fernandez-Rodriguez *et al.*, 2017; Lo *et al.*, 2017; Fomicheva *et al.*, 2019).

The most common application of optogenetics is to control gene expression. Light is inexpensive, non-invasive and offers much greater spatial and temporal control of gene expression than inducible promoters (Lalwani *et al.*, 2021a). Optogenetic tools can be rapidly switched on and off to reversibly modulate gene expression. This allows the optimisation of metabolic pathways through precise control over the timing of enzyme expression. In addition, multiple optogenetic tools can be used in combination with numerous pairs of wavelengths to reversibly control the expression of multiple target genes in a cell.

Optogenetic control of gene expression is often elicited through light-controlled binding of transcriptional activators or repressors to promoters. An example of the light-regulated control of a transcriptional activator is the CcaS/CcaR system, which originates from the cyanobacterium Syn6803. Green light absorption by the CcaS photoreceptor triggers phosphorylation and activation of the CcaR transcription factor, which binds to and activates transcription from the pCpcG2 promoter (discussed further in section 3.1.2) (Hirose *et al.*, 2008). An example of a light-controlled transcriptional repressor is the LightOff system, which enables blue Light-regulated repression of transcription in *E. coli*. This system comprises the Vivid photoreceptor of the fungus *Neurospora crassa*, which contains a blue light-absorbing LOV domain, fused to the DNA-binding domain of the LexA repressor. The system is inactive as a monomer in the dark, however, blue light absorption triggers reversible homodimerisation of the fusion protein, allowing subsequent binding to the target promoter region to inhibit transcription via steric hindrance (Chen *et al.*, 2016).

Another strategy to control gene expression is via light-regulated reconstitution of RNAP. The Opto-T7RNAP system uses a split T7 RNAP to control gene expression (Baumschlager *et al.*, 2017). Each half of the RNAP is fused to the heterodimeric "Magnets" domains, which originate from the blue light-absorbing Vivid photoreceptor. Upon blue light absorption, the Magnet domains dimerise, which allows the reconstitution of the two halves of the RNAP and subsequent activation of transcription from the T7 promoter. The absence of blue light triggers the dissociation of the magnets and RNAP, switching transcription off.

Through controlling gene expression, optogenetics can significantly enhance product titres under certain bioproduction scenarios. When using optogenetics to control isobutanol production in *Saccharomyces cerevisiae,* providing regular pulses of inducing light resulted in triple the isobutanol titre versus the use of a constant light induction phase (Zhao *et al.*, 2018). This type of dynamic regulation would not be possible when using chemical inducers due to the inability to remove the inducer from the culture without extensive processing steps.

Post-transcriptional regulation via optogenetics is also possible, with light-controlled RNA binding proteins able to influence mRNA translation efficiency. The pCrepusculo system is based on the PAL photoreceptor, which consists of a LOV domain fused to a sequence-specific RNA-binding domain (Ranzani *et al.*, 2022). In darkness, the RNA-binding domain is tightly bound to the LOV domain, preventing RNA binding. Blue light absorption triggers a conformational change in the protein which results in the RNA-binding domain being unwound from the LOV domain. This allows the protein to now bind to its specific aptamer sequence. By placing this specific aptamer within the RBS of target genes, their expression can be significantly reduced through the steric hindrance of RNAP upon blue light illumination.

Optogenetic systems have also been developed which enable post-translational control of protein degradation. There are fewer examples of this type of regulation in bacteria than in eukaryotes. However, a system has been developed based on the ability of the ClpAPX protease to cleave bacterial proteins which contain the ssrA peptide tag. Similar to the strategies discussed above, the fusion of the blue light-absorbing Vivid photoreceptor to a split Tobacco Etch Virus (TEV) protease results in a chimeric protein in which cleavage activity is very low in darkness. Upon blue light absorption, the fusion protein homodimerises, causing reconstitution of the split TEV protease domains and triggering cleavage of ssrA-tagged proteins (Komera *et al.*, 2022).

Finally, optogenetics can be used to control the formation and dissociation of intracellular compartments. By fusing target enzymes with protein domains which enable reversible oligomerisation, different enzymes can be brought into close proximity to enhance substrate channelling and product formation. The

Cry2 photoreceptor from *Arabidopsis thaliana* undergoes reversible oligomerisation upon blue light absorption. This has been exploited by fusing Cry2 to the intrinsically disordered region of the FUS protein to generate Cry2-FUS (Shin *et al.*, 2017). The fusion of Cry2-FUS to key enzymes in the deoxyviolacein pathway in *S. cerevisiae* enabled the blue light-regulated formation and dissociation of clusters containing these enzymes. This significantly enhanced product synthesis, presumably by reducing the accumulation of toxic intermediates and increasing the efficiency of substrate channelling between the enzymes (Zhao *et al.* 2019). A summary of the mechanisms by which optogenetics can be used to control protein abundance and function is shown in Figure 1-13.

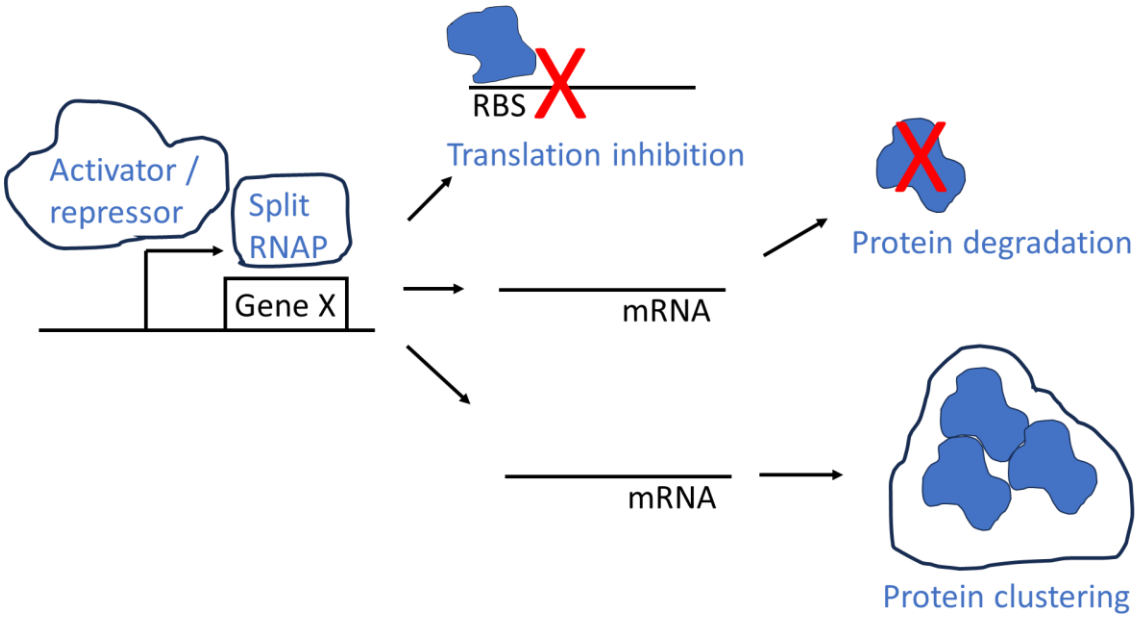


Figure 1-13. Optogenetic mechanisms to control target protein abundance and function. Points of control are highlighted in blue text. RBS – ribosome binding site; RNAP – RNA polymerase.

While optogenetic tools have clear utility, their use also has certain disadvantages. Firstly, most systems require the supply of a chromophore for the photoreceptor to function correctly. If the chosen organism does not produce this chromophore endogenously, it must be supplied exogenously, or the associated biosynthetic enzymes must be expressed. Also, optogenetic systems are not always completely portable between species, even when chromophore production is accounted for. Light may also be toxic to some organisms, especially UV/violet light.

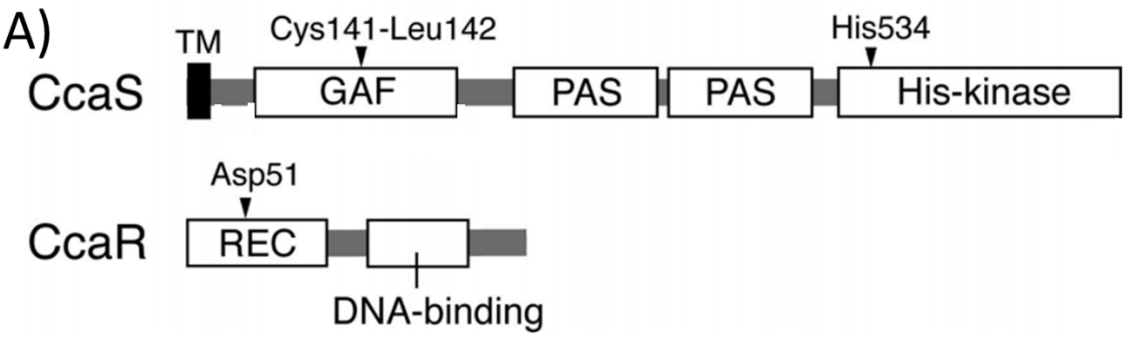
To date, only one optogenetic tool, the CcaS/CcaR system, has been applied in cyanobacteria and it has been limited to a small number of species. Several issues contribute to the limited number of optogenetic tools deployed in cyanobacteria. Firstly, the use of optogenetics in cyanobacteria is complicated as they rely on light for growth. This means that the wavelengths used for photosynthesis should not overlap with the wavelengths required for optogenetic control. Removing specific wavelengths from the photosynthetic light source, usually white light, could have negative effects on culture growth. Secondly, optogenetic tools that have been well-characterised in other organisms may function differently in a cyanobacterial system that contains complex light-controlled regulatory mechanisms. Thirdly, the cyanobacterial genetic engineering toolbox lags significantly behind model heterotrophic organisms. Therefore, limited research has focused on the application of optogenetics in cyanobacteria.

Developing systems that respond to different light wavelengths than the green/red light responsive CcaS/CcaR system would expand the optogenetic toolkit in cyanobacteria. Systems which use blue light are particularly interesting, as blue light appears to be much less important for the fast growth of some cyanobacterial species than red or orange light (Luimstra *et al.*, 2018). The addition of another optogenetic tool to the cyanobacterial toolkit could also enable light-regulated control of multiple genes in metabolic pathways and help realise the full potential of cyanobacteria for industrial biotechnology.

1.7 CcaS/CcaR origin, protein domains and mechanism

CcaS is a CBCR that controls the chromatic acclimation response to green/red light in Syn6803 (Hirose *et al.*, 2008). The monomeric CcaS photoreceptor contains a putative transmembrane domain, a single GAF domain, two PAS domains, and a histidine kinase domain (Figure 1-14a). The GAF domain binds PCB as its chromophore and is responsible for photoperception. The histidine kinase domain phosphorylates the downstream target CcaR. The role of the PAS domains is unclear. In *E. coli*, a truncated version of CcaS with both PAS domains removed still responds to green and red light in a similar manner as the full Length protein (Nakajima *et al.*, 2016). PAS domains are present across all kingdoms of-life and have a wide range of functions, for example, enabling

protein-protein interactions or as sensory domains for ligands and oxygen (Lindebro *et al.*, 1995; Gong *et al.*, 1998; Jiang *et al.*, 1999). CcaS PAS domains could mediate interactions with other unknown proteins *in vivo* to allow additional signal input other than green and red light. A crystal structure is not available for CcaS, but a high-confidence predicted structure is available on AlphaFold, albeit it is unclear whether this is the green or red light absorbing form (Figure 1-14b) (Jumper *et al.*, 2021). While the mechanism of action of CcaS has not been confirmed experimentally, one hypothesis is that it operates via the coiled-coil mechanism like YF1 (discussed later). This hypothesis is based on previous work, which showed the deletion of varying numbers of amino acids in the linker region between the second PAS domain and the histidine kinase domain inverted the response of CcaS from green light kinase activity to red light kinase activity (Nakajima *et al.*, 2016).


A)
TM
Cys141-Leu142
His534
CcaS
GAF
PAS
PAS
His-kinase
Asp51
CcaR
REC
DNA-binding

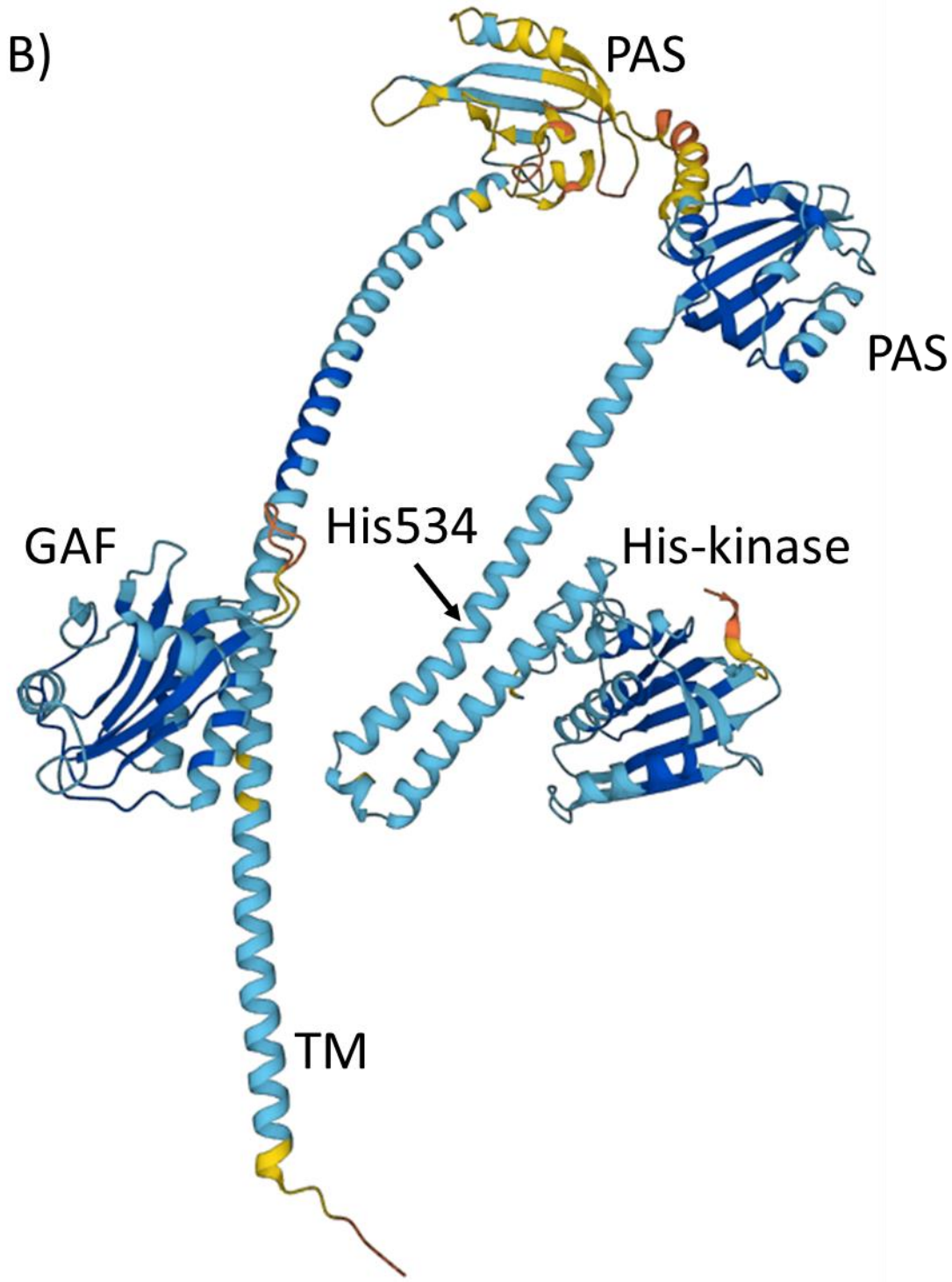

B)
PAS
PAS
GAF
His534
His-kinase
TM

Figure 1-14. Protein domains of CcaS and CcaR and predicted structure of CcaS. A) Domain structure of CcaS and CcaR. B) Predicted structure of CcaS available on AlphaFold (https://alphafold.ebi.ac.uk/entry/Q06FY7 - accessed 21 Feb 24). Residues important to the function of specific domains are highlighted. TM – transmembrane domain; GAF - cGMP-phosphodiesterase/adenylate cyclase/FhlA domain; PAS – Per-Arnt-Sim domain; His-kinase – histidine kinase; REC – receiver domain. Adapted from Hirose *et al.*, 2008 and Jumper *et al.*, 2021.

The GAF domain of CcaS is responsible for photoperception and exists in two forms. The dark-state ^{15Z}Pg form (Pg-CcaS) of the GAF domain binds deprotonated PCB in the C15-Z conformation and absorbs light wavelengths between 470-600 nm, with maximal absorption at 535 nm. Conversely, the photoproduct ^{15E}Pr form (Pr-CcaS) binds protonated PCB in the C15-E conformation and absorbs light wavelengths between 600-700 nm, with maximal absorption at 672 nm (Hirose *et al.*, 2008). In both forms, the chromophore ligation motif is located at Cys141 (Figure 1-14a), where the cysteine forms a covalent adduct with PCB. The leucine likely plays a role in indirectly stabilising PCB binding through non-covalent interactions; however, this has not been confirmed experimentally. Substitution of Cys141 with alanine prevents the binding of PCB to the GAF domain (Hirose *et al.*, 2008).

Absorption of green light by Pg-CcaS, the dark state of CcaS, causes a switch to Pr-CcaS. This triggers autophosphorylation of CcaS at His-534 in the histidine kinase domain. The phosphate group is then transferred from CcaS to Asp-51 in the receiver domain of CcaR. Phosphorylated CcaR then binds to the promoter sequence pCpcG2, activating transcription of the *cpcG2* gene, also known as *CpcL*. From studies of the CcaS and CcaR orthologs of *N. punctiforme*, CcaR is thought to bind to a region of pCpcG2 known as the G-box, a double repeat of the sequence CTTTNCNATTT (Hirose *et al.*, 2010). The G-box sequence is identical between Syn6803 and *N. punctiforme* pCpcG2.

Absorption of red light by Pr-CcaS triggers a switch back to Pg-CcaS. This reduces CcaS autophosphorylation and may cause dephosphorylation of CcaR, which is hypothesised to occur via CcaS phosphatase activity (Hirose *et al.*, 2008). Dephosphorylated CcaR has low affinity for pCpcG2, thus under red light transcription of *CpcG2* is reduced. The mode of action of the CcaS/CcaR system can be measured by replacing *CpcG2* with *GFP* as shown in Figure 1-15.

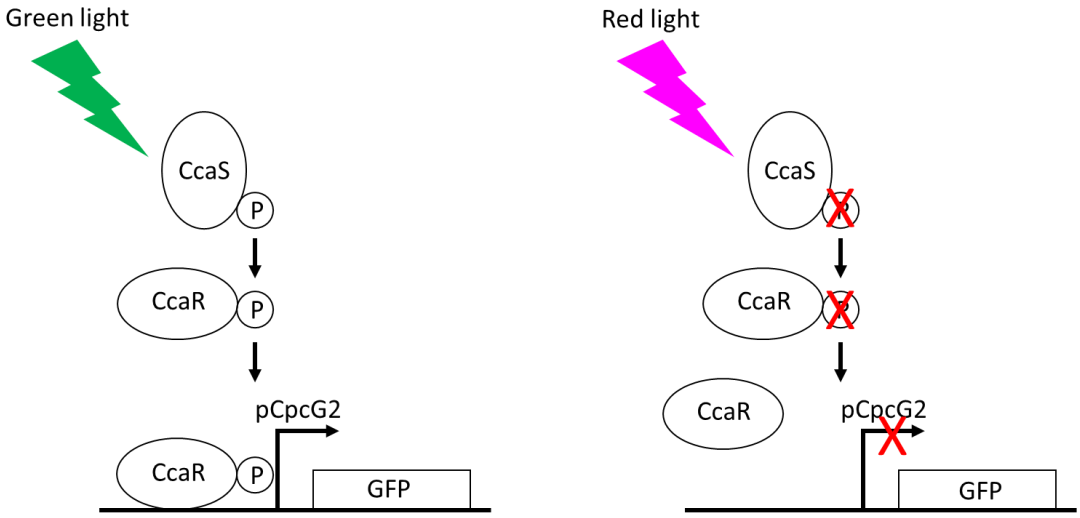


Figure 1-15. CcaS system function under green and red light. Upon green light absorption (λ_{max} = 535 nm), CcaS becomes autophosphorylated and phosphorylates the transcriptional activator CcaR. Phosphorylated CcaR binds to the G-box region of the pCpcG2 promoter, activating the transcription of the target gene. For ease of measurement, the native CpcG2 can be replaced with the sequence for Green Fluorescent Protein (GFP) as shown here. Upon red light absorption (λ_{max} = 672 nm), CcaS becomes dephosphorylated, dephosphorylates CcaR and transcription from pCpcG2 is reduced.

The native CcaS system controls the expression of CpcG2, a rod Linker protein involved in the formation of the CpcG2-phycobilisome. The CpcG2-phycobilisome preferentially transfers excitation energy to PSI and lacks the typical allophycocyanin core, whereas the CpcG1-phycobilisome preferentially transfers energy to PSII and contains allophycocyanin (Kondo *et al.*, 2007). It has been hypothesised that the accumulation of the CpcG2-phycobilisome compensates for reduced PSI excitation caused by the poor absorption of green light by PSI reaction centre chlorophyll (Hirose *et al.*, 2008).

The rate of dark reversion, the speed at which the photoreceptor returns from the photoproduct state to the dark state in the absence of absorption of activating wavelengths, has been calculated for many photoreceptors. However, this data is unavailable yet for CcaS. One study resulted in the production of a model to predict the output response of the CcaS system in *E. coli* using various light intensities (Olson *et al.*, 2017). While the authors were unable to calculate an exact dark reversion rate from their experimental setup, it was clear from the data that the half-life was likely on the scale of hours rather than minutes. This contrasted with the red/far-red light absorbing Cph8 photoreceptor, which demonstrated a dark reversion half-life of 5.5 minutes in the same study. Cph8

has an ON state half-life of 4 minutes upon far-red light absorption (Olson *et al.*, 2014).

1.8 Blue light-regulated YF1/FixJ optogenetic system

The YF1/FixJ system, as opposed to being a naturally existing two-component system, was designed specifically for optogenetics. The blue light-repressible YF1/FixJ system was developed by replacing the heme-binding, oxygen-responsive PAS domain of the histidine kinase FixL of *Bradyrhizobium japonicum* with the blue light-sensitive LOV domain from *B. subtilis* YtvA (Möglich *et al.*, 2009). In their native organisms, FixL controls the expression of aerobic respiration and nitrogen fixation genes while YtvA is thought to initiate the *B. subtilis* general stress response upon absorption of blue light (Jurk *et al.*, 2013). The fusion of these two domains results in a chimeric protein in which histidine kinase activity is now controlled by blue light rather than oxygen concentration. In the absence of blue light, YF1 phosphorylates the transcription factor FixJ, which then binds to and activates transcription from the pFixK2 promoter. Blue light absorption by YF1 triggers its phosphatase activity and thus FixJ becomes dephosphorylated, resulting in a decrease in transcription from pFixK2 (Figure 1-16).

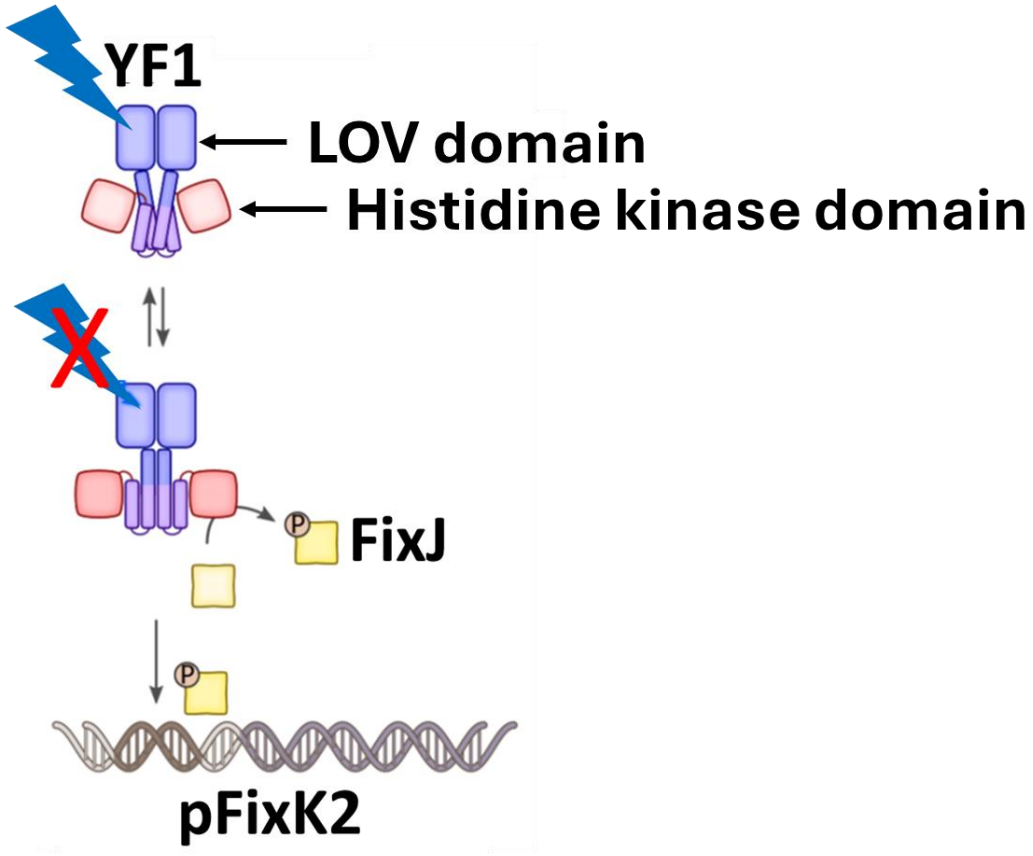


Figure 1-16. YF1/FixJ system for blue light-regulated gene expression. In the absence of blue light, YF1 phosphorylates the transcription factor FixJ, allowing activation of expression from the promoter pFixK2. Blue light prevents phosphorylation of FixJ and switches off gene expression from pFixK2. Adapted from Losi *et al.*, 2018.

The YF1 protein is a homodimer that is connected by the interaction of the short N-terminal A'α helices (Figure 1-17). Each monomer contains a photosensory LOV domain which binds flavin mononucleotide (FMN) as its chromophore. The LOV photosensory domain is connected via the Jα-helix to the output dimerisation and histidine phosphotransfer domain (DHp). Blue light absorption by FMN triggers the formation of an adduct with a conserved cysteine residue located within the LOV domain. Within 2 μs of cysteine adduct formation, both LOV domains pivot away from each other. This initiates left-handed supercoiling of the two Jα-helices and after 250 ms subsequent rearrangement of the interaction between the catalytic and TAP binding domain (CA) and DHp domains. The domain rearrangement sequesters His-161 from the protein surface to inside the core of the DHp domain. It has been proposed that this sequestering of H161 and the subsequent uncovering of residues important for phosphatase activity allows the switch from dark-activated kinase activity to

blue light-induced phosphatase activity (Berntsson *et al.*, 2017). Previous work also provides evidence that left-handed supercoiling of the Jα-helix linker is responsible for regulating kinase and phosphatase activity. Insertion of a single residue into the Jα-helix linker was able to switch YF1 from blue light-induced phosphatase activity to blue light-induced kinase activity (Ohlendorf *et al.*, 2016). Insertion of residues into the linker alters the orientation of the DHp domain so that H161 is sequestered under darkness rather than blue light and is uncovered in blue light rather than darkness.

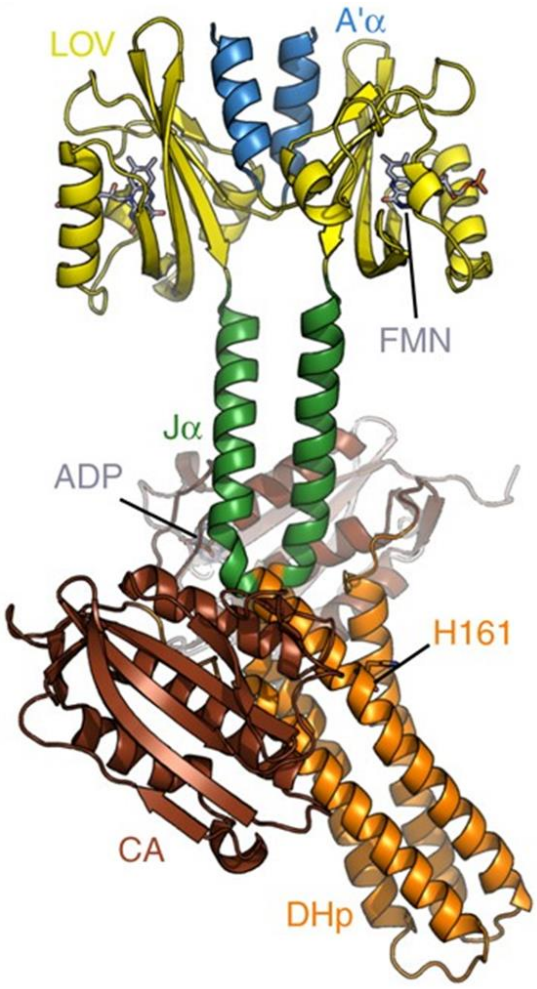


Figure 1-17. Structure of dark-state YF1 as determined by X-ray crystallography. Domains, binding sites and key residues are highlighted. LOV – light, oxygen, voltage domain, FMN – flavin mononucleotide, ADP – adenosine diphosphate, CA – catalytic and TAP binding domain, DHp - dimerisation and histidine phosphotransfer domain. Figure from Berntsson *et al.*, 2017.

1.9 The potential of optogenetics for L-alanine production

As discussed above, cyanobacteria have the potential to be exploited for biochemical production. While a range of commercially valuable compounds can be produced by engineered cyanobacteria, one which has not yet been reported

is alanine, an aliphatic, non-essential amino acid. Alanine exists as two stereoisomers: L-alanine and D-alanine. L-alanine is incorporated into proteins, whereas D-alanine is commonly found in the peptidoglycan cell wall of bacteria (Nagata, 1999). In bacteria, L-alanine is commonly synthesised via two routes. The first is NADH-dependent direct conversion of pyruvate and ammonia to L-alanine by an alanine dehydrogenase. Alternatively, L-aspartic acid can be directly converted to L-alanine via aspartate 4-decarboxylase.

L-alanine has many commercial applications including as a food supplement due to its sweet taste, as a precursor for the synthesis of pharmaceuticals like antibiotics and in cosmetic products where it acts as a natural skin moisturiser (Ikeda and Takano, 2013). The global L-alanine market was estimated at $300 million in 2020 and is expected to rise to $344.6 million by 2026 (360 Research Reports, 2020), while current global annual demand is estimated at 50,000 tonnes per year (Liu *et al.*, 2022). Historically, L-alanine was produced at industrial scale using a cell biotransformation process. *Pseudomonas dacunhae* and *E. coli* cells were immobilised on κ-carrageenan and converted ammonium fumarate substrate, via L-aspartic acid, to L-alanine (Sato *et al.*, 1982). However, ammonium fumarate is a fossil fuel-derived feedstock, which called for the development of a more environmentally sustainable fermentation process. Now, at least 60% of L-alanine annual output is produced via microbial fermentation of simple sugars, demonstrating that fermentation is an economically viable method for L-alanine production (Liu *et al.*, 2022). Conceptually, cyanobacteria could improve the sustainability of L-alanine production even further, by requiring only CO_2 as a carbon input. However, the significant challenges of lower growth rates and much lower product titres achieved by cyanobacteria versus heterotrophs would need to be overcome.

Several bacterial fermentation processes for L-alanine production have been reported. *E. coli* has been engineered to produce a final titre of 114 g L^{-1} of L-alanine from glucose feedstock, with a productivity of 2.37 g L^{-1} h^{-1} and a yield of 0.95g L-alanine per gram of glucose (Zhang *et al.*, 2007). This approach involved the constitutive expression of the alanine dehydrogenase *alaD* from *Geobacillus stearothermophilus* XL-65-6, as well as the knockout of seven genes from competing metabolic pathways. A metabolic evolution approach was also

used to increase productivity using a growth-based selection method. The resulting strain was named XZ132 and is currently being used to produce 23,000 tons of L-alanine per year. This fermentation takes place in 250 m^3 vessels using a mineral salt medium and is operated by Anhui Huaheng Biotechnology Co., Ltd. (Liu *et al.*, 2022).

More recently published work achieved an improved final titre of 121 g L^{-1} L-alanine via fermentation and an improved productivity of 3.1 g L^{-1} h^{-1} in *E. coli* (Zhou *et al.*, 2016). Here, a temperature-inducible promoter was used to control *G. stearothermophilus* alaD expression and thus separate culture growth from L-alanine synthesis. This led to faster growth throughout the fermentation process as it prevented L-alanine synthesis from depleting pyruvate during culture growth. Subsequently, the culture reached the stationary phase faster than when using a strong constitutive promoter for alaD expression, leading to improved productivity.

There are no reports describing the use of cyanobacteria or microalgae for L-alanine production. However, several cyanobacteria species have been engineered to produce and secrete other amino acids using only CO_2 as a carbon source. Syn7002 has been engineered for the production of L-Valine, albeit at a productivity of only 0.003 g L^{-1} h^{-1} (Korosh *et al.*, 2017). Similarly, *Synechococcus* UTEX 2973 (UTEX 2973) has been used to produce L Lysine to a final titre of 0.56 g L^{-1}, demonstrating a productivity of 0.005 g L^{-1} h^{-1} (Dookeran and Nielsen, 2021). This level of L Lysine production is roughly three orders of magnitude lower than the reported productivity of 4 g L^{-1} h^{-1} demonstrated during *C. glutamicum* fermentation (Becker *et al.*, 2011).

As L-alanine synthesis via *alaD* would deplete pyruvate and likely reduce cyanobacterial growth, optogenetics provides an opportunity to optimise the balance between biomass accumulation and L-alanine synthesis. Using the CcaS system to control *alaD* expression would allow the separation of growth and product synthesis phases without the drawbacks of using a chemical inducer. This could also enable the use of dynamic metabolic pathway control strategies by reversibly modulating alaD expression under red/green light in response to the effect on cell growth. As has been proven in *S. cerevisiae* for the production of various compounds which affect cell growth, dynamic control via optogenetics

could improve productivity when compared to a simple two-phase production process (Zhao *et al.*, 2021).

Syn7002 also appears well suited for L-alanine production when compared to other model cyanobacterial strains. Metabolite profiling of Syn7002, Syn6803 and Syn7942 during the mid-exponential phase of growth showed pyruvate concentrations in Syn7002 were two and three times higher than in Syn6803 and Syn7942 respectively (Dempo *et al.*, 2014). Furthermore, the intracellular pyruvate concentration of Syn7002 was 6.5-fold higher than aspartic acid, the other potential precursor of L-alanine synthesis if expressing a native aspartate 4-decarboxylase. While informative, these experiments must be taken with caution, as they were performed on cultures grown under constant light.

1.10 The potential of optogenetics for mannitol biosynthesis

Another product of interest is D-mannitol, a sugar alcohol produced by various prokaryotes and eukaryotes. Mannitol has many functions including in the regulation of cell turgor, as a reactive oxygen species (ROS) scavenger and for energy storage (Iwamoto and Shiraiwa, 2005). However, mannitol also has various commercial uses, including as a low-calorie sweetener that does not affect blood sugar levels. Furthermore, mannitol has multiple uses in the pharmaceutical industry as a bulking agent or excipient due to it being chemically inert, improving drug dissolution rates and reducing the negative effect of ROS on certain drugs *in vivo* (Martínez-Miranda *et al.*, 2022). The global market value of mannitol was estimated at $420 million in 2022 and is expected to rise to $564 million by 2028 (Expert Market Research, 2024).

Mannitol is typically produced at industrial scale via the chemical hydrogenation of sugars such as fructose or glucose (Martínez-Miranda *et al.*, 2022). These methods use metal catalysts, high temperatures and high pressures, providing an opportunity for more environmentally sustainable production processes to be developed. Several reports have detailed the use of microbial fermentation processes to convert fructose or glucose to mannitol. The microbes used include lactic acid bacteria such as those of the genus *Lactobacillus* and *Leuconostoc,* and genetically modified *E. coli* (Saha, 2006; Heuser et al. 2009; Zhang *et al.*, 2017). Reported mannitol titres typically exceed 100 g L^{-1}, with productivity rates of at least 2 g L^{-1} h^{-1}. While fermentation of sugar feedstocks is a more

sustainable method than chemical hydrogenation, it still raises concerns over the use of sugars extracted from valuable food crops as feedstocks for biochemical synthesis.

A more environmentally sustainable method of production would be to use cyanobacteria to convert CO_2 to mannitol. Mannitol is not produced endogenously by the model cyanobacterial species Syn7002, Syn6803 and Syn7942. However, each of these species has been engineered to produce mannitol via the expression of heterologous enzymes (Madsen *et al.*, 2018; Wu *et al.*, 2020; Pritam *et al.*, 2023). The highest mannitol titre achieved in cyanobacteria to date was 1.1 g L^{-1} after 12 days of cultivation of engineered Syn7002 (Jacobsen and Frigaard, 2014). This required the expression of mannitol-1-phosphate dehydrogenase (*mtlD*) from *E. coli* and mannitol-1-phosphatase (*mlp*) from *Eimeria tenella* in a glycogen synthase mutant Syn7002 strain. An alternative approach to engineer mannitol production in Syn7002 involved the expression of a fusion protein containing mannitol-1-phosphate dehydrogenase (M1PDH) and mannitol-1-phosphatase (M1Pase) from the microalga *Micromonas pusilla* using the strong constitutive promoter pCpcB594 (Madsen *et al.*, 2018). The use of a fusion enzyme should improve substrate channelling between the two enzymes and reduce the buildup of intermediates, thus improving reaction efficiency. The codon-optimised version of this fusion gene was named *mpusfus* and converts fructose-6-phosphate, via mannitol-1-phosphate, to mannitol (Figure 1-18). A final mannitol titre of 100 mg L^{-1} after 52 days of cultivation was reported in this work.

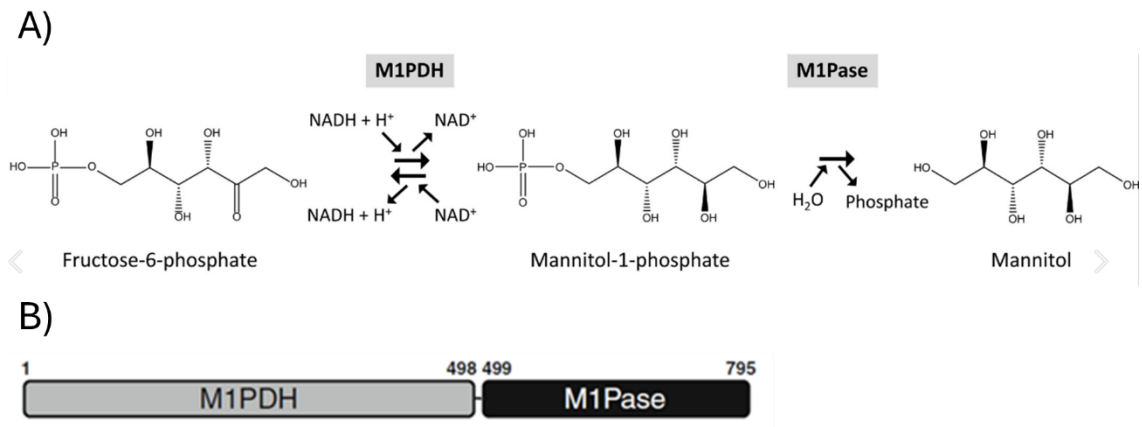


Figure 1-18. Mannitol synthesis via M1PDH/M1Pase fusion enzyme. A) Mannitol synthesis pathway via mannitol-1-phosphate dehydrogenase/mannitol-1-phosphatase (M1PDH/M1Pase) from *Micromonas pusilla*. B) M1PDH/M1Pase gene structure. Figure adapted from Tonon *et al.*, 2017.

1.11 Aims and objectives

This project had four key aims. To address the absence of optogenetic tools which have been characterised in Syn7002, the first aim was to express and characterise the activity of the CcaS/CcaR optogenetic system in Syn7002 using the GFP reporter gene. Characterisation was performed under a range of relevant light conditions using both fluorescence spectrometry and qRT-PCR. The second aim was to express a second optogenetic tool, the YF1/FixJ system in Syn7002. Various troubleshooting steps were also performed due to the initial poor performance of the system. The third aim was to apply optogenetic tools to control the synthesis of industrially relevant products in Syn7002. To this end, L-alanine and mannitol were chosen as the products to be synthesised. The final aim was to use RNAseq to identify wavelength-specific transcripts, namely under blue and orange LED light. This would potentially aid in the future identification and characterisation of wavelength-specific signalling pathways in Syn7002, which could lead to the discovery of new optogenetic tools.

2 Materials and methods

2.1 Cyanobacteria culture conditions

2.1.1 Growth cabinet cultures

Synechococcus sp. PCC 7002 (Syn7002) was originally obtained from Prof Linda Lawton (Robert Gordon University) and had been continuously cultivated and maintained as DMSO stocks by the Amtmann group for the last 10 years. Syn7002 was grown photo-autotrophically in modified artificial seawater (Medium A, Stevens Jr *et al.*, 1973) supplemented with cycloheximide to a final concentration of 25 µg/mL to inhibit fungal contamination. Cryopreserved DMSO stocks were used to inoculate 20 mL pre-cultures which were grown to an optical density at 730 nm (OD_{730}) of 1-3 inside glass tubes with a rubber stopper. 0.5 mL of pre-culture diluted to OD_{730} = 1 was then added to 150 mL Medium A in 250 mL Bijou Bottles. Cultures were grown at 30°C under 150 µmol m^{-2} s^{-1} warm white LED light (Valoya Oy C65 NS12) and a 16/8 hour light/dark cycle inside a Fitotron SGC-2 growth chamber (Weiss Technik). The emission spectrum of the Valoya warm white LED lighting is shown below (Figure 2-1). Cultures were sparged with humidified air via an aquarium pump and a flask containing distilled water. A Lambda 45 UV/VIS Spectrophotometer (PerkinElmer) was used for monitoring culture growth by measuring OD_{730}.

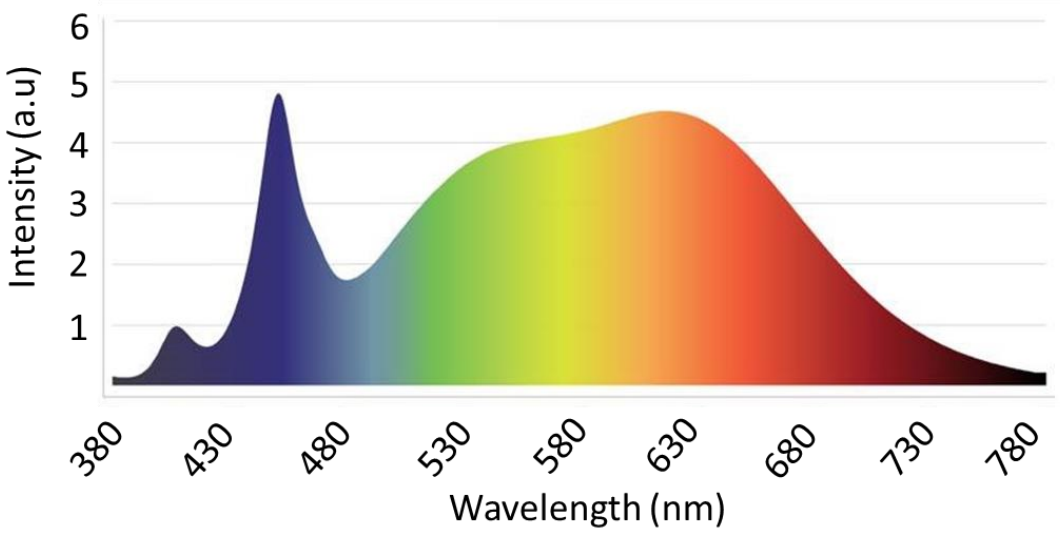


Figure 2-1. Emission spectra of Valoya Oy C65 NS12 white LED lights. Spectral data was obtained and visualised by Stiina Kotiranta (Valoya Ltd).

For some experiments, light-transmitting filters were used to adapt and modify the wavelengths of light reaching the cultures (Figure 2-2). The light treatments for all 150 mL culture experiments are described as follows. Red light (and removal of blue light for YF1/FixJ system control) was provided by illuminating cultures with warm white LED light and covering the cultures with the red light transmitting filter Fire 019 (LEE Filters). Green light was provided by illuminating cultures with warm white LED light and covering the cultures with the green light transmitting filter Dark Green 124 (LEE Filters). The intensity of the white LED light source was adjusted to result in a final measured intensity of 30 μmol m^{-2} s^{-1} under both the red and green filters.

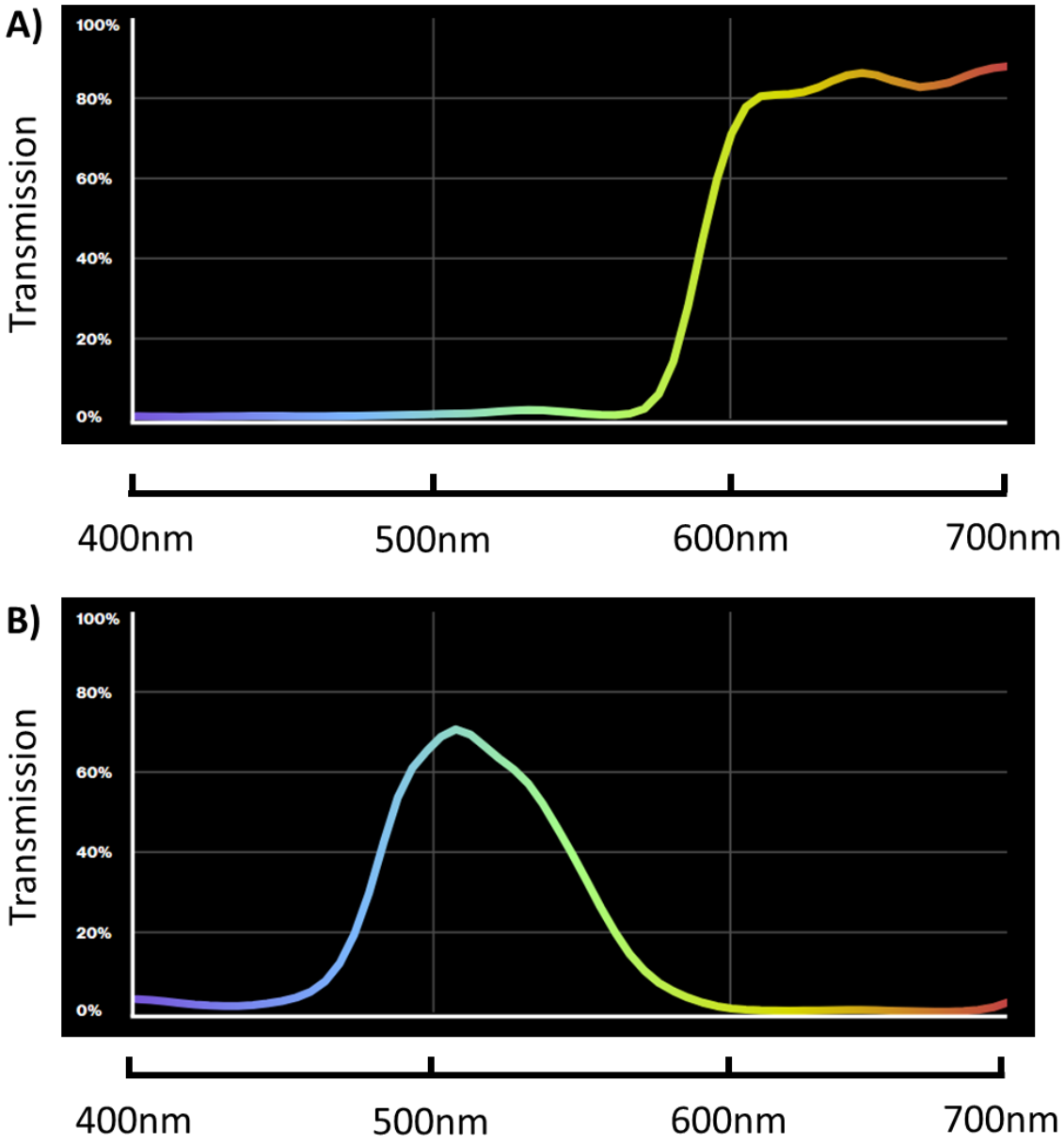


Figure 2-2. Light transmission spectra for filters. A) LEE Filters Fire 019 B) LEE Filters Dark Green 124. Adapted from LEE Filters, 2023.

2.1.2 Photobioreactor cultures

For some experiments, custom-made 1 L MicroPharos PBR photobioreactors (Xanthella, UK) were used. Temperature, light intensity and photoperiod in the reactor were controlled by the "Zeus II" control units, which allow the use of user-defined programmes. Light was provided by the custom-made 'rainbow tiles' comprised of narrowband LED panels (Figure 2-3). These panels contain 5 LED colours: blue (λmax = 465 nm), cyan (λmax = 500 nm), green (λmax = 520 nm), orange (λmax = 590 nm) and red (λmax = 640 nm) (Figure 2-4). Aluminium foil was taped across the uncovered parts of the bioreactor vessel to prevent

penetration of ambient light. As the Rainbow light tiles were attached directly to the left and right sides of the vessel and the LEDs were spaced across the tiles, the conventional method of reporting light intensity at the culture surface could not be applied. Instead, the light intensity as measured at 5 cm from each tile inside the reactor is described for each light treatment. To grow Syn7002 cultures inside the photobioreactors, 20 mL pre-cultures were grown to an OD_{730} = 1-3 inside the Fitotron growth cabinet. 3 mL of pre-culture diluted to OD_{730} = 1 was then added to 1 L of Medium A inside the bioreactor. 2 L of air per minute was injected using a vacuum pump which generated sufficient bubbling for culture mixing. Cultures were grown at 30°C under various wavelength protocols.

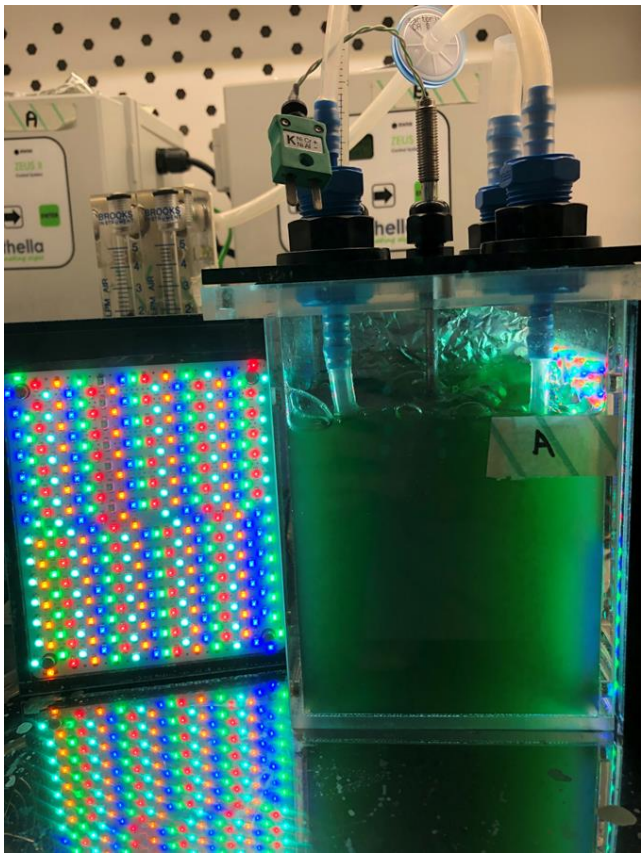


Figure 2-3. Photograph of the Xanthella photobioreactor and 'rainbow tile' LED panels.

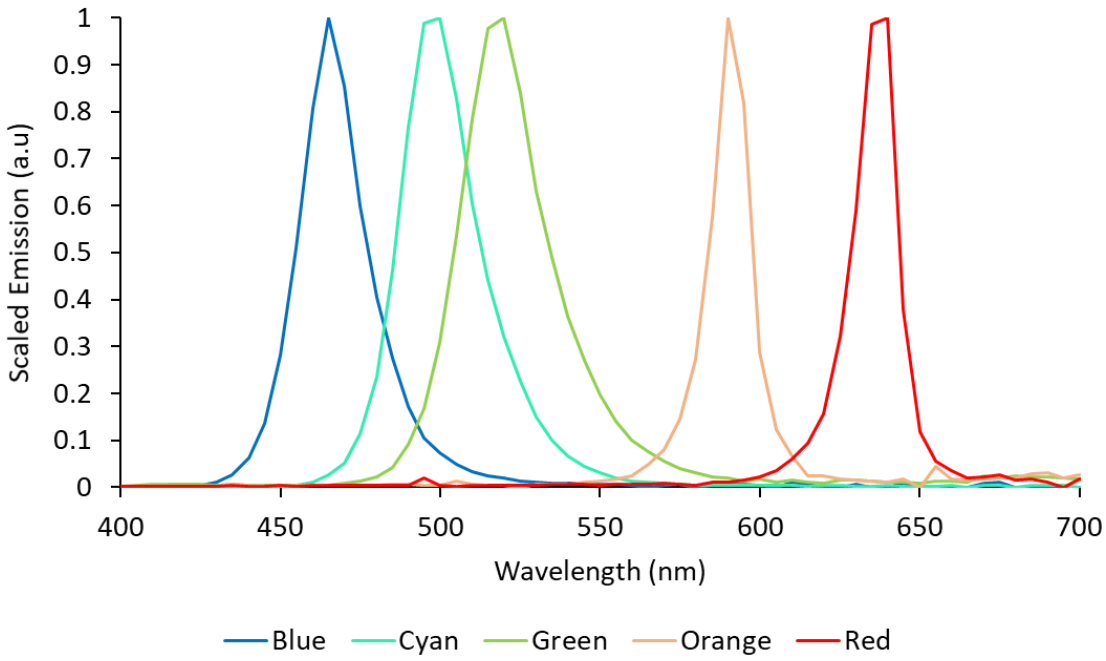


Figure 2-4. Wavelength spectrum provided by Xanthella photobioreactor LED panels. Spectral data was obtained and analysed by Dr Aleksander Mihnev (PhD thesis 2020, University of Glasgow).

2.1.3 Cyanobacteria cryopreservation

For long-term preservation of cyanobacterial strains, 2 mL of culture was harvested at OD_{730} = 1-3 and pelleted at 13000 x rpm for 2 minutes. The supernatant was removed and the cell pellet was resuspended in 2 mL of half-strength Medium A containing 8% (v/v) DMSO. The resuspended culture was then added to screw-cap tubes which were placed inside a Mr Frosty container (Merck, C1562) filled with 100% (v/v) isopropanol and stored at -80°C for 1 day. The tubes were then transferred to cardboard boxes for long-term storage at -80°C.

2.2 Molecular cloning

2.2.1 Primers

Primers were designed *in silico* using either Primer3 software (Untergasser *et al.*, 2012) or Primer-BLAST (Ye *et al.*, 2012) and were chemically synthesised by Integrated DNA Technologies (Table 2-1).

Table 2-1. Primers used in this project.

Primer Name	Sequence	Use
YF1v2 F1 F	tatggcttcattcagctccggttcccaac	Generation of pAQpAQ1BB-Blue-v2
YF1v2 F1 R	Gtaaaatcgttccttacttatataaatgaa ttaatctcctacttgactttatgagttg	Generation of pAQpAQ1BB-Blue-v2
YF1v2 F2 F	Tttatataagtaaggaacgattttacgtggc gagttttcagtcttttg	Generation of pAQpAQ1BB-Blue-v2
YF1v2 F2 R	cggagctgaatgaagccataccaaacgac	Generation of pAQpAQ1BB-Blue-v2
CcaS F	Tttctgcagttacataagcttgggcatcacg catggtatgg	Generation of pAQ1BB-CcaSR-GFP
CcaS R	Aaactgcagaagtaacgtacgctagtatttc tcctctttagataaagttag	Generation of pAQ1BB-CcaSR-GFP
CcaR F	ttcaataagctttactcaggagagcgttcacc	Generation of pAQ1BB-CcaSR-GFP
CcaR R	Aagtaacgtacgctagtatttctcctctttaga taaagttag	Generation of pAQ1BB-CcaSR-GFP
172-F1 F	aaagaggagaaatactagcg	Generation of pUC-172-GFP
172-F1 R	gtatttggtatctgcgctc	Generation of pUC-172-GFP

172-F2 F	Agagcgcagataccaaatactgttcttctagtgtagcc	Generation of pUC-172-GFP
172-F2 R	Cgctagtatttctcctctttttaaaaatgcgatcctaacaaag	Generation of pUC-172-GFP
172-10 BB F	acttggttgagtactcac	Generation of pUC-172-10-GFP
172-10 BB R	ggcaagccattagtcaatg	Generation of pUC-172-10-GFP
172-10 F1 F	Ccattgactaatggcttgccgcaccctatctaattcag	Generation of pUC-172-10-GFP
172-10 F1 R	Tctcctctttttaaaaatgcgattataacaaagtaaaattg	Generation of pUC-172-10-GFP
172-10 F2 F	gcatttttaaaaagaggagaaatactagcg	Generation of pUC-172-10-GFP
172-10 F2 R	Tggtgagtactcaaccaagtcattctgagaatagtgtatgc	Generation of pUC-172-10-GFP
172-35 BB F	tctcagaatgacttggttgagtactcac	Generation of pUC-172-35-GFP
172-35 BB R	gatagggtgcggcaagccattagtcaatg	Generation of pUC-172-35-GFP
172-35 F1 F	atggcttgccgcaccctatctaattcag	Generation of pUC-172-35-GFP
172-35 F1 R	Gatgtctgggggagaaaatcgtaaagaaatc	Generation of pUC-172-35-GFP
172-35 F2 F	Attttctccccccagacatcaattttactttgttaggatcg	Generation of pUC-172-35-GFP
172-35 F2 R	tcaaccaagtcattctgagaatagtgtatgc	Generation of pUC-172-35-GFP
35-10 BB F	tctcagaatgacttggttgagtactcac	Generation of pUC-172-35-10-GFP
35-10 BB R	gatagggtgcggcaagccattagtcaatg	Generation of pUC-172-35-10-GFP
35-10 F1 F	atggcttgccgcaccctatctaattcag	Generation of pUC-172-35-10-GFP
35-10 F1 R	gatgtctgggggagaaaatcgtaaagaaatc	Generation of pUC-172-35-10-GFP
35-10 F2 F	Attttctccccccagacatcaattttactttgttataatcgc	Generation of pUC-172-35-10-GFP
35-10 F2 R	tcaaccaagtcattctgagaatagtgtatgc	Generation of pUC-172-35-10-GFP
CcaS SDM F1	agacgtacggccattgggatatatcaacggtg	Generation of pUC-172-GFP, pUC-172-10-GFP, pUC-172-35-GFP and pUC-172-35-10-GFP
CcaS SDM R1	agacctagggtacgctagtatttctcctcttt	Generation of pUC-172-GFP, pUC-172-10-GFP, pUC-172-35-GFP and pUC-172-35-10-GFP
GFPq F	agcgttcaactagcagacca	qRT-PCR - GFP
GFPq R	cgaaagggcagattgtgtgg	qRT-PCR - GFP

A1356 F	tccacacctttatcgccacc	qRT-PCR - A1356
A1356 R	caacacatccttgcggcttt	qRT-PCR - A1356

2.2.2 Ribosome binding site calculator

All RBS strength calculations were performed using Salis Lab RBS calculator v2.2 (Reis and Salis, 2020).

2.2.3 Vectors

The pAQ1BB cyanobacterial transformation vector used in this work was kindly provided by Dr Mary Ann Madsen (University of Glasgow, UK). pAQ1BB contains two flank sites that allow homologous recombination and integration into the native PAQ1 plasmid of Syn7002 (Madsen MA, PhD thesis, University of Glasgow). The pJT119b vector containing the DNA parts required for the function of the CcaS optogenetic system was kindly provided by Prof Jeffrey Tabor (Addgene plasmid #50551; http://n2t.net/addgene:50551). The pDusk vector containing the DNA parts required for the function of the YF1 optogenetic system was kindly provided by Andreas Moeglich (Addgene plasmid # 43795; http://n2t.net/addgene:43795). A list of all vectors used is provided (Table 2-2). Plasmids maps are provided in Appendix I.

Table 2-2. Plasmids used in this project.

Plasmid Name	Function	Source
pAQ1BB	Syn7002 transformation vector containing the GFP gene	Dr Mary Ann Madsen (University of Glasgow)
pDusk	Enables expression of the YF1/FixJ system in *E. coli*	Andreas Moeglich (Addgene plasmid # 43795; http://n2t.net/addgene:43795)
pJT119b	Enables expression of the CcaS/CcaR system in *E. coli*	Prof Jeffrey Tabor (Addgene plasmid #50551; http://n2t.net/addgene:50551

pAQpAQ1BB-Blue	Enables expression of GFP under control of the YF1/FixJ system in Syn7002	Generated in this project
pAQpAQ1BB-Blue-v2	Enables expression of GFP under control of the YF1/FixJ system with stronger RBS in Syn7002	Generated in this project
pAQ1BB-Green	Enables expression of GFP under the control of the CcaS/CcaR system in Syn7002	Generated in this project
pUC-172-GFP	GFP under control of CcaS system with truncated 172bp pCpcG2 promoter for expression in Syn7002	Generated in this project
pUC-172-10-GFP	GFP under control of CcaS system with truncated 172bp pCpcG2 promoter and edited -10 hexamer for expression in Syn7002	Generated in this project
pUC-172-35-GFP	GFP under control of CcaS system with truncated 172bp pCpcG2 promoter and edited -35 hexamer for expression in Syn7002	Generated in this project
pUC-172-35-10-GFP	GFP under control of CcaS system with truncated 172bp pCpcG2 promoter and edited-35 and -10 hexamers for expression in Syn7002	Generated in this project
pUC-YF1-GFP	Enables expression of GFP under control of cyanobacterial codon	Generated in this project

	optimised YF1/FixJ system in *E. coli*	
pUC-172-10-alaDE	Enables optogenetic control of L-alanine production from the optimised version of the CcaS system (pCpcG2-172-10) in Syn7002	Generated in this project
pUC-cpc-alaDE	Enables L-alanine production via constitutive expression of *alaD* and *alaE* in Syn7002	Generated in this project
pUC-172-10-mpusfus	Enables optogenetic control of *mpusfus* expression from the optimised version of the CcaS system (pCpcG2-172-10) in Syn7002	Generated in this project
pAQ1BB-pCpcB594-mpusfus	Enables constitutive expression of *mpusfus* in Syn7002	Generated in this project

2.2.4 PCR amplification and *de novo* synthesis of DNA fragments

For amplification of DNA fragments for cloning, 50 µL PCR reactions were performed using New England Biolabs (NEB) Phusion polymerase according to the manufacturer's instructions. PCR reactions were performed using the peqSTAR 96X universal gradient thermocycler (VWR). The PCR products were visualised using 1.5% (w/v) agarose gel electrophoresis stained with ethidium bromide. The appropriate DNA bands were extracted and purified using the QIAGEN QIAquick Gel Extraction Kit (QIAGEN). In some cases when DNA fragments could not be amplified using PCR, they were instead de-novo synthesised by Integrated DNA Technologies.

2.2.5 Restriction enzyme-based DNA assembly

DNA fragments were cloned into plasmids using NEB restriction enzymes. 1-2 µg of plasmid or PCR-amplified DNA was incubated with 10 units of each of the required enzymes for 1 hour at 37°C as per the manufacturer's instructions. Digested fragments were separated using agarose gel electrophoresis and the required fragments were purified using the QIAquick Gel Extraction Kit (QIAGEN). Digested fragments were then inserted into the digested plasmid using T4 DNA Ligase (NEB) using a 3:1 insert:vector ratio as per the manufacturer's instructions. T4 DNA ligase was inactivated by incubation of the completed reaction mixtures at 65°C for 10 minutes before *E. coli* transformation.

2.2.6 HiFi DNA assembly

Multiple linear DNA fragments were assembled into plasmids simultaneously using the NEBuilder HiFi DNA Assembly Cloning Kit (NEB) as per the manufacturer's instructions. In brief, DNA fragments were PCR-amplified using primers incorporating 20-30 bp homologous overlaps. An equimolar amount of each fragment was mixed and incubated at 50°C for 15-60 minutes with NEBuilder Mastermix. The reaction mix was then transformed into *E. coli* DH5α cells as described below. All HiFi primers were designed using the NEBuilder Assembly Tool.

2.2.7 Site-directed mutagenesis

Site-directed mutagenesis (SDM) was performed using NEBuilder HiFi DNA Assembly Cloning Kit as described above. Primers were designed which allowed either insertion, deletion or substitution of desired bases by incorporating mutations within the primers. For deletions, primers were designed which bound outside of the region targeted for deletion, thus preventing its amplification.

2.2.8 *E. coli* transformation and plasmid purification

For *E. coli* transformation, chemically competent Top10 cells (Invitrogen) were used in the following heat shock transformation reaction. A 20 µL ligation or HiFi reaction was added to a 50 µL aliquot of Top10 or DH5α cells and incubated on ice for 30 minutes. The cells were then incubated at 42°C for 30 seconds in a water bath before being placed on ice for 2 minutes. The cells were then added

to 350 mL Luria-Bertani (LB) media (Top10) or 950 µL SOC media (DH5α) and incubated at 37°C for 1 hour with shaking at 180 x rpm. The reaction was then plated on an LB agar plate containing the appropriate antibiotic for selection. After drying, the plate was stored in a 37°C incubator overnight.

After overnight incubation, single colonies from the transformation plate were picked and added to 6 mL LB media containing the appropriate antibiotic. These cultures were incubated overnight at 37°C with shaking at 180 x rpm. After overnight incubation, the plasmid was purified from a 5 mL culture using the QIAprep Spin Mini Kit (QIAGEN). Correct plasmid formation was verified by restriction fragment analysis. The plasmid sequence was then verified by Sanger or whole plasmid Nanopore sequencing performed by Source BioScience Genomics.

2.2.9 Syn7002 natural transformation

For the natural transformation of Syn7002 (uptake of exogenous DNA from cell culture media and integration via homologous recombination), 1.5 mL of culture grown to OD_{730} = ~1 was incubated with 1-20 µg of DNA for 3 days under 20 µmol m^{-2} s^{-1} white light at 30°C with very slow bubbling. After 3 days, the cultures were pelleted at 13000 x rpm for 2 minutes, resuspended in 100 µL of supernatant, and spread on plates containing Medium A and 50 µg/mL spectinomycin for selection. Plates were then incubated under 30 µmol m^{-2} s^{-1} white light at 30°C for 7-10 days. Successful transformants were verified using colony PCR. Colonies were continuously re-streaked on Medium A plates of successively higher antibiotic concentration until full segregation was achieved where possible. Colonies were then used to inoculate a 20 mL liquid culture.

2.3 RNA sequencing and qRT-PCR

2.3.1 RNA extraction

Total RNA was extracted from Syn7002 using the RNeasy Mini Kit (QIAGEN) with on-column DNase digestion. In brief, 20 mL culture diluted to OD_{730} = 1 was harvested and pelleted by centrifugation at 4500 x rpm for 10 minutes. The cell pellet was resuspended in 450 µL of QIAGEN RLT lysis buffer with added ß-Mercaptoethanol (1:100). The cells were then lysed inside screw cap tubes

containing 0.3 g of acid-washed glass beads (425-600 µm diameter) using the QIAGEN TissueLyser at 30 Hz for 5 minutes. The RNA was then purified using the RNeasy Mini Kit as per the manufacturer's instructions and stored at -80°C. RNA concentration and purity were determined using the NanoDrop One spectrophotometer (Thermo Fisher Scientific).

2.3.2 qRT-PCR and qPCR

For qRT-PCR analysis of gene expression, total RNA extracted from Syn7002 cultures was transcribed into cDNA using the QuantiTect Reverse Transcription Kit (QIAGEN) and random primers as per the manufacturer's instructions. Primers were designed to amplify 80-150 bp regions of the cDNA target genes in the qRT-PCR and were also used for the generation of a standard curve. To generate the standards, PCR reactions were performed using GoTaq G2 DNA polymerase (Promega) as per the manufacturer's instructions. Reactions were visualised by gel electrophoresis and the correct DNA band was extracted. Purified DNA concentration was measured using the NanoDrop One spectrophotometer and diluted to 1 pg/µL. The diluted DNA was then used to generate the other standards using 10-fold serial dilutions in water. qRT-PCR reactions were performed using the StepOnePlus Real-Time PCR System (Life Technologies) and visualised with Brilliant III Ultra-Fast SYBR Green QPCR Master Mix (Agilent). After each qRT-PCR run, a melting curve analysis was performed to check for primer dimers. The qRT-PCR thermal cycling protocol is shown (Table 2-3).

Table 2-3. qRT-PCR thermocycling conditions.

Stage	Temperature	Time (s)
1. Initial denaturation	95°C	180
2. Denaturation 3. Annealing Repeat for 40 cycles	95°C 60°C	10 20
4. Melt curve Increasing from 60°C to 95°C in 0.3°C increments	95°C 60°C 95°C	60 10 10

The absolute quantity of target cDNA molecules in sample wells was calculated by comparing Ct values with those of the serially diluted standards of known concentration. Technical replication of each reaction was performed in triplicate and the average Ct value of the three replicates was used for quantitation. Reference gene selection was performed using the RefFinder web tool, to identify which gene showed the least variation across the experimental conditions tested (Xie *et al.*, 2023). qPCR was performed as described above but using genomic DNA (gDNA) extracted from Syn7002 cells rather than cDNA. All primers used for qRT-PCR are shown (Table 2-4).

Table 2-4 qRT-PCR primers used in this work.

Gene / Region	Primer Name	Sequence (5' to 3')
GFP	GFPq F	AGCGTTCAACTAGCAGACCA
	GFPq R	CGAAAGGGCAGATTGTGTGG
SYNPCC7002_A0956	A0956 F	CCGTGTACCGTCAGTTCCAA
	A0956 R	ATCCAGAGGTTGGCTTGCTC
SYNPCC7002_A1356	A1356 F	TCCACACCTTTATCGCCACC
	A1356 R	CAACACATCCTTGCGGCTTT
SYNPCC7002_A1784	A1784 F	GTGTCAACGGGCAACAAGTC
	A1784 R	TCGGGACATCAGCAATCACC
SYNPCC7002_A2364	A2364 F	CTTCCCAATCAACTCGCCCT
	A2364 R	CTCATGGGGTTGGAGTTGCT
pAQ1 (non-gene region)	pAQ1G F	CGCTAAATCTGCCCCGGTAT
	pAQ1G R	CAAACAGGGACGGCAAAACC
YF1	YF1q F	CGCCACTTGCGTTAGATTCG
	YF1q F	GACACTGGAATGGGAGTGGG
FixJ	FixJq F	ATAACTCGATGCCGTCCAGC
	FixJq R	GGCAGGATTCGGAGTTACCC
16S rRNA	16S F	CGGGTTTGATGAGATTCGCTTGC
	16S R	AGTTGGGCACTCTAGGGAGACTG

2.3.3 Genomic DNA extraction

Genomic DNA (gDNA) was extracted from Syn7002 cells as follows. 20 mL culture at OD_{730} = ~1 was harvested and pelleted by centrifugation at 4500 x rpm for 10 minutes. The supernatant was removed and the cell pellet was frozen in liquid nitrogen. gDNA was then extracted using the DNeasy Blood and Tissue Kit (QIAGEN) as per the manufacturer's instructions for extraction from gram-negative bacteria.

2.3.4 RNA sequencing

For sequencing of RNA extracted from Syn7002, total RNA was extracted as described. Subsequent library prep, sequencing and bioinformatic analysis were performed by Glasgow Polyomics. Sequencing libraries were prepared from total RNA using the TruSeq Stranded mRNA Sample Preparation Kit (Illumina). Libraries were sequenced in 75 bases, paired-end mode on the NextSeq 500 platform (Illumina). Raw sequence reads were trimmed for contaminating sequence adapters and poor-quality bases using the program FastP. Bases with an average Phred score lower than 15 were trimmed. Reads that were trimmed to less than 54 bases were discarded. The quality of the reads was checked using the Fastqc program before and after trimming. The sequenced reads were checked for other species contamination using the Fastq-Screen program, which selects a sample of the reads and aligns them to a selection of reference genomes that are routinely sequenced at Glasgow Polyomics. The quality trimmed reads were aligned to the reference genome using the program Hisat2 and the read counts using HtSeq Count. Differential expression analysis was performed using the Bioconductor package DESeq2.

2.3.5 Clustering analysis

Hierarchical clustering of gene expression profiles from the RNAseq dataset was performed using various functions within R studio. The expression data was scaled to remove the effect of highly expressed genes. The LRT test was used to identify genes for which the expression changed versus the zero-hour timepoint at any time point across the time course using an adjusted p-value of < 0.005 as a cutoff. Dynamic tree cutting was applied to the produced dendrogram which takes account of the shape of the branches when cutting the tree and adapts the cut to the height of each branch, with the output being the number of clusters. The sequence 500 bp upstream of genes within each identified cluster was extracted using a custom-made R script written by Dr Graham Hamilton.

2.4 Protein extraction and analysis

2.4.1 Protein extraction from Syn7002

20 mL of Syn7002 cells grown to OD_{730} = 5-10 were diluted to OD_{730} = 5 and lysed using 5 mL Millipore BugBuster Mastermix per gram of wet cell paste with one-quarter of a protease inhibitor tablet added (#11836170001, Roche). Cells resuspended in BugBuster Mastermix were mixed for 20 minutes on a shaking platform at 150x rpm at room temperature. Cell debris was removed by centrifugation at 21,000 x g for 15 seconds at 4°C and the supernatant was transferred to a fresh tube. The supernatant was then pelleted by centrifugation at 21,000 x g for 10 minutes at 4°C. The resulting supernatant contained the soluble protein fraction, and the pellet contained the crude membrane fraction. The membrane fraction was resuspended in an equivalent volume of BugBuster Mastermix.

2.4.2 SDS-PAGE

Cell lysate samples were added to an equivalent amount of 1x Laemmli buffer and incubated at 70°C for 5 minutes. The samples were then pelleted at 14,000 x rpm for 2 minutes and the supernatant was then run on an SDS-PAGE gel. The separating gel contained 10% (w/v) acrylamide, 0.267% (w/v) bisacrylamide, 375 mM Tris-Cl (pH 8.8), 0.1% (w/v) sodium dodecyl sulphate (SDS) and 0.2% (w/v) ammonium persulfate. The stacking gel contained 3.9% (w/v) acrylamide, 0.1% (w/v) bisacrylamide, 130 mM Tris-Cl (pH 6.8), 0.1% (w/v) SDS and 0.1% (w/v) ammonium persulfate. 3 µL of PageRuler Protein Ladder (Thermo Scientific) was used in each gel. 15 µL of protein sample was loaded per well and the gel was run at 150 V for 1 hour. Gels were visualised using Coomassie staining unless being used for Western Blotting.

2.4.3 Western Blotting

Western Blotting was performed using the semi-dry method. Protein bands were transferred from the SDS-PAGE gel to a nitrocellulose membrane wet with BJJERUMS buffer (48 mM Tris, 39 mM glycine, pH 9.2, 20% (v/v) methanol) at 11 V for 1 hour. Membranes were then blocked by incubating in TBS-T (137 mM NaCl, 2.7 mM KCl, 19 mM Tris-Cl (pH 7.4), 0.1% (v/v) Tween 20) + 5% (w/v) milk

buffer for 1 hour at room temperature with shaking. The blocking solution was then discarded, and the membrane was incubated in 25 mL of 1:2500 dilution anti-His (#H1029, Sigma-Aldrich) or 1:3000 dilution anti-HA (#ROAHAHA, Sigma-Aldrich) primary antibody solution overnight at 4°C. Primary antibody solutions were diluted in TBS-T + 2.5% (w/v) milk buffer. The membrane was then washed in TBS-T and incubated with shaking for 10 minutes a total of three times. The membrane was then incubated in either 25 mL anti-mouse IgG (ab97046, Abcam) or anti-rat IgG (Ab6734, Abcam) secondary antibody solution diluted in TBST + 2.5% (w/v) milk buffer for 1 hour at room temperature. The membrane was washed in TBST buffer and incubated with shaking for 10 minutes a total of three times. Bands were then visualised using SuperSignal West Dura Extended Duration Substrate (Thermo Fisher Scientific) as per the manufacturer's instructions.

2.5 Product quantitation

2.5.1 Syn7002 and *E. coli* GFP fluorescence assays

GFPmut3b reporter gene expression was quantified by measuring the sample fluorescence of 1 mL of culture with the Perkin Elmer LS 55 Luminescence Spectrophotometer. The excitation and emission wavelengths used were 480 nm and 514 nm respectively, with a 10 nm path length. Measured GFP fluorescence was normalised to OD_{730} for Syn7002 and OD_{600} for *E. coli* to normalise for cell number, where OD_{730}/OD_{600} are used as a proxy for cell number. For measurement of sfGFP expression in *E. coli*, 1 mL samples were taken and pelleted at 13,000 x rpm for 2 minutes before being washed once with 1 X phosphate-buffered saline (PBS) and resuspended in 1 mL PBS to remove LB media which produces high background signal. GFP fluorescence was then measured in 1 mL of the resuspended culture.

In experiments where flavin mononucleotide (FMN) was added to Syn7002 cultures, 1 mL culture was pelleted at 13,000 x rpm for 2 minutes before being washed once with 1 mL Medium A and then resuspended in 1 mL fresh Medium A. This prevented the interference of FMN with GFP fluorescence measurements.

2.5.2 L-alanine and mannitol quantitation

L-alanine concentrations in samples were measured using the Abcam L-Alanine Assay Kit (ab83394, Abcam). In brief, alanine in the test sample is converted via an alanine-converting enzyme to pyruvate which is specifically detected, leading to proportional colour development that can be measured via absorbance at 570 nm. The absorbance measured in background control samples lacking the alanine converting enzyme is subtracted and alanine concentration is then derived from a standard curve.

Mannitol concentration was measured using the D-mannitol Assay Kit (K-MANOL, Megazyme) as per the manufacturer's instructions. In brief, mannitol in the test sample is converted via a mannitol dehydrogenase to D-fructose, producing NADH in the process. The amount of NADH produced in this reaction is stoichiometric with the amount of D-mannitol. The NADH which is measured by the increase in absorbance at 340 nm, which allows the calculation of mannitol concentration.

3 Characterisation of the CcaS/CcaR system

3.1 Introduction

The CcaS/CcaR optogenetic system enables greed/red light control of gene expression. Upon green light illumination, the CcaS photoreceptor phosphorylates the transcription factor CcaR. Phosphorylated CcaR binds to and activates transcription from the pCpcG2 output promoter. Upon red light illumination, CcaS becomes dephosphorylated, CcaR is dephosphorylated and transcription from pCpcG2 is reduced (discussed in Introduction).

As the CcaS system is well-characterised in comparison to other natural two-component systems it has been exploited extensively for optogenetics. The system has been used to control gene expression in several organisms including *E. coli*, *Bacillus subtilis* and *Pseudomonas putida* (Fernandez-Rodriguez *et al.*, 2017; Castillo-Hair *et al.*, 2019; Hueso-Gil *et al.*, 2020). The system has also been used to control GFP expression in the cyanobacteria Syn6803 and *Synechococcus* NKBG 15041c (Syn15041c) (Abe *et al.*, 2014; Badary *et al.*, 2015). The CcaS system has also been exploited in a green light-activated cell lysis system, for cell surface display of proteins and production of isobutanol and 3-methyl-1-butanol in Syn6803 (Miyake *et al.*, 2014; Ferri *et al.* 2015; Kobayashi *et al.*, 2022).

While the system is useful for controlling gene expression in its original form, improving its maximal activity under green light whilst reducing activity under red light is desirable for most applications. To this end, the system has been modified to improve its performance in heterotrophic bacteria. In *E. coli*, the truncation of pCpcG2 from 238 bp to 172 bp (pCpcG2-*172*) and altering the expression levels of CcaS and CcaR improved the dynamic range from 6-fold to 117-fold (Schmidl *et al.*, 2014). Transcription from pCpcG2 under red light was reduced by its truncation, presumably by removing a predicted weak constitutive promoter identified in the 66 bp at the 3' end of pCpcG2. The performance of the system was then improved further by deleting both CcaS PAS domains, increasing the dynamic range to 600-fold (Ong and Tabor, 2018). However, this large increase in dynamic range from 6-fold to 600-fold can mostly be attributed to reducing CcaS system activity in red light to negligible levels,

rather than an increase in activity under green light. This ability to tightly repress gene expression under red light enables improved control over metabolic pathways.

While reducing CcaS system activity under red light is desirable, if high levels of gene expression are required, an increased transcriptional response in green light would be beneficial. In *B. subtilis*, as well as reducing this 'leakiness' in red light by CcaS PAS domain deletion and pCpcG2 truncation, maximal system activity in green light has also been increased (Castillo-Hair *et al.*, 2019). This result was achieved by focusing on the modification of the pCpcG2-172 promoter. Altering the sequence of the -10 hexamer to match the *B. subtilis* promoter consensus sequence increased maximal activity under green light 3-fold. Then, by replacing the region downstream of the consensus -10 hexamer with that of a strong *B. subtilis* constitutive promoter, maximal activity was increased by another 43-fold. This resulted in a final optimised promoter that had a maximal output over 100-fold higher than pCpcG2-172.

The CcaS system has also been modified to improve its performance or modify its colour sensitivity in cyanobacteria. A miniaturised CcaS termed CcaS#11 has been applied in Syn6803, which is activated by red light and repressed by green light (Kobayashi *et al.*, 2019). In addition to inverting the response to green and red light, this system demonstrated a 2-fold greater dynamic range than the original. The CcaS system has also been adapted to control transcription via the T7 RNA polymerase from the T7 promoter in Syn6803 (Shono *et al.*, 2020). This resulted in a 10-fold increase in activity under both red and green light versus the original system, whilst maintaining a similar dynamic range. However, the CcaS system has yet to be characterised in Syn7002.

3.2 Chapter aims

The aim of this chapter was to express and characterise the performance of the Syn6803 original native CcaS system in Syn7002. The original system was chosen as opposed to any optimised versions as we assumed this had the highest chance of initial success and that any optimised versions could be tested later in the project. First, the function of the system in *E. coli* was confirmed as per previous research. Next, the system was expressed in Syn7002 and used to

control the expression of the GFP reporter gene. The system was confirmed as functional, and its activity was demonstrated under a variety of lighting conditions using GFP fluorescence assays. Then, qRT-PCR was used to better understand the transcriptional kinetics of the system in response to red and green light. Finally, an improved variant of the CcaS system was produced, which showed a 3-fold greater transcriptional response to green light than the original system whilst maintaining a similar level of activity under red light, thus resulting in a superior dynamic range of 13.3-fold versus 8.5-fold in the original system.

3.3 Results

3.3.1 Confirming CcaS system function in *E. coli*

The CcaS system has been previously used to control gene expression in *E. coli* and displayed a 6-fold dynamic range (Schmidl *et al.*, 2014). However, the activity of the same genetic components may differ between the varying sub-strains contained within different laboratories. To confirm that the obtained CcaS construct was functional, the pJT119b plasmid was expressed in *E. coli*. pJT119b contains the CcaS and CcaR genes expressed from their native Syn6803 promoters as well as the *GFP* gene immediately downstream of pCpcG2. *E. coli* was transformed with pJT119b and grown on selective media. The presence of the correct DNA in selected clones was confirmed by restriction digest analysis and sequencing. Cultures expressing pJT119b were grown to OD_{600} = ~0.2 under cool white LED light and covered with a red light-transmitting filter. The cultures then either remained under the red filter or were covered with a green light-transmitting filter (for a detailed description of the light conditions see Material and Methods). GFP fluorescence was measured at several time points over a 24-hour period and normalised to OD_{600} as a proxy of cell number.

Under red light illumination, GFP fluorescence remained constant over 24 hours (Figure 3-1). However, GFP fluorescence was not equal to zero indicating some level of GFP transcription from pCpcG2. Upon green light illumination, GFP fluorescence doubled within 1 hour of exposure. After 24 hours, the cultures exposed to green light displayed 5-fold higher GFP fluorescence than the

cultures exposed to red light, confirming that the CcaS system was functional in *E. coli*.

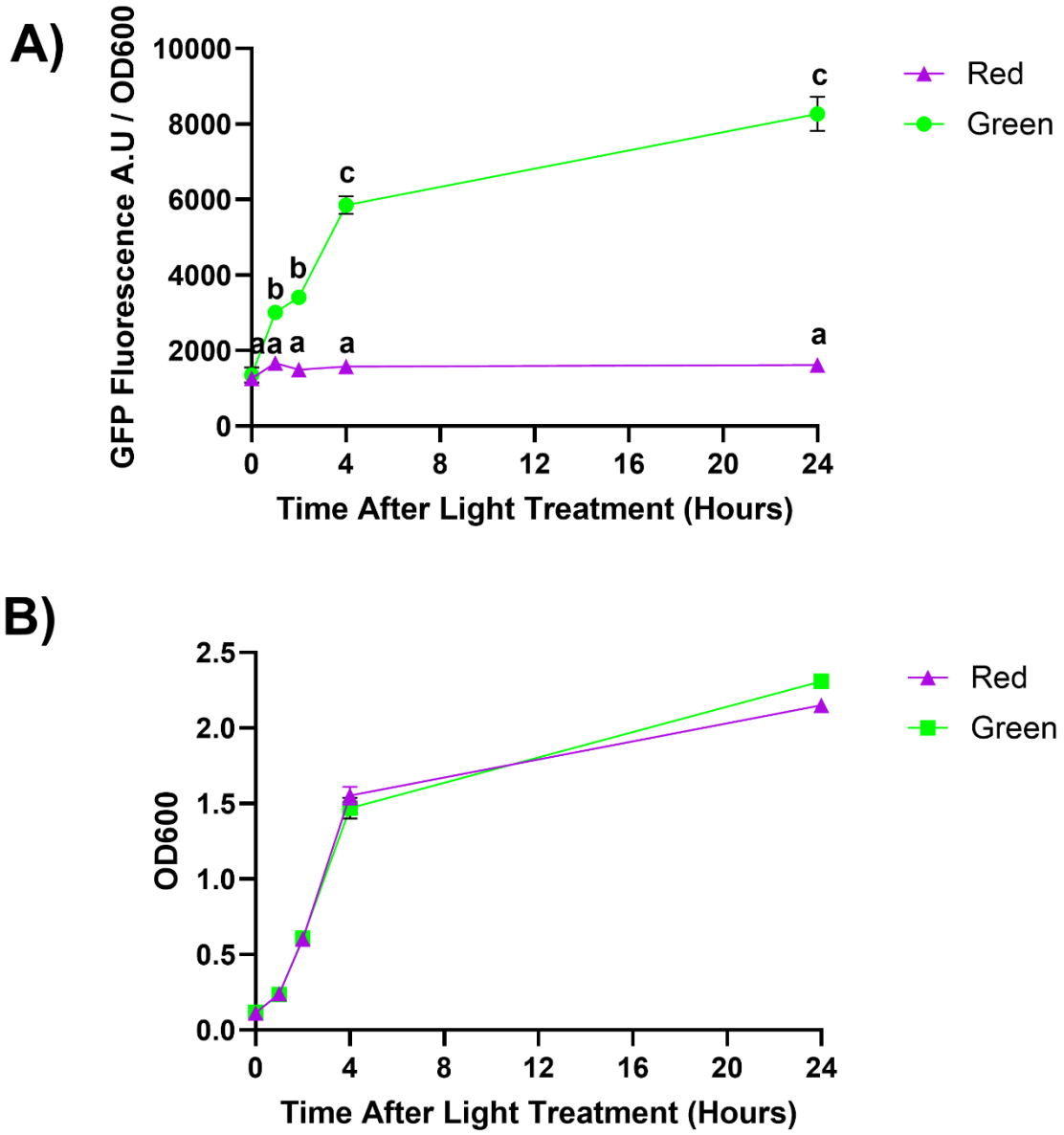


Figure 3-1. CcaS system activity in *E. coli*. A) GFP fluorescence and **B)** OD_{600} of *E. coli* cultures expressing pJT119b. 150 mL cultures were grown to OD_{600} = ~0.2 under red light. At t = 0 hours, the cultures were then maintained under red light illumination (purple triangles) or were illuminated with green light (green circles). Samples were harvested on ice at 1, 2, 4 and 24 hours, and GFP fluorescence was measured and normalised to OD_{600}. Data are means ± S.E.M of three independent cultures. Significance at $p < 0.05$ as determined by two-way ANOVA with Tukey's multiple comparison test.

3.3.2 Growth of Syn7002 expressing the CcaS system

Next, the performance of the CcaS system was tested in Syn7002. To this purpose, the pAQ1BB-CcaSR-GFP suicide plasmid was cloned and transformed into Syn7002. In this plasmid (Figure 3-2) the CcaS-CcaR-pCpcG2-GFP fragment was placed between two flank sites, allowing homologous recombination and integration into the high-copy number Syn7002 endogenous plasmid pAQ1. After sequence verification and confirmation of homozygousity, a transformed clone was used as inoculate for 1 L cultures in the photobioreactor (Xanthella). The cultures were illuminated with an equivalent intensity of red or green light provided by narrowband LEDs. In separate experiments, 150 mL cultures of both wild-type Syn7002 and Syn7002 expressing pAQ1BB-CcaSR-GFP were grown in 250 mL Bijou bottles under either red or green light as provided by white light LEDs with red or green filters (see Materials and Methods for a description of the light treatments). Both the 1 L bioreactors with LEDs and the 150 mL cultures with light transmitting filters as some research groups may not have access to photobioreactors with narrowband LEDs.

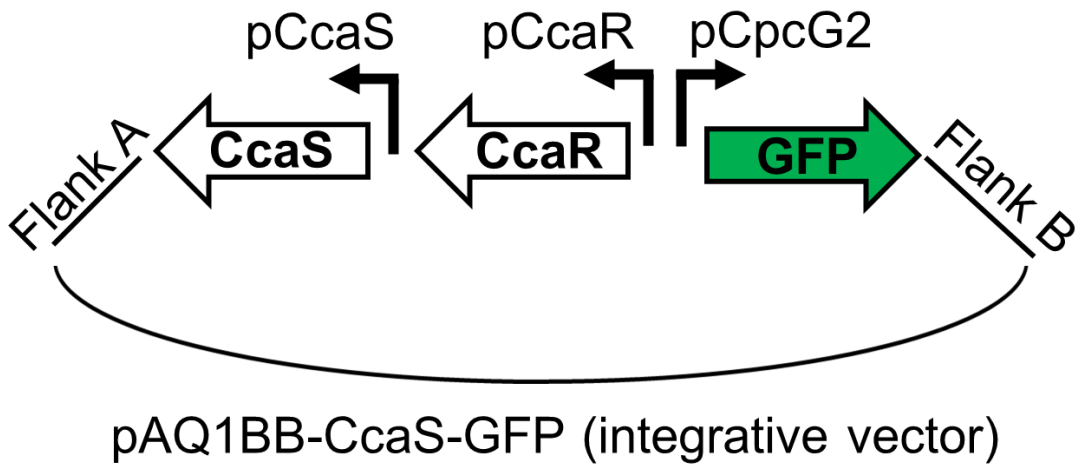


Figure 3-2. pAQ1BB-CcaS-GFP plasmid design. The target region between Flank A and Flank B is integrated into the Syn7002 endogenous plasmid pAQ1 via homologous recombination.

OD_{730} measurements in the 150 mL cultures showed that growth curves of Syn7002 expressing pAQ1BB-CcaSR-GFP were similar to those of wild-type Syn7002 under both red and green light (Figure 3-3a). Thus, the expression of the CcaS system did not impact the growth of Syn7002 under the conditions and at the optical densities tested. Culture growth was generally faster under red

light than under green light as would be expected given red/green light usage for photosynthesis (Figure 3-3a and Figure 3-3b).

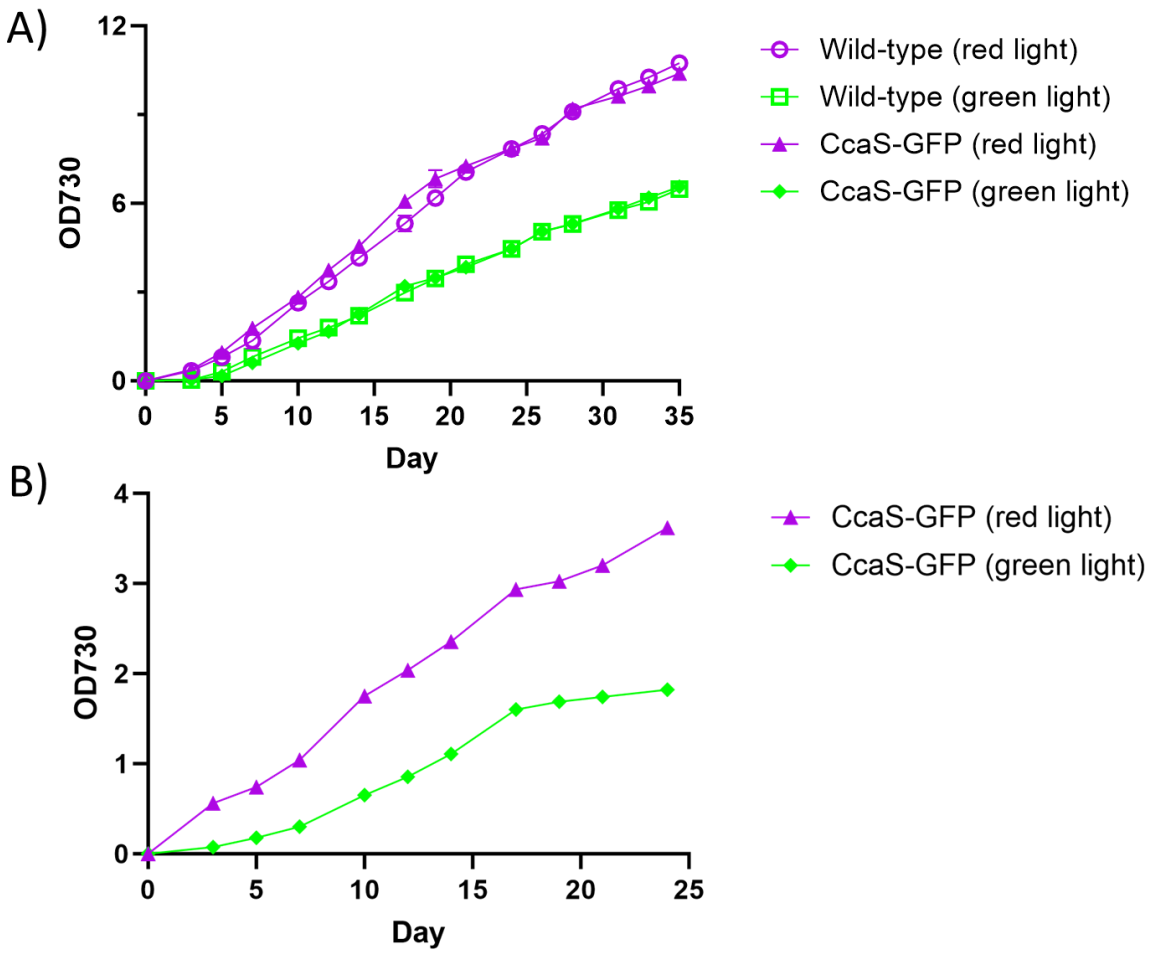


Figure 3-3. Growth of *Synechococcus* sp. PCC 7002 expressing the CcaS system. Culture density (OD_{730}) of wild-type Syn7002 (open symbols) or Syn7002 expressing pAQ1BB-CcaSR-GFP (CcaS-GFP, filled symbols). **A)** 150 mL cultures were grown in Bijou bottles under illumination with 30 µmol m^{-2} s^{-1} red light (purple symbols) or green light (green symbols) provided by white LED light with the red or green light-transmitting filters. **B)** Syn7002 cultures expressing CcaS-GFP were grown in 1 L photobioreactors and illuminated with an equivalent intensity of red (purple symbols) or green LED light (green symbols). All cultures were grown under 16h/8h light/dark cycles. Data are means ± S.E.M of three independent cultures.

3.3.3 Characterisation of CcaS system activity using GFP fluorescence assays

To test and characterise the function of the CcaS system in Syn7002, light-control of GFP expression was characterised by measuring GFP fluorescence under different light treatments. Syn7002 cultures expressing pAQ1BB-CcaSR-GFP were grown in 1 L photobioreactors under constant red light for 4 days to OD_{730} = ~1. The cultures were then given different light treatments consisting of

either red light only, green light only, an equivalent intensity of combined red and green light (Red + Green), green light of 2.5-fold increased intensity, or darkness (for a detailed description of the light treatments see Materials and Methods). Samples were harvested at the indicated timepoints before GFP fluorescence was measured and normalized to OD_{730} as a proxy of cell number.

Under red light illumination, GFP fluorescence remained constant over 48 hours but was not zero (Figure 3-4). GFP fluorescence also remained constant in the dark-incubated cultures, with similar values as measured in the red light treatment. Under both green light only and red plus green light illumination, GFP fluorescence rose steadily over the 48 hours. At 24 hours and 48 hours, GFP fluorescence was roughly 2-fold and 3.5-fold higher, respectively, than under red light only. Increasing the green light intensity further (2.5x green) did not increase GFP fluorescence compared to the values measured under the standard green light treatment. Collectively, this data shows that green light either alone or added to red light, increases the activity of the CcaS system. GFP fluorescence remained low, but above zero, under red light and in darkness, indicating a constant 'leak' in transcription even when CcaS is in the dark state.

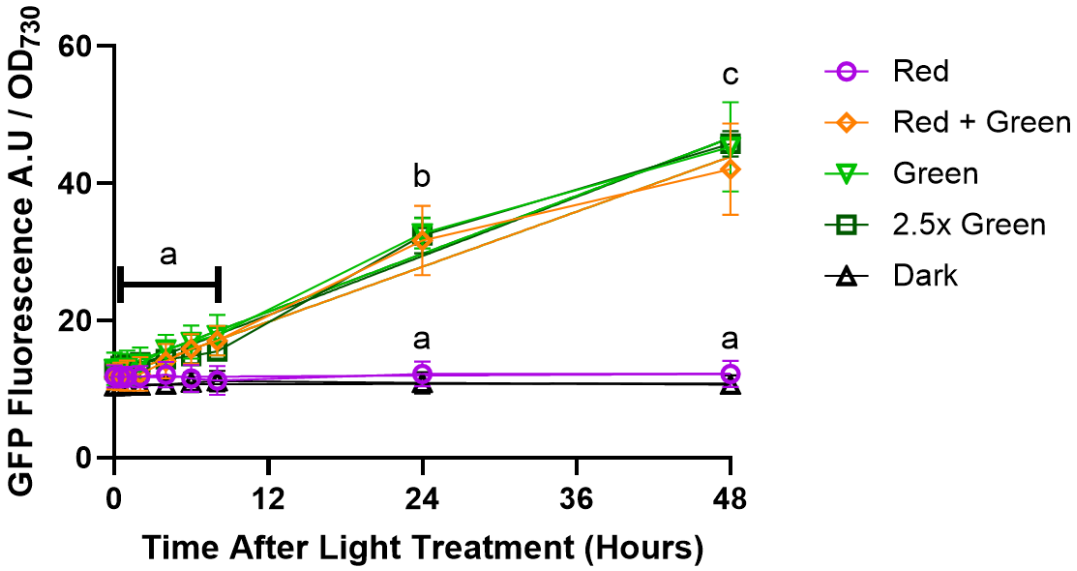


Figure 3-4. CcaS system activity under constant light conditions. 1 L cultures of Syn7002 expressing pAQ1BB-CcaSR-GFP were grown for 4 days in photobioreactors under constant red light. Cultures were then illuminated with the following light treatments: Red = 20 µmol m^{-2} s^{-1} red LED (purple symbols); Red + Green = 20 µmol m^{-2} s^{-1} red LED + 20 µmol m^{-2} s^{-1} green LED (orange symbols); Green = 20 µmol m^{-2} s^{-1} green LED (pale green symbols); 2.5x green = 50 µmol m^{-2} s^{-1} green LED (dark green symbols); Dark = darkness (black symbols). GFP fluorescence was measured, normalised to OD_{730} and background fluorescence measured in cells expressing empty pAQ1BB vector was subtracted. Data are means ± S.E.M of three independent cultures. Groups sharing the same letter are not significantly different from each other ($p > 0.05$) based on Tukey's multiple comparison test following two-way ANOVA.

Cyanobacteria can be grown under constant light conditions which allows fast initial growth. However, PSII suffers irreversible damage under lengthy periods of constant illumination and thus fast growth is unable to be sustained over longer time periods. Alternatively, culturing cyanobacteria under diurnal light/dark cycles allows photosystem repair during darkness, which results in sustained growth over long time periods and more closely mimics natural environmental conditions (Nixon *et al.*, 2010). Previous characterisation of CcaS activity in cyanobacteria has only been performed under constant light conditions and it is not known whether or how light/dark cycles impact CcaS performance. To test this, 150 mL Syn7002 cultures expressing pAQ1BB-CcaSR-GFP were grown under red light for 5 days under a 16/8-hour light/dark cycle. The cultures were then given different light treatments of either red, green or warm white LED light while continuing the diurnal light/dark cycles. GFP fluorescence was measured at several time points over 48 hours.

In the cultures illuminated with red light, GFP expression remained constant over the 48 hours (Figure 3-6). After 4 hours, GFP expression was around 1.5-fold higher in green light than in red light. After 48 hours in both the green and the white light treatments, GFP fluorescence was 4.5-fold higher than in red light. Differences in GFP fluorescence between green and white light treatments within the same timepoint were not statistically significant. This experiment confirmed that the expression of the target gene is still increased by green light when cultures are grown in diurnal light/dark cycles. Additionally, it was found that warm white LED light results in similar levels of CcaS activation as green light alone. However, GFP fluorescence was not measured at the beginning of the dark periods to determine the impact of the dark cycle. This was addressed in a later experiment (Figure 3-13).

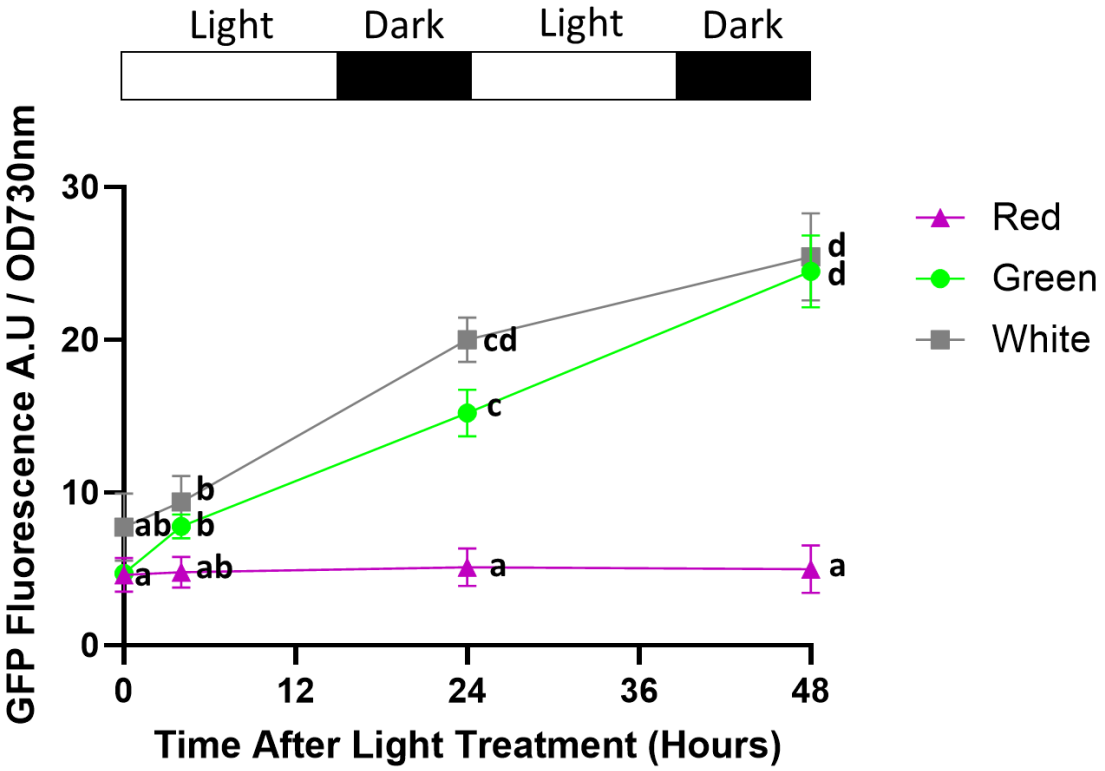


Figure 3-5. CcaS system activity under light/dark cycles. 150 mL Syn7002 cultures expressing pAQ1BB-CcaSR-GFP were grown under illumination with red light for 5 days to OD_{730} = ~1 under a 16/8-hour light/dark cycle. Cultures then either remained under red light (purple symbols) or were illuminated with green light (green symbols) or white light (grey symbols) and samples were harvested at the indicated timepoints. GFP fluorescence was measured, normalised to OD_{730} and background fluorescence from cells expressing empty pAQ1BB vector was subtracted. Data are means ± S.E.M of three independently grown cultures. Groups sharing the same letter are not significantly different from each other ($p > 0.05$) based on Tukey's multiple comparison test following two-way ANOVA.

After demonstrating that green light causes CcaS system activation, experiments were performed to measure the rate at which the system deactivates under red light after a period of green light illumination. Cultures expressing pAQ1BB-CcaSR-GFP were grown in 1 L photobioreactors for 4 days under constant red light until reaching an OD_{730} =~1. After the 4 days, the light was switched to green for 24 hours, before switching back to red for another 24 hours. GFP fluorescence and OD_{730} were measured at several time points.

GFP fluorescence increased 3-fold after 24 hours of green light, in accordance with previous experiments (Figure 3-6). After the switch to red light, GFP fluorescence decreased over 24 hours, albeit at a slower rate than the increase measured under green light. After 24 hours of red light, GFP fluorescence decreased by 20% from the peak observed at the end of the 24-hour green light treatment. No statistically significant differences were detected between any of the measured timepoints. This data suggested that CcaS deactivation under red light in Syn7002 is slow. However, this apparent slow deactivation could be due to high GFP protein stability. Therefore, even if the CcaS system was deactivated at the transcriptional level, GFP fluorescence would remain high until GFP protein was naturally degraded. Therefore, another method of measuring CcaS activity was required to better understand the kinetics of system deactivation.

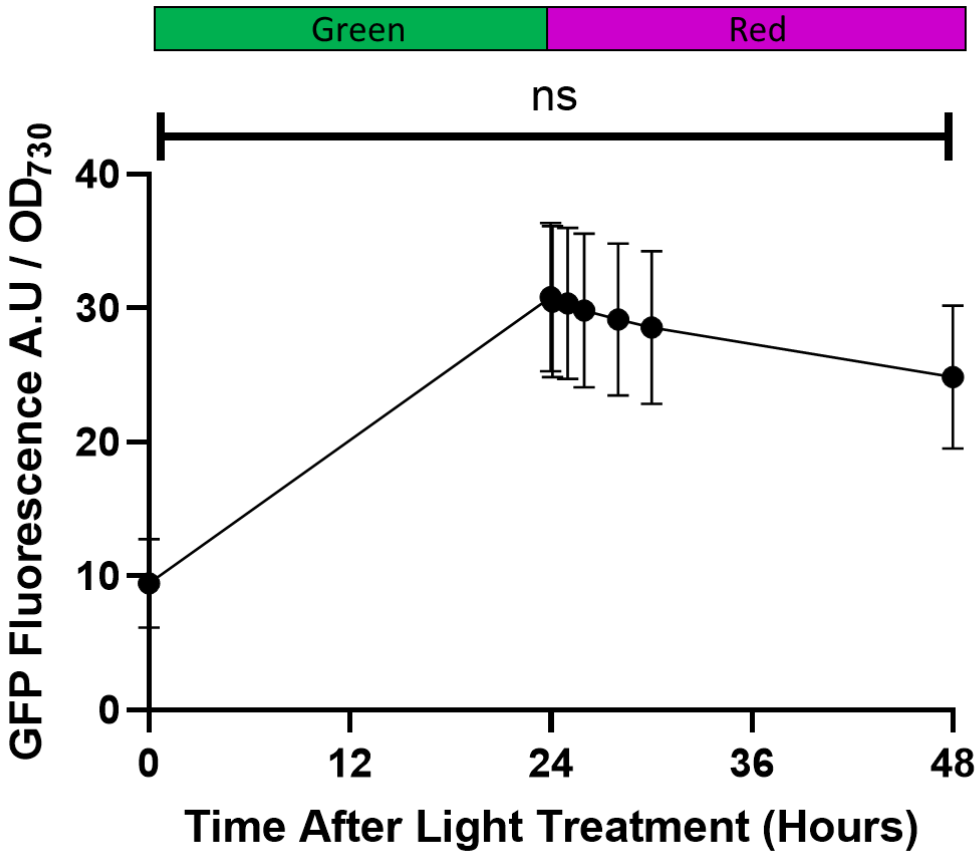


Figure 3-6. CcaS system OFF rate as measured by GFP fluorescence. Syn7002 cultures expressing pAQ1BB-CcaSR-GFP were grown in 1 L photobioreactors to OD_{730} = 1 for 4 days under constant red LED light. Cultures were then illuminated with green light only for 24 hours before then being illuminated with red light only for another 24 hours. GFP fluorescence was measured at several time points, normalised to OD_{730} and background fluorescence from cells expressing empty pAQ1BB vector was subtracted. Data are means ± S.E.M of three independently grown cultures. Significance at $p < 0.05$ was calculated using one-way ANOVA with Tukey's multiple comparison test.

3.3.4 Characterisation of CcaS system activity using qRT-PCR

qRT-PCR is routinely used to quantitatively measure gene expression. Measuring GFP RNA transcript levels via qRT-PCR would likely be more suited to measuring the kinetics of CcaS system deactivation than GFP fluorescence. However, before qRT-PCR could be used, a suitable reference gene had to be validated as demonstrating stable expression under the chosen experimental conditions. There are no previous examples of confirmed constitutive genes in Syn7002 cultures grown under different light wavelengths. One study analysed the most suitable reference genes for several conditions including UV-B treatment but did not use any other light wavelengths (Szekeres *et al.*, 2014). Indeed, there are very few qRT-PCR studies in Syn7002 which use different lighting conditions as

an experimental parameter. One study used 16S or 23S rRNA as reference genes for studying high light stress in Syn7002, without these genes appearing to have been previously validated under their chosen experimental conditions (Xiong *et al.*, 2015).

In Chapter 4 (Identification of Wavelength-specific Transcripts), RNAseq data obtained from wild-type Syn7002 cultures grown under blue, orange, and white light is presented. From this dataset, a panel of candidate qRT-PCR reference genes was identified which showed low variation in expression level across all lighting conditions and time points. From this panel, four genes were chosen for analysis via qRT-PCR: *SYNPCC7002_A0956*, *SYNPCC7002_A1356*, *SYNPCC7002_A1784* and *SYNPCC7002_A2364*. To validate the expression levels of these genes, samples were harvested from two separate experiments. First, wild-type Syn7002 cultures grown in 1 L photobioreactors were grown under white light for 4 days before illumination was switched to red or green LED light only. RNA was extracted from samples harvested at 0, 2 and 8 hours after red or green illumination. The second experiment involved growing 150 mL Syn7002 cultures expressing pAQ1BB-CcaSR-GFP under red light for 4 days. Cultures then either continued under red light illumination or were illuminated with green light as provided by white LED light combined with green light-transmitting filters. RNA was extracted from samples harvested at 0, 4, 24 and 48 hours post-illumination and subjected to qRT-PCR with primers for the candidate reference genes. The resulting Ct values were used as the input to the RefFinder web tool (Xie *et al.*, 2023).

RefFinder identifies the most suitable candidate reference gene from qRT-PCR data via four different methods. These methods and software packages are known as comparative delta CT, BestKeeper, NormFinder and geNorm. Each of these methods uses different statistical techniques to calculate the stability of gene expression across different samples and treatments. Various different statistical methods are used across the packages including calculating standard deviation, covariance, power, and pairwise variation. While each method has limitations, the combination of four methods should confer a robust process for reference gene selection. It has been stated that the use of RefFinder may produce slightly less accurate results than when using each software package

separately as the online tool only uses raw Ct values as input, whereas the individual packages also incorporate reaction efficiency (De Spiegelaere *et al.*, 2015). However, while using RefFinder may not be as accurate as using the four methods separately, it is a quick and easy method for validating a reference gene. It is impractical for most researchers to have access to paid software packages for all four methods, which has led to RefFinder being the most popular choice (cited in over 100 publications to date).

RefFinder provides a comprehensive gene stability ranking after analysing the variation of each gene through the individual software packages and combining the scores. The comprehensive ranking is the mean of the rankings of each gene (i.e. 1, 2, 3 or 4) provided by each software package. Of the tested genes, RefFinder identified *SYNPCC7002_A1356* as being the most stable gene, with an average ranking of 1.189 (Figure 3-7). The least stable gene was *SYNPCC7002_A2364*, which was designated a gene stability ranking of 4. Therefore, *SYNPCC7002_A1356* was selected as the reference gene for future experiments. In studies where a pair of reference genes are preferred, the pair of *SYNPCC7002_A1356* and *SYNPCC7002_A0956* would be most suitable, as *SYNPCC7002_A0956* demonstrated the second lowest stability ranking of 1.861.

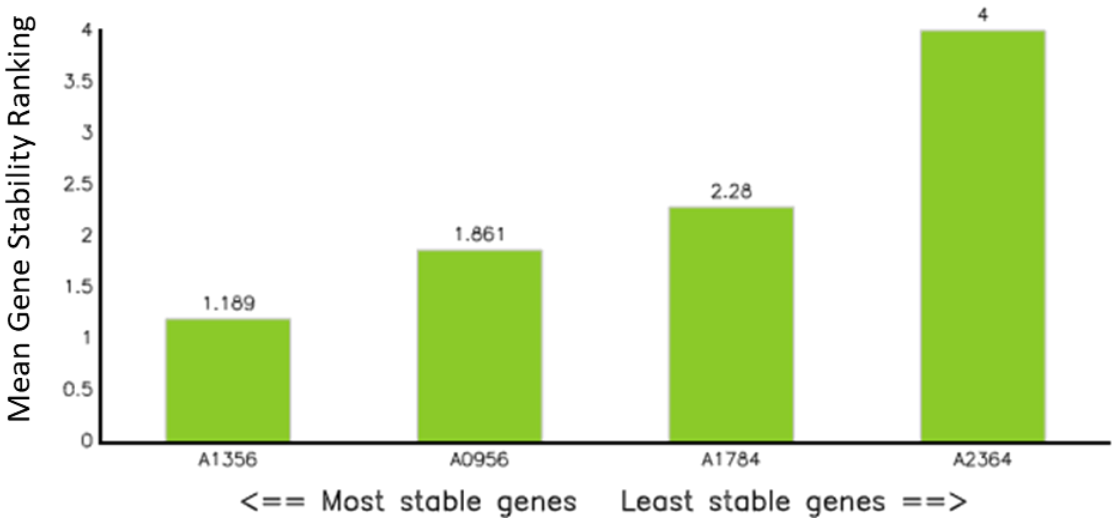


Figure 3-7. Average comprehensive gene stability ranking of candidate qRT-PCR reference genes. RefFinder Web tool output for ranking candidate qRT-PCR reference gene stability. The expression of four candidate reference genes *SYNPCC7002_A0956*, *SYNPCC7002_A1356*, *SYNPCC7002_A1784* and *SYNPCC7002_A2364* was measured in three independently grown Syn7002 cultures via qRT-PCR. Cultures were grown under either red or green LED light and the raw Cq values were inputted into the RefFinder tool. RefFinder integrates the results from Normfinder, BestKeeper, geNorm and the comparative delta CT method (Xie *et al.*, 2023).

After a suitable reference gene had been identified, qRT-PCR was used to measure GFP RNA transcript levels in Syn7002 cultures expressing pAQ1BB-CcaSR-GFP. The samples for RNA extraction were taken from the same cultures and at the same time points as used for the GFP fluorescence measurements shown in Figure 3-5. The RNA was reverse transcribed into cDNA and qRT-PCR was performed with GFP gene-specific primers.

After 4 hours, GFP transcript levels increased in cultures illuminated with red and green light, with a larger increase observed in those illuminated with green (Figure 3-8). The 24 and 48-hour time points were sampled at the end of an 8-hour dark cycle just before the lights were switched back on. At the 24-hour timepoint, GFP transcript had decreased slightly from the 4-hour timepoint but remained higher than the RNA levels measured at the 0-hour timepoint in both red and green light. Despite considerable variation in RNA levels measured at the 24-hour timepoint under green light for both the 24 and 48-hour time points, GFP RNA transcript levels were significantly higher in green light than in red light. This data suggested that an 8-hour dark cycle was not sufficient to reduce

GFP RNA transcript levels back to baseline, as was also observed in the GFP fluorescence data (Figure 3-5).

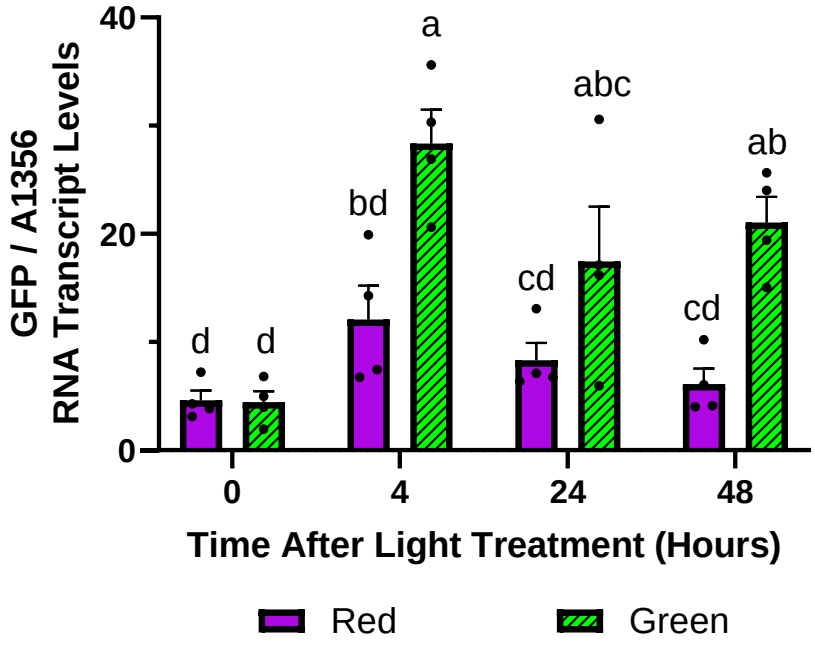


Figure 3-8. CcaS system activity under light/dark cycles as measured by qRT-PCR. 150 mL Syn7002 cultures expressing pAQ1BB-CcaSR-GFP were grown under red light for 5 days to OD_{730} = 1 under a 16/8-hour light/dark cycle. Cultures then either remained under red light or were illuminated with green light and samples as provided by combining warm white LED light with red or green light transmitting filters and were harvested at the indicated timepoints. At the 24 and 48-hour time points, samples were harvested from the cultures at the end of the 8-hour dark cycle before the lights switched back on. GFP transcript levels were measured by qRT-PCR and normalised to *SYNPCC7002_A1356*. Data are means ± S.E.M of four independently grown cultures. Groups sharing the same letter are not significantly different from each other ($p > 0.05$) based on Tukey's multiple comparison test following two-way ANOVA.

As the CcaS system was expressed from the pAQ1 endogenous plasmid, fluctuations in plasmid copy number could affect the measured values through a gene dosage effect. Furthermore, the fraction of pAQ1 plasmid containing the GFP transgene could potentially fluctuate throughout culture growth as well as differ between replicates. This could explain some of the variation observed in the qRT-PCR data (Figure 3-9). To address this, as well as measuring GFP RNA transcript levels via qRT-PCR, qPCR was used to measure the copy number of the GFP transgene and a neighbouring region on the pAQ1 plasmid. GFP RNA transcript levels were then normalised to various combinations of transgene and plasmid copy number.

First, the ratio of GFP transgene to pAQ1 plasmid was measured (Figure 3-9a). While significant variation was observed between replicates, the average remained consistent across each of the colour conditions and time points. Four independent cultures were used for each colour condition and timepoint as opposed to three in an attempt to reduce variation. This data suggests that the fraction of pAQ1 plasmids containing GFP transgene is unlikely to cause any significant differences between mean GFP RNA transcript levels, although it could contribute to inter-replicate variation. Next, pAQ1 plasmid copy number was calculated from the qPCR data (Figure 3-9b). A similar pattern was observed as in Figure 3-10a, with considerable variation between replicates but similar average values across the conditions. Normalising GFP / *SYNPCC7002_A1356* RNA transcript levels to pAQ1 plasmid copy number (Figure 3-9c) resulted in a similar pattern of mean expression levels as without the normalising to pAQ1 copy number (Figure 3-9a). Finally, GFP / *SYNPCC7002_A1356* RNA transcript levels were normalised to the fraction of pAQ1 plasmids containing GFP transgene (Figure 3-9d). Again, a similar pattern of expression was observed as that of measuring GFP / *SYNPCC7002_A1356* RNA transcript levels only shown in Figure 3-8. Collectively, this data shows that variations in pAQ1 plasmid copy number and the fraction of pAQ1 plasmid containing GFP transgene do not to explain the differences between mean GFP RNA transcript levels measured in red versus green light. These experiments validated the use of qRT-PCR to monitor light regulation of the target gene by the CcaS system.

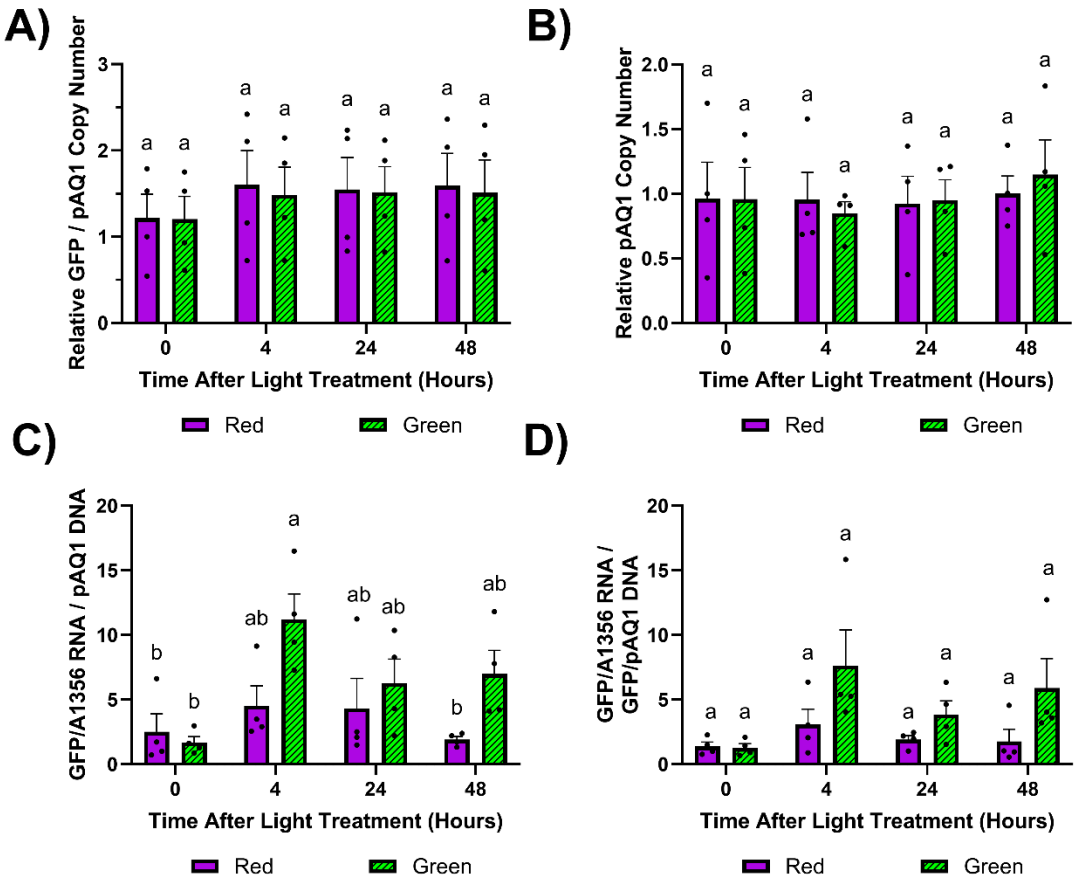


Figure 3-9. CcaS system activity as measured by qRT-PCR and normalised to DNA copy number. 150 mL Syn7002 cultures expressing pAQ1BB-CcaSR-GFP were grown under red light for 5 days to OD_{730} = 1 under a 16/8-hour light/dark cycle. Cultures then either remained under red light or were switched to green light and samples were harvested at the indicated timepoints. GFP RNA transcript levels were measured by qRT-PCR and transgene/plasmid copy number was measured by qPCR. **A)** GFP transgene DNA copy number relative to pAQ1 copy number; **B)** pAQ1 plasmid copy number; **C)** GFP RNA transcript (normalised to *SYNPCC7002_A1356*) normalised to pAQ1 plasmid copy number; **D)** GFP RNA transcript (normalised to *SYNPCC7002_A1356*) normalised to GFP transgene DNA copy number relative to pAQ1 copy number. All data points are shown relative to the first replicate of the zero-hour timepoint in red light. Data are means ± S.E.M of four independently grown cultures. Groups sharing the same letter are not significantly different from each other ($p > 0.05$) based on Tukey's multiple comparison test following two-way ANOVA.

Next, qRT-PCR was used to better resolve the kinetics of the CcaS system. The GFP fluorescence data shown in Figure 3-4 suggested that functional GFP protein levels increased very slowly across the first 8 hours of green light activation. To better understand the kinetics of CcaS activation in response to green light without the confounding effects of GFP translation and protein folding, GFP RNA transcript levels were measured in cultures grown under green light using qRT-

PCR. Cultures expressing pAQ1BB-CcaSR-GFP were grown in photobioreactors under constant red light for 4 days until they reached OD_{730} = ~1. Cultures were then illuminated with green light only and GFP RNA transcript levels were measured at several time points.

GFP RNA transcript levels increased 6-fold after 30 minutes of green light illumination (Figure 3-10). Mean GFP transcript levels appeared to increase slightly further to reach a maximum at 6 hours but due to variation, any differences were not significant. However, no statistically significant differences were detected between any of the timepoints. This is likely due to the significant variation observed between replicates. Regardless, this data shows that the CcaS system is activated rapidly in response to green light.

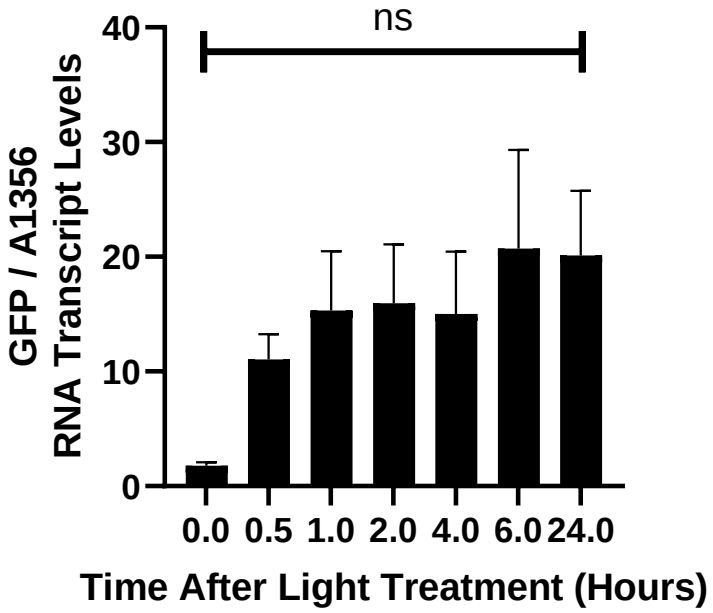


Figure 3-10. The rate of CcaS system activation under green light as measured by qRT-PCR. 1 L cultures of Syn7002 expressing pAQ1BB-CcaSR-GFP were grown for 4 days in photobioreactors under constant red LED light. Cultures were then illuminated with green light only for 24 hours and samples were harvested for RNA extraction at the indicated timepoints. GFP RNA transcript levels were measured using qRT-PCR and normalised to *SYNPCC7002_A1356* RNA transcript levels. Data are means ± S.E.M of three independent cultures. Significance at $p < 0.05$ was calculated using two-way ANOVA with Tukey's multiple comparison test.

After demonstrating the kinetics of CcaS activation at the transcriptional level under green light, the speed at which GFP RNA transcript levels decreased under red light was tested. Cultures expressing pAQ1BB-CcaSR-GFP were grown in photobioreactors under constant red light for four days until they reached OD_{730}

= ~1. The cultures were then illuminated with green light only for 24 hours to activate the CcaS system, before switching back to red light and measuring GFP RNA transcript levels using qRT-PCR.

GFP RNA transcript levels after 24 hours of green light illumination were approximately 5 times higher than at time point 0 (Figure 3-11). Upon switching back to red light only, GFP RNA transcript levels appeared to continue to increase in the first hour (24-25h), albeit the difference was not statistically significant. GFP RNA transcript then decreased across the remaining timepoints, although the difference between 24 and 48 hours was not statistically significant. Minimal further reduction was seen in GFP RNA transcript after 24 hours of red light (48 h time point). At this point, GFP RNA transcript was still around 2-fold higher than at the 0-hour timepoint although the difference was not significant. In summary, by measuring GFP RNA transcript levels instead of protein fluorescence, effective red Light deactivation of CcaS was confirmed. Nevertheless, the shape of the curve may reflect the kinetics of transcription and RNA degradation rather than the light-dependent dephosphorylation of the photoreceptor.

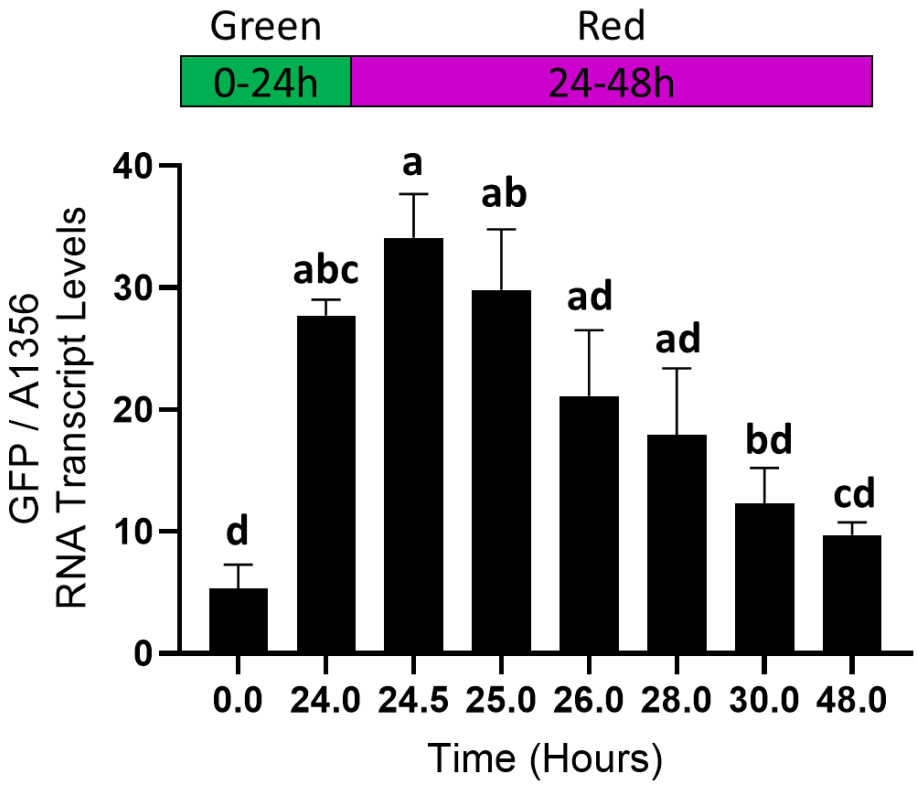


Figure 3-11. The rate of CcaS system deactivation under red light. 1 L cultures of Syn7002 expressing pAQ1BB-CcaSR-GFP were grown for 4 days in photobioreactors under constant red LED light. Illumination was switched to green light for 24 hours, before switching back to red light illumination for 24 hours. RNA was extracted from samples harvested at the indicated time points, GFP RNA transcript was measured by qRT-PCR and normalised to *SYNPCC7002_A1356* RNA transcript levels. Data are means ± S.E.M of three independent cultures. Groups sharing the same letter are not significantly different from each other ($p > 0.05$) based on Tukey's multiple comparison test following one-way ANOVA

To enable the use of dynamic metabolic pathway control, an optogenetic system must be capable of repeated reversible activation and deactivation in response to different light wavelengths. However, this ability has not been demonstrated in any previous applications of the CcaS system in cyanobacteria. To test the response of the CcaS system to alternating periods of green and red light, cultures expressing pAQ1BB-CcaSR-GFP were grown in photobioreactors for 4 days under constant red light. The cultures were then illuminated with alternating 24-hour periods of green light or red light and GFP RNA transcript levels were measured at several time points using qRT-PCR.

CcaS up- and down-regulated transcript levels of the target gene over repeated green/red light cycles. GFP RNA transcript levels rose significantly over 24 hours of green light (Figure 3-12). GFP transcript levels then decreased in the

subsequent 24 hours. The second 24-hour period of green light resulted in GFP transcript levels that were close to double that of the first 24-hour period of green light, a difference which was deemed to be statistically significant. GFP transcript levels decreased after the second 24-hour period of red light back to baseline levels. After the third 24-hour period of green light, GFP transcript levels increased to a similar level as the second 24-hour period of green light (not a statistically significant difference). Finally, the third 24-hour period of red light resulted in a return of GFP transcript levels of baseline. This data confirms that the CcaS system can be reversibly activated and deactivated under alternating 24-hour periods of green and red light. The transcriptional response increases over the first two cycles, which could be due to slightly higher base levels at the end of the red light periods, and/or indicate a 'primed state' of either photoreceptor or transcriptional machinery.

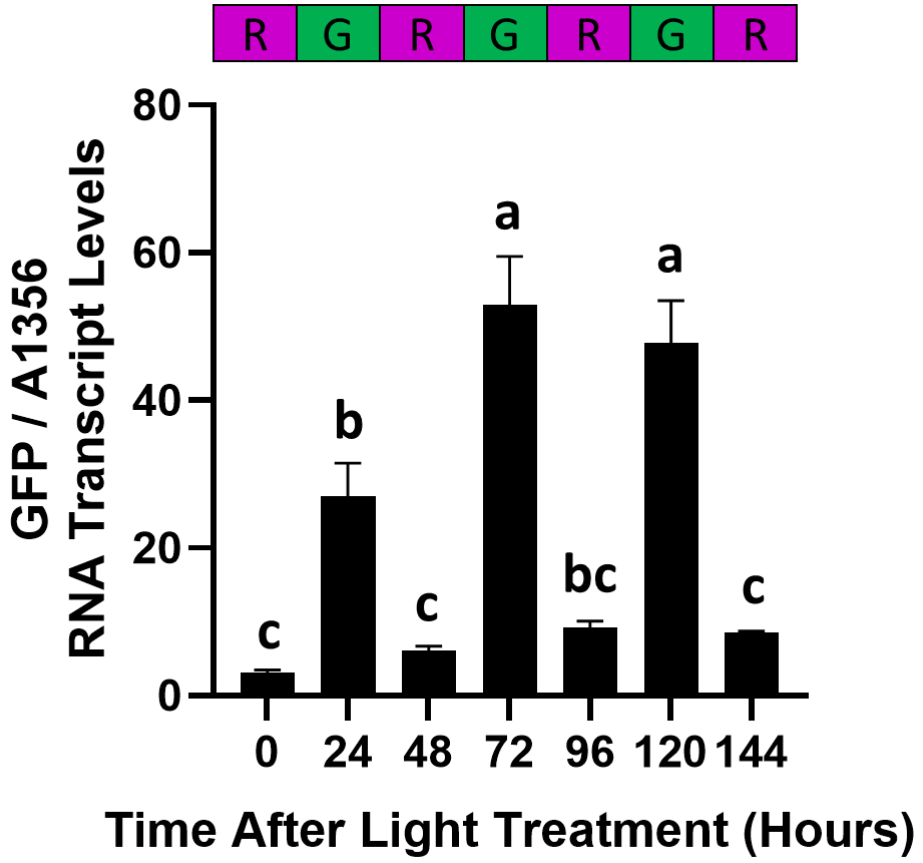


Figure 3-12. CcaS system activity during alternating periods of red or green light. 1 L cultures of Syn7002 expressing pAQ1BB-CcaSR-GFP were grown for 4 days in photobioreactors under constant red LED light. Cultures were then illuminated with alternating 24-hour periods of green or red light, beginning with green light between 0-24 hours. Symbols above each bar represent whether at the specified time point the culture had been illuminated with 24 hours of red (R) or green (G) light. RNA was extracted from samples harvested at the indicated time points, and GFP RNA transcript was measured by qRT-PCR and normalised to *SYNPCC7002_A1356* RNA transcript levels. Data are means ± S.E.M of three independent cultures. Groups sharing the same letter are not significantly different from each other ($p > 0.05$) based on Tukey's multiple comparison test following one-way ANOVA

The previous experiment which measured the effect of light/dark cycles on CcaS system activity measured GFP RNA transcript levels at the end of the 8-hour dark cycle but not at the beginning (Figure 3-5). This prevented the reduction in GFP transcript specifically across this 8-hour dark period from being quantified. However, this data was still presented as it was part of the initial validation of the qRT-PCR method of measuring CcaS system transcriptional output. To better measure the effect of the 8-hour dark cycle on CcaS system activity, pAQ1BB-CcaSR-GFP cultures were grown in photobioreactors under red light using a 16h/8h light/dark photoperiod for 5 days until cultures reached OD_{730} ~1. Illumination was then switched to green light only whilst maintaining the 16h/8h

light/dark photoperiod and GFP RNA transcript levels were measured at several time points using qRT-PCR.

GFP RNA transcript levels increased during the first 16-hour green light period before decreasing by around 30% at the end of the first dark period (Figure 3-13). GFP transcript levels increased again during the second light cycle to a slightly higher level than at the end of the first green light period. However, the relative increase was lower during this second light cycle and the difference in GFP RNA transcript levels at the end of each green light period (16h vs 40h) was not deemed to be statistically significant. GFP transcript levels then decreased again in the second dark period at a similar rate to the first dark cycle. This data shows that an 8-hour dark period is not sufficient to decrease GFP RNA transcript levels close to baseline and that red light allows faster deactivation (Figure 3-11).

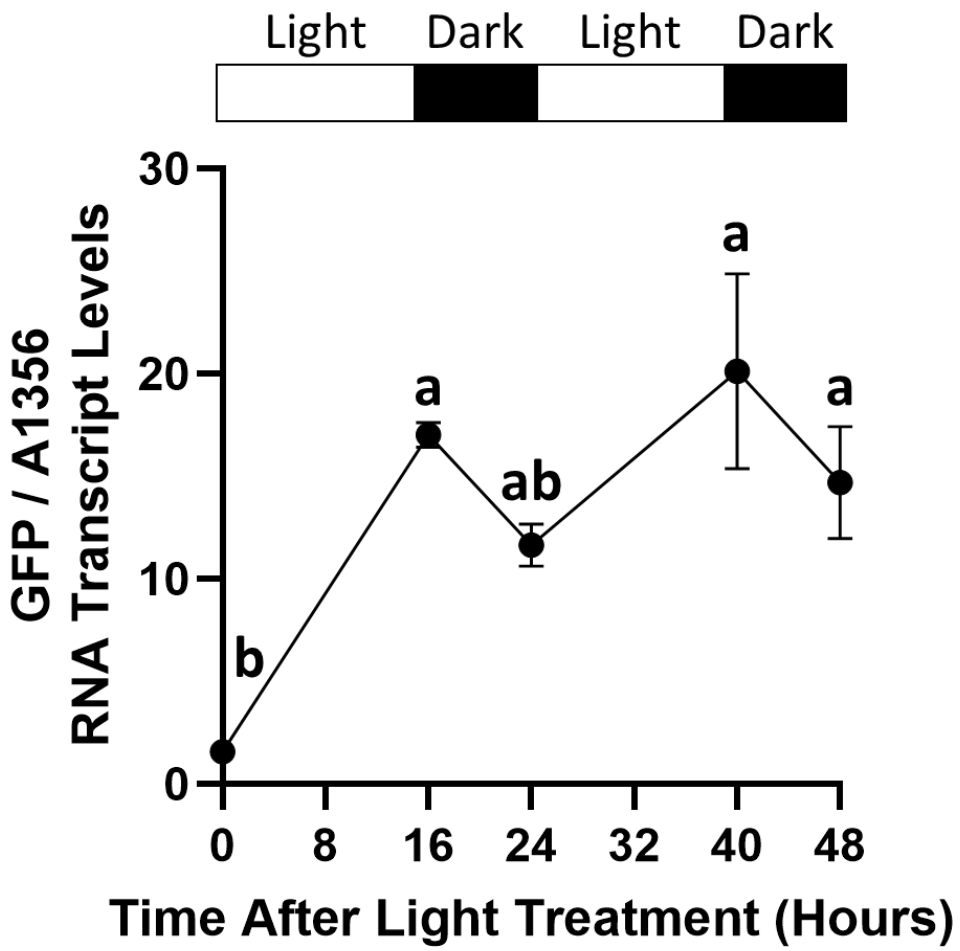


Figure 3-13. CcaS system activity under light/dark cycles. 1 L cultures of Syn7002 expressing pAQ1BB-CcaSR-GFP were grown for 5 days in photobioreactors under red light using a 16h/8h light/dark cycle. Whilst maintaining the same light/dark photoperiod, cultures were then illuminated with green light rather than red light during the subsequent 48 hours. RNA was extracted from samples harvested at the indicated time points, GFP RNA transcript was measured by qRT-PCR and normalised to *SYNPCC7002_A1356* RNA transcript levels. Data are means ± S.E.M of three independent cultures. Groups sharing the same letter are not significantly different from each other ($p > 0.05$) based on Tukey's multiple comparison test following one-way ANOVA.

3.3.5 Optimisation of CcaS system performance via engineering of pCpcG2

Two factors which affect the utility of an optogenetic system are the dynamic range, i.e., the fold-increase in system activity between the ON/OFF states, and the maximal level of transcriptional output achieved in the ON state. The system component which was deemed most likely to affect these factors was the pCpcG2 output promoter. It was hypothesised that truncation of the pCpcG2 promoter would reduce the activity of the promoter in red light in Syn7002. Truncation of pCpcG2 in *E. coli* reduced promoter activity in red light,

presumably due to the removal of a weak constitutive promoter present in the truncated region (Schmidl *et al.*, 2014). It was also hypothesised that modifying the -35 and -10 hexamers of pCpcG2 to match those of the strong Syn7002 constitutive promoter pCpct223 (Markley *et al.*, 2015) could increase the maximal activity of the CcaS system under green light. However, it was unknown whether modifying the -35 or -10 hexamers alone, or in combination, would have the greatest impact on the dynamic range and maximal transcriptional output of the CcaS system under green light. Site-directed mutagenesis (SDM) was used to generate the suicide plasmids pUC-172-GFP, pUC-172-10-GFP, pUC-172-35-GFP and pUC-172-35-10-GFP, which each enabled expression of GFP from the modified pCpcG2 promoters (Figure 3-14a). All constructs contained the 172 bp truncated pCpcG2 promoter combined with various combinations of -10 and -35 hexamers from the pCpct223 strong promoter. While the constructs facilitated integration into the same region of the Syn7002 endogenous plasmid pAQ1 as pAQ1BB-CcaSR-GFP, they were now cloned into a pUC19 backbone plasmid which improved cloning efficiency. Furthermore, the homology arms were increased from 250-350 bp to 1000 bp to increase the transformation efficiency. 150 mL cultures expressing the constructs were grown under red light using a 16h/8h light/dark photoperiod for 5 days until cultures reached OD_{730} = 1. The cultures were then switched to illumination with green light only for 4 hours and GFP RNA transcript levels were measured using qRT-PCR.

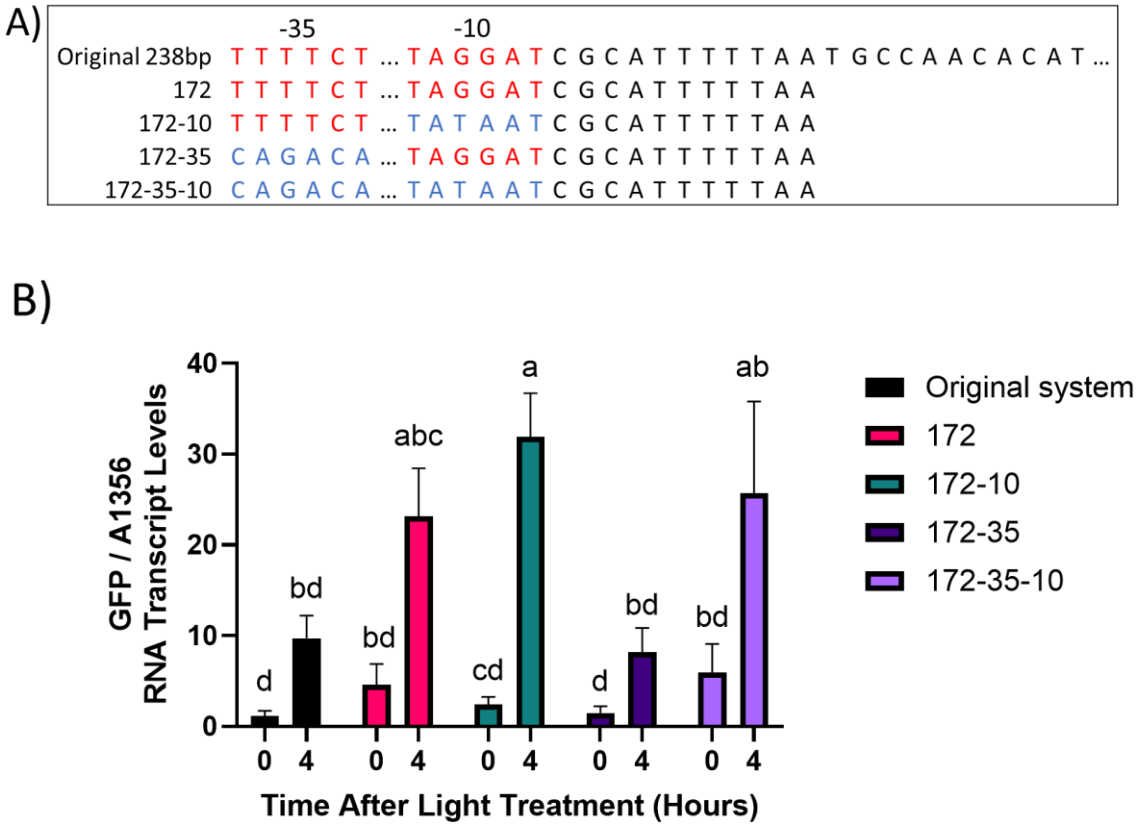


Figure 3-14. CcaS activity after modification of pCpcG2. A) Schematic showing the sequence of pCpcG2 in the following constructs: pAQ1BB-CcaSR-GFP (Original system), pUC-172-GFP (172), pUC-172-10-GFP (172-10), pUC-172-35-GFP (172-35) and pUC-172-35-10-GFP (172-35-10). The -35 and -10 hexamers are highlighted in red, with bases differing from the original pCpcG2 sequence highlighted in blue. Ellipses indicate regions of sequence omitted for simplicity. Constructs labelled with "172" indicate the truncation of bases 173-238 from the 3' end of pCpcG2. **B)** 150 mL cultures of Syn7002 cultures expressing the above constructs were grown for 4 days under illumination with red light as provided by red light transmitting filters combined with warm white LED lights. The cultures were then switched to green light as provided by green light transmitting filters. RNA was extracted from samples harvested at the indicated time points, GFP RNA transcript was measured by qRT-PCR and normalised to *SYNPCC7002_A1356* RNA transcript levels. Data are means ± S.E.M of three independent cultures. Significance at $p < 0.05$ was determined using two-way ANOVA with Tukey's multiple comparison test.

The original CcaS system demonstrated a dynamic range of 8.5-fold after 4 hours of green light, which was similar to previous experiments (Figure 3-14b). In cultures expressing pUC-172, which contained a truncated 172 bp version of the original 232 bp pCpcG2 promoter, GFP RNA transcript levels were higher under both red light and green light than the original system. However, the dynamic range was only 5-fold due to the higher baseline level of transcription under red light, indicating lower performance than the original system. In cultures

expressing pUC-172-10, which contained truncated pCpcG2 with a modified -10 hexamer, GFP RNA transcript levels were 3-fold higher under green light versus the original system. This resulted in an improved dynamic range of 13.3-fold. In cultures expressing pUC-172-35, which contained truncated pCpcG2 with a modified -35 hexamer, GFP RNA transcript levels under green light were lower than that of the original system, with a reduced dynamic range of 5.5-fold. In cultures expressing pUC-172-35-10, which contained truncated pCpcG2 with both -35 and -10 hexamers modified, GFP RNA transcript levels under both red and green light were higher than the original system, resulting in the lowest dynamic range of 4.3-fold due to increased baseline transcription under red light. This data showed that the combination of pCpcG2 truncation and modification of the -10 hexamer resulted in a CcaS system variant with improved properties compared to the original.

3.4 Discussion

3.4.1 Comparison of CcaS system performance in Syn7002 and other organisms

While the potential of optogenetics has been demonstrated in many microbes, application in cyanobacteria has been scarce. Furthermore, previous applications of the CcaS system in cyanobacteria have not explored the transcriptional kinetics of the system in detail and have focused on a small number of constant lighting conditions. This work addresses these issues and is the first to express the CcaS system in Syn7002, which has certain advantages over the previously used Syn6803 (see Introduction). The system has also been optimised for improved performance, and this improved variant has been characterised.

First, it was confirmed that the CcaS system is functional in *E. coli*, with a 5-fold increase in GFP fluorescence observed between red and green light illumination (Figure 3-1). These results agree with previous research which has demonstrated a 6-fold increase in reporter gene expression in *E. coli* between red and green light (Schmidl *et al.*, 2014). Next, the CcaS system was expressed in Syn7002, with a 3.5-fold increase in GFP expression observed after 48 hours of constant green light or with red plus green light (Figure 3-4). These results contrast with those seen when expressing CcaS in Syn6803, where 39 hours of green light

(alone and with red light) resulted in a 10 to 11-fold increase in reporter gene expression. (Abe *et al.*, 2014). Similarly, in Syn15041c, 20 µmol m^{-2} s^{-1} green plus 20 µmol m^{-2} s^{-1} red light caused a 7-8-fold increase in reporter gene expression compared to red light alone after 48 hours (Badary *et al.*, 2015). Differences in pigment composition between these species may affect the exact amounts of light absorbed by CcaS and thus affect its activity. It is also possible that the relative amounts of CcaS and CcaR protein in the cell could influence this, as the CcaS and CcaR promoters from Syn6803 used may have different strengths across the species. Furthermore, the subtraction of background fluorescence can have a significant impact on the perceived "fold change" in expression levels. Subtle errors or deviations in measurements, especially background fluorescence measurements, can alter fold changes significantly.

The application of optogenetics in a photosynthetic organism is complicated by the trade-off between the light wavelengths required for optimal photosynthesis and the application or removal of wavelengths for controlling gene expression. Previous work in Syn6803 and Syn15041c has shown that illumination with red plus green light resulted in a similar CcaS system output as under green light alone (Abe *et al.*, 2014; Badary *et al.*, 2015). A similar result was obtained here for Syn7002, where red plus green light resulted in similar levels of GFP fluorescence as green light alone (Figure 3-4). This property is very useful as it allows cultures to be grown under red light illumination, and green light can then be applied or removed as desired to modulate gene expression. This avoids the negative impact on growth that is observed under illumination with green light illumination alone (Figure 3-3).

Previous studies in other strains have often used qRT-PCR to measure CcaS system output. In Syn15041c, GFP mRNA levels measured 3, 5 and 8 hours post-green light illumination were 3-fold, 20-fold and 4-fold higher than in the red light control (Badary *et al.*, 2015). However, significant variation between replicates was also observed, particularly at the 5-hour timepoint which showed the largest fold increase. In Syn7002, we found a 6-fold increase in GFP RNA transcript levels within 30 minutes of green light illumination (Figure 3-10). However, variation was also high and therefore fold changes measured by qRT-PCR may be imprecise. Badary and colleagues also observed fluctuation in GFP

transcript levels under red light conditions, although levels were always lower than in green light. Fluctuation of mRNA levels has also been observed when using the CcaS system to control transcription of the T4 holin gene in Syn6803 (Miyake *et al.*, 2014).

This work is the first to characterise the de-activation kinetics of the CcaS system in cyanobacteria by measuring the transcriptional output. An effort was made to characterise the kinetics of both activation and de-activation of another potential optogenetic system, the far-red light-activated promoter $P_{chlFJSC1}$ from the cyanobacterium *Leptolyngbya* sp. JSC-1 (Liu *et al.*, 2023). When used to control the expression of EYFP in *Chlorogloeopsis fritschii* PCC 9212, far-red light illumination caused a 30-fold increase in EYFP fluorescence, although this level was not sustained and dropped steadily over several days. It is unclear how quickly the system activated as samples were taken every 24 hours and EYFP transcript levels were not measured. EYFP fluorescence was significantly reduced by switching the light source back to white light, with a 60-70% decrease within 12 hours. It remains to be investigated whether the deactivation is caused by a specific light wavelength within the white light or simply by the absence of far-red light. It is thought that the response of $P_{chlFJSC1}$ to far-red light is mediated by the action of the photoreceptor RfpA and related response regulators RfpB and RfpC, which are part of the native *Chlorogloeopsis fritschii* PCC 9212 far-red light photoacclimation (FaRLiP) response. This was hypothesised after showing that transcription from $P_{chlFJSC1}$ could not be activated in Syn7002 which lacks the FaRLiP gene cluster.

The present work is also the first to demonstrate the effect of multiple red and green light periods on CcaS system output in cyanobacteria (Figure 3-12). This has been demonstrated in *E. coli* and is a critical characteristic if an optogenetic system is to be used for dynamic control of metabolic pathways (Fernandez-Rodriguez *et al.*, 2017). In this context, it is also important to understand whether similar levels of activation and deactivation of gene expression can be achieved over repetitive cycles of ON/OFF switching. In measuring GFP mRNA transcript levels over three 24-hour green-red cycles, two interesting observations were made (Figure 3-12). Firstly, the red light treatment significantly reversed the green light activation but did not reach the baseline

expression level recorded before the initial green light treatment. Secondly, the level of green light activation increased from the first to the second treatment but not any further in the third treatment. The initial 24 hours of green light may have a 'priming' effect, allowing the system to respond stronger to subsequent green light illumination. Interestingly, the priming effect was not apparent during the light/dark cycle experiment, where the difference between the 16 and 40-hour time points was not statistically significant (Figure 3-13).

The molecular mechanisms behind a potential priming effect are unclear. One possibility is that it is simply due to higher residual GFP transcript after the 24-hour red light period. However, if this was true, an even greater increase in GFP expression after the third period of green light could be expected as residual GFP expression was higher after the second period of red light than the first. Other possible explanations are that after the first periods of red and green light, CcaR, other general transcriptional initiation factors, or RNAP remain near pCpcG2 and/or in a pre-activated state. This characteristic of the CcaS system could potentially be exploited during industrial applications. By providing an initial period of green light activation to 'prime' the system this would allow higher levels of target gene expression during subsequent periods of green light illumination.

Previous work in cyanobacteria has only characterised CcaS system activity under constant light conditions. However, cyanobacteria are often grown under light/dark cycles which is an important consideration for the deployment of optogenetics. The present work is thus the first to characterise the effect of light/dark cycles on the CcaS system in cyanobacteria. Although a direct comparison was not made within the same experiment, the reduction in GFP RNA transcript in darkness appeared to be slower than under red light (Figure 3-13 compared to Figure 3-11). In the darkness, CcaS will remain active until it naturally reverts to the dark state, and the observed reduction of 25-30% suggests that the ON state half-life if the green activated CcaS system is more than 8 hours. Thus, product synthesis will continue at least to some extent during the night period. If fast deactivation of the system is desired, red light should be used.

3.4.2 qRT-PCR versus protein fluorescence spectrometry for characterising optogenetic tools

Measuring GFP fluorescence via fluorescence spectrometry is a quick and straightforward way to measure promoter output. However, from the GFP fluorescence data, it appears as though the CcaS optogenetic system is slow in both activation and deactivation in Syn7002. GFP fluorescence increased much more slowly in response to green light than GFP RNA transcript as measured by qRT-PCR (Figure 3-4 and Figure 3-10). Translation speed varies but it is in the range of 20 amino acids per second and usually matches the speed of transcription. Longer time scales for achieving protein activity are likely to be due to the folding and targeting of proteins and the amount of protein needed to achieve measurable fluorescence. Regardless of the rate limiting factor, measuring the transcriptional response rather than protein activity provides a more accurate reflection of the dynamics of an optogenetic system.

The GFP variant used in this work was GFPmut3 (Cormack *et al.*, 1996), which was chosen due to its regular use as a reporter gene in Syn7002 in the Amtmann lab. However, the data collected in this study suggests that GFPmut3b may exhibit high stability in Syn7002, as GFP fluorescence remained elevated even when GFP RNA transcript levels had reduced close towards baseline levels under red light (Figure 3-6 and Figure 3-11). In future work, a less stable GFP variant could be used to more accurately reflect the transcriptional dynamics of the CcaS optogenetic system in Syn7002. Several degradation tags of differing strengths have been tested in Syn6803, however these have yet to be characterised in Syn7002 (Landry *et al.*, 2013). The combination of optogenetic tools which control the speed of protein degradation with tools which control transcriptional activation may be key for tight control over metabolic pathways in Syn7002.

While using qRT-PCR was more suitable for measuring the rate of CcaS system activation and deactivation, this method also has drawbacks. When measuring the decrease in CcaS system activity under red light, the stability of the GFP RNA transcript is likely to significantly influence the results. This would mean that the transcriptional output of the system would present different dynamics depending on the target gene expressed. Several cellular mechanisms can affect

the mRNA transcript stability, one of which is the cleavage of AU-rich regions by cyanobacterial RNase E (Horie *et al.*, 2007). The lower the frequency of AU-rich regions encoded within a gene, the longer the potential half-life of the mRNA. Some genes also encode antisense RNAs on the opposite strand, which when transcribed can bind to these AU-rich regions of the single-stranded mRNA upstream of the RBS. This prevents RNase E cleavage whilst still allowing mRNA translation, as has been shown for *PsbA* expression in Syn6803 (Sakurai *et al.*, 2012). Therefore, when using optogenetics to control the expression of a target gene, it is important to consider the presence of AU-rich regions and promoters which may drive antisense RNA expression. Modulating these factors via gene synthesis could be a valuable method for modifying target gene mRNA stability and therefore improving the transcriptional kinetics of the optogenetic system. Alternatively, other optogenetic tools such as the 'Lockdown' system, which allows blue light-activated knockdown of RNA expression using CRISPR/Cas13b could be used to actively degrade mRNA (Bloemeier *et al.*, 2021). However, these types of tools have yet to be demonstrated in cyanobacteria.

Another drawback of using qRT-PCR to characterise the transcriptional dynamics of the CcaS system is that significant variation between replicates was observed in several of the experiments, even when four independently grown cultures were assayed. As the variation observed in the GFP fluorescence assays was much lower, this suggests that the variation in the qRT-PCR data has a significant technical component. GFP fluorescence did not increase across 48 hours of red light, whereas fluctuating levels of GFP RNA transcript were measured under red light using qRT-PCR in other experiments (Figure 3-4 and Figure 3-8). This discrepancy could be due to ambient light being absorbed by samples during the cell harvesting process. However, this is unlikely as great care was taken to harvest and process samples under red light.

One final consideration is that the use of fluorescence spectrometry and qRT-PCR for analysis of optogenetic system output only allows the analysis of expression levels in bulk cell populations. It is therefore unclear whether there is variability in the levels of expression of GFP at the single cell level. Future experiments could use flow cytometry to better characterise this potential

variability. Another method could be to use fluorescence microscopy and manually count the percentage of cells expressing GFP.

3.4.3 Plasmid copy number variance and gene dosage effects

Several approaches were taken to normalise the effect of potential fluctuations in transgene and plasmid copy number on CcaS system activity. First, the fraction of pAQ1 plasmids containing the GFP transgene was measured (Figure 3-9a). Whilst significant variation was present between replicates, no significant differences were present between time points or light treatments. This suggests that any differences in GFP RNA transcript levels between the light treatments and across time were not due to variations in the fraction of plasmids containing the GFP transgene. Next, an attempt was made to measure pAQ1 plasmid copy number only (Figure 3-9b). Again, whilst there was significant variation between replicates for some of the time points, the mean pAQ1 plasmid copy number was similar across all time points and light treatments. One drawback of these experiments was that no reference gene could be used for qPCR measurement of transgene or plasmid copy number. In contrast, a validated internal reference gene *SYNPCC7002_A1356* was used when measuring GFP RNA transcript levels. Whilst the same amount of gDNA was used as a template in each qPCR reaction, pipetting errors, gDNA integrity and discrepancies between the measured and actual DNA concentration could all have influenced the results.

Next, GFP RNA transcript levels (in addition to being normalised to the internal control *SYNPCC7002_A1356*) were normalised to the pAQ1 plasmid copy number (Figure 3-9c). This resulted in a very similar pattern as in Figure 3-8 where only *GFP*/*SYNPCC7002_A1356* RNA transcript levels were measured. Finally, GFP RNA transcript levels (in addition to being normalised to the internal control *SYNPCC7002_A1356*) were normalised to the fraction of pAQ1 plasmid containing GFP transgene (Figure 3-9d). Whilst there were some differences in the absolute values compared to normalising to pAQ1 plasmid copy number only, a very similar pattern in expression was observed. Collectively, this data suggests that fluctuations in pAQ1 plasmid copy number or the fraction of pAQ1 plasmids containing GFP transgene did not significantly affect measured GFP RNA transcript levels.

3.4.4 CcaS system optimisation

Previous work has focused on optimising the performance of the CcaS system in organisms other than cyanobacteria. Truncation of pCpcG2 in *E. coli* from 238 bp to 172 bp reduces promoter activity under red light by removing a predicted weak constitutive promoter, whilst only resulting in a modest reduction in activity under green light (Schmidl *et al.*, 2014). This reduces the 'leakiness' of the system and improves the dynamic range. In contrast, the present work showed that the truncation of pCpcG2 resulted in increased system activity under both red and green light (Figure 3-14). The removed 66 bp region may contain a sequence which influences promoter activity under both red and green light in Syn7002. This underscores the differences of transcriptional regulation between different bacteria which limit the transferability of tools and systems between species.

pUC-172-10-GFP, which contained truncated pCpcG2 in addition to a modified -10 hexamer, showed both the largest dynamic range and the highest activity under green light of the five constructs tested, 3-fold higher than the original system (Figure 3-14). Modifying the -10 hexamer to match that of the consensus sequence also resulted in a 3-fold increase in pCpcG2 activity under green light in *B. subtilis* (Castillo-Hair *et al.*, 2019). Optogenetic tools which enable high levels of transgene expression are advantageous in metabolic engineering as expression can always be lowered by reducing light intensity or duration.

4 Troubleshooting the YF1/FixJ system in Syn7002

4.1 Introduction

The chimeric YF1/FixJ system of bacterial origin enables gene expression in the absence of blue light, with transcription repressed under blue light. In the absence of blue light, YF1 phosphorylates the transcription factor FixJ, which then binds to and activates transcription from the pFixK2 promoter. Blue light absorption by YF1 triggers its phosphatase activity and thus FixJ becomes dephosphorylated, decreasing transcription from pFixK2 (discussed in Introduction).

The initial purpose of the YF1 system was to allow light-regulated control of gene expression. Indeed, YF1 has been used for this purpose in several organisms including *E. coli*, *Pseudomonas aeruginosa* and *Vibrio natriegens* (Fernandez-Rodriguez *et al.*, 2017; Pu *et al.*, 2018; Tschirhart *et al.*, 2019). When the YF1 system was used to control the expression of the lacZ reporter gene in *E. coli*, a 70-fold increase in lacZ expression was observed upon switching from blue light irradiation to darkness (Möglich *et al.*, 2009). The YF1 system has also been used to re-engineer the lac operon to be induced by blue light rather than IPTG in *E. coli* (Lalwani *et al.*, 2021a). This takes advantage of the most commonly used inducible gene expression system and widens its applicability by making it reversible via optogenetics. The YF1 system has also been used to control protein expression in an *E. coli*-based cell-free expression system (Zhang *et al.*, 2020). When the YF1 system was used to control DsRed expression in *E. coli*, the maximum level of fluorescence was reached after 1 hour of darkness (Ohlendorf *et al.*, 2012). This shows that the system can be rapidly activated, which highlights its potential for controlling metabolic pathways.

The LOV domain present in YF1 has been exploited for various functions beyond controlling gene expression in both prokaryotes and eukaryotes, including regulation of cytoskeleton remodelling, subcellular protein localisation, ion channels and protein degradation (Wu *et al.*, 2009; Niopek *et al.*, 2014; Usherenko *et al.*, 2014; Cosentino *et al.*, 2015). The molecular basis of many of these applications is the blue light-activated unfolding of the LOV domain Jα-helix. For example, for blue light-regulated control of protein degradation in *S.*

cerevisiae, target proteins have been fused to the LOV2-Jα photoswitch derived from *Arabidopsis thaliana* Phototropin1. Blue light absorption by the LOV domain triggers unfolding of the Jα-helix which exposes a previously sequestered degron motif. This causes subsequent ubiquitination and protein degradation via the ubiquitin-proteasome system, reducing the average protein half-life by 6-fold (Usherenko *et al.*, 2014). Protein degradation systems allow for faster downregulation of enzymes in metabolic pathways than relying on the slower natural degradation of mRNA transcripts. However, while these types of protein degradation systems have been used extensively in eukaryotes, they are not widespread in bacteria.

More complex multichromatic optogenetic approaches have also been developed. *E. coli* have been engineered to express three optogenetic tools, allowing them to respond to red, green and blue light. These cells expressed the blue Light responsive YF1 system, red/green responsive CcaS system and red/far-red responsive chimeric histidine kinase Cph8. Using these switches to control the expression of the red, green and blue fluorescent protein reporter genes, the authors showed that there was minimal spectral overlap between the three systems. The multichromatic system was then used to control the expression of three CRISPRi sgRNAs targeted to genes within the *E. coli* biosynthetic pathway for converting glucose to acetate (Fernandez-Rodriguez *et al.*, 2017). Optogenetic downregulation of these genes resulted in reduced acetate production, a commonly produced by-product unwanted in many industrial production scenarios. This demonstrates the feasibility of using different wavelengths for dynamic control of the expression of multiple genes in a metabolic pathway. The YF1/FixJ system has yet to be characterised in Syn7002.

4.2 Chapter aims

The aim of this chapter was to express and characterise the activity of the YF1/FixJ system in Syn7002. First, it was confirmed that the original system worked in *E. coli* as has been reported in the literature. Next, it was found that the YF1/FixJ system performs poorly in Syn7002, even after codon optimisation. Expression of the YF1 and FixJ genes was confirmed in the Syn7002 strain expressing the construct, as well as the functionality of the codon-optimised

version in *E. coli*. Finally, troubleshooting steps including supplying exogenous FMN chromophore and increasing the predicted strength of the YF1/FixJ operon RBS 20-fold were applied in an attempt to improve system performance.

4.3 Results

4.3.1 Confirming YF1/FixJ function in *E. coli*

To confirm that the original YF1 system is functional in *E. coli*, the GFP gene was inserted into the multiple cloning site of the pDusk plasmid (Ohlendorf *et al.*, 2012) to generate pDusk-GFP. pDusk-GFP contains the YF1 and FixJ genes which are expressed as a bicistronic operon from the *E. coli* constitutive promoter LacIq and also expresses GFP from pFixK2. *E. coli* was transformed with pDusk-GFP and grown on selective media. The presence of the correct DNA sequence in selected clones was confirmed by restriction digest analysis and sequencing. A successfully transformed clone was used to inoculate 150 mL cultures containing pDusk-GFP. These were grown to OD_{600} = ~0.3 under 100 µmol m^{-2} s^{-1} white LED light and were then either wrapped in foil (Dark) or left uncovered (Light). GFP fluorescence was measured 3 hours after the respective light treatments, normalised to OD_{600} and background fluorescence of the empty pDusk vector was subtracted. After 3 hours, GFP fluorescence was 45-fold higher in cultures incubated in darkness than those incubated under white light (Figure 4-1). This confirmed that pDusk-GFP, and thus the original YF1 system, is functional in *E. coli* as has been previously reported.

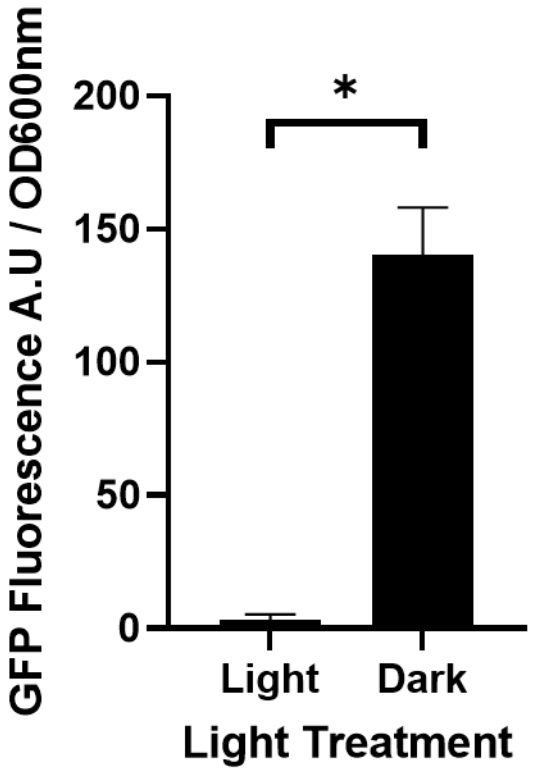


Figure 4-1. YF1 system activity in *E. coli* in white light versus darkness. 150 mL cultures of *E. coli* expressing the pDusk-GFP plasmid were grown to OD_{600} = ~0.3 under 100 µmol m^{-2} s^{-1} white LED light. The cultures were then either wrapped in foil (Dark) or left uncovered (Light). Samples were harvested from each culture 3 hours after the respective light treatments and immediately placed on ice to stop growth. GFP fluorescence was measured, normalised to OD_{600} and background fluorescence measured in cells expressing empty pDusk vector was subtracted. Data are means ± S.E.M of three independently grown cultures. Significant differences at $p < 0.05$ as determined by unpaired t-test are shown by asterisk.

4.3.2 Expression of YF1/FixJ in Syn7002

After confirming that the YF1 system is functional in *E. coli*, the system was then expressed in Syn7002. A DNA fragment containing the YF1/FixJ operon expressed from a medium-high strength cyanobacterial constitutive promoter pCpc350 and pFixK2 was chemically synthesised and inserted into pAQ1BB-RBS9-GFP using restriction enzymes. pCpc350 is a truncated version of the strong cyanobacterial constitutive promoter pCpcB594 generated in the Amtmann lab (Madsen, unpublished). The new construct, designated pAQ1BB-Blue, contains GFP immediately downstream of pFixK2 (Figure 4-2). pAQ1BB-Blue is a suicide vector, allowing DNA insertion into integration into the Syn7002 endogenous PAQ1 plasmid. Syn7002 were transformed with pAQ1BB-Blue and grown on selective media. The presence of the correct DNA in selected clones and

obtaining of a homozygous strain was confirmed by colony PCR and sequencing. As a control, Syn7002 were also transformed with empty pAQ1BB vector (1BB). 150 mL Syn7002 cultures expressing pAQ1BB-Blue or pAQ1BB were grown for 4 days to OD_{730} = ~1 under white LED light. Cultures were then either covered with a blue light-removing filter (No BL) (see Materials and Methods) or left uncovered (BL). GFP fluorescence and OD_{730} were measured at several time points.

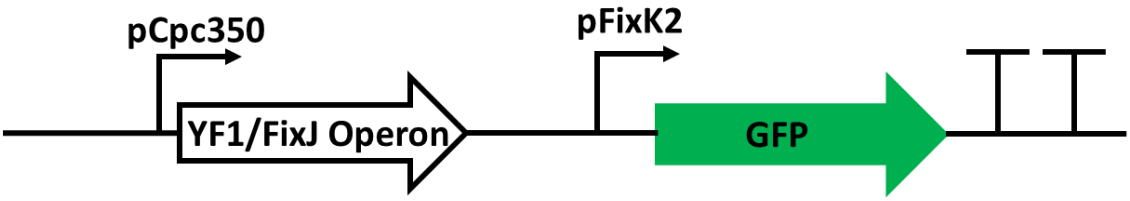


Figure 4-2. Schematic showing design of pAQ1BB-Blue construct. Double terminator after the GFP gene.

Minimal differences in GFP fluorescence were observed between the two light treatments across most of the time points in the pAQ1BB-Blue strain (Figure 4-3). After 32 hours, the No BL pAQ1BB-Blue cultures produced slightly higher fluorescence than the BL pAQ1BB-Blue cultures, with a p-value of 0.06. A similar difference between the BL and No BL treatments was also observed in the pAQ1BB cultures containing empty vector suggesting that the slight increase in fluorescence may not be from increased GFP production. Under both light treatments, pAQ1BB-Blue cultures showed similar levels of fluorescence as pAQ1BB cultures. As the cultures were grown under a 16/8 hour light/dark cycle, the cultures were in the dark in the 16-24 hours after the light treatment began. This meant that the No BL pAQ1BB-Blue cultures would have been in the 'induced' state for 32 hours. These results demonstrate the poor performance of the YF1 system in Syn7002.

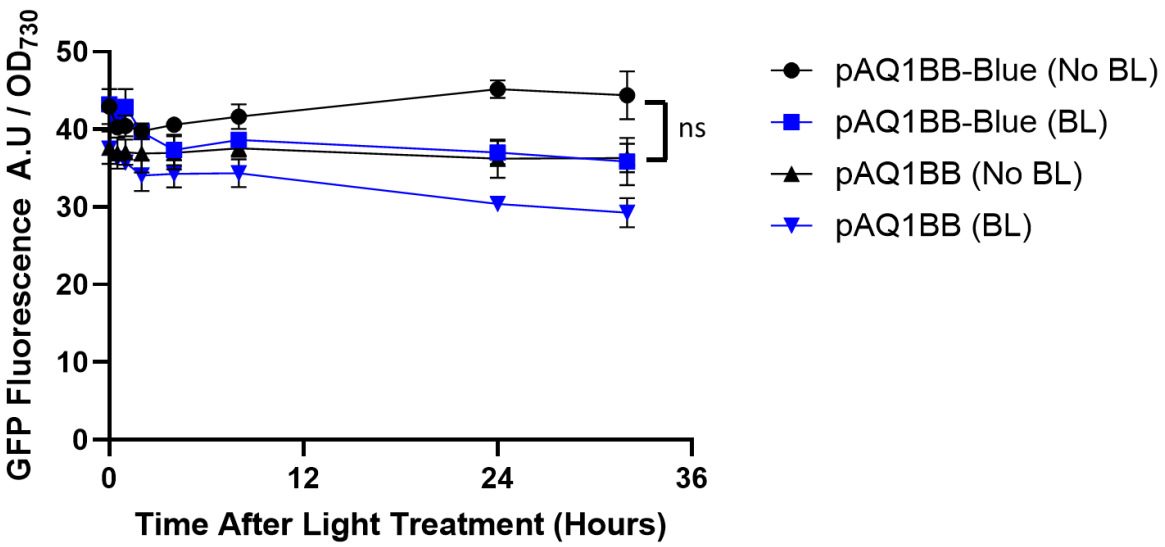


Figure 4-3. YF1 system activity in Syn7002 in the presence and absence of blue light. 150 mL Syn7002 cultures expressing either GFP under control of the YF1 system (pAQ1BB-Blue) or empty pAQ1BB vector (pAQ1BB) were grown to OD_{730} = ~1 under 150 µmol m^{-2} s^{-1} white LED light. The cultures were then either covered with a blue light-removing filter (No BL) (black symbols) or left uncovered (BL) (blue symbols). As cultures were grown under a 16/8 hour light/dark cycle, the cultures were in the dark between 16-24 hours after the light treatment began. GFP fluorescence was measured and normalised to OD_{730}. Data are means ± S.E.M of three independent cultures. No significant differences at $p < 0.05$ were determined by unpaired t-test.

4.3.3 Troubleshooting the poor performance of YF1/FixJ in Syn7002

It was hypothesised that the poor performance of the YF1 system in Syn7002 may be due to a lack of YF1 and FixJ gene expression in the pAQ1BB-Blue strain. To investigate this, RNA was extracted from pAQ1BB-Blue and pAQ1BB cultures containing empty pAQ1BB vector and used to produce cDNA which was used as a template for RT-PCR analysis and was visualised by agarose gel electrophoresis. Primers were designed to amplify short 100-200 bp regions of the YF1 and FixJ genes and to amplify the constitutively expressed 16S rRNA as a control. Clear bands were observed in the pAQ1BB-Blue cDNA YF1 and FixJ lanes (Figure 4-4). The absence of bands in the no template control (NTC) and pAQ1BB samples highlighted that amplification in the pAQ1BB-Blue cDNA samples was not due to primer dimers or non-specific amplification from the Syn7002 genome. RNA samples used as a template for cDNA synthesis in the absence of reverse transcriptase (No-RT) also demonstrated that any amplification observed in the

pAQ1BB-Blue cDNA samples did not arise from contaminating gDNA. The results show that the YF1 and FixJ genes are expressed in the pAQ1BB-Blue strain.

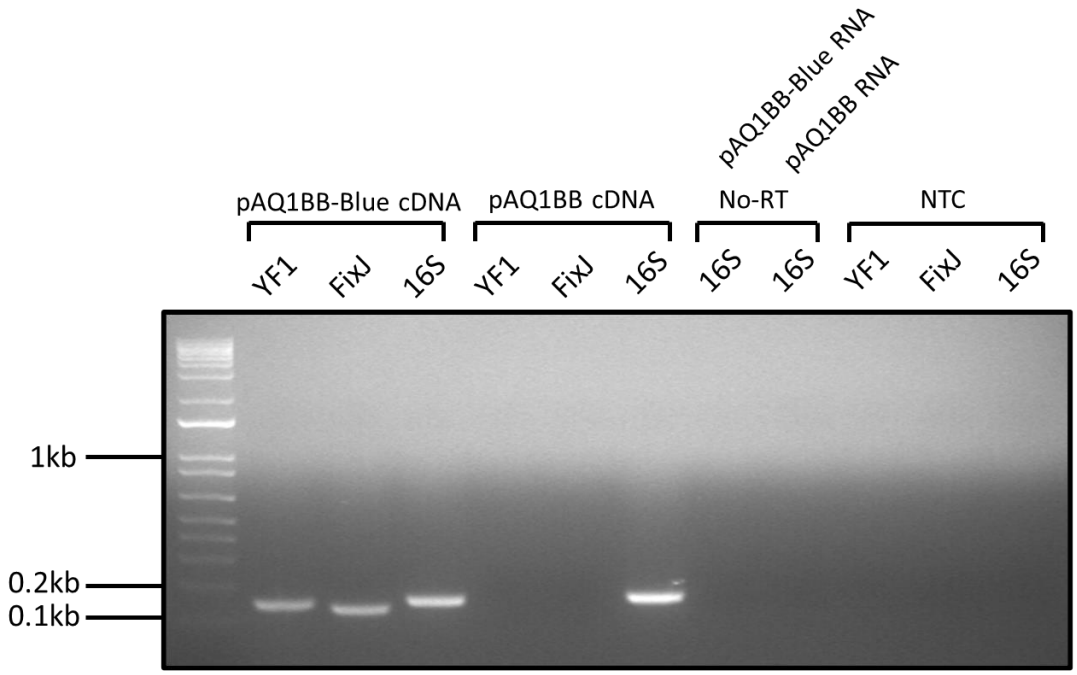


Figure 4-4. Gel image of YF1 and FixJ RT-PCR reactions from pAQ1BB-Blue culture cDNA. pAQ1BB-Blue and pAQ1BB cultures were grown to OD_{730} = ~1 under standard conditions, RNA was extracted, and cDNA was synthesised from 0.5 µg RNA. 0.5 µl undiluted cDNA was then used in 25 µL PCR reactions containing either YF1, FixJ, or 16S rRNA primers. Reactions were run for 30 cycles and visualised on a 1.5% agarose gel stained with ethidium bromide. No-RT = no reverse transcriptase, NTC = no template control.

After confirming that the *YF1* and *FixJ* genes were expressed in Syn7002, it was hypothesised that the poor performance of the YF1 system could be due to a requirement for an increased intracellular level of the YF1 co-factor flavin mononucleotide (FMN). To investigate this, 150 mL pAQ1BB-Blue cultures were grown to OD_{730} = ~1 under white light. Cultures were then either covered with a blue light-removing filter (No BL) or left uncovered (BL) and riboflavin 5′-monophosphate sodium salt hydrate (FMN-Na) was added to a final concentration of 1 µM. GFP fluorescence and OD_{730} were measured at several time points. 48 hours after the addition of FMN-Na, no significant differences in GFP fluorescence were observed between the two light treatments (Figure 4-5). This suggests that the availability of FMN co-factor was not limiting the performance of the YF1 system in Syn7002.

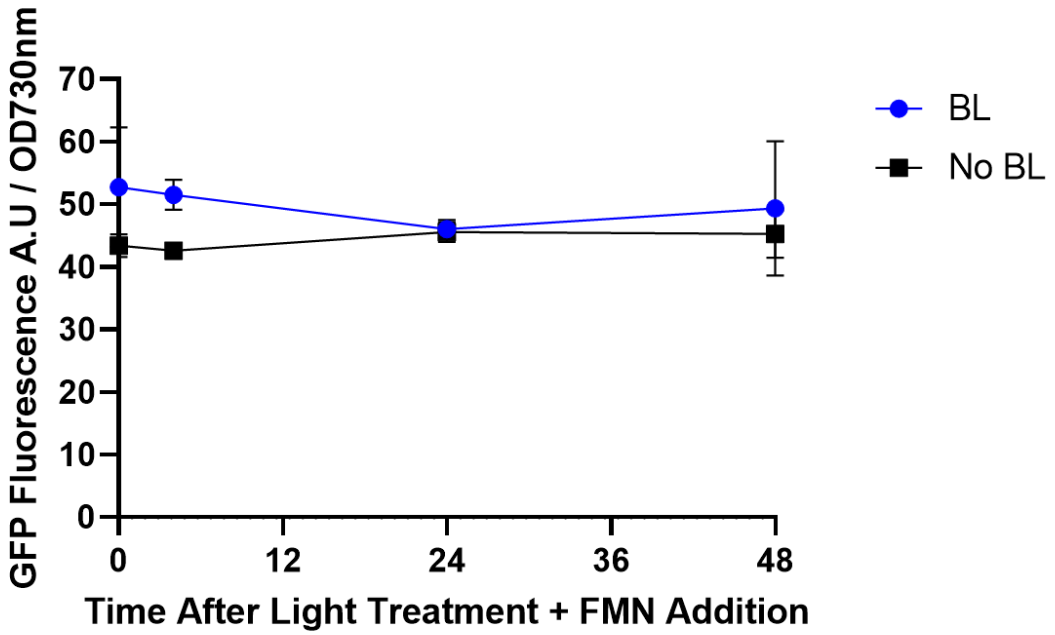


Figure 4-5. YF1 system activity in Syn7002 with the addition of FMN. 150 mL Syn7002 pAQ1BB-Blue cultures expressing GFP under the control of the YF1 system were grown under white LED light to OD_{730} = ~1. Cultures were then either covered with a blue light-removing filter (No BL) (black symbols) or left uncovered (BL) (blue symbols) and supplemented with 1 µM FMN-Na. As the cultures were grown under a 16/8 hour light/dark cycle, the cultures were in the dark between 16-24 hours after the light treatment began. GFP fluorescence was measured, and normalised to OD_{730} and background fluorescence measured in cells expressing empty pAQ1BB vector was subtracted. Data points represent the mean ± S.E.M of three independently grown cultures.

Next, it was hypothesised that the poor performance of the YF1 system could result from the codon optimisation of the YF1 and FixJ genes for expression in Syn7002. For example, a mistake in codon optimisation could result in poor translation of YF1/FixJ protein. To test this, the codon-optimised YF1/FixJ operon was expressed from the constitutive lacIq promoter in *E. coli* with GFP under the control of pFixK2. This construct was assembled in the pUC19 vector and was designated pUC-YF1-GFP. *E. coli* were transformed with pUC-YF1-GFP and grown on selective media. The presence of the correct DNA in selected clones was confirmed by restriction digest analysis and sequencing. 150 mL cultures containing pUC-YF1-GFP were grown to OD_{600} = ~0.3 under white LED light and then were either wrapped in foil (Dark) or left uncovered (Light). GFP fluorescence was measured 3 hours after the respective light treatments. After 3 hours, GFP fluorescence was 6-golf higher in cultures incubated in darkness than the cultures incubated under white light (Figure 4-6). These results confirmed that pUC-YF1-GFP is functional in *E. coli* and that the codon optimisation of the

YF1/FixJ operon had not completely diminished its function. However, this version of the system showed significantly higher overall fluorescence in both light and darkness versus the original pDusk plasmid, whilst also displaying a narrower dynamic range.

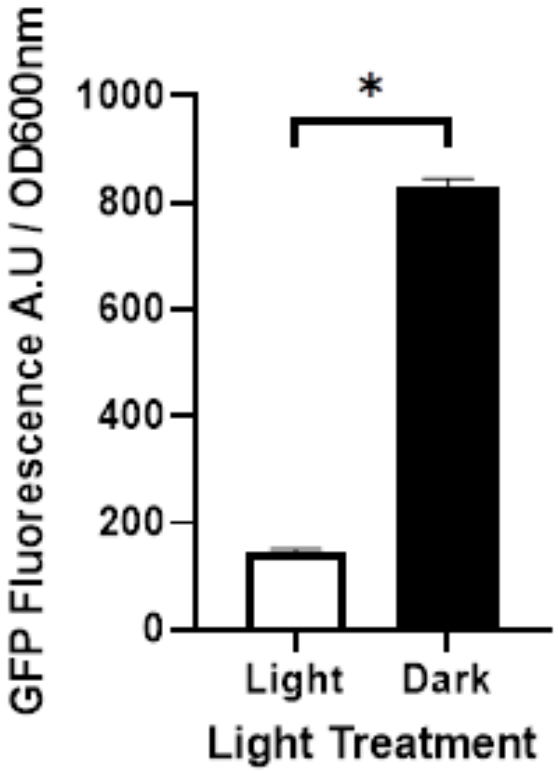


Figure 4-6. Codon optimised YF1 system activity in *E. coli* in white light versus darkness. 150 mL cultures of *E. coli* expressing GFP under control of the YF1 system, where the YF1 and FixJ genes were codon optimised for Syn7002, were grown to OD_{600} = ~0.3 under 100 µmol m^{-2} s^{-1} white LED light. The cultures were then either wrapped in foil (Dark) or left uncovered (Light). Samples were harvested 3 hours after the respective light treatments and immediately placed on ice to stop growth. GFP fluorescence was measured, and normalised to OD_{600} and background fluorescence measured in cells expressing empty pUC19 vector was subtracted. Data are means ± S.E.M of independently grown cultures. Significant differences at $p < 0.05$ as determined by students t-test are shown by asterisk.

Finally, it was hypothesised that the poor performance of the YF1 system could be due to poor YF1 or FixJ translation. The Salis RBS calculator v2.2 was used to calculate a predicted translation initiation rate of 100 A.U at the YF1 and 1000 A.U FixJ start codons (Reis and Salis, 2020). Subsequently, a new YF1 RBS was designed using the RBS calculator with a predicted translation initiation rate of 2000 A.U. This stronger RBS was then inserted directly upstream of YF1 in the pAQ1BB-Blue construct, replacing the weaker RBS. This new construct was designated as pAQ1BB-Blue v2. Syn7002 were transformed with pAQ1BB-Blue v2

and grown on selective media. The presence of the correct DNA in selected clones was confirmed by restriction digest analysis and sequencing. 150 mL Syn7002 cultures expressing pAQ1BB-Blue v2 or empty vector (1BB) were grown to OD_{730} = ~1 under white light. Cultures were then either covered with foil (Dark) or left uncovered (Light) and GFP fluorescence and OD_{730} were measured at several time points across 48 hours.

After 48 hours, GFP fluorescence in pAQ1BB-Blue v2 cultures incubated in darkness was 1.5-fold higher than in pAQ1BB-Blue v2 cultures incubated in light and 3-fold higher than in the pAQ1BB cultures (Figure 4-7). Similarly, the GFP fluorescence in light-treated pAQ1BB-Blue v2 cultures was 2-fold higher than in the pAQ1BB cultures. These differences were deemed statistically significant. While the performance of pAQ1BB-Blue v2 was marginally better than the original pAQ1BB-Blue (Figure 4-3), the data again highlight the poor performance of the YF1 system in Syn7002 and show that increasing the predicted RBS strength of the YF1 gene has minimal effect on the overall system performance.

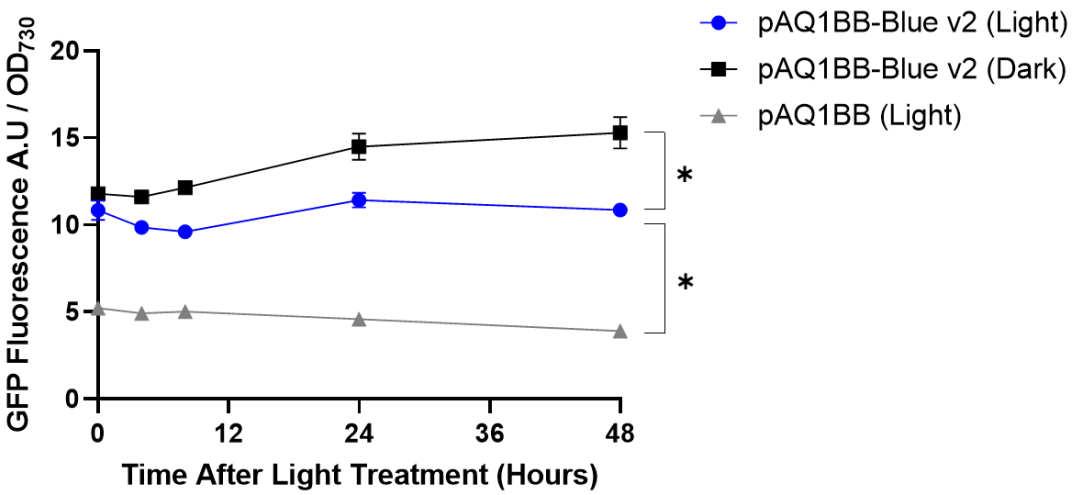


Figure 4-7. YF1-v2 system activity in Syn7002 in blue light and darkness. 150 mL Syn7002 cultures expressing either GFP under control of the YF1 system (pAQ1BB-Blue-v2) (blue and black symbols) or empty pAQ1BB vector (pAQ1BB) (grey symbols) were grown to OD_{730} = ~1 under 150 µmol white LED light. The cultures were then either covered with foil (Dark) or left uncovered (Light). As cultures were grown under a 16/8 hour light/dark cycle, the cultures were in the dark between 16-24 hours after the light treatment began. GFP fluorescence was measured and normalised to OD_{730}. Data are means ± S.E.M of three independent cultures. Significant differences at $p < 0.05$ as determined by unpaired t-test are shown by asterisk.

4.4 Discussion

4.4.1 Possible reasons for the poor performance of YF1/FixJ

Collectively, the results show that the YF1 system demonstrates poor performance when expressed in Syn7002. No significant differences in GFP fluorescence were observed between pAQ1BB-Blue cultures in the presence or absence of blue light, with fluorescence levels barely above those measured in the pAQ1BB empty vector control (Figure 4-3). The slight differences in fluorescence observed between pAQ1BB-Blue/pAQ1BB-Blue-v2 and the pAQ1BB cultures towards the end of the time course may be due to chromatic acclimation rather than increased expression of GFP. The fluorescence spectrophotometer laser excites the sample at 480 nm and measures emission at 514 nm. Therefore, a gradual change in pigment content may occur around this range of the light spectrum. This would have been triggered by the change in light conditions perceived by the cultures which then increases the measured emission at 514 nm.

The addition of the YF1 co-factor had no clear impact on YF1 performance in Syn7002 (Figure 4-5), suggesting FMN limitation was not an issue. However, the FMN supply added to the culture was FMN-Na, as pure FMN is not available to purchase. It appears unlikely that the addition of a sodium atom would affect FMN uptake and use by the cell. Regardless, the Syn7002 genome encodes the RibF gene which converts riboflavin to FMN, so it is unlikely that they would be unable to endogenously produce FMN.

When the YF1 system was expressed in *E. coli*, a 46-fold increase in fluorescence was observed in the dark-incubated cultures versus the light-incubated cultures (Figure 4-1), confirming that the system was functional in *E. coli* and showed similar blue light sensitivity as previously reported (70-fold increase, Möglich *et al.*, 2009). The cyanobacterial codon-optimised version of the system was also expressed in *E. coli*. While the tool was still functional, GFP expression in both the light and dark states was several-fold higher than with the original pDusk plasmid. The only two differences between these constructs were the codon optimisation of the YF1 and FixJ genes and the backbone vector used. pDusk is a low copy number plasmid whereas the pUC-YF1-GFP plasmid expressing the codon optimised version is a high copy number plasmid. This demonstrates how the performance of an optogenetic tool can be significantly affected by the expression levels of the photoreceptor and transcription factor genes.

RT-PCR analysis confirmed that both the YF1 and FixJ genes were expressed in Syn7002 (Figure 4-4). In the future, it may be beneficial to express YF1/FixJ from different promoters of varying strengths to further troubleshoot performance. As the RT-PCR used is only semi-quantitative, expression levels of YF1 and FixJ were not quantified. Nevertheless, the results indicate that the genes were reasonably well-expressed as the gel bands are only slightly fainter than those of 16S rRNA which is very highly expressed.

The poor performance of the YF1 system could also be due to a lack of translation, production of misfolded or insoluble YF1/FixJ protein or rapid degradation. Increasing the predicted RBS strength of the YF1 gene 20-fold only had a marginally positive effect on system performance (Figure 4-7). For context, the translation initiation rate of GFP in the 1BB-Green construct used in results chapter one was around 5000 A.U, whilst the newly designed YF1/FixJ

RBS was 2000 A.U. Future work could generate tagged YF1 and FixJ versions for Western Blot analysis of proteins from Syn7002 cell lysate. Various tags exist for this purpose with commercial antibodies available such as 6xHis, FLAG or c-myc tags. In the context of the bicistronic operon used to express YF1 and FixJ in this work, tagging at the C-terminus of YF1 and the N-terminus of FixJ would be difficult without disrupting this bicistronicity. However, there is no obvious reason why expressing each of these genes from a separate promoter would be an issue. Also, using separate promoters could be useful for system optimisation by altering the balance of expression levels of the two genes.

Another potential issue could be that in Syn7002 YF1 is unable to phosphorylate FixJ, preventing system activation in response to darkness. Future work could use tagged YF1 and FixJ proteins extracted from Syn7002 in *in vitro* radiolabelled phosphorylation assays to measure FixJ phosphorylation. However, success in this assay would not rule out the possibility that YF1 is unable to phosphorylate FixJ due to interference by the Syn7002 intracellular environment. It is possible that YF1 can phosphorylate FixJ *in vivo*, but that an unknown endogenous protein phosphatase dephosphorylates FixJ at a much faster rate.

If functional YF1 and FixJ proteins are being expressed in Syn7002 and YF1 can phosphorylate FixJ, FixJ may be unable to bind to pFixK2 and initiate transcription. It is conceivable that an unknown endogenous repressor protein binds to pFixK2 in Syn7002, blocking access by FixJ. Alternatively, a repressor protein or allosteric inhibitor may bind directly to FixJ, preventing its interaction with pFixK2. It is also possible that an additional unknown protein, such as a transcriptional activator, is necessary for the initiation of transcription from pFixK2. This protein could be present in species in which the YF1 system has been successfully deployed such as *E. coli* and *V. natriegens* but absent from Syn7002. Future experiments could use chromatin immunoprecipitation coupled with mass spectrometry on samples from both *E. coli* and Syn7002 to investigate whether FixJ binds to pFixK2 in Syn7002. This would also identify any additional transcriptional regulators that bind to pFixK2 in *E. coli* which are absent in Syn7002.

Another possibility is that, in the context of Syn7002, the YF1 LOV domain is non-functional, preventing YF1 from responding to changes in light irradiation. Several cyanobacteria species have been shown to contain functional LOV domains including *Anabaena* sp. PCC 7120 and *N. punctiforme* (Narikawa *et al.*, 2006). However, it is unclear whether this is also true in Syn7002. A protein BLAST search using the LOV domain present in YF1 against the Syn7002 genome returned a single hit. This protein was annotated as "sensor domain-containing diguanylate cyclase" with a 43% sequence match to the YF1 LOV domain. This diguanylate cyclase contains key sub-motifs which are required for blue Light sensitivity such as FXXXT(G/E)Y and N(Y/F)XXX(G/D)XX(F/L)XN (Zayner *et al.*, 2013). However, the diguanylate cyclase contains a single residue substitution within the conserved GX(N/D)C(R/H)(F/I)L(Q/A) sub-motif (Figure 4-8). It is unclear whether this would affect LOV domain function, especially when the highly conserved cysteine required for adduct formation is present.

Score	Expect	Method	Identities	Positives	Gaps
107 bits(267)	4e-29	Composition-based stats.	51/119(43%)	79/119(66%)	6/119(5%)

```
Query  13   LEVIKKALDHVRVGVVITDPALEDNPIVYVNQGFVQMTGYETEEILGKNCRFLQGKHTDP  72
            + ++ ++++ +R GVVITD   E+NPI+YVNQGF ++TGY  +EILG+NC FLQG+
Sbjct  359  MNLLHRSINVIRQGVVITDAREENNPIIYVNQGFEKITGYTADEILGRNCNFLQGRDRRQ  418

Query  73   AEVDNIRTALQNKEPVTVQIQNYKKDGTMFWNELNIDPMEIEDK----TYFVGIQNDIT  127
               +  +R A+   +   V ++NY+K+G  FWN L+I P  + D     T+F+GIQ DI+
Sbjct  419  PPLTELRAAILKGQECNVTLRNYRKNGEPFWNALHIFP--VRDPAGYLTHFIGIQTDIS  475
```

Figure 4-8. Sequence similarity between YF1 LOV and a Syn7002 diguanylate cyclase. A protein BLAST search using the LOV domain (residues 1-127) of the YF1 protein as a query identified a potential LOV domain-containing protein encoded by the Syn7002 genome, a sensor domain-containing diguanylate cyclase. Sub-motifs commonly conserved in LOV domains are underlined, with those in red representing a complete match to a conserved LOV motif and in blue a single residue mismatch to a conserved motif.

A summary of some of the possible mechanisms which could be causing the poor performance of the YF1 system in Syn7002 is shown below (Figure 4-9).

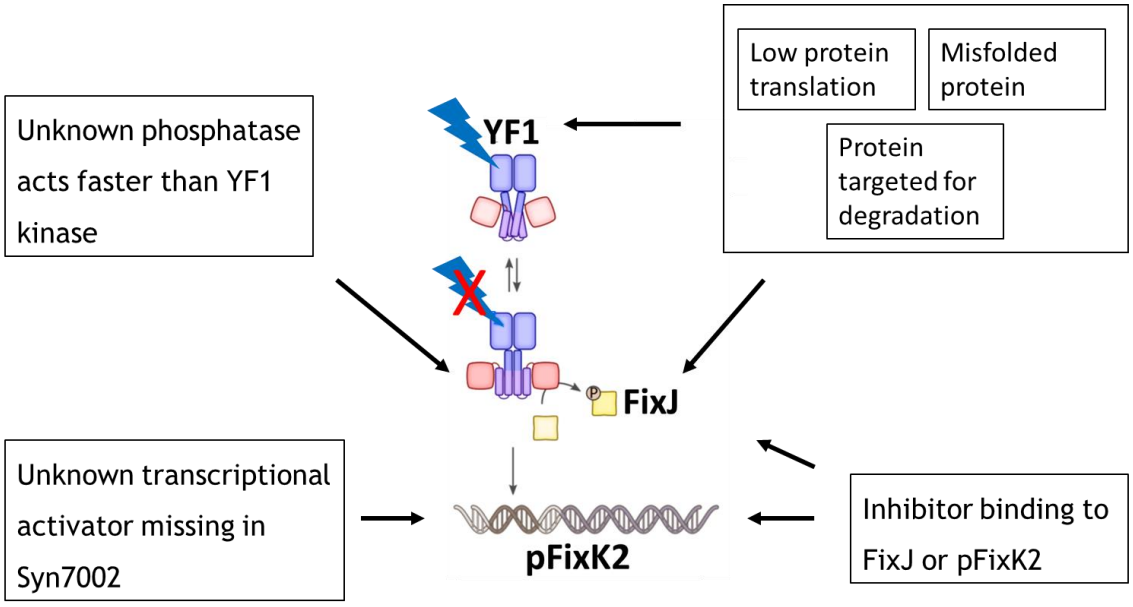


Figure 4-9. Possible issues causing the poor performance of the YF1 system in Syn7002.

5 Applying optogenetics for product synthesis

5.1 Introduction

5.1.1 Using optogenetics for L-alanine and mannitol production

The functionality of the CcaS system in Syn7002 was shown using GFP as a readout (Chapter 3). The next step was to test whether the optogenetic tool could be used for producing industrially relevant products via metabolic engineering. The two products chosen for this were L-alanine and mannitol (see Chapter 1 Introduction). L-alanine was the initial product upon which this chapter was focused. This choice was based upon the goal of demonstrating green/red light regulated production as proof-of-concept using a simple one-step enzymatic reaction with a ubiquitous precursor and a simple product quantitation assay. The simple one-step reaction converting pyruvate and ammonia into L-alanine using the alaD enzyme and measuring L-alanine concentration with a commercially available kit and spectrophotometer seemed like a strong candidate. However, upon failure to detect L-alanine production even after substantial troubleshooting, we pivoted to mannitol production as an alternative. Mannitol production using the mpusfus enzyme with a constitutive promoter had previously been demonstrated in Syn7002 and thus we were confident of detecting mannitol production using the CcaS system instead (Madsen *et al.*, 2018). Using optogenetics as opposed to a strong constitutive promoter like pCpcB594 to control mpusfus expression could potentially lead to increased mannitol titres. The expression of mpusfus reduces the growth rate of Syn7002 (Madsen *et al.*, 2018). The CcaS system could be used to toggle ON/OFF mpusfus expression with different light regimens and thus potentially better balance mannitol synthesis and culture growth to achieve maximal mannitol titres.

5.2 Chapter aims

5.2.1 Engineering Syn7002 to produce L-alanine and mannitol using the CcaS system

The main aim of this chapter was to use the optimised version of the CcaS system to enable red/green light-regulated production of L-alanine and

mannitol, products with significant commercial applications. The pUC-172-10 construct generated in Chapter 3 was used instead of the original CcaS system due to its improved dynamic range and significantly higher transcriptional output under green light. The L-alanine and mannitol biosynthetic genes were inserted into the pUC-172-10 plasmid directly downstream of the optimised pCpcG2 promoter, replacing GFP. The resulting constructs were designated as pUC-172-10-alaDE for L-alanine synthesis and pUC-172-10-mpusfus for mannitol synthesis. Each construct was expressed separately in Syn7002 and L-alanine and mannitol production was measured in successful transformants. After failing to detect L-alanine production in a strain expressing pUC-172-10-alaDE, Western Blotting was used to confirm the production of the necessary proteins under both red and white light. Mannitol production in a strain expressing pUC-172-10-mpusfus was measured under a variety of relevant lighting conditions and compared to that of a strain which expresses mpusfus from the strong constitutive promoter pCpcB594.

5.3 Results

5.3.1 Failure to detect L-alanine synthesis in Syn7002

To engineer L-alanine production in Syn7002, alanine dehydrogenase (*alaD*) from *Geobacillus stearothermophilus* was codon optimised, *de novo* synthesised and inserted into pUC-172-10, replacing GFP and thus placing it under the control of the optimised pCpcG2 promoter. The L-alanine exporter alaE from *E. coli* was also codon optimised and inserted into pUC-172-10 under the control of the medium-strength constitutive promoter J23110 (Markley *et al.*, 2015). In *E. coli*, native expression of *alaE* enables the secretion of L-alanine into the cell culture media, preventing potentially toxic excess intracellular accumulation of L-alanine (Katsube *et al.*, 2019). However, the function of alaE in Syn7002 has not been previously reported. The corresponding construct was named pUC-172-10-alaDE (Figure 5-1). Another construct named pUC-cpc-alaDE was also generated which enabled the expression of alaD from the strong constitutive promoter pCpct223, as well as the expression of alaE from the J23110 promoter (Markley *et al.*, 2015). Both constructs facilitated integration into the endogenous pAQ1 plasmid as described in previous chapters. L-alanine production and OD_{730} were

measured in strains separately expressing each construct to compare the effect of optogenetic control on L-alanine synthesis versus constitutive expression.

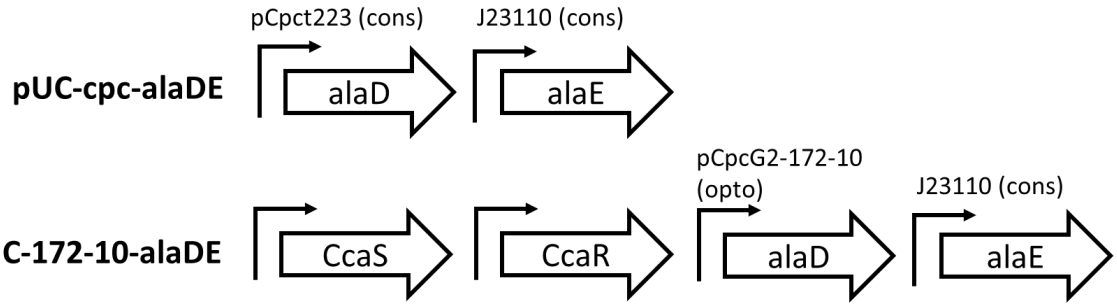


Figure 5-1. Construct design for L-alanine production in Syn7002. pCpct223 and J23110 are constitutive (cons) promoters (Markley *et al.*, 2015) and pCpcG2-172-10 is an optogenetic (opto) controlled promoter (see Chapter 1).

150 mL cultures expressing pUC-172-10-alaDE or pUC-cpc-alaDE were grown under white light using a 16h/8h light/dark photoperiod for 35 days. White light was used as opposed to green as white light facilitated similar levels of CcaS activation as green light without resulting in the slow growth of Syn7002 under green light observed previously (Figure 3-4 and Figure 3-5). The cultures were also supplemented with ammonium chloride to a final concentration of 0.2 g L^{-1} to ensure sufficient ammonia precursor supply for alaD to convert pyruvate to L-alanine. This concentration of ammonium chloride was chosen as it was the maximum concentration which did not significantly inhibit culture growth (data not shown). Samples were harvested at weekly intervals and the L-alanine concentration was measured in both lysed cell pellets and cell culture supernatants. However, no L-alanine could be detected in any of the samples when using an L-alanine assay kit (see Materials and Methods), regardless of ammonium chloride concentration, genotype or light conditions. The kit was confirmed as functional as L-alanine could be easily detected in cell culture supernatant samples spiked with L-alanine standards. The assay kit has a reported lower limit sensitivity of 1.7 mg L^{-1} L-alanine.

To troubleshoot the L-alanine production, the pUC-172-10-alaDE construct was modified to add an N-terminal 6x-His tag to alaD (alaD-His) and a C-terminal HA-tag to alaE (alaE-HA), resulting in the construct pUC-172-10-alaDEHAHIS. Wild-type cultures and a strain expressing pUC-172-10-alaDEHAHIS were grown under white light for 10 days before samples were harvested and pelleted by centrifugation. The cells were then lysed and the proteins were extracted and

separated into separate soluble and crude membrane fractions by centrifugation. The fractions were then assayed by Western Blot using anti-HA and anti-His antibodies to detect the alaD-His and alaE-HA proteins.

The alaD-His protein has a predicted molecular weight of 45 kDa, whilst alaE-HA has a predicted molecular weight of 18 kDa. Expression of alaD was detected in both the soluble and crude membrane fractions of cell lysate from cultures expressing pUC-172-10-alaDEHAHIS (Figure 5-2a). alaE expression was also detected in the crude membrane fraction of cultures expressing pUC-172-10-alaDEHAHIS (Figure 5-2b). No bands of the expected sizes were observed in the wild-type control cell lysate samples in either blot. Additional bands of other sizes were also present on the anti-HA blot, which are most likely due to non-specific binding of antibody or background luminescence due to the long image exposure time of 80 seconds. The bands observed at 10 kDa in the soluble fraction lanes of the anti-HA blot are most likely chlorophyll pigments, which were visible by eye as green bands on the Western Blot membrane. The results show that the lack of L-alanine production in cultures expressing pUC-172-10-alaDE is not due to a lack of alaD or alaE protein expression.

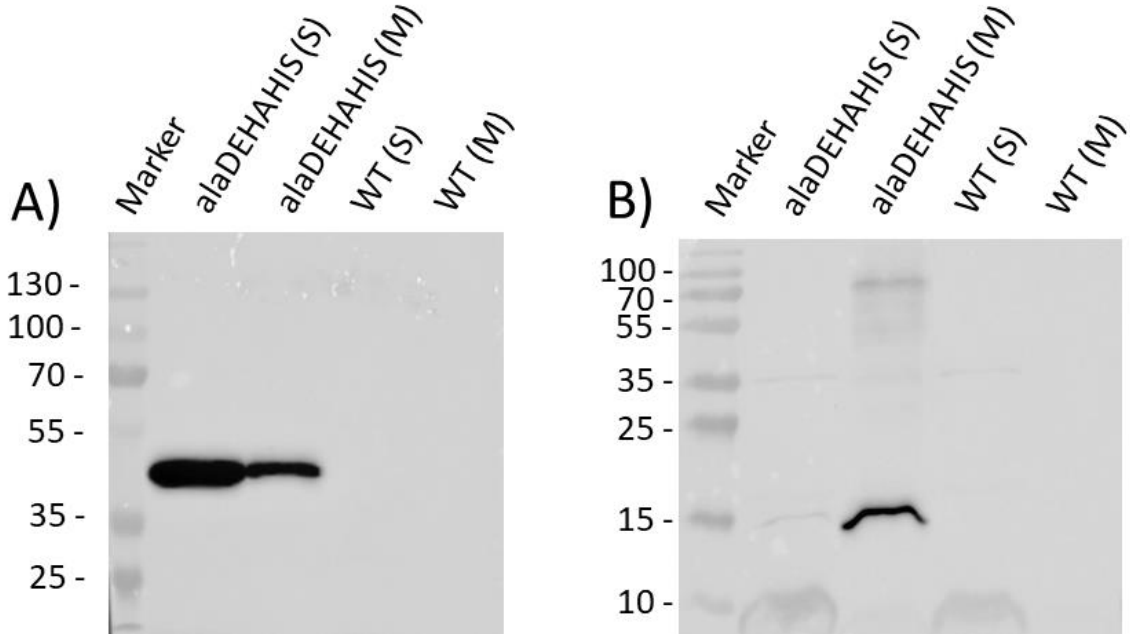
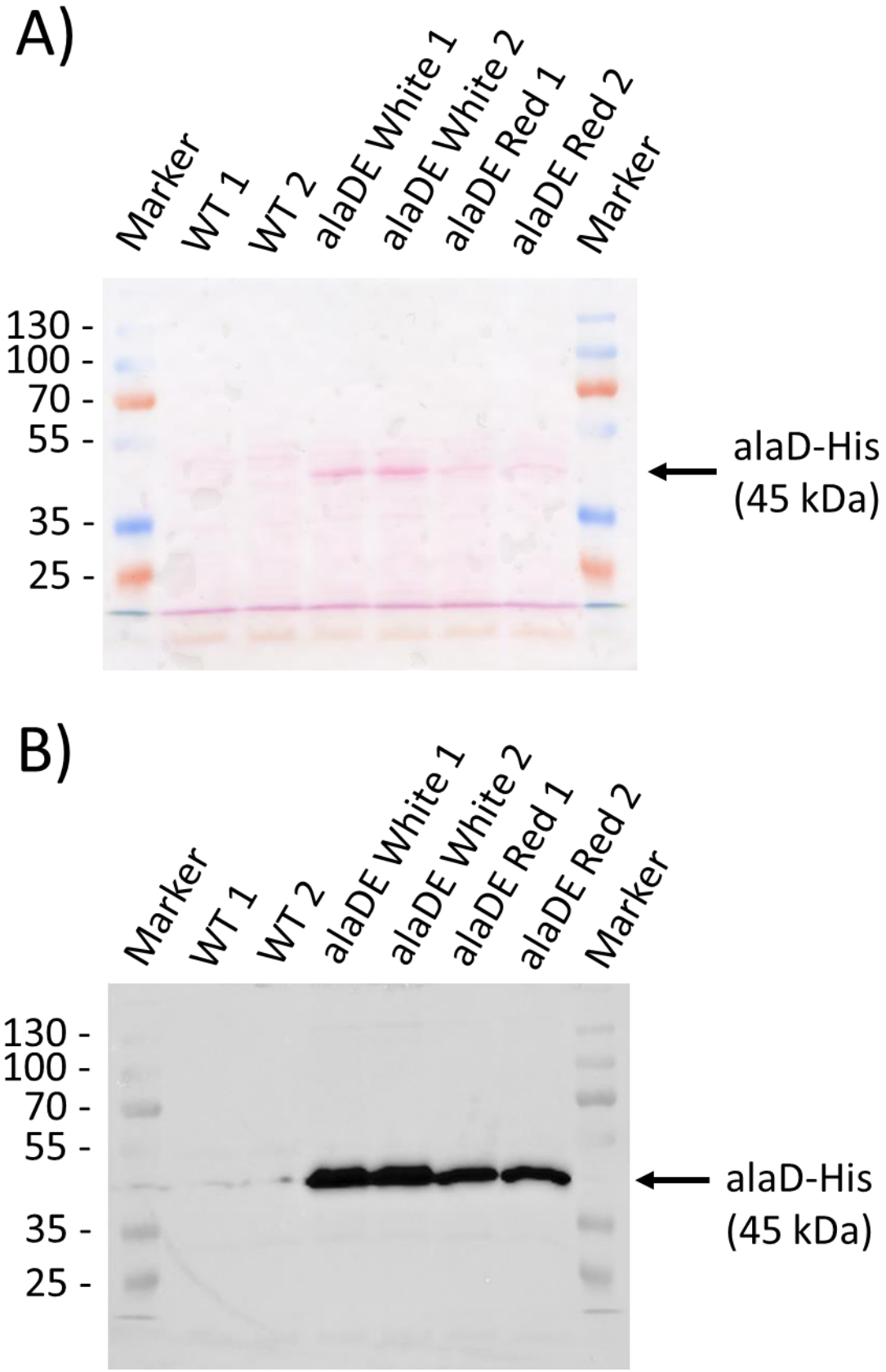


Figure 5-2. Expression of alaD-His and alaE-HA proteins in Syn7002. Western blot analysis of cell lysates from wild-type (WT) Syn7002 cultures and cultures expressing pUC-172-10-alaDEHAHIS. **A)** alaD-His probed with anti-His antibody using an exposure time of 3 seconds. **B)** alaE-HA probed with anti-HA antibody using an exposure time of 80 seconds. Each lane is labelled with the strain genotype and whether the soluble (S) or membrane (M) fraction was used. The numbers to the left of each blot indicate the size (kDa) of ladder proteins.

After confirming alaD and alaE expression in the pUC-172-10-alaDEHAHIS strain, the effect of red versus green light on alaD protein expression was compared. Wild-type cultures and a culture expressing pUC-172-10-alaDEHAHIS were grown under either red or white light for 10 days before cell samples were harvested and pelleted by centrifugation. The cells were then lysed, the proteins were extracted and the cell lysate samples were assayed by SDS-PAGE and Western Blot using anti-His antibodies.


A)
Marker
WT 1
WT 2
alaDE White 1
alaDE White 2
alaDE Red 1
alaDE Red 2
Marker
130 -
100 -
70 -
55 -
35 -
25 -
alaD-His
(45 kDa)
B)
Marker
WT 1
WT 2
alaDE White 1
alaDE White 2
alaDE Red 1
alaDE Red 2
Marker
130 -
100 -
70 -
55 -
35 -
25 -
alaD-His
(45 kDa)

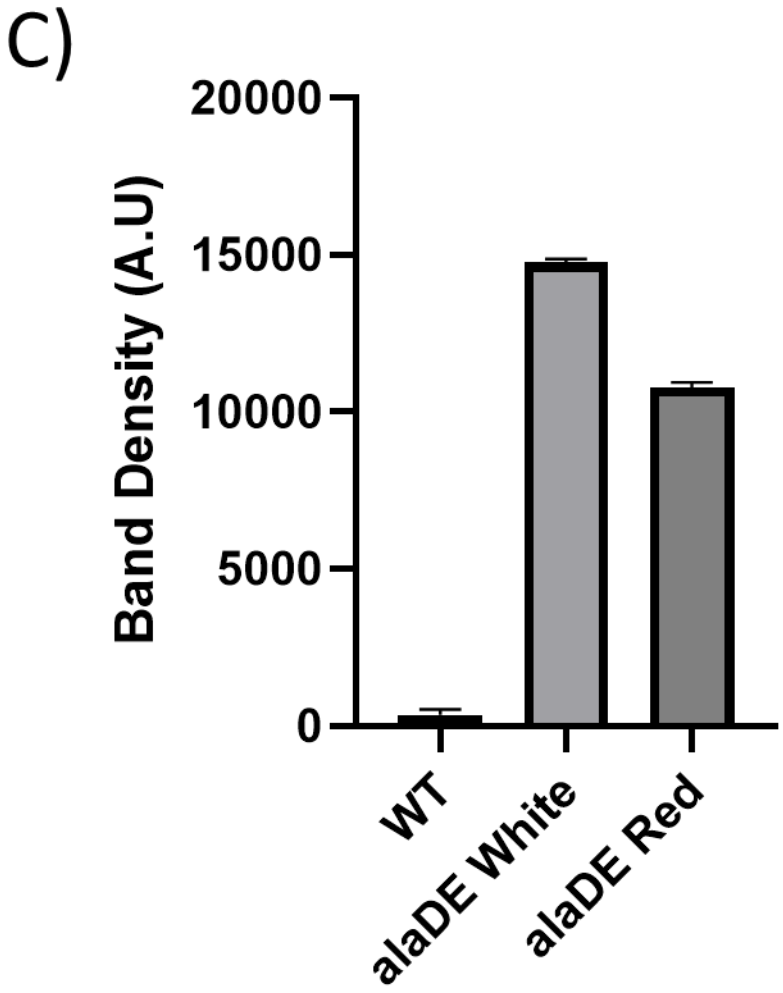


Figure 5-3. Expression of alaD protein in red versus white light. A) Ponceau S-stained nitrocellulose membrane after SDS-PAGE. **B)** Western blot analysis of cell lysates from wild-type (WT) Syn7002 cultures and those expressing pUC-172-10-alaDEHAHIS. **C)** Densitometry of bands post Western blot measured using ImageJ (mean ± S.E.M of 2 bands per condition). alaD-His probed with anti-His antibody using an exposure time of 2 seconds. Each lane is labelled with the genotype, replicate number and light treatment where relevant. Numbers on the left indicate the size (kDa) of proteins in the ladder. 20 µg total protein was loaded per lane.

In cultures expressing pUC-172-10-alaDEHAHIS, alaD protein bands were stronger in the white light samples than in the red light samples (Figure 5-3a and Figure 5-3b) indicating that alaD expression was increased by green light. Some protein was also detectable under red light, highlighting the previously observed 'leakiness' of the CcaS system under red light. A faint band of the expected size of alaD was also detected in the wild-type samples. However, this band is also present in the marker lane, suggesting a background signal or bleed-through from the other lanes. Ponceau S staining of the nitrocellulose membrane before antibody incubation shows even amounts of protein loaded per lane (Figure 5-

5a). This data confirms that the CcaS optogenetic system allows a wavelength-dependent increase of alaD protein expression in white light versus red light. Whether enzyme activity is inhibited in the cellular environment of Syn7002 or whether L-alanine produced is immediately transformed into other metabolites needs now to be investigated (see discussion).

5.3.2 Mannitol production in Syn7002

To engineer mannitol production in Syn7002, codon optimised mpusfus from *M. pusilla* was PCR amplified from the pAQ1BB-pCpcB594-mpusfus plasmid and inserted into pUC-172-10, replacing GFP and thus placing it under the control of the optimised pCpcG2 promoter. The corresponding construct was named pUC-172-10-mpusfus (Figure 5-4). pAQ1BB-pCpcB594-mpusfus was also used to compare the effect of optogenetic control of mpusfus expression versus constitutive expression of mpusfus on mannitol production.

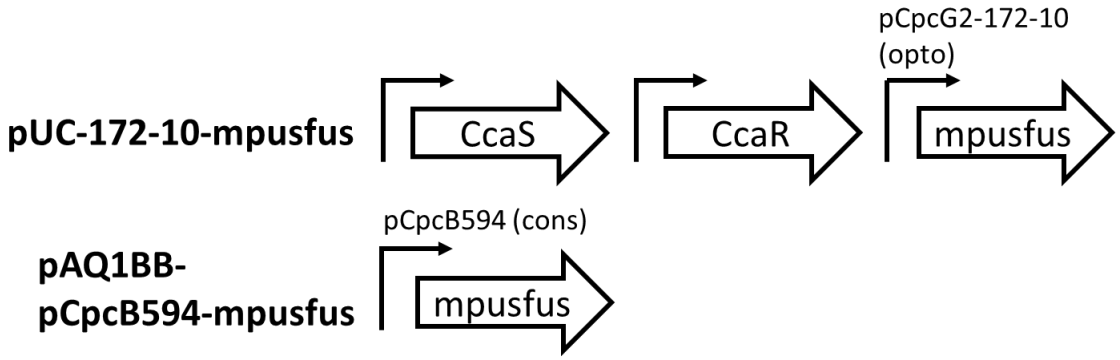


Figure 5-4. Construct design for mannitol production Syn7002 strains. pCpcB594 is a strong constitutive (cons) promoter and pCpcG2-172-10 is an optogenetic (opto) controlled promoter.

Fully segregated homozygous strains were unable to be achieved when using either pAQ1BB-pCpcB594-mpusfus or pUC-172-10-mpusfus, but spectinomycin was included in cultures to maintain selective pressure. 100 mL cultures of wild-type Syn7002 and Syn7002 expressing pUC-172-10-mpusfus or pAQ1BB-pCpcB594-mpusfus were grown under various light regimens using a 16h/8h light/dark photoperiod. These consisted of either red light for 28 days, white light for 28 days, red light for 14 days then white light for 14 days or red light for 14 days and alternating 48-hour periods of red or white light for the remaining 14 days. At weekly intervals, samples were harvested to measure OD_{730} and the extracellular mannitol concentration using a mannitol assay kit (see Materials and Methods). For the first 14 days, cultures expressing pUC-172-10-mpusfus

grew at a similar rate to wild-type cultures regardless of the light regimen used (Figure 5-5a). However, wild-type cultures continued to grow to an average OD_{730} = 20-21 by day 28, with the differences between wild-type and all other cultures being statistically significant at day 21 and day 28. Cultures expressing pUC-172-10-mpusfus reached a maximal OD_{730} = 10.5-12.5 between day 14 and day 21. Cultures expressing pAQ1BB-pCpcB594-mpusfus grew more slowly than both those expressing pUC-172-10-mpusfus and wild-type, whilst also reaching a lower maximal average OD_{730} = 8.5 (Figure 5-5a). Whilst the differences in OD_{730} between cultures expressing pAQ1BB-pCpcB594-mpusfus and pUC-172-10-mpusfus were statistically significant at day 7 and day 14, the differences at day 21 and 28 were not statistically significant (Figure 5-5b). No statistically significant differences in OD_{730} were observed at any timepoints between the cultures expressing pUC-172-10-mpusfus across the different light regimens applied.

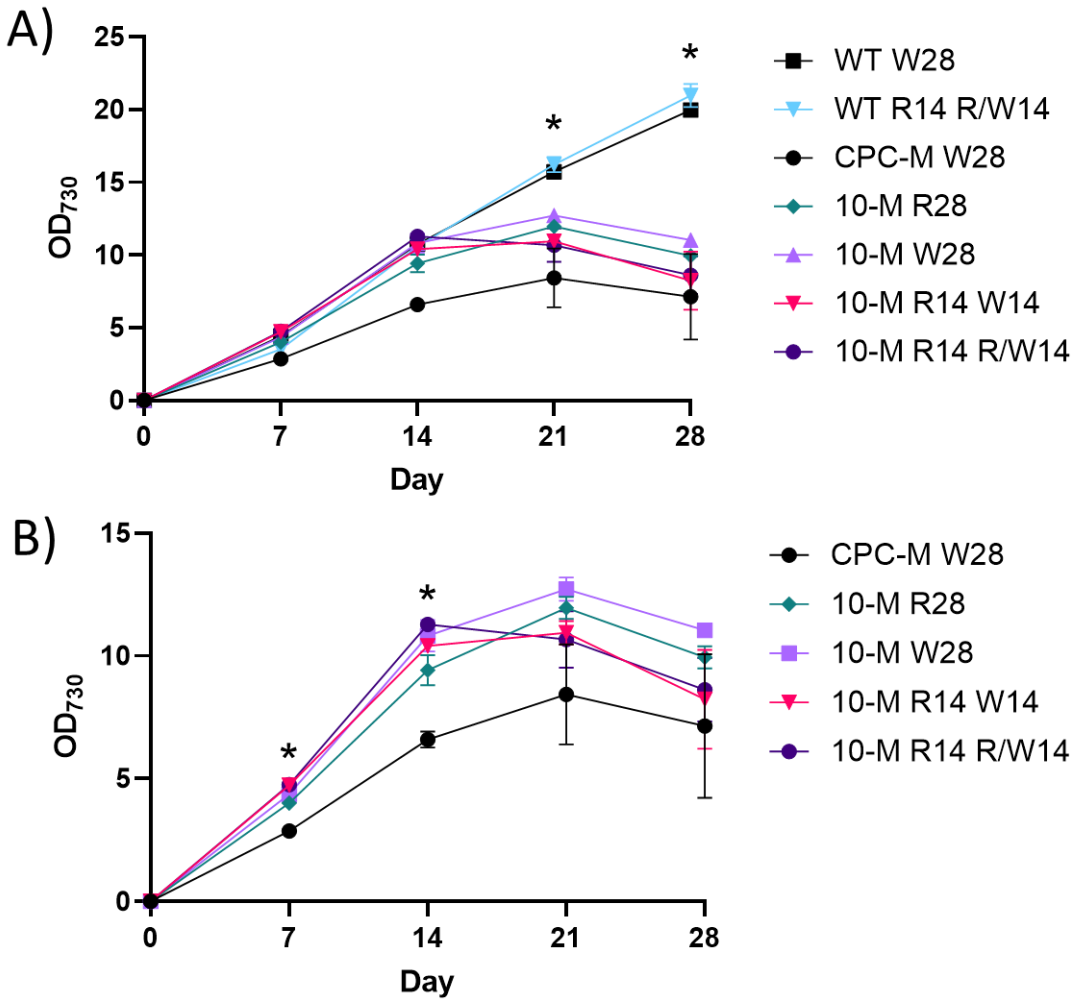


Figure 5-5. Growth of mannitol-producing Syn7002 strains. Culture density (OD_{730}) of wild-type Syn7002 (half-filled symbols), Syn7002 expressing pAQ1BB-pCpcB594-mpusfus (CPC-M, filled symbols) and Syn7002 expressing pUC-172-10-mpusfus (10-M, open symbols). **A)** Culture density of all genotypes and **B)** Culture density minus wild-type. 100 mL cultures were grown in 250 mL Bijou bottles under illumination with 100 µmol m^{-2} s^{-1} red light as provided by white LED light with red light-transmitting filters or 100 µmol m^{-2} s^{-1} white LED light. Cultures designated as "R28" or "W28" were grown under red or white light respectively for 28 days. Cultures designated as "R14 W14" were grown under red light until day 14 when illumination was switched to white light. Cultures designated as "R14 R/W14" were grown under red light until day 14 when illumination was switched to 48-hour alternating periods of red or white light. All cultures were grown under 16h/8h light/dark cycles. Data are means ± S.E.M of three independent cultures. Asterisks represent significance at $p < 0.05$ as determined using one-way ANOVA with Tukey's multiple comparison test.

Next, the extracellular mannitol concentration was measured in samples harvested from cultures expressing pUC-172-10-mpusfus and pAQ1BB-pCpcB594-mpusfus. Similar final mannitol titres were measured in cultures expressing pUC-172-10-mpusfus and pAQ1BB-pCpcB594-mpusfus (Figure 5-8a). Average mannitol

concentrations in samples harvested at day 7 were 0.034 g L^{-1} in cultures expressing pAQ1BB-pCpcB594-mpusfus and between 0.005-0.011 g L^{-1} in cultures expressing pUC-172-10-mpusfus. At day 14, mannitol concentrations had increased to 0.97 g L^{-1}, which was 4-5 times higher than in cultures expressing pUC-172-10-mpusfus. Statistically significant differences in mannitol titre were observed between cultures expressing pAQ1BB-pCpcB594-mpusfus and those expressing pUC-172-10-mpusfus at day 7 and day 14 but not at day 21 or day 28. No significant differences were detected when limiting the statistical analysis to only the cultures expressing the optogenetic construct at day 14 (10-M). Significant variation in mannitol concentrations was observed at day 21 and day 28 across both genotypes and light regimens. No statistically significant differences were detected between cultures subjected to any of the four light regimens at any of the time points. The largest increase in mannitol titre in cultures expressing pAQ1BB-pCpcB594-mpusfus was between days 7-14 whereas for those expressing pUC-172-10-mpusfus, it was between days 14-21.

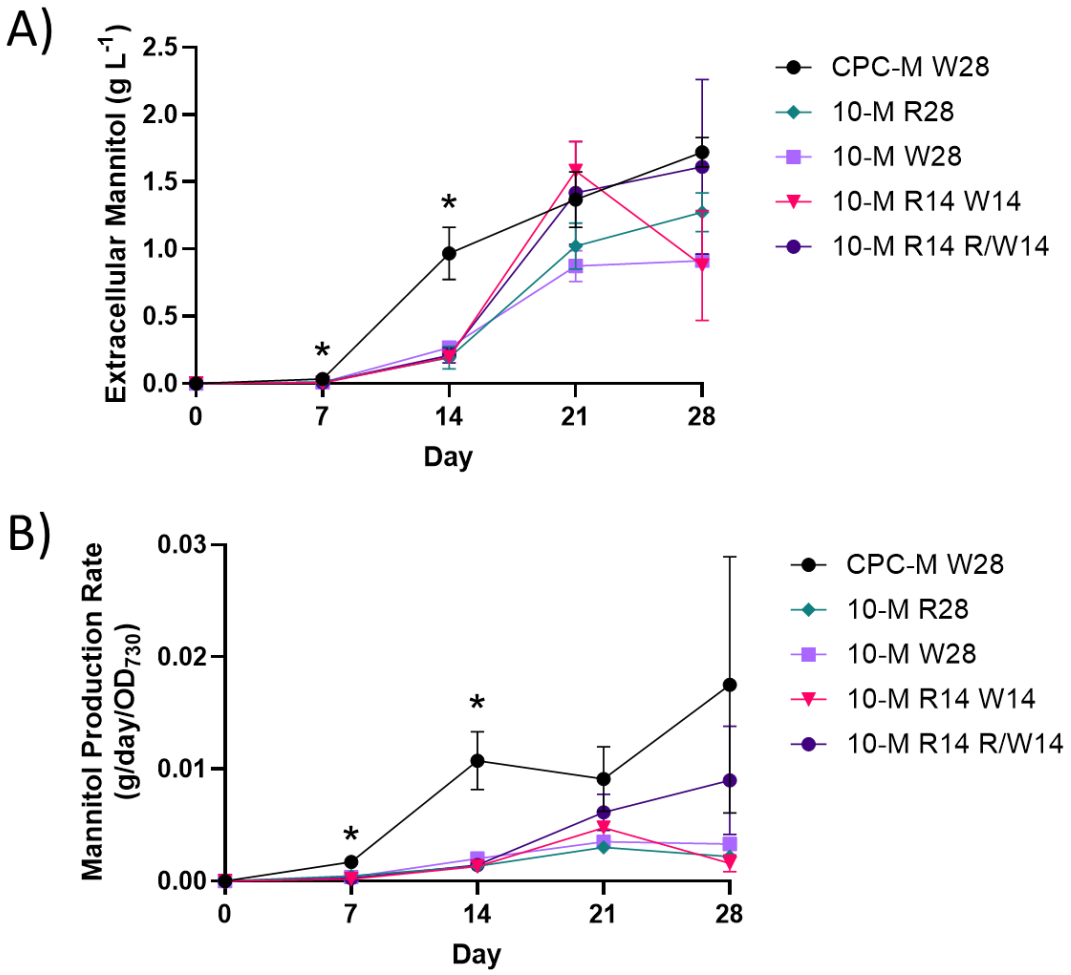


Figure 5-6. Mannitol titres of mannitol-producing Syn7002 strains. A) Extracellular mannitol titre and **B)** mannitol production rate of wild-type Syn7002 (half-filled symbols), Syn7002 expressing pAQ1BB-pCpcB594-mpusfus (CPC-M, filled symbols) and Syn7002 expressing pUC-172-10-mpusfus (10-M, open symbols). 100 mL cultures were grown in 250 mL Bijou bottles under illumination with 100 µmol m^{-2} s^{-1} red light as provided by white LED light with red light-transmitting filters or 100 µmol m^{-2} s^{-1} white LED light. Cultures designated as "R28" or "W28" were grown under red or white light respectively for 28 days. Cultures designated as "R14 W14" were grown under red light until day 14 when illumination was switched to white light. Cultures designated as "R14 R/W14" were grown under red light until day 14 when illumination was switched to 48-hour alternating periods of red or white light. All cultures were grown under 16h/8h light/dark cycles. Data are means ± S.E.M of three independent cultures. Asterisks represent significance at $p < 0.05$ between CPC-M and all 10-M cultures as determined using one-way ANOVA with Tukey's multiple comparison test.

Mannitol production rate, as calculated by dividing the mannitol titre by both the number of days of culture growth and normalising to OD730, showed a similar pattern to the mannitol titre data alone. Cultures expressing pAQ1BB-pCpcB594-mpusfus demonstrated a higher mannitol production rate on day 7 and day 14 versus cultures expressing pUC-172-10-mpusfus (Figure 5-6b). However,

the differences on day 21 and day 28 were not deemed statistically significant. The very large standard error shown in mannitol production rate in cultures expressing pAQ1BB-pCpcB594-mpusfus at day 28 is due to the cultures having entered the death phase. Mean mannitol titres are shown in Table 6-1.

Table 5-1. Mean mannitol titre measured in each culture condition.

Condition/Genotype	Mannitol Titre at Day 28 (Mean ± SD)
CPC-M W28	1.72 ± 0.19
10-M R28	1.27 ± 0.25
10-M W28	0.91 ± 0.06
10-M R14 W14	0.88 ± 0.71
10-M R14 R/W14	1.61 ± 1.12

5.4 Discussion

5.4.1 Failure to detect heterologous L-alanine production

The white light-dependent increase of protein levels of metabolic enzyme alaD (Figure 5-3) proved that the CcaS system can be used for wavelength-dependent heterologous protein expression. The failure to detect any L-alanine is likely to reflect typical issues with metabolic engineering. One possibility is that either of the proteins may be non-functional in the intracellular context of Syn7002. Future work could increase the expression of the native Syn7002 alanine dehydrogenase (alD). This would circumvent uncertainty around the function of the heterologous alaD used in this work. However, the protein might still be subject to endogenous negative feedback regulation. An alternative could be the use of alaD from *Bacillus sphaericus*, which has been used successfully for L-alanine production in *E. coli, C. glutamicum, Zymomonas mobilis* and *Lactococcus lactis* (Uhlenbusch *et al.*, 1991; Lee *et al.*, 2004; Wardani *et al.*, 2007; Jojimi *et al.*, 2010).

Another possibility explaining the lack of L-alanine production is that synthesised L-alanine is immediately consumed by other metabolic pathways. Knockout of

genes that either consume L-alanine directly or direct carbon flux away from pyruvate could increase L-alanine production. The Syn7002 genome encodes an alanine:glyoxylate aminotransferase which converts L-alanine to pyruvate. This enzyme is unlikely to be essential for Syn7002 survival and could be a target for knockout or knockdown. Furthermore, Syn7002 contains an alanine racemase (alrA), which converts L-alanine to D-alanine. However, D-alanine is an important constituent of the bacterial cell wall and knockout could prove lethal. Similarly, knockout out of TCA-cycle enzymes which consume pyruvate are likely to have severe detrimental effects on cell growth and metabolism. A successful example of this type of approach in Syn7002 describes the production of citric acid (Zhang *et al.*, 2022). Overexpression of citrate synthase did not increase citric acid production above the level observed in wild-type cultures. However, combining citrate synthase overexpression with knockout of genes in competing metabolic pathways enabled a 100-fold increase in citric acid production versus wild-type.

Another possibility is an issue with L-alanine export out of the cell. In *E. coli*, alaE is located within the inner cytoplasmic membrane, exporting L-alanine to the periplasm (Kim *et al.*, 2017). L-alanine is then exported across the outer plasma membrane by the action of an unknown pore protein or exits by diffusion through the plasma membrane. Syn7002 contains a thick cellulose layer between the cytoplasmic and outer membranes which could potentially prevent L-alanine from diffusing out of the cell (Zhao *et al.*, 2015). The L-alanine could also be trapped in the glycocalyx, the polysaccharide layer which surrounds the outer cell membrane. As this membrane consists of mostly cellulose in Syn7002, cellulase digest could release potentially bound L-alanine (Zhao *et al.*, 2015).

It is possible that the L-alanine assay kit used to quantify L-alanine concentrations was not sensitive enough to detect low quantities of L-alanine. However, the lower sensitivity range of the assay would equate to a product titre of 1.7 mg L^{-1}. For context, previous work engineering overexpression and secretion of amino acids in cyanobacteria resulted in product titres of 400 mg L^{-1} of L Lysine in Syn7002 and 560 mg L^{-1} of L Lysine in Syn2973 (Korosh *et al.*, 2017; Dookeran and Nielsen, 2021). It is also possible that intracellular L-alanine could not be detected due to the cell lysis protocol resulting in L-alanine degradation.

Due to the interference of native cellular proteins and pigments with the assay, the extracted intracellular fractions were treated with TCA which removed both proteins and pigmentation. It is possible that this also interfered with the detection of L-alanine or led to its degradation. This could be addressed in the future by using a gas chromatography or high-performance liquid chromatography-based method to measure L-alanine concentration.

5.4.2 Optogenetic control of mannitol production in Syn7002

No statistically significant differences were detected at any timepoint for culture density, mannitol titres or mannitol production rates in cultures expressing pUC-172-10-mpusfus across the four light regimens tested (Figure 5-5 and Figure 5-6). It had been hypothesised that mannitol concentration would have increased at a higher rate in cultures grown under white light than in red light due to increased mpusfus expression. However, there may be 'true' differences in the mannitol production rates between cultures grown under different light regimens which are masked by the significant variation observed across replicates. In future, using the CcaS system to control the synthesis of another product with less titre variation would be useful to make a better comparison between the effect of different light regimens on product formation.

The difference in mannitol titre between cultures expressing pAQ1BB-pCpcB594-mpusfus and those expressing pUC-172-10-mpusfus at day 7 and day 14 are most likely to be explained by higher levels of mpusfus transcription (Figure 5-6a). pCpcB594 is known to be among the strongest constitutive promoters tested in Syn7002 (Markley et al., 2015). Therefore, higher levels of mpusfus protein would enable faster mannitol synthesis. To better understand the amount of mannitol produced by each cell available for production up to each harvested time point, an area under the growth curve equation could be used. However, this would require sampling at more frequent time points than every 7 days.

5.4.3 Comparison of mannitol titres and growth rates with previous research

The mannitol titres achieved in this work appear to match or surpass the maximum titres achieved previously in Syn7002 (Figure 5-6a) (Jacobsen and Frigaard, 2014). However, making direct comparisons is challenging due to the large variation in mannitol production across replicates reported in both studies.

Titres of 0.1-0.6 g L^{-1} were reported across 3 different substrains after 6.5 days of cultivation, resulting in productivity rates of 0.01-0.09 g L^{-1} per day (Jacobsen and Frigaard, 2014). Furthermore, the same study reported one biological replicate achieving a mannitol titre of 1.1 g L^{-1} after 12 days, representing a productivity rate of 0.15 g L^{-1} per day. Final mannitol titres observed in the current work ranged between 0.39–2.12 g L^{-1} across all individual replicates, corresponding to a productivity rate of 0.01-0.07 g L^{-1} per day (Figure 5-8a). The highest reported mannitol titres in other cyanobacteria species are 701 mg L^{-1} for Syn7942, 401 mg L^{-1} for *Synechococcus elongatus* PCC 11801 and 537 mg L^{-1} for *Synechococcus elongatus* PCC 11802 (Pritam *et al.*, 2023).

It is unclear why the mannitol titres achieved in the current work using pAQ1BB-pCpcB594-mpusfus are 10-20 times higher than previously reported (Madsen *et al.*, 2018). The same mannitol assay kit was used in both studies. The assay kit explicitly states that there is no need to run standard curves and instead sample mannitol concentration is calculated directly using the extinction coefficient of NADH produced in the reaction. As a precaution, the assay kit was regularly used in conjunction with standards of known mannitol concentrations to ensure that the final calculated mannitol concentrations aligned with the known amount of mannitol. Possibilities explaining the differences in results between the two studies include differences in culture volumes, light sources and the level of gene segregation achieved. Whilst homozygous strains were unable to be achieved when using either pAQ1BB-pCpcB594-mpusfus or pUC-172-10-mpusfus, colonies had been repeatedly streaked on plates containing successively higher concentrations of antibiotic to increase gene copy number. It is unclear if successive colony re-streaking was performed in the study by Madsen and colleagues. An inability to obtain fully homozygous mannitol-producing Syn7002 strains has previously been reported (Jacobsen and Frigaard, 2014).

One notable observation during the mannitol experiments was that on several occasions cultures expressing either pAQ1BB-pCpcB594-mpusfus or pUC-172-10-mpusfus would fail to produce mannitol or demonstrate extremely low mannitol titres. A similar phenomenon was also previously observed when engineering Syn7002 for mannitol production. Sequencing of one of these cultures which had lost the ability to produce mannitol revealed the presence of a frameshift

mutation in the mannitol biosynthesis gene *mtlD* production (Jacobsen and Frigaard, 2014). Interestingly, the loss of mannitol production was not reported in other work which involved the expression of pAQ1BB-pCpcB594-mpusfus in Syn7002 (Madsen *et al.*, 2018). However, the study by Madsen and colleagues resulted in significantly lower mannitol titres reached in both the current work and that of Jacobsen and Frigaard. The growth of wild-type Syn7002 is unaffected by exogenous mannitol concentrations of up to 20 g L^{-1} (Jacobsen and Frigaard, 2014). Therefore, it is likely that the loss of mannitol production and the often Large variation in mannitol titre across replicates can be explained by genomic instability. The high metabolic burden incurred due to excess mannitol biosynthesis would result in the selection of cells with frameshift mutations in the mannitol biosynthetic enzymes.

Expression of mpusfus under both constitutive and optogenetic control leads to significantly reduced final culture densities versus wild-type (Figure 5-5a). This may be due to cellular resources typically used for cell growth and replication being diverted to mannitol biosynthesis. This effect was also present, but less pronounced, in the work by Madsen and colleagues. In contrast, severe growth impairments in mannitol-production strains have been reported in other work (Jacobsen and Frigaard, 2014). The differences between the studies can most likely be explained by two factors. The first is the use of 1% CO_2 for sparging in the work by Jacobsen and Frigaard, versus bubbling with atmospheric air (0.04% CO_2) in the present work and that of Madsen and colleagues. CO_2-enriched air increases cell growth rates, which is likely to make a growth impairment effect in mannitol-production strains versus wild-type more pronounced. The second factor is that the up to 20-fold higher mannitol titres achieved in the present work and that of Jacobsen and Frigaard led to a higher level of intermediate accumulation or product toxicity than in the study by Madsen and colleagues.

5.4.4 Potential future improvements to mannitol production

To improve mannitol production further, cultures expressing pAQ1BB-pCpcB594-mpusfus or pUC-172-10-mpusfus could be sparged with 1-3 % CO_2 rather than atmospheric air, thus increasing the growth rate and the carbon availability for metabolism. This could lead to both the maximal mannitol titre being achieved in less time and allow a better comparison to be made between the effect of

expressing each mannitol biosynthetic enzyme (mtlD and mlp) separately versus expressing the fusion enzyme mpusfus. Furthermore, the use of a glycogen-deficient mutant as previously reported could also increase mannitol titre by directing the flux of excess carbon away from natural storage and towards mannitol production (Jacobsen and Frigaard, 2014).

Large variations in mannitol titre or cessation of mannitol production (Figure 5-7c) were not reported in other cyanobacteria species engineered for mannitol production including Syn7942, Syn11801 and Syn11802 (Pritam *et al.*, 2023). However, while this phenomenon has also been reported in Syn6803, a natural selection-based approach was used to overcome it taking advantage of mannitol function as a compatible osmolyte. Applying salt stress to a Syn6803 mutant strain which cannot produce the compatible osmolytes glucosyl-glycerol and sucrose, whilst also expressing the mannitol biosynthetic enzymes mtlD and mlp resulted in stable mannitol production over time in a turbidostatic continuous culture (Wu *et al.*, 2020). In contrast, the removal of salt stress or the use of a strain expressing fully functional biosynthetic pathways for the endogenous compatible osmolytes resulted in the eventual cessation of mannitol production. This approach provides selection pressure towards cells which maintain functional copies of mltD and mlp. The synthesis of the compatible osmolyte mannitol protects the cells from salt-induced osmotic stress. A similar approach in Syn7002 could result in more stable mannitol production.

Only extracellular mannitol concentrations were measured in the present work, as it has been previously reported that Syn7002 secretes 99% of synthesised mannitol (Madsen *et al.*, 2018). The mechanism by which Syn7002 secretes mannitol is unknown. No specific mannitol transporter has been identified in Syn7002; however, mannitol may be exported through channels functioning in the transport of other metabolites (Jacobsen and Frigaard, 2014). Syn7002 secretes some sugars through outer membrane vesicles, which could be a mannitol export mechanism (Xu *et al.*, 2013). Mannitol-specific transporters have been identified in *E. coli* and *Vibrio cholerae*; however, these are understood to function as importers rather than exporters (Opačić *et al.*, 2010; Mustachio *et al.*, 2012). Identifying an endogenous mannitol transporter and

increasing its expression could lead to improved growth rates by alleviating the strain of high intracellular mannitol concentration.

Insights into the potential future improvement of mannitol production can be obtained by studying the production of other sugar alcohols in cyanobacteria. Syn6803 has been engineered to produce sorbitol by heterologous expression of S6PDH from apple (*Malus domestica*), which converts fructose-6-phosphate to sorbitol-6-phosphate by using NADPH (Chin *et al.*, 2018). Sorbitol titre was 2.5-fold by heterologous expression of fructose-1,6-bisphosphatase, which converts fructose-1,6-bisphosphate to the sorbitol precursor fructose-6-phosphate. Sorbitol titre was increased further upon overexpression of Syn6803 pyridine nucleotide transhydrogenases, which increased the supply of NADPH. Mannitol, like sorbitol, can be synthesised from fructose-6-phosphate, and therefore modulating the expression of Syn7002 fructose-1,6-bisphosphatase may enhance carbon flux towards mannitol. Furthermore, increasing the NADH supply through the overexpression of relevant native enzymes could increase mpusfus activity and increase mannitol titre.

6 Identification of wavelength-specific transcripts

6.1 Introduction

As discussed in the Introduction (Section 1.3 – Wavelength-specific Signalling Pathways), many cyanobacterial photoreceptors have been identified but associated signalling pathways are not well understood. Identification of the associated transcriptional regulators and output promoters bound by these regulators is key to characterising wavelength-specific signalling pathways. This could also allow the identification of new optogenetic systems.

To identify components of these pathways, a first step is identifying wavelength-specific transcripts using RNAseq. These can be pinpointed by analysing the transcriptome of cultures grown under different colours of narrowband LED light and comparing them to another colour. Blue and orange light were chosen as a starting point because there are very few optogenetic tools in the orange part of the spectrum. Furthermore, previous work in the Amtmann lab identified a novel blue/orange responsive CBCR named 7GP-3 in Syn7002 (Mihnev *et al.*, 2023). Whilst this photoreceptor does not contain an obvious effector domain for signalling like a histidine kinase or DNA-binding domain, this photoreceptor may bind to another signalling protein to modulate its function.

While many optogenetic tools which respond to wavelengths across the UV-Vis light spectrum have been characterised, some gaps exist (Figure 6-1). In particular, there are no tools which are activated by yellow or orange light. In contrast, a large range of optogenetic tools are available which respond to blue light, many of them containing LOV domains (Möglich *et al.*, 2009; Zhao *et al.*, 2021). Identifying a tool absorbing in the yellow/orange region would increase the scope for applying multiple optogenetic tools within the same strain. This is currently limited due to spectral overlap between the green/red responsive CcaS and many of the red/far-red systems available (Figure 6-1). While the yellow/teal light-absorbing photosensory kinase DpxA has been identified in *F. diplosiphon*, it has not been applied as an optogenetic tool (Wiltbank and Kehoe, 2016). This is most likely because a transcriptional regulator phosphorylated by DpxA has not yet been identified. Furthermore, not every optogenetic tool is guaranteed to function in a given organism, even after optimisation. This was

evident in this work with the attempted application of the YF1/FixJ system in Syn7002, even after confirming its function in *E. coli* (discussed in Chapter 4).

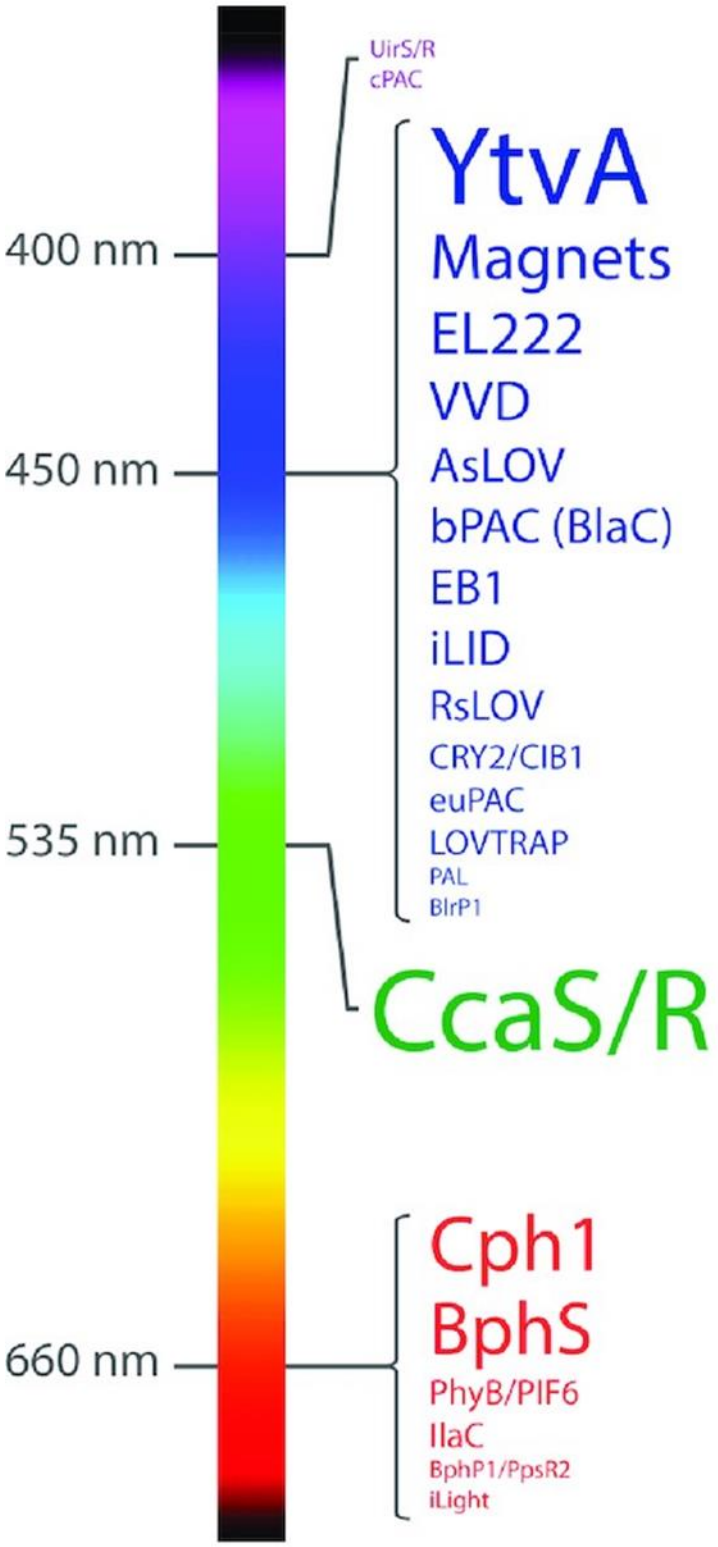


Figure 6-1. Optogenetic tools and their activating wavelengths. Systems are organised based on their activation wavelength. The size of each word is proportional to the number of published applications which apply to each system as of August 2021 (Lindner and Diepold, 2022).

6.2 Chapter aims

The main aim of this work was to identify wavelength-specific transcripts. This would aid the future identification of potential components of wavelength-specific signalling pathways. First, the effect of blue versus orange light on Syn7002 growth was measured. Next, RNAseq was used to compare genome-wide transcript levels in Syn7002 cultures grown under blue, orange and white LED light. A list of constitutive wavelength-independent genes was identified from the dataset for use as reference genes in subsequent qRT-PCR experiments. The RNAseq expression profiles of various commonly used qRT-PCR reference genes were also examined. The expression profiles of genes differentially expressed under blue or orange light were then individually analysed, especially if they were annotated with functions related to cell signalling. Finally, clustering analysis was performed on the RNAseq dataset to identify clusters of genes which showed similar expression patterns. The upstream regions of genes in the identified clusters, as well as potential 'blue-specific' and 'orange-specific' genes, were then evaluated using promoter analysis algorithms to search for enriched DNA motifs which could indicate a wavelength-specific promoter element.

6.3 Results

6.3.1 Growth of Syn7002 in blue versus orange light

To measure the effect of blue versus orange LED light on Syn7002 growth, wild-type cultures were grown in 1 L photobioreactors (MicroPharos PBR, Xanthella) under a 16h/8h light/dark cycle. Cultures were illuminated with either 20 µmol m^{-2} s^{-1} orange LED light, or 20 µmol m^{-2} s^{-1} blue LED light and growth was determined by measuring OD_{730} at regular time points. 20 µmol m^{-2} s^{-1} was the maximum intensity achievable with the orange LED. Syn7002 cultures grew significantly faster under orange light than in blue light (Figure 6-2). Cultures grown under blue light required 19 days to reach a similar OD_{730} as the cultures grown under orange light had reached on day 5.

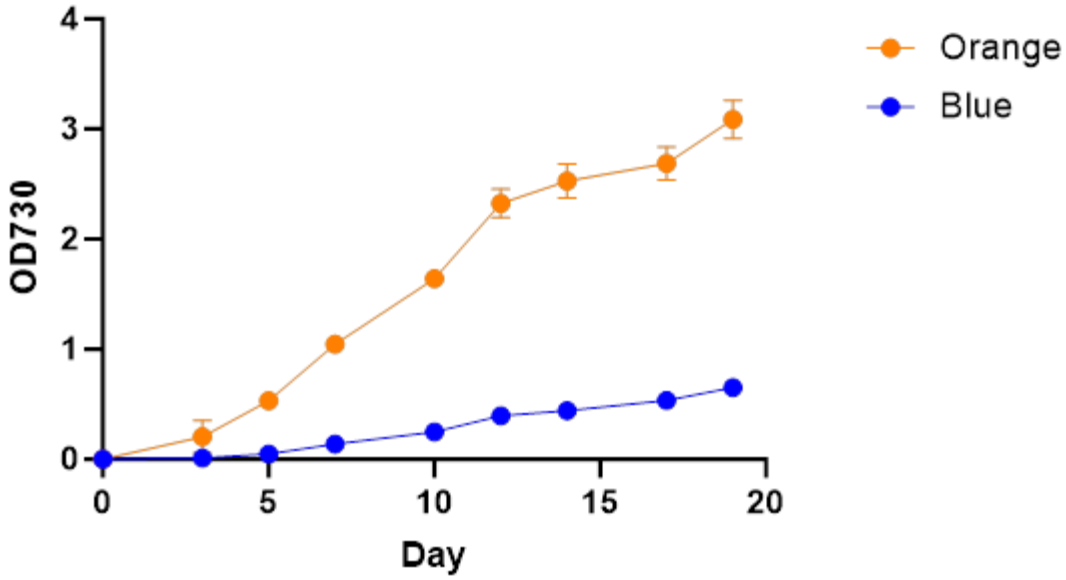


Figure 6-2. Growth of Syn7002 under blue versus orange light. Wild-type Syn7002 cultures were grown in 1 L photobioreactors under the equivalent intensity of either 20 µmol m^{-2} s^{-1} blue (blue symbols) or 20 µmol m^{-2} s^{-1} orange LED light (orange symbols) under a 16h/8h light/dark cycle. Data are means ± S.E.M of three independent cultures.

6.3.2 RNAseq principal component analysis and setting a low counts threshold

Next, RNAseq was used to identify blue light-specific and orange light-specific transcripts. Wild-type Syn7002 cultures were illuminated with an equivalent intensity of light from each of the five Xanthella photobioreactor LEDs (blue, cyan, green, orange and red) under a 16h/8h light/dark cycle which produced a white light effect. After 4 days, at the end of the 8-hour dark cycle, before the lights switched back on, the cultures were given a specific light treatment of either orange only, blue only or a continuation of 'white' light from the 5 LEDs. All light treatments used an intensity of 20 µmol m^{-2} s^{-1}. At several time points across 26 hours, samples were harvested and RNA extracted. A subset of these samples shown in red were submitted for RNAseq, with samples from 3 independently grown cultures sequenced per condition (4 timepoints x 3 light conditions x 3 replicates = 36 samples (Figure 6-3). Due to budget constraints, the 0, 1, 2 and 4 hour timepoints were chosen to be sequenced as we reasoned some wavelength-specific transcriptional responses may be short-lived and could be missed by instead measuring the later timepoints.

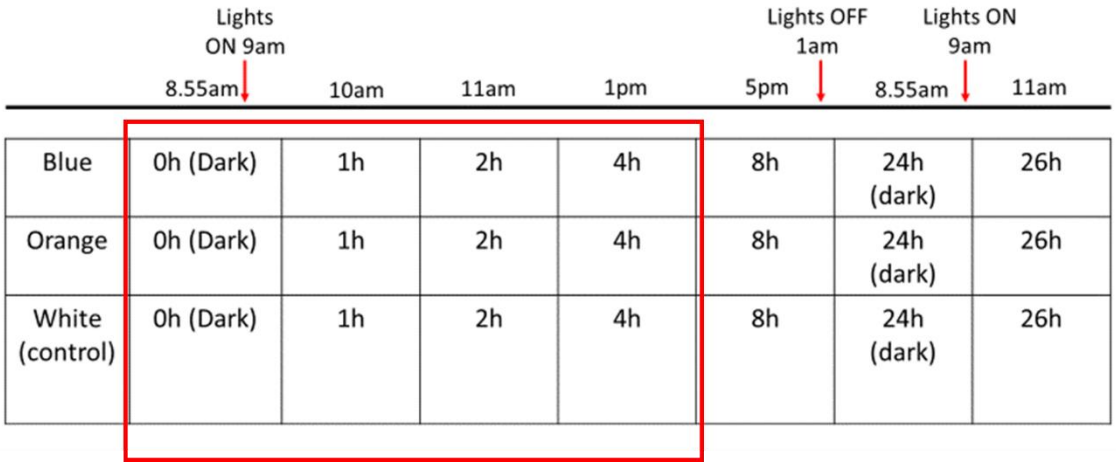


Figure 6-3. Experimental design for RNAseq experiments. Cultures were grown in Xanthella photobioreactors under white light until the 0-hour time point. The cultures were then switched to either 20 µmol m^{-2} s^{-1} blue, 20 µmol m^{-2} s^{-1} orange or continued in white light and samples were collected at the indicated timepoints. Samples in the red box were submitted for RNAseq. 3 biological replicates were sequenced for each timepoint/colour condition (36 samples total).

RNAseq was performed as described in the Methods section. First, the variance between samples was assessed using principal component analysis (PCA). This method reduces the dimensionality of the RNAseq data which contains thousands of genes, each with a different expression value across many samples, into a small number of principal components (PC) which can account for the variation between samples. Each PC contains a set of genes and visualising these on a PCA plot enables positioning of the samples based on their transcript levels. The PCA plot shows PC1 and PC2 which contribute the most variation across all of the samples. PC1 explained 68% of the variance between the samples while PC2 explained 15% (Figure 6-4). The nine 0-hour timepoint samples formed a cluster as expected. Each subsequent light treatment/timepoint formed a separate clearly defined cluster, showing that inter-replicate variation was low. A clear separation was observed between the 0-hour samples and all other samples which had been exposed to light. This highlighted a strong effect of light on the transcriptome. At the 1-hour, 2-hour and 4-hour time points, the orange and white samples grouped together whereas the blue samples were located at a further distance.

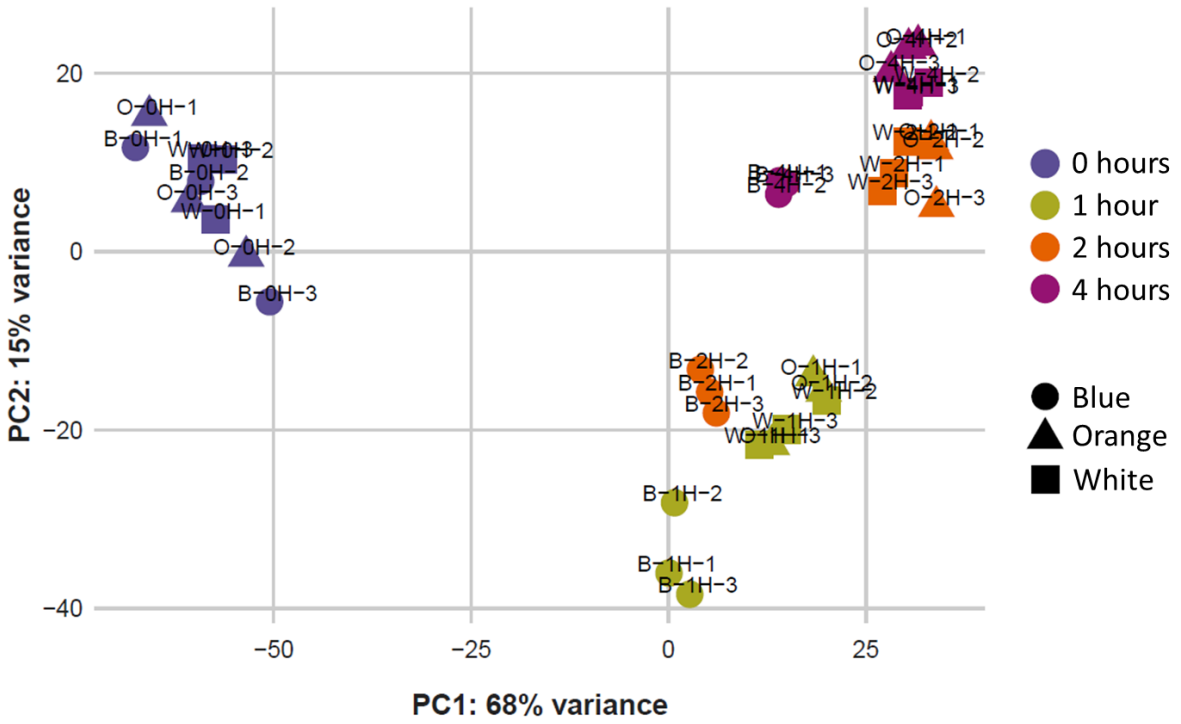


Figure 6-4. Principal component analysis (PCA) of Syn7002 RNA transcript levels. RNA transcript levels of Syn7002 cultures irradiated with blue, orange or white LED light after an 8-hour dark treatment were measured by RNAseq. 36 samples were sequenced in total, comprising of three colour treatments (blue, orange and white), four time points (0, 1, 2 and 4 hours) and three independently grown cultures. Each sample is labelled by colour treatment, timepoint and replicate number. Colour treatments were differentiated by shapes while timepoints were differentiated by colour on the plot. The PCA was performed by Dr Graham Hamilton of Glasgow Polyomics.

Next, a low count threshold was applied to exclude genes with very low expression values from further analyses. Two of the least similar inter-replicate samples as identified by PC1 and PC2 were B-OH-1 and B-OH-3 (Figure 6-4). These two samples were used in the setting of a low counts threshold, by plotting the normalised transcript levels for each gene against each other (Figure 6-5). While most of the genes at moderate to high expression values show a positive correlation, several genes with a very low number of counts (e.g. <10 normalised counts) do not appear to be well correlated. Thus, any gene with average normalised counts across all samples of <10 was excluded from further analysis. Out of 3235 genes, 3171 genes were used in the subsequent analysis. The plot also highlighted a single gene with an expression level two orders of magnitude higher than that of the second most highly expressed gene: *23S rRNA* (Figure 6-6).

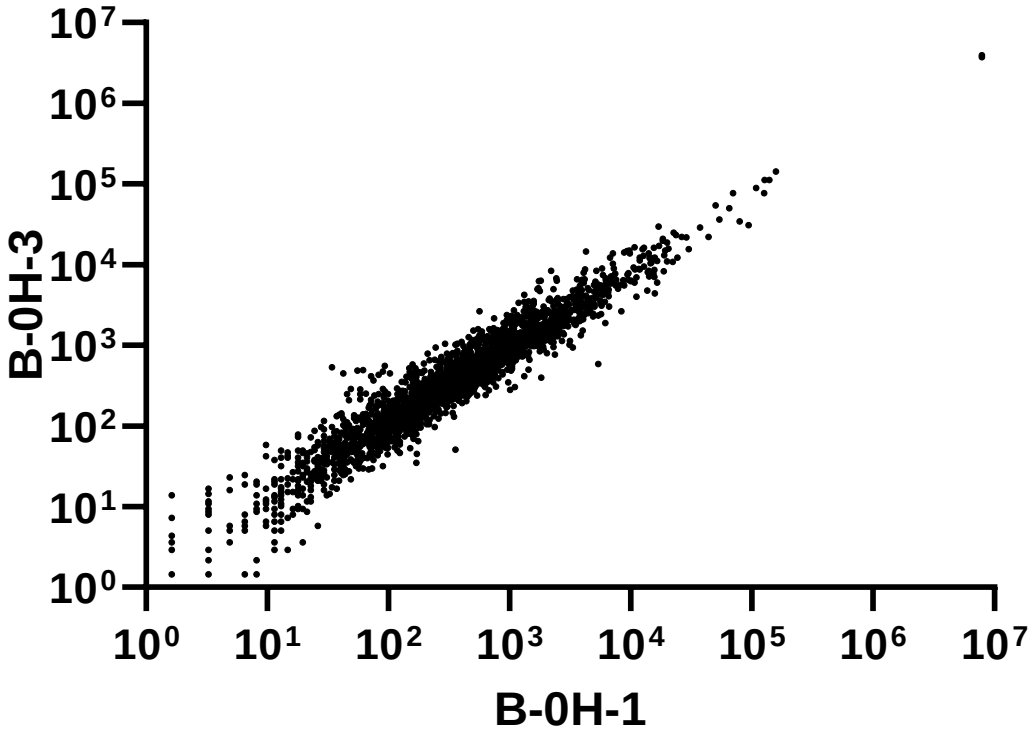


Figure 6-5. Normalised transcript levels of the least correlated inter-replicate samples. The normalised RNA transcript levels of all genes from the blue 0-hour replicate 1 sample (B-0H-1) were plotted against those from the blue 0-hour replicate 3 sample (B-0H-3). These were identified as being the two least similar inter-replicate samples due to the distance between them identified by PCA (Figure 6-4).

6.3.3 Differential expression analysis

After setting a low counts threshold, differential expression analysis was performed to identify genes which are differentially expressed between the different light treatments. Pairwise comparisons were made between the 0-hour timepoints and each of the other timepoints (0h/1h, 0h/2h, 0h/4h) within each colour treatment. Genes were classified as differentially expressed if they demonstrated a log2fold change > 1.5 and adjusted p-value < 0.05 in at least one of the pairwise comparisons. Adjusted p-values were used rather than raw p-values to account for multiple comparison testing. 430 genes were identified as being differentially expressed in blue light, 638 genes were identified as being differentially expressed in orange light and 537 were identified as differentially expressed in white light. 298 genes were identified as differentially expressed in all three light treatments. However, 49 genes were differentially expressed in blue light only, 201 genes were differentially expressed in orange light only and

65 genes were differentially expressed in white light only (Figure 6-6). The number of genes differentially expressed under both blue and white light but not orange light was around half of that observed under orange and white light but not blue light. Due to the presence of both blue and orange LEDs in the 'white' light treatment, genes which were differentially expressed in blue and white light, but not orange, were described as 'blue-specific'. Tables highlighting genes differentially expressed in blue only and blue plus white as well as orange only and orange plus white are shown in Appendix II and Appendix III.

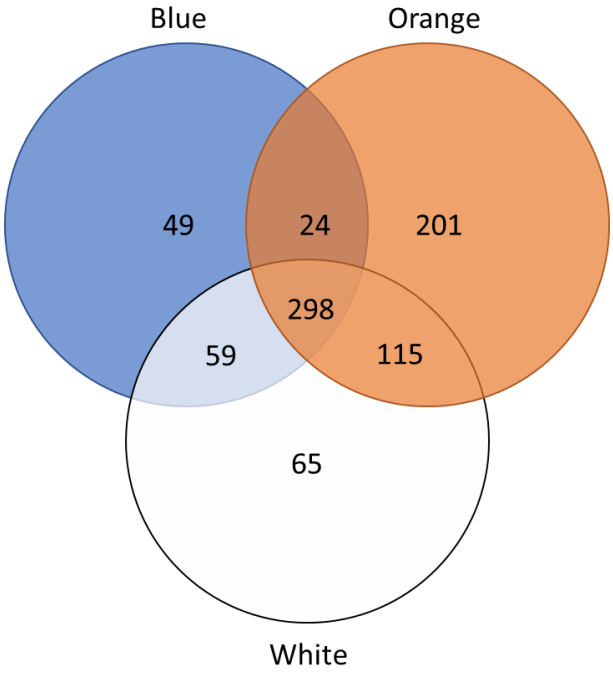


Figure 6-6. Identifying overlap of differentially expressed genes between light treatments. Genes were considered differentially expressed if they demonstrated a log2fold change > 1.5 and adjusted p-value < 0.05 in either the 1-hour, 2-hour or 4-hour timepoint versus the 0-hour timepoint within each colour treatment.

When using the above criteria, a gene identified as being differentially expressed in, for example, blue light may still have a similar expression pattern in orange light. If the log2fold change is slightly below the applied threshold of > 1.5, this gene may be identified as being potentially 'blue-specific', even when it displays a similar expression profile under orange light. An example of this nuance from the dataset is the expression profile of gene *SYNPCC7002_A1285* (Figure 6-7). Using only the criteria of log2fold change > 1.5 and adjusted p-value < 0.05, this gene was identified as being 'blue-specific'. However, manual inspection clearly shows a very similar pattern of expression across all three light treatments. Therefore, manual inspection of gene expression profiles was used for the identification of individual genes of interest which may demonstrate wavelength-specific expression patterns.

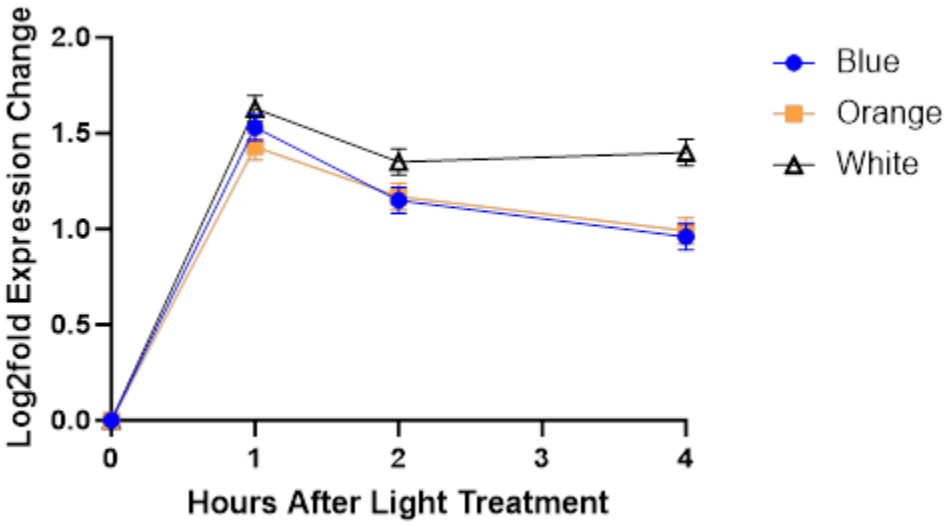


Figure 6-7. RNA transcript levels of gene *SYNPCC7002_A1285*. The log2fold change in expression of *SYNPCC7002_A1285* as calculated by DEseq2 in blue (blue symbols), orange (orange symbols) and white (black and white symbols) light as determined by RNAseq. Data points show the mean transcript levels of 3 independently grown cultures ± S.E.M.

6.3.4 Identification of candidate qRT-PCR reference genes

A panel of candidate reference genes for qRT-PCR was also identified from the RNAseq dataset. 389 genes were identified as demonstrating a log2fold change between -0.5 and 0.5 across all of the 0h/1h, 0h/2h and 0h/4h pairwise comparisons in blue, orange and white light. Thus, the expression of these genes remains fairly constant across all samples. Using a stricter criterion of a log2fold change between -0.2 and 0.2, 10 genes were identified (Table 6-1). As these 10

genes contained a basemean below 1000 (the average normalised expression value across all samples for this gene as calculated using DESeq2), another 2 genes were chosen from the -0.5 to 0.5 group. These were *SYNPCC7002_A0956* and *SYNPCC7002_A1356*, which contained higher basemeans of 10076 and 13108 respectively (Table 6-2). These genes would be more suitable as reference genes for highly expressed transcripts.

Table 6-1. Candidate qRT-PCR reference genes identified using the RNAseq dataset. Genes demonstrated a log2fold change between -0.2 and 0.2 across all time points and light treatments.

Gene ID	Cyanobase Annotation	Gene Name	baseMean
SYNPCC7002_A0028	conserved hypothetical protein	N/A	321.95
SYNPCC7002_A0368	Lipid A disaccharide synthetase	*lpxB*	261.76
SYNPCC7002_A0491	conserved hypothetical protein	N/A	467.75
SYNPCC7002_A1408	DNA polymerase III alpha subunit	*dnaE-C*	708.34
SYNPCC7002_A1566	photosystem I assembly protein ycf3	N/A	551.18
SYNPCC7002_A1638	SpoIID Like protein	N/A	626.39
SYNPCC7002_A1784	NH(3)-dependent NAD(+) synthetase	*nadE*	775.27
SYNPCC7002_A2185	glycosyl transferase, group 2 family protein	N/A	286.57
SYNPCC7002_A2364	Two-component response regulator; GGDEF domain protein	N/A	914.43
SYNPCC7002_A2682	conserved hypothetical protein	N/A	631.90

Table 6-2. Highly expressed candidate qRT-PCR reference genes. Selected genes which demonstrated a log2fold change between -0.5 and 0.5 across all time points and light treatments plus a basemean >10000.

Gene ID	Cyanobase Annotation	Gene Name	Basemean
SYNPCC7002_A1356	2-isopropylmalate synthase	*leuA*	13108.76
SYNPCC7002_A0956	conserved hypothetical protein TIGR00244, ATP cone domain	N/A	10076.65

Rather than plotting the log2fold change of genes on the y-axis, normalised counts as calculated by DEseq2 are shown. This provides a more accurate visual representation of the expression pattern of certain genes, as differences in the 0-hour measurements between replicates can skew log2fold changes. An example of this is shown in Figure 6-8, with many other similar examples being contained within the dataset.

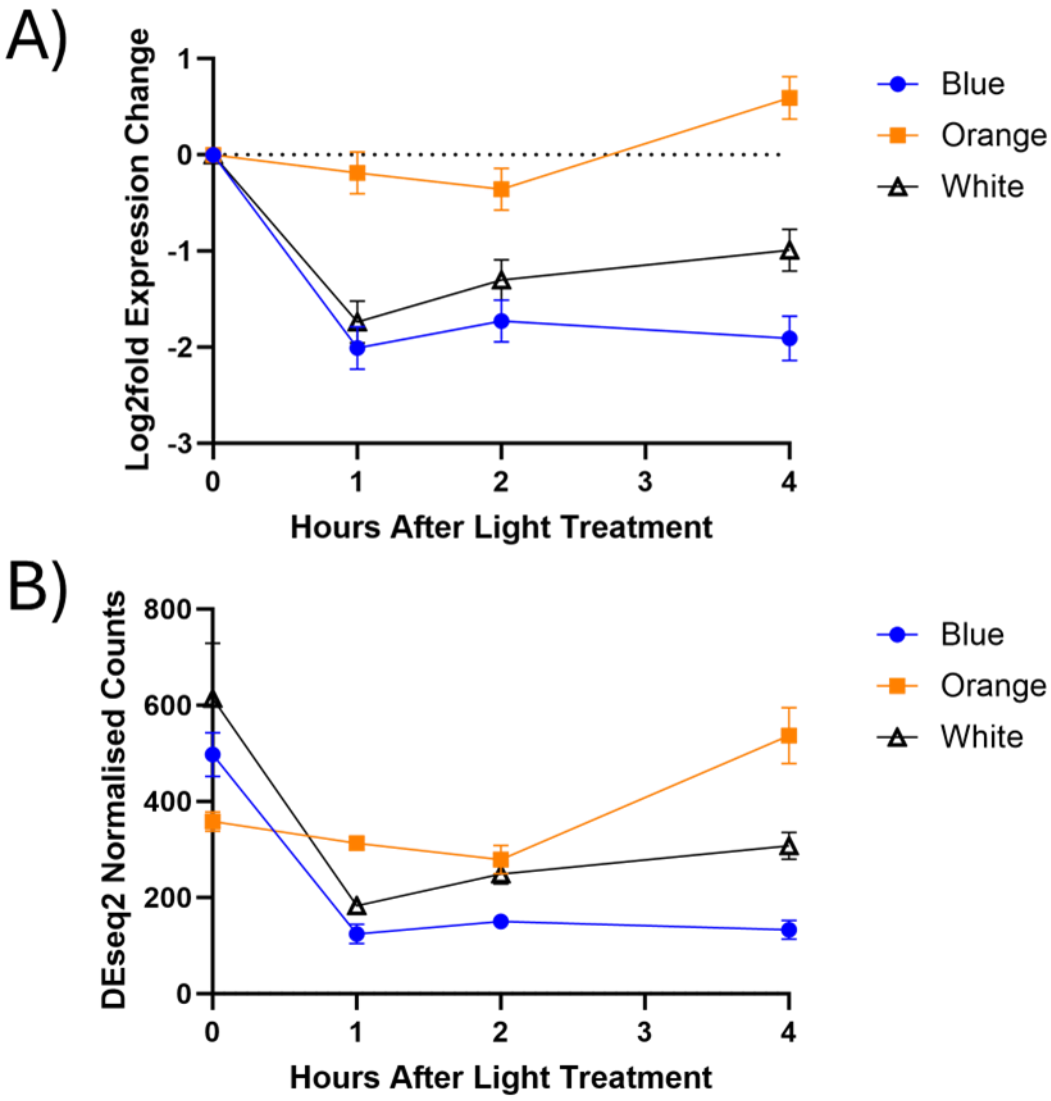


Figure 6-8. Normalised expression values versus log2foldchange for *SYNPCC7002_G0148*. A) Log2fold expression change and B) DEseq2 normalised counts for gene *SYNPCC7002_G0148* in blue (blue symbols), orange (orange symbols) and white (black and white symbols) light as determined by RNAseq. Data points show the mean transcript levels of 3 independently grown cultures ± S.E.M.

Four genes in total were selected from Table 6-1 and Table 6-2 and their normalised expression values in blue, orange and white light were plotted (Figure 6-9). The four genes showed reasonably stable expression patterns across time points and light treatments. As expected, based on their selection criteria, the lower expressed *SYNPCC7002_A1784* and *SYNPCC7002_A2364* demonstrated more stable expression patterns than *SYNPCC7002_A0956* and *SYNPCC7002_A1356*. These four genes were assayed using qRT-PCR under blue, orange, green and red light and the most stable gene as identified by RefFinder was *SYNPCC7002_A1356* which was subsequently used for all qRT-PCR experiments (discussed in depth in Chapter 3).

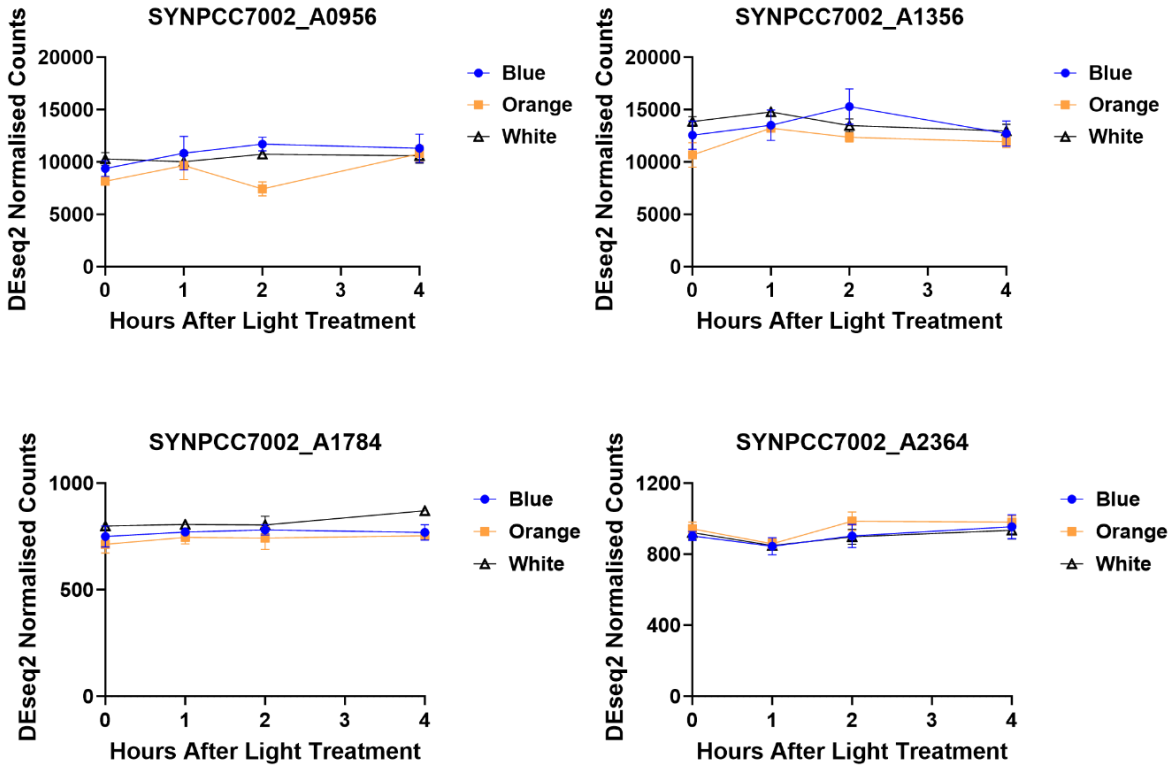


Figure 6-9. RNAseq expression profiles of four selected candidate qRT-PCR reference genes. DEseq2 normalised counts of four selected genes from cultures grown in blue (blue symbols), orange (orange symbols) and white (black and white symbols) light as determined by RNAseq. Data points show the mean transcript levels of 3 independently grown cultures ± S.E.M.

Significant variation in expression levels of several commonly used qRT-PCR reference genes was observed across the light conditions tested. Common reference genes in Syn7002 include *rnpA*, *secA*, *16S rRNA* and *23S rRNA*. *rnpA* and *secA* have previously been validated as suitable reference genes under certain conditions such as UV-B treatment and dark incubated cultures (Szekeres *et al.*, 2014). However, the present work clearly shows that these genes show

significant increases in expression upon the onset of light, making them unsuitable for timecourse-based experiments (Figure 6-10). *16S rRNA* and *23S rRNA* both demonstrate variability in expression between light treatments and across the time course. Furthermore, these genes are very highly expressed in Syn7002, both ranking among the top 10 most abundant transcripts in all samples. This would make these genes unsuitable as reference genes for low-abundance transcripts. This highlights the importance of verifying the expression levels of a chosen reference gene under the relevant experimental conditions to obtain accurate qRT-PCR results.

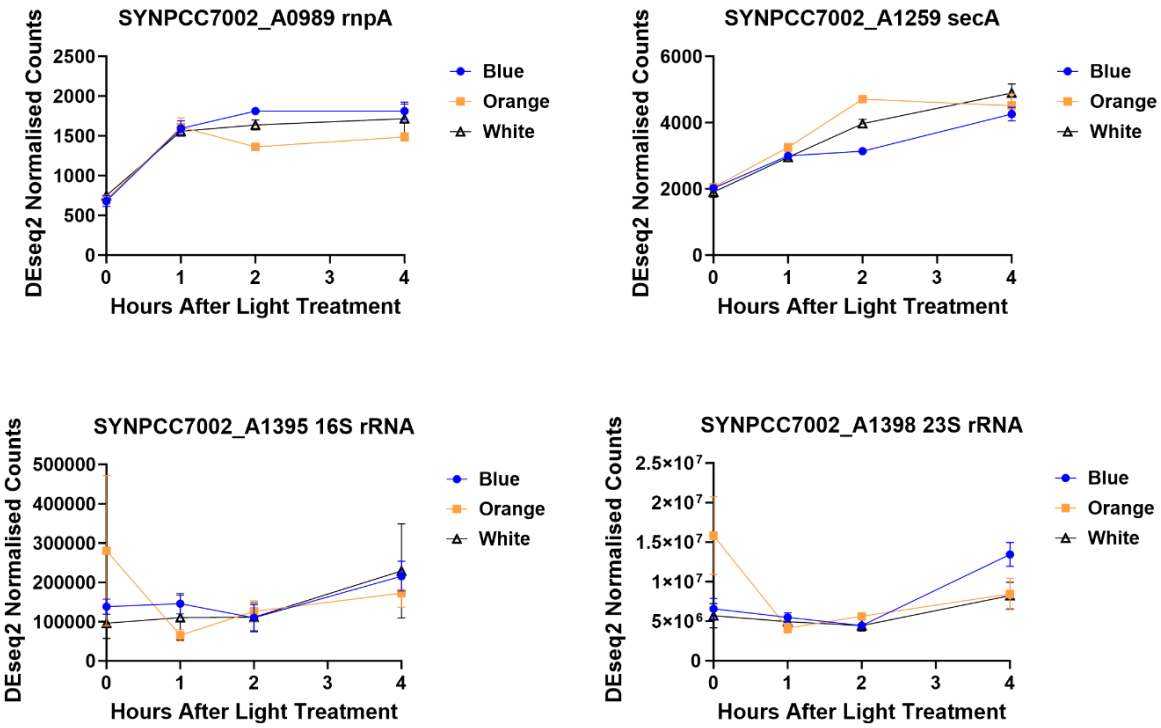


Figure 6-10. RNAseq expression profiles of commonly used qRT-PCR reference genes. DEseq2 normalised counts of four selected genes from cultures grown in blue (blue symbols), orange (orange symbols) and white (black and white symbols) light as determined by RNAseq. Data points show the mean transcript levels of 3 independently grown cultures ± S.E.M.

6.3.5 Identification of wavelength-specific transcripts

Next, the lists of genes identified as being 'blue-specific' or 'orange-specific' were manually inspected to identify wavelength 'reporter genes'. A range of genes were selected based on the presence of strongly differing expression profiles between blue and orange light. Genes showing 'blue-specific' expression patterns are shown in Figure 6-11. The potential function of these genes is discussed later.

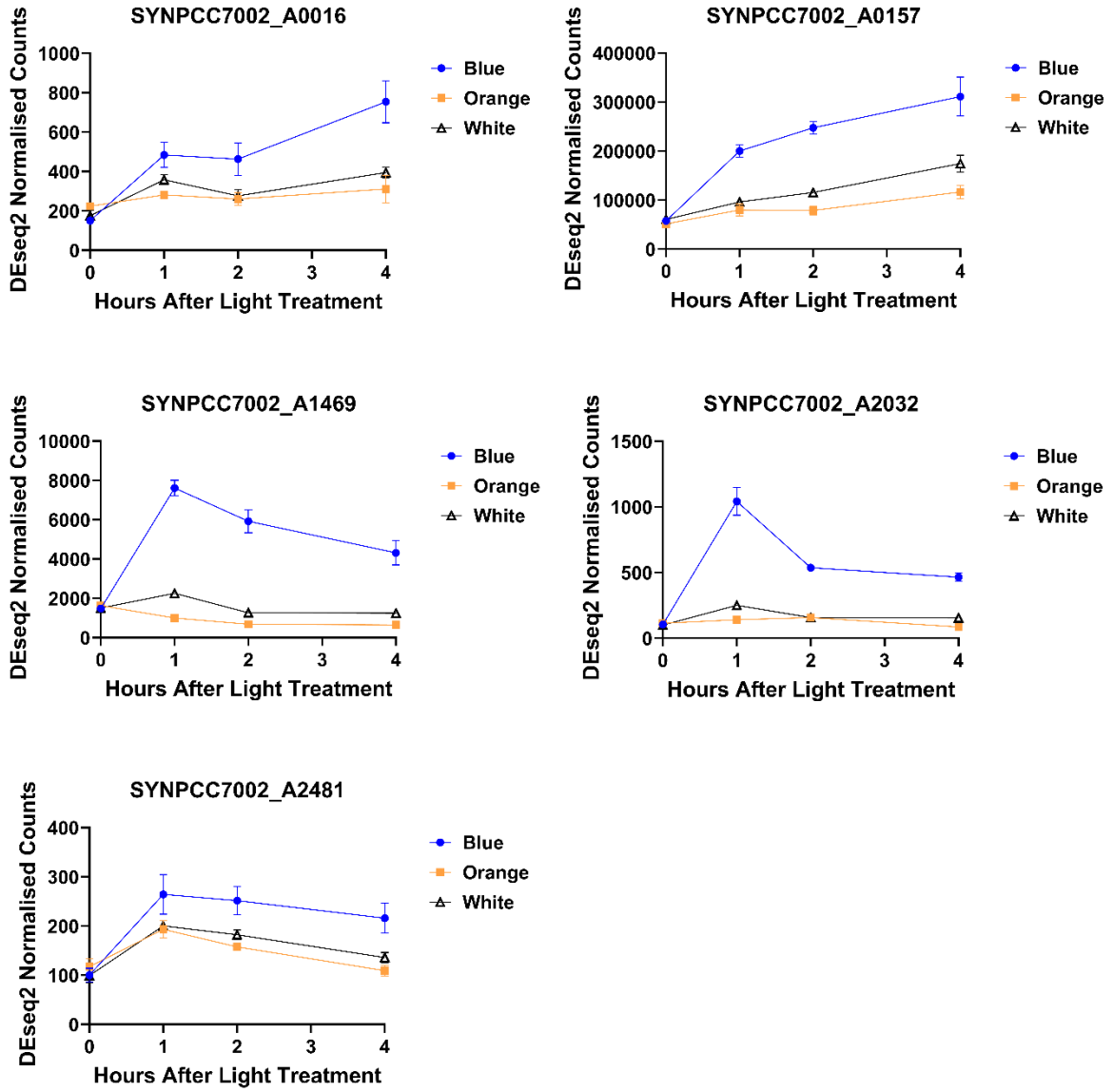


Figure 6-11. RNAseq expression profiles of genes showing 'blue-specific' expression patterns. DEseq2 normalised counts of five selected genes from cultures grown in blue (blue symbols), orange (orange symbols) and white (black and white symbols) light as determined by RNAseq. Data points show the mean transcript levels of 3 independently grown cultures ± S.E.M.

Several interesting expression profiles were also identified from genes categorised as being differentially expressed in orange light. Many of these 'orange-specific' genes showed similar patterns in both orange and white light (Figure 6-12). This contrasted with the selected 'blue-specific' genes, most of which showed up or downregulation in blue light but not white light (Figure 6-11). For example, *SYNPCC7002_A0267*, *SYNPCC7002_A1000*, and *SYNPCC7002_A1509* all showed highly similar expression profiles under orange and white light (Figure 6-12). Similarly, *SYNPCC7002_A0172 and*

SYNPCC7002_A2347 show no increase in expression under blue light, a moderate increase under white light and a large increase under orange light (Figure 6-12).

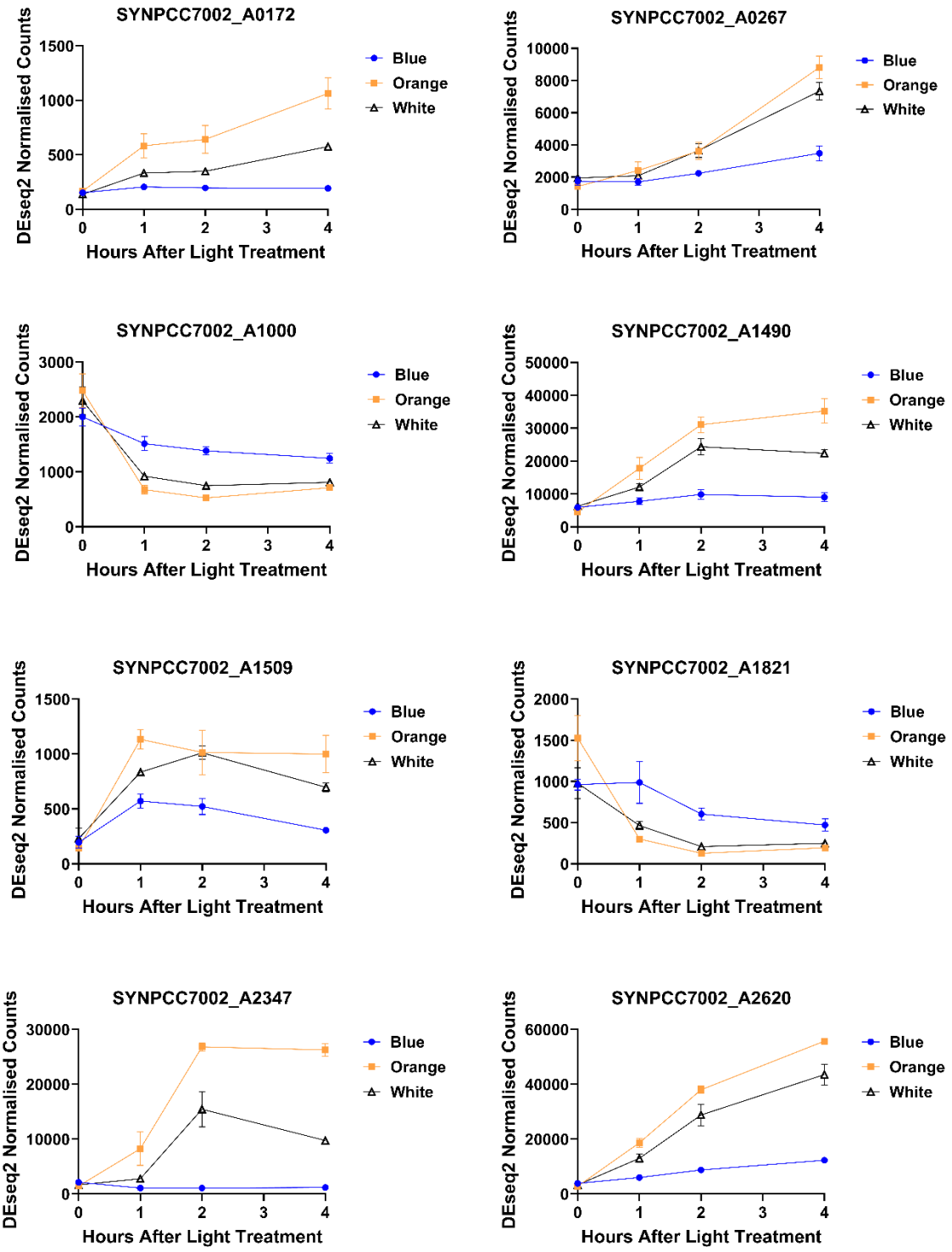


Figure 6-12. RNAseq expression profiles of genes showing 'orange-specific' expression patterns. DEseq2 normalised counts of eight selected genes from cultures grown in blue (blue symbols), orange (orange symbols) and white (black and white symbols) light as determined by RNAseq. Data points show the mean transcript levels of 3 independently grown cultures ± S.E.M.

The ‘blue-specific’ and ‘orange-specific’ genes shown above could be used as palettes of potential ‘reporter genes’ as discussed. Another question of interest is whether any of the wavelength-regulated genes could function as kinases or transcriptional regulators. Two genes were identified as possessing kinase functions *SYNPCC7002_A0049 and SYNPCC7002_A0900*, both of which showed strong upregulation in response to orange light (Figure 6-13).

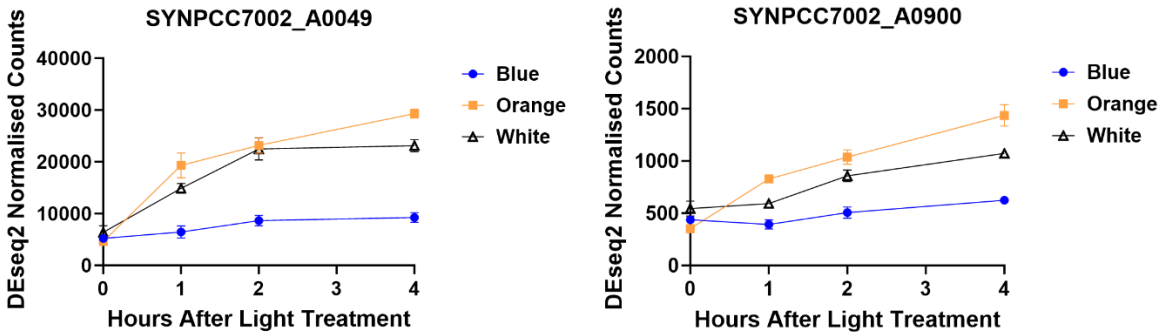


Figure 6-13. RNAseq expression profiles of genes containing kinase functions. DEseq2 normalised counts of two selected genes from cultures grown in blue (blue symbols), orange (orange symbols) and white (black and white symbols) light as determined by RNAseq. Data points show the mean transcript levels of 3 independently grown cultures ± S.E.M.

In addition to protein kinases, several genes which were annotated as either transcriptional regulators or contained phosphorylation-accepting domains were identified as showing ‘colour-specific’ expression patterns. *SYNPCC7002_A0171, SYNPCC7002_A0364* and *SYNPCC7002_A1224* were strongly upregulated in orange light, whereas *SYNPCC7002*_A1832 was strongly upregulated in blue light (Figure 6-14).

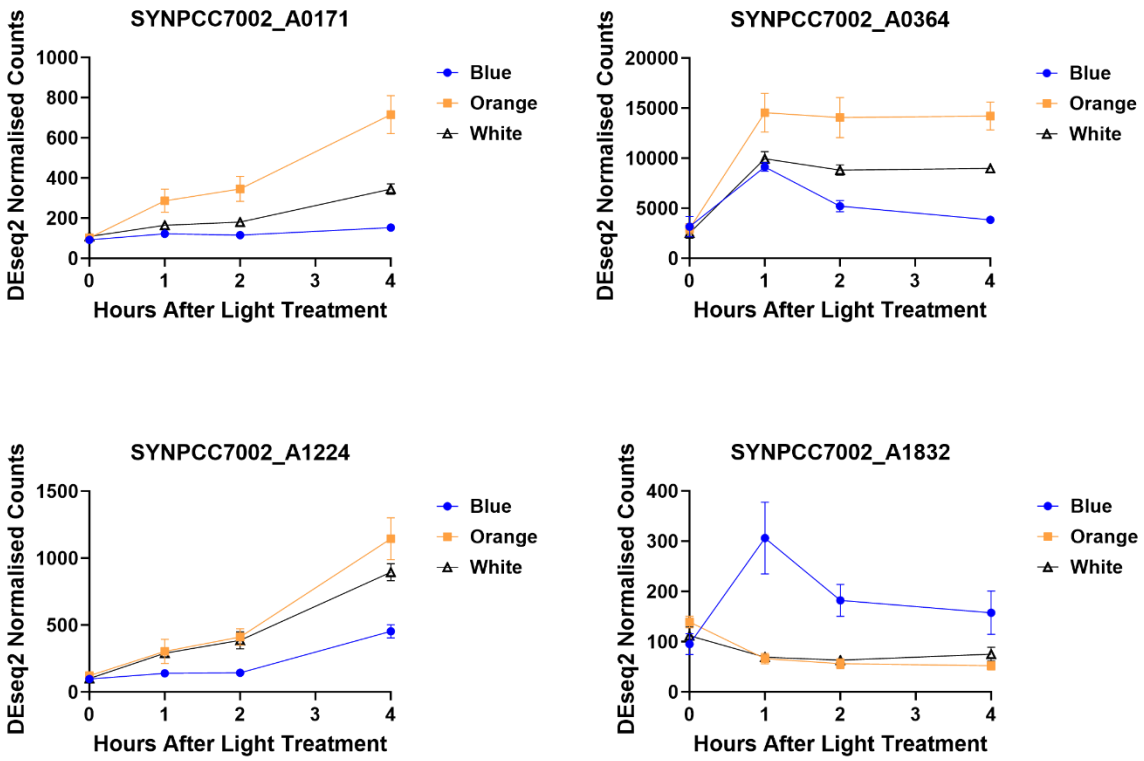


Figure 6-14. RNAseq expression profiles of genes containing transcriptional regulatory functions. DEseq2 normalised counts of four selected genes from cultures grown in blue (blue symbols), orange (orange symbols) and white (black and white symbols) light as determined by RNAseq. Data points show the mean transcript levels of 3 independently grown cultures ± S.E.M.

6.3.6 Identification of wavelength-specific promoter motifs

To identify groups of potentially co-regulated genes with similar expression patterns, hierarchical clustering with dynamic tree cutting was performed on the RNAseq dataset. For this approach, the Likelihood Ratio Test (LRT) was used to identify differentially expressed genes. This approach is especially useful for analysing data from timecourse experiments. LRT tests for statistical significance differences at any level of the factor, i.e. any light treatment or time point. As this approach only uses p-values, a more stringent adjusted p-value of < 0.005 cut-off was applied to reduce the number of genes identified. The dynamic tree-cutting method results in an optimal number of clusters to allow inter-cluster variability to be greater than intra-cluster variability. This resulted in the division of the differentially expressed genes into 9 clusters ranging between 21 to 62 genes. The LRT test and subsequent scaling and clustering of the data were performed by Dr Graham Hamilton in Glasgow

Polyomics. The expression profiles of all genes in cluster 7 are shown as an example (Figure 6-15).

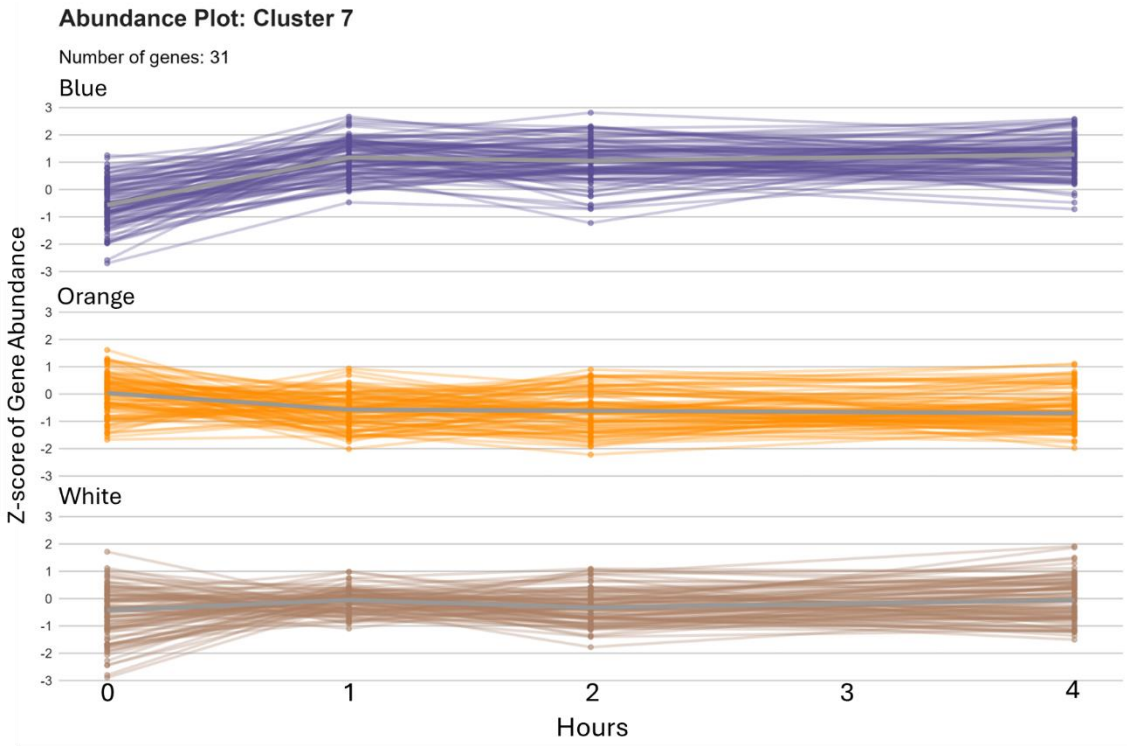


Figure 6-15. Gene expression profiles from cluster 7 generated by hierarchical clustering. Gene expression profiles of 31 selected genes from cultures grown in blue (blue symbols), orange (orange symbols) and white (brown symbols) light as determined by RNAseq. Z-scores are provided as the data was scaled before clustering to remove the skewing effect of highly expressed genes. This plot was generated by Dr Graham Hamilton from Glasgow Polyomics.

Next, the 9 clusters were used in a motif analysis to identify whether any common DNA motifs were present within each cluster. In addition, the lists of genes identified as being differentially expressed in blue or blue and white light (108 genes) or orange and orange or white light (316 genes) were also analysed. The analysis was performed using two tools from the Multiple EM for Motif Elicitation (MEME) suite. From a user-defined set of DNA sequences, MEME identifies novel un-gapped motifs that are enriched in comparison to a set of background sequences (Bailey and Elkan, 1994). The standard settings of the algorithm will identify 3 motifs of between 6-50 bases and work best when providing less than 50 test sequences. Sensitive, Thorough, Rapid, Enriched Motif Elicitation (STREME) performs the same function but works best when providing more than 50 test sequences and is more sensitive than MEME (Bailey, 2021).

The 500 bp region directly upstream of genes of interest was extracted from the relevant test sets of genes and several random sets using a custom R script developed by Dr Graham Hamilton. 500 bp was selected as most Syn7002 intergenic regions ranged between 50-500 bp. While this 500 bp will also include the 3' regions of some genes and their terminator regions, these will also be present in the control sequences and should not affect motif reporting. Choosing a shorter upstream region would risk missing potential motifs contained in longer promoters. These upstream regions were then used as input into both MEME and STREME. MEME allows motif discovery to be performed using two methods, discriminative mode, or differential enrichment mode, each of which uses a different type of algorithm for motif discovery. All test sets were analysed using both methods. The test sets included the 9 clusters generated by hierarchical clustering as well as the groups of 'blue-specific' genes and 'orange-specific' genes. The control sets consisted of randomly chosen genes from the Syn7002 genome. Control sets were used which were of similar size to each of the test sets i.e. 50, 100 or 300. Furthermore, all test sets were also run against the background control of all Syn7002 genes.

None of the tests resulted in the identification of a motif with an E-value < 0.05. The only motif which showed any degree of statistical significance was discovered in the 'orange-specific' gene set. This motif, ATCTCGAA, produced a p-value of 0.026 (Figure 6-16). However, the E-value, which is the p-value multiplied by the number of motifs returned by the algorithm, was 0.11. Furthermore, this motif only appeared when tested against one control set. Testing against a second random control set, as well as against all Syn7002 genes did not result in repeated identification of the motif. Therefore, it is likely that this motif was a false positive.

Motif	Logo	RC Logo	P-value	E-value	Sites
1-ATCTCGAA			2.6e-002	1.1e-001	24 (7.6%)

Figure 6-16. Identification of a possible 'orange-specific' promoter motif using STREME. The ATCTCGAA motif was identified when comparing the 'orange-specific' gene set identified by RNAseq to a random control set. RC logo – reverse complement logo; Sites – the number of occurrences of the motif in the test set. This figure was generated using the Sensitive, Thorough, Rapid, Enriched Motif Elicitation (STREME) tool (Bailey, 2021).

6.4 Discussion

6.4.1 Blue versus orange light effect on Syn7002 growth

First, the effect of blue and orange light on Syn7002 growth was assessed before performing RNAseq. Orange light resulted in significantly faster growth, possibly due to the overlap in the emission and absorption spectra between the orange LEDs used (λmax = 590 nm) and C-phycocyanin (Figure 6-2). C-phycocyanin is the most abundant pigment produced by Syn7002, which absorbs strongly between 550-650 nm. Similarly, Syn6803 grow significantly faster in orange (625 nm) or red light (660 nm) than in blue light (450 nm) (Luimstra *et al.*, 2018). Thus, Syn6803 cannot use blue light as efficiently for photosynthesis. This is due to excess energy being generated at PSI versus PSII under blue light, which is not absorbed by phycobilisomes. In Syn6803, most phycobilisomes associate with PSII, which counteracts the fact that the cells generally contain 2-5 times more PSI than PSII. Also, on average, PSI contains around three times more chlorophyll *a* molecules than PSII. Thus, the funnelling of photons in the orange/red region to chlorophyll *a* via phycocyanin maintains a balance in energy generation between PSI and PSII and preserves linear electron flow. It seems likely that this would also explain the results observed in Syn7002.

6.4.2 'Blue-specific' and 'Orange-specific' genes

With regards to the identified 'blue-specific' and 'orange-specific' genes, it is unclear whether their differential expression is controlled directly by wavelength-specific light absorption or indirectly by the effects of blue and orange light on photosynthetic efficiency and energy generation. For example,

orange light results in significantly faster growth in Syn7002 than blue light (Figure 6-2). Therefore, some genes which are differentially expressed in orange light may be regulated by increased concentration of components of carbon metabolism such as 3-PGA. Similarly, decreased photosynthetic efficiency and energy generation in blue light, rather than blue light absorption by a photoreceptor, may trigger increased expression of PSII genes as is observed in Syn6803 (Luimstra *et al.*, 2020). In the present work, cultures were grown in white light for four days before initiation of the blue or orange light treatments. The short duration of these light treatments (4 hours) is unlikely to cause indirect effects from slow growth or long-term energy deficiency.

Many of the identified 'blue-specific' or 'orange-specific' genes displayed expression profiles which could be reasonably expected based on gene function. For example, *SYNPCC7002_A2032* encodes gamma-carotene 1' hydroxylase (*cruF*), an enzyme which catalyses myxoxanthophyll biosynthesis. *SYNPCC7002_A2032* was strongly upregulated in blue light but not in orange or white light (Figure 6-11). Myxoxanthophyll is a blue Light-absorbing carotenoid which plays a role in photoprotection (Zhu *et al.*, 2010). Another example is *SYNPCC7002_A0157*, which encodes photosystem II D1 subunit PsbA-II (Qb protein) (psbA-II). *SYNPCC7002_A0157* is strongly upregulated in blue light but not orange or white light (Figure 6-12). As discussed above, in Syn6803 blue light illumination creates an energy excess at PSI which reduces growth rates (Luimstra *et al.*, 2018). Therefore, it is likely that this upregulation in blue light in Syn7002 is in part an effort to alleviate the energy imbalance between PSI and PSII.

One gene with a possible interesting link between its expression pattern and function is *SYNPCC7002_A1469*, a Secreted and surface protein containing fasciclin Like repeats. Fascicilin Like repeats, also known as FAS1 domains, are ancient domains involved in cell adhesion. *SYNPCC7002_A1469* expression was strongly upregulated upon blue light illumination, with no change under orange or white light (Figure 6-11). It could be hypothesised that *SYNPCC7002_A1469* is involved in a blue light-induced cell aggregation or flocculation process. Indeed, previous work in the Amtmann lab has implicated the role of blue and green LED light in enhancing Syn7002 flocculation (Mihnev, unpublished). Furthermore,

blue light has also been demonstrated to enhance flocculation in Syn6803 through the action of the Cph2 photoreceptor and type IV pili (Conradi *et al.*, 2019). A role of *SYNPCC7002_A1469* in blue light-controlled flocculation should be tested experimentally in the future.

Many of the identified orange-specific genes have roles in photosynthesis. *SYNPCC7002_A1000,* strongly downregulated under orange light (Figure 6-12), encodes magnesium-chelatase subunit H which is responsible for the insertion of magnesium into the protoporphyrin IX ring of chlorophyll *a*. In contrast, *SYNPCC7002_A2347, which* encodes a light-independent protochlorophyllide reductase (*chlL*) involved in chlorophyll biosynthesis, is upregulated under orange light. *SYNPCC7002_A1821,* which encodes the phycobilisome degradation protein (*NblA*), is downregulated under orange light. *SYNPCC7002_A2620* encoding PSI reaction center subunit XI *(psaL) and* demonstrates strong upregulation under orange light. Finally, *SYNPCC7002_A0172* encodes NADH2 dehydrogenase (plastoquinone) chain 5 (*ndhF3*) which is involved in high-affinity CO_2 uptake (Klughammer *et al.*, 1999).

Several of the genes identified as showing differential expression in blue light have no currently known function. *SYNPCC7002_A0016* encodes a glyoxalase family metalloenzyme with no known function. *SYNPCC7002_A2481, SYNPCC7002_A0267 SYNPCC7002_A1490* and *SYNPCC7002_A1509* are all hypothetical proteins. Future improvements to gene annotation in Syn7002 may yield interesting insights regarding the function of these genes.

6.4.3 Protein kinases and transcriptional regulators

Several of the 'blue-specific' and 'orange-specific' genes were annotated as protein kinases or transcriptional regulators. *SYNPCC7002_A0049,* which encodes a two-component hybrid sensor and regulator was strongly upregulated under orange light (Figure 6-13). *SYNPCC7002_A0049* contains a histidine kinase domain, a CheY Like superfamily domain which accepts phosphorylation and CheW domains which are often involved in chemotaxis. *SYNPCC7002_A0049* appears to be contained in a gene cluster with four other genes also implicated in chemotaxis *(SYNPCC7002_A0045-SYNPCC7002_A0048)*. However, only *SYNPCC7002_A0049* showed a colour-specific expression pattern. One reasonable

hypothesis is that *SYNPCC7002_A0049* could be involved in regulating Syn7002 phototaxis in response to orange or blue light. However, *SYNPCC7002_A0049* is not predicted to encode a GAF domain and thus would require an upstream photoreceptor. Indeed, *SYNPCC7002_A0048* is predicted to encode 6 GAF domains, a very similar structure to the phototaxis-controlling photoreceptor PixJ of Syn6803 (Burriesci and Bhaya, 2008).

Like *SYNPCC7002_A0049*, *SYNPCC7002_A0900* also encodes a two-component hybrid sensory kinase but additionally contains a GAF domain. *SYNPCC7002_A0900* expression remained constant across 4 hours of blue light but increased significantly under orange light (Figure 6-13). A computational analysis in the Amtmann lab, which built a model to try and predict the colour sensitivity of GAF domains in Syn7002 based on their sequence, identified *SYNPCC7002_A0900* as most likely belonging to that of the *CikA* Like family (Mihnev and Amtmann, 2022). This family of proteins is involved in circadian clock regulation, but not all are photoreceptors (Narikawa *et al.*, 2008). *SYNPCC7002_A0900* also contains a CheY Like superfamily domain suggesting its activity can be modulated by phosphorylation. In the case of both *SYNPCC7002_A0049* and *SYNPCC7002_A0900*, it is interesting that proteins which contain CheY Like domains to modulate their activity in response to phosphorylation also appear to be regulated transcriptionally in orange light.

Another transcriptional regulator identified was *SYNPCC7002_A0171*, which encodes a transcriptional regulator of the rubisco and carboxysome operon (*RbcR*) and belongs to the *LysR* family of transcriptional regulators. *SYNPCC7002_A0171* expression increases significantly under orange light (Figure 6-15). *LysR*-type transcriptional regulators in other cyanobacteria are known to bind small molecules of metabolism as ligands (Bolay *et al.*, 2022). Interestingly, *SYNPCC7002_A0172*, one of the identified potential 'orange-specific' reporter genes, shows a very similar expression pattern and appears to be part of the same operon (Figure 6-12).

Two sigma factors were identified as showing interesting potential 'colour-specific' expression patterns. *SYNPCC7002_A0364* encodes a Group II sigma-70 type sigma factor (*SigC*) of the *RpoD* family. *SYNPCC7002_A0364* expression is strongly upregulated after 1 hour of orange light and remains elevated across

the four-hour time course (Figure 6-14). SigC expression is significantly elevated within 2 hours of onset of carbon or nitrogen stress before decreasing back to baseline after around 16 hours (Caslake *et al.*, 1997). It is therefore interesting that this sigma factor, which regulates the expression of a large number of genes, appears to be strongly upregulated by orange light. *SYNPCC7002_A1832* encodes an alternative sigma factor *RpoD*. *SYNPCC7002_A1832* expression is strongly upregulated by blue light illumination (Figure 6-14). No literature is available discussing the function of this gene in Syn7002, however, *RpoD* is known to be the essential sigma factor in bacteria which controls the expression of a range of housekeeping genes.

Finally, *SYNPCC7002_A1224* encodes a response regulator receiver domain protein which belongs to the CheY Like superfamily. *SYNPCC7002_A1224* expression increases to a greater extent under orange and white light than blue light. As it contains a REC domain, SYNPCC7002_A1224 could be controlled by kinase-mediated phosphorylation in a phosphor-signalling pathway. However, the function of this protein is unknown. *SYNPCC7002_A1224* would be an interesting target for gene knockout and subsequent pathway characterisation. Although it does not have a DNA-binding domain, *SYNPCC7002_A1224* could potentially bind to other transcriptional regulators to modulate their function.

6.4.4 Failure to identify wavelength-specific promoter motifs

Many genes have been identified in this work which display 'colour-specific' expression profiles. However, the mechanism by which groups these genes may be regulated is unclear. The expression of groups of genes may be controlled by the binding of a transcriptional regulator to a 'colour-specific' motif contained within their promoters. Thus, groups of genes which are co-regulated would show similar expression profiles to each other. Hierarchical clustering was applied to the RNAseq dataset in an attempt to identify groups of genes which showed similar expression patterns. However, analysing these clusters and the identified 'blue-specific' and 'orange-specific' genes using the MEME and STREME motif discovery algorithms failed to identify any common motifs with high confidence. The ATCTCGAA motif was identified in the 'orange-specific' gene set, but identification could not be replicated when testing against a second random gene set or all Syn7002 genes. Identification of a wavelength-

specific promoter motif would be useful for identifying new wavelength-specific signalling pathways. The failure to identify any relevant motifs could be due to the algorithms being used not being sensitive enough to identify a very low number of motifs. Wavelength-specific promoter motifs may be contained in the Syn7002 genome but are present at too low a frequency to be detected by the MEME or STREME algorithms.

6.4.5 Future work

Due to time constraints, the next phase of the proposed method of identifying components of wavelength-specific signalling pathways using qRT-PCR could not be completed. Nevertheless, the new RNAseq dataset provides a solid foundation to approach this in the future. Target genes such as the protein kinases and transcriptional regulators identified in Figure 6-13 and Figure 6-14 could be knocked out or knocked down using CRISPRi and the expression of the 'reporter genes' identified in Figure 6-11 and Figure 6-12 measured by qRT-PCR. Loss of responsiveness to the specific wavelength in the mutant compared to the wildtype would suggest that the gene is required for the response. Then, considerable thought would be required to decide on a threshold or cut-off by which the expression patterns between the knockout and wild-type strains differed enough to be considered significant.

As RNAseq analysis is expensive, only 36 samples were submitted for RNAseq in this study: 3 colour treatments x 4 timepoints x 3 replicates. However, culture samples were also harvested at the 8 hour, 24 hour and 26 hour timepoints. Future work could perform RNAseq on these remaining samples which may identify further wavelength-specific transcripts. As the 24-hour and 26-hour time points are at the same point in the circadian cycle as the 0 and 2-hour time points (Figure 6-3), it would be interesting to make a comparison between these samples. It is possible that the first period of orange or blue light 'primes' some genes to demonstrate even higher expression in the 26-hour samples. qRT-PCR could also be used to measure the expression of genes of interest in the 8, 24 and 26-hour samples.

Future work could also measure transcript levels in samples grown under different intensities of orange and blue light. This could help us better

understand specifically how ‘blue-specific’ and ‘orange-specific’ genes are regulated. For example, the expression of some genes may be regulated in a binary ON/OFF manner by the presence or absence of a light wavelength and thus show similar expression profiles under a range of orange light intensities. However, the expression of other genes could be regulated in a dose-response manner, with a positive correlation between light intensity and gene expression.

Another interesting experiment would be to perform RNAseq on cultures grown under different light wavelengths. The Xanthella photobioreactors used in this work contain three other colours of LED: cyan, green and red. While the colour sensitivities of most of the GAF domains in Syn7002 are unknown, work in the Amtmann lab confirmed the presence of a blue/orange photoreceptor, as well as a predicted orange/red photoreceptor (Mihnev and Amtmann, 2022). It is likely that the colour sensitivities of some of the other 20 predicted GAF domains overlap with the cyan, green or red LEDs. Furthermore, GAF domains could also respond to other wavelengths including UV, violet, yellow, far-red and near-infrared. Commercially available LEDs can emit any of these wavelengths.

7 General discussion

7.1 Project summary

Cyanobacteria present a low-carbon alternative to both biochemical synthesis from fossil fuel-derived feedstocks and the use of valuable food crops to power microbial fermentation. However, significant improvements in culture growth rates and product titres are required to make cyanobacterial bioprocesses economically competitive. One avenue to improving these product titres is through the development and application of better tools for controlling transgene expression and thus the ability to dynamically regulate flux through metabolic pathways by precisely controlling the timing of enzyme expression. This work addressed this issue, by applying existing optogenetic tools in the industrially relevant strain Syn7002, characterising their performance in depth, and applying them for product synthesis. A foundation was also laid for the future discovery of new optogenetic components through analysing the Syn7002 transcriptome under blue and orange light. A summary of the new insights generated and the questions remaining after completion of this project is shown in Table 7-1. The following sub-chapters will explore the future opportunities and challenges for optogenetics and industrial biotechnology in cyanobacteria.

Table 7-1. Summary of new insights and unknowns after project completion. (overleaf)

Chapter	New insights from this work	Unknowns
CcaS/CcaR	- CcaS/CcaR system is functional in Syn7002. - Characterisation of CcaS transcriptional activation. - System characterisation under multiple red/green light periods (i.e. reversibility). - System characterisation under light/dark cycles. - Natural variances in plasmid copy number have minimal effect on system performance. - Truncation of pCpcG2 plus modification of the -10 hexamer improves system performance.	- Why do multiple red/green light periods result in higher transcriptional output than the first green light period? - How can CcaS system performance be improved further (altering CcaS/CcaR expression, circuit amplifiers etc.)? - What are the kinetics of CcaS/CcaR phosphorylation? - Does CcaS expression affect Syn7002 physiology? - Which light intensity is required to maintain strong CcaS activation in dense stationary phase cultures?
YF1/FixJ	- YF1/FixJ system exhibits poor performance in Syn7002. - YF1/FixJ genes are transcribed in Syn7002. - FMN co-factor availability is not a limiting factor. - Increasing YF1/FixJ predicted RBS strength 20x has minimal effect.	- Are functional YF1/FixJ proteins expressed in Syn7002? - Can YF1 phosphorylate FixJ in Syn7002 *in vivo*? - Are there additional unknown regulatory factors required for transcriptional output from pFixK2? - Are LOV domains functional in Syn7002?
Applying optogenetics for product synthesis	- CcaS system can be used to control mannitol production in Syn7002 via mpusfus expression. - L-alanine production could not be achieved using alaD/alaE. - alaD protein expression is increased under white light versus red light.	- Is mannitol titre variance due to genetic instability? - Can varying light regimens improve the synthesis of products other than mannitol via optogenetics? - Why is L-alanine undetectable when alaD and alaE proteins are expressed under both constitutive and optogenetic control?
Identification of wavelength-specific transcripts	- Identification of genes differentially expressed in blue and orange light including kinases and transcriptional regulators. - Identification of candidate qRT-PCR reference genes.	- Are there promoter elements which cause specific expression in blue/orange light? - Can blue/orange-specific genes be used with gene knockouts to identify wavelength-specific signalling components?

7.2 Potential future improvements to the CcaS/CcaR system

To further improve the performance of the CcaS system in Syn7002, additional modifications could be made based on previous successes in other organisms. One such modification includes altering the expression level of CcaS and CcaR in E. coli (Schimdl et al., 2014). Another approach could be to delete the two PAS domains from CcaS, which was shown to improve the dynamic range in both E. coli and Syn15041c (Schimdl et al., 2014; Kobayashi et al., 2019). Another potential avenue for future work could be to combine the optimised pCpcG2 promoter characterised in the present work with that of the CcaS T7-based optogenetic system used in Syn6803 (Shono et al., 2020). This approach used the CcaS system to control the expression of the T7 RNA polymerase from the pCpcG2 promoter. GFP was placed under the control of the T7 promoter, enabling green light activation of GFP expression via the T7 RNA polymerase. This system resulted in a 10-fold increase in GFP expression compared to the original CcaS system under both red and green light, whilst maintaining a 6-fold dynamic range. Therefore, combining the pCpcG2 optimised promoter developed in the current work with the T7-based CcaS system could result in even greater system activation under green light.

Another modification which could result in improved system performance or altered properties is modular transcription factor domain swapping. In E. coli, the DNA-binding domains of OmpR family transcription factor domains can be swapped with other OmpR family members, whilst retaining similar or improved levels of transcriptional output (Schmidl et al., 2019). Indeed, the DNA binding domain of CcaR was fused to the REC domain of the transcriptional regulator TorR. This allowed transcription from pCpcG2 to now be controlled by the action TMAO-sensing histidine kinase TorS, which normally exhibits kinase and phosphatase activity towards the transcription factor TorR. As the present work has shown the CcaS/CcaR/pCpcG2 system is functional in Syn7002, this strategy could be adapted to allow transcription from pCpcG2 to be controlled by other photoreceptor histidine kinases which respond to light wavelengths other than green or red. This would allow the use of two-component systems for optogenetics for which either the output promoter is unknown or is non-functional in Syn7002. Alternatively, the REC domain of CcaR could be fused to the DNA-binding domain of other OmpR family response regulators and paired

with the corresponding promoter. This would still allow green/red light control of gene expression but potentially with improved dynamic range.

For optimisation of CcaS system performance, this work focused solely on the engineering of the pCpcG2 promoter. Due to limited resources, this consisted of targeted mutations which were hypothesised to have beneficial effects on performance. However, while this approach resulted in one construct with improved performance, in future a better strategy may be to make random mutations in the promoter. This would involve high throughput random mutagenesis and screening to test a much larger panel of potential promoter variants. Another strategy could be to place multiple copies of pCpcG2 upstream of the target gene, which could potentially increase the maximal output of the system. However, even if successful, this could trigger genetic instability due to the presence of repeated sequences and would increase the genetic footprint size of the system.

One disadvantage of using the CcaS optogenetic system is that the genetic components of the system have a combined length of around 4 Kb. This is not an issue when controlling the expression of one or two genes, however, when assembling complex metabolic pathways, the large size of the system could pose problems with cyanobacterial transformations. As was experienced during this work, the larger the size of the region to be inserted into the cyanobacterial genome, the less efficient the transformation reaction. Inserting sequences of greater than 7 Kb often resulted in transformation reactions yielding either no colonies or very few. However, there is scope to reduce the size of the system, as it has been shown that miniaturised CcaS with a double PAS domain deletion is functional in cyanobacteria, which would reduce the system size from 4 Kb to 3.1 Kb (Kobayashi et al., 2019). Furthermore, the native CcaS promoter region is large in respect to many bacterial promoters at 0.65 kb. Truncation of this promoter could reveal the minimal promoter size required for sufficient CcaS expression, or it could be replaced with a smaller heterologous promoter. Indeed, a constitutive promoter library containing an over 100-fold dynamic range has been characterised in Syn7002 in which all promoters are under 40 bp (Markley et al., 2015).

When using the CcaS system to control expression of GFP, after 24 hours GFP fluorescence had only decreased by around 20% (Figure 3-7). In practical applications, a more rapid decrease in target protein abundance would be desirable. To increase the speed of the downregulation of the target protein in red light, a degradation tag could be added. However, while this would increase the speed of protein turnover upon switching to red light, it would also likely reduce the maximal amount of protein that could be synthesised under green light. A better approach could be to use an additional optogenetic system which allows light-activated protein degradation under any light colour other than green. Several blue Light activated protein degradation systems have been applied in eukaryotes such as S. cerevisiae, which contain well-understood ubiquitin-proteasome systems. This enables the pairing of the target protein with light-controlled degron motifs. However, very few light-induced protein degradation systems have been applied in bacteria. One system which has been used in E. coli involves the blue light-regulated reconstitution of a TEV protease, which cleaves target proteins containing a specific target sequence (discussed in section 1.5.3 Optogenetics) (Komera et al., 2022). Indeed, a TEV protease system has been adapted for use in Syn6803, which means it could potentially also be applied to Syn7002 (Zhang et al., 2021).

It is unclear whether the expression of the CcaS system has any effect on cellular physiology by regulating endogenous promoters other than pCpcG2. The CcaS system controls chromatic acclimation in Syn6803. A BLAST search for similar green/red light photoreceptors did not result in the identification of a homologous system in Syn7002. However, a green/red chromatic acclimation system likely exists in Syn7002. Indeed, viewing Syn7002 cells grown under red versus green light results in visible colour differences between the two cultures, most likely due to a chromatic acclimation response. If a homologous system does exist, it raises the possibility of cross-talk. There is very little data available regarding cross-talk with native systems in optogenetic applications. Future work could use RNAseq to determine the effect of CcaS expression on the Syn7002 transcriptome versus wild-type cells.

One unanswered question is whether the CcaS system can be activated to a similar level in very dense cultures as in the early exponential phase cultures used mostly in this work. It would be useful for potential industrial applications

to understand the minimum light dose required at high culture densities to allow strong system output under green light. This work demonstrated that in Syn7002, green light resulted in significantly slower growth than red light (Figure 3-4). This might be due to red light being more efficiently harvested for photosynthesis than green light. As a consequence, green light might penetrate deeper into the culture, as fewer photons will be absorbed by cells at the culture perimeter and converted to energy or quenched. The exact dark reversion half-life of CcaS has not been calculated but appears to be at least several hours in duration (Olson et al., 2017). Therefore, it would be reasonable to assume that the CcaS system has a much longer dark reversion half-life than many other photoreceptors commonly used in optogenetic systems. The combination of poor green light absorption by Syn7002 and the long activation state half-life of CcaS would still allow effective activation of gene expression under the light Limiting conditions of a dense stationary phase culture. Furthermore, targeted modifications to residues close to the chromophore binding site often can affect the activation half-life of the photoreceptor (Zhao et al., 2021). This approach could be used to increase or decrease the half-life of the CcaS-activated signalling state.

7.3 Considerations and opportunities for applying optogenetics in cyanobacteria

7.3.1 The advantage of cyanobacterial cultivation systems for optogenetics

Applying optogenetics in industrial-scale cultivation systems relies on all cells within the culture absorbing the necessary wavelengths. Heterotrophic microbes are usually grown in conventional steel stirred-tank bioreactors, which would have to be redesigned to allow lights to be inserted inside the vessel for adequate light penetration. As well as introducing additional design and manufacture costs, this may be incompatible with the stirring mechanisms and internal design of these bioreactors. In contrast, cyanobacteria are often grown in tubular or flat-panel photobioreactors at industrial scale. These systems maximise the surface area to volume ratio required for optimal light penetration and are therefore well suited for the level of light penetration required for optogenetics. Filters which allow manipulation of the wavelength spectra are

cheap and could be used to cover flat panel bioreactors or open pond systems to switch on and off optogenetic tools as required.

7.3.2 The complexity of applying multiple optogenetic systems simultaneously

There are various factors to consider for future applications of optogenetics in cyanobacteria. One is the use of multiple optogenetic tools within the same cell, which is often required for advanced control of multiple enzymatic steps in metabolic pathways. While expressing one optogenetic system is reasonably straightforward, combining two or three tools in one organism increases complexity. Cross-talk between the optogenetic system and the native regulatory circuits can already impact the functioning of a single heterologous system. Introducing multiple systems not only compounds this risk but also introduces the possibility of cross-talk between the heterologous optogenetic systems. This could have undesirable effects on cell physiology or the efficiency of optogenetic control. Some optogenetic systems, such as those based upon the plant phytochrome PhyB, also respond to additional signals such as temperature which must be considered (Legris *et al.*, 2016). Furthermore, photoreceptors generally have fairly wide absorption spectra and can have primary and secondary absorption peaks. Thus, combining photoreceptors with overlapping absorption spectra should generally be avoided. The lack of optogenetic systems which respond to wavelengths in the yellow/orange area of the UV-Vis spectrum further limits the use of multiple systems in one organism.

Another complication when expressing multiple optogenetic systems in one organism is the large number of genetic components required. For the expression of three optogenetic systems in *E. coli*, 18 genes each with different promoters, RBS and terminators were expressed across 4 plasmids (Fernandez-Rodriguez *et al.*, 2017). This level of genetic engineering can place a high metabolic burden on cells, which can lead to genetic instability and culture heterogeneity during industrial fermentation. One-component optogenetic systems contain a clear advantage over two-component systems in this context. To sustain the desired levels of performance, maintaining the stoichiometry of these genetic parts is important. However, this could easily be disturbed during cell division, with daughter cells containing non-optimal ratios of the photoreceptors and transcription factors.

7.3.3 Low transcriptional output from optogenetic systems

Perhaps the key disadvantage of many currently available optogenetic tools is that they often induce lower levels of transcriptional output than the strongest constitutive or chemically inducible promoters. Some engineered tools offer dynamic ranges on the order of several hundred-fold between activated and deactivated states (Ong *et al.,* 2018). However, this is often due to extremely low transcriptional output under non-inducing light conditions rather than an increase under inducing conditions. A high level of transcription is required under the inducing condition to facilitate protein accumulation and maximal product synthesis. This problem is compounded by the effect of high culture density on optogenetic tool performance, where light cannot penetrate deep enough into the culture for a sustained period to trigger homogenous activation or deactivation of the photoreceptor. While several photobioreactors have been designed for effective light penetration, future strain improvements which facilitate increased cell densities are likely to make this an issue.

Several optogenetic systems have been developed which enable high levels of target gene expression, even at high culture densities. One strategy is to place a transcriptional repressor, rather than an activator, under optogenetic control, causing transcriptional repression in light and activation in the dark (Zhao *et al.*, 2018). This allows much greater transcriptional activation when culture density is high, although it hinders the ability to reduce transcription using light. This complicates the application of dynamic gene expression control strategies, which is the unique advantage of optogenetics over chemically inducible systems.

An alternative solution is the use of optogenetic circuit amplifiers. The EL222 one-component optogenetic system is activated by blue light and repressed in darkness. EL222 has been used to control the expression of the strong transcriptional activator Gal4p in *S. cerevisiae*, which then binds to the pGAL promoter to drive the expression of GFP (Zhao *et al.*, 2021). This circuit amplifier resulted in a transcriptional response 7-fold higher than when using the EL222 system to control the expression of GFP alone without Gal4p. The maximal transcriptional output achieved by this approach was greater than that of some of the strongest constitutive and inducible promoters used for *S.*

cerevisiae metabolic engineering. Furthermore, in this work, the authors performed targeted mutagenesis of EL222 residues known to increase the photoreceptor-activated state half-life from 30 seconds to 300 seconds. This was combined with the expression of fusion protein consisting of the Gal80p repressor, which binds to the pGAL promoter, fused to a blue light-induced photosensitive degron motif to maintain low pGAL activity in the dark (Figure 7-1). The new system, named OptoAMP, with all the combined improvements, resulted in high GFP expression even at very high cell densities. In a situation where brief light pulses of 5s ON / 95s OFF were used and only 7% of the bioreactor surface area was illuminated, OptoAMP enabled high GFP expression. In contrast, no GFP expression could be detected when using only EL222 to control GFP expression without the circuit amplifier or additional modifications. These results at very promising for applying optogenetics in large-scale cultures with very high cell densities. Circuit amplifiers could be applied to the CcaS system to increase the transcriptional output in Syn7002.

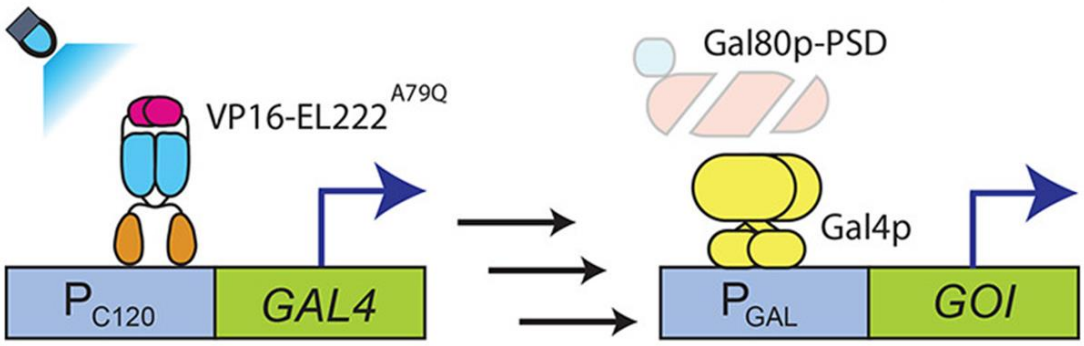


Figure 7-1. OptoAMP optogenetic circuit amplifier. Upon blue light illumination, the Gal80p repressor is degraded by the action of its blue light photosensitive degron (PSD). EL222, boosted by the activation half-life extending A79Q mutation, activates transcription of the Gal4p transcription factor. Gal4p transcribes the gene of interest (GOI) downstream of the pGAL promoter. Figure from Zhao *et al.*, 2021.

7.3.4 Exploiting optogenetics for advanced metabolic pathway control strategies

Several recent studies have highlighted the application of optogenetics for advanced metabolic pathway control strategies. The OptoAMP system described above, which incorporates a circuit amplifier, has been used to increase the

titre of several different products (Zhao *et al.*, 2021). The use of optogenetics facilitated the splitting of the fermentation protocol into three distinct phases: a growth phase with no light induction of target gene expression, an induction phase with very brief light pulses and a production phase with constant light activation of gene expression. Applying this to the synthesis of three different products, the authors demonstrated that the optimal titre for each product was achieved using a different light regimen. For example, maximal lactic acid titre was achieved by illuminating the culture with light pulses in phase 1, followed by constant light in phases 2 and 3. In contrast, the maximum titre of the flavonoid naringenin was achieved using light pulses in phase 1, constant light in phase 2 and no light in phase 3. This type of dynamic control of gene expression enabled by optogenetics shows great promise for increasing product titres.

A more unconventional application of optogenetics for increasing product titres is in the control of cell division. Applying optogenetics in *E. coli* to regulate the expression of DNA replication and cell division-related genes improved biochemical production significantly (Ding *et al.*, 2020). A system was devised to enable a light-regulated increase in cell surface area and a corresponding decrease in cell volume, or vice versa. Shortening the cell cycle resulted in a 39% increase in acetoin titre to 67 g L^{-1}, possibly through smaller cells improving the mass transfer of nutrients. Conversely, lengthening the cell cycle doubled poly(lactate-co-3-hydroxybutyrate) (PLH) production to 14 g L^{-1}, possibly by increasing the mean cell volume and thus allowing more space for intracellular accumulation of PLH granules.

Optogenetics also opens potential avenues for advanced bioproduction techniques. Until recently a hypothetical technology, metabolic cybergenetics uses computers and feedback systems to maintain automated control over culture growth and production (Carrasco López *et al.*, 2020). Enzyme expression is controlled by optogenetics, whilst genetically encoded biosensors provide real-time feedback on aspects of cellular state such as metabolic burden and product toxicity. The fluorescence output from the biosensor is detected and fed back to a computer algorithm which then adjusts the intensity of light being used to illuminate the culture (Figure 7-2). This type of closed Loop control over cultures can result in superior product titres by near-perfectly balancing growth

and production. As a novel technique, metabolic cybergenetics has not been widely applied. However, one recent example is in the optimisation of α-amylase production in *S. cerevisiae* (Benisch *et al.*, 2023). The blue Light responsive one-component system EL222 was used to control α-amylase expression, whilst a genetically encoded biosensor was used to measure the unfolded protein stress response in cells. Using a closed Loop feedback system, the rate of α-amylase expression was modulated by a cybergenetic control system to maintain the unfolded protein stress response at a pre-defined optimal level. This resulted in a 60% improvement in final α-amylase titres versus illuminating the culture with constant blue light.

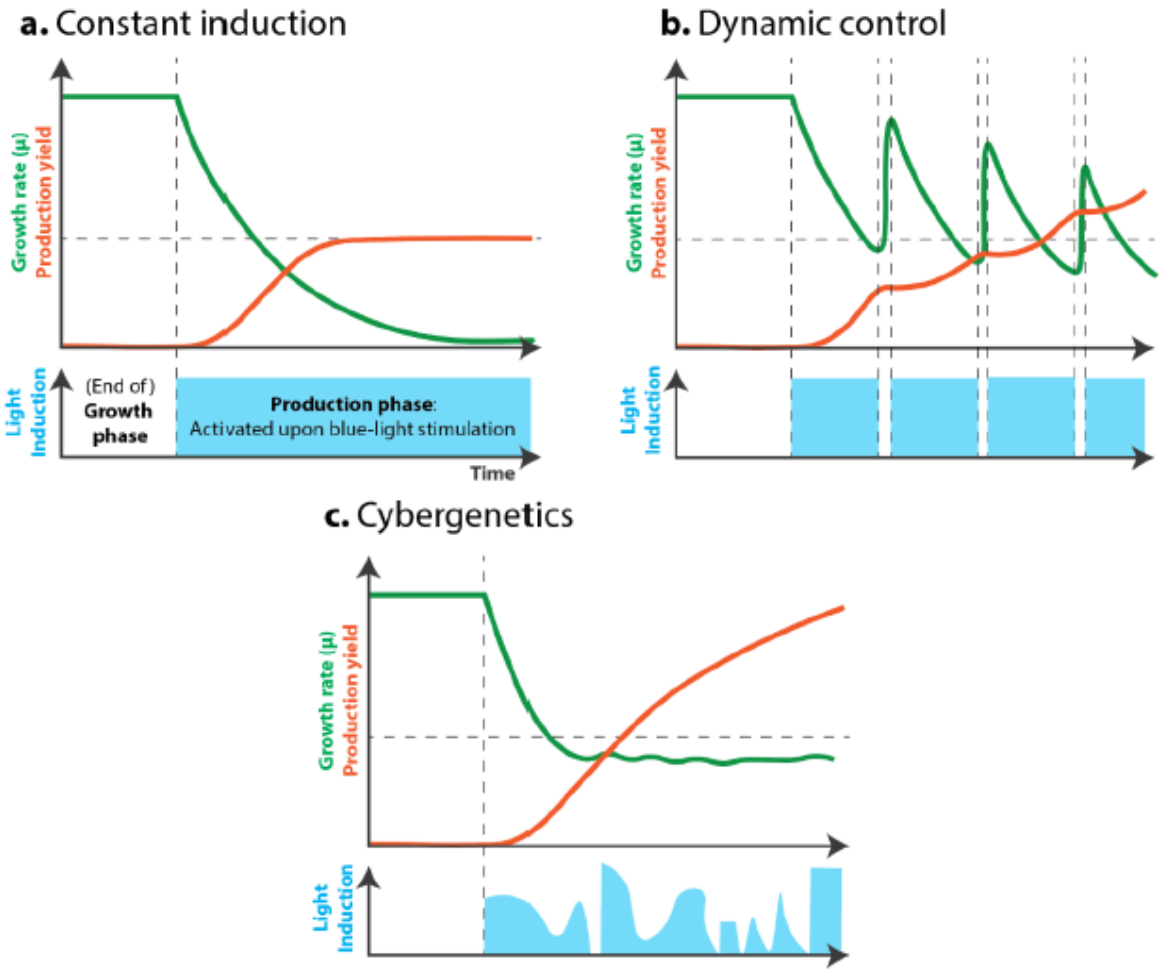


Figure 7-2. Using optogenetics to optimise product titres. As opposed to a dark growth phase and constant light induction phase (a), optogenetics enables reversible dynamic control of target gene expression (b) to increase product titres. Cybergenetics (c) enables even higher product titres by finely balancing growth and production rates through closed Loop feedback systems. Figure from Pouzet *et al.*, 2020.

7.3.5 Optogenetics for controlling growth rates in microbial co-cultures

Recent approaches have exploited microbial co-cultures to improve biochemical product titres. Some compounds with commercial applications require the expression of very complex non-native biosynthetic pathways. Expressing all the required genes and knocking out others in one host can cause a high metabolic burden and poor productivity. In addition, enzymes expressed early in the pathway may interfere with those expressed subsequently by consuming precursors or reaction products. Furthermore, enzymatic reactions can either fail or demonstrate varying levels of efficiency in different species. Thus, splitting the biosynthetic pathway between two or more species to take advantage of each species' strengths can lead to improved product titres (Yuan *et al.*, 2020). However, maintaining control over the growth of these species can prove difficult within a co-culture, with one species often outcompeting the other. Optogenetics is uniquely suited for fine-tuning co-culture composition. For example, in co-cultures of *E. coli* and *S. cerevisiae*, blue light has been used to control the expression of a toxin-antitoxin system in the faster-growing *E. coli* (Lalwani *et al.*, 2021b). This enabled the *E. coli* growth rate to be modulated by blue light intensity which improved co-production of isobutyl acetate and naringenin.

7.4 Facilitating cost-effective, low-carbon production of biochemicals

7.4.1 The relationship between production volume and product value

Currently, industrial production of cyanobacteria and microalgae is mostly limited to low-volume, high-value products such as the pigments astaxanthin, fucoxanthin and phycocyanin. Their use in food products is also often in niche markets, with many products being offered on the market but the market size is small overall (Hachicha *et al.*, 2022). Much work over the previous two decades has focused on photosynthetic production of biofuels. However, due to these being very high-volume, low-value products, economically viable production processes have never been developed. Growing photosynthetic organisms at the vast scale required for biofuel production is challenging in terms of infrastructure and land use, especially when considering the much lower product titres and growth rates versus heterotrophs. In addition, unless using wastewater

or saltwater-based culture media, the freshwater demand at this scale of production would be enormous. For example, global bioethanol production from yeast totals around 110 billion litres per year (Jacobus *et al.*, 2021). Whilst cyanobacteria could provide a more environmentally sustainable alternative for biofuel production in the future, there are huge hurdles to be addressed.

Like biofuels, bioplastics are another high-volume, low-value product for which production has been engineered in cyanobacteria. However, a techno-economic analysis (TEA) identified that under the most optimistic scenario, the PHB production cost in cyanobacteria was still 2-fold higher than the PHB market value (Price *et al.*, 2022). This scenario included a system where multiple other products such as pigments were extracted in a biorefinery approach, wastewater was used as media, anaerobic digestion, a solar power purchase agreement and CO_2 was supplied at cost. Production costs must be reduced significantly to make these types of processes economically viable.

High production costs have resulted in a shift to targeting low-volume, high-value markets for cyanobacterial and algal products. Some products are naturally produced in high volumes in certain species, whilst their production can often be increased by modulating the environmental conditions. One such example is C-phycocyanin production by species of the *Arthrospira* genus, also known as Spirulina. The market size for phycocyanin in 2022 was estimated at $754 million (Future Market Insights, 2023). While the market value can vary 100-fold depending on its purity, lower-value food-grade phycocyanin can cost $130/g (Manirafasha *et al.*, 2017). The use of a two-stage cultivation system in Spirulina, using red and blue LEDs to increase biomass in phase one and then blue LEDs only to stimulate phycocyanin production in phase two, resulted in a final yield of 1.28 g L^{-1} phycocyanin (Lee *et al.*, 2016). Very limited success has been achieved when using heterotrophs such as *E. coli* to produce large amounts of functional phycocyanin with fully attached chromophores (Chen and Jiang, 2019). Furthermore, many synthetic pigments are known to have negative effects on human health and consumers are moving towards 'natural' sources of compounds. Collectively, this has led to the success of cyanobacteria and microalgae in commercial pigment production.

7.4.2 Gas fermentation

Beyond photosynthetic microbes, another competing technology which has the potential to convert greenhouse gases into valuable biochemicals is gas fermentation. Gas fermentation utilises microbes which can convert 1-carbon feedstocks such as CO and CO_2 to biochemicals. These feedstocks are often present in high quantities in flue gases emitted from heavy industries such as steel mills. In a process developed by LanzaTech, the anaerobic acetogen *Clostridium autoethanogenum* converts gas containing 50% CO, 10% H_2, 20% CO_2 and 20% N_2 into acetone and isopropanol in a carbon-negative process (Liew *et al.*, 2022). This process was developed in 125 L bioreactors, with productivity reaching 3 g L^{-1} h^{-1} for both products. This level of productivity is on par with those exhibited by common industrial workhorses for many products like *E. coli* and *C. glutamicum*.

While gas fermentation appears to be another promising method for low-carbon biochemical production, it is also a new and fairly unproven technology at scale. The only biochemical which is currently being produced via gas fermentation at industrial scale is ethanol, which is being produced by *C. autoethanogenum* at a scale of around 90,000 metric tons per annum (Liew *et al.*, 2022). Anaerobic bacteria may be unable to synthesise certain biochemicals in high quantities. Working with anaerobic bacteria requires expensive specialised equipment, including anaerobic chambers and bioreactors which can facilitate gas solubility and distribution. Furthermore, gas-fermenting microbes such as *C. autoethanogenum* require hydrogen as an electron donor, which is not always present in high quantities in flue gases. Flue gases can contain significant quantities of impurities such as sulphur dioxides and nitrogen oxides, which can inhibit bacterial growth (Takors *et al.*, 2018). In the absence of adequate hydrogen in the flue gas, the expensive process of electrolysis using renewable power is the only current clean hydrogen production method. Therefore, it seems sensible to invest in the development of both photosynthetic and gas-fermenting microbes for low-carbon biochemical production. Ideally, in the future, each production system can be applied in different scenarios to best utilise its unique capabilities.

7.5 Opportunities to make photosynthetic production of biochemicals economically feasible

7.5.1 Identifying new species and improving growth rates

To significantly reduce the cost of using photosynthetic microbes for low-carbon biochemical production, a multi-faceted approach is required. One opportunity lies in the identification of novel species which display rapid growth. Currently, Syn11901 appears to be the species which accumulates the largest amount of cell biomass within the shortest timeframe when bubbling cultures with CO_2 (Włodarczyk *et al.*, 2020). However, there are likely other species yet to be isolated which demonstrate even faster growth rates. Isolating new strains from high CO_2 environments such as volcanic springs could result in the identification of species which have rapid carbon assimilation rates.

As well as identifying new species, existing species can be genetically engineered to increase photosynthetic efficiency and carbon assimilation. For example, overexpression of the sodium-dependent bicarbonate transporters SbtA and BicA in Syn7002 increased the biomass production rate by 50% compared to wild type under 1% CO_2 concentration, with a 70% increase under 3% CO_2 concentration (Gupta *et al.*, 2020). Overexpression of these genes in Syn11901 could potentially further improve the fast growth demonstrated by this species. However, the rapid growth of Syn11901 could already be due to improved carbon uptake mechanisms, and therefore the expression of these transporters may not improve growth rates further. Improving carbon uptake strategies and dissecting the mechanisms underlying rapid growth phenotypes in fast-growing species may inform future engineering of other cyanobacterial species for maximal productivity. Some of the mechanisms implicated in the fast growth of Syn7002 compared to many other cyanobacterial species include rapid photosynthetic electron transport and increased transcription of protein translation machinery (Bernstein *et al.*, 2016).

7.5.2 Improving cultivation systems and downstream processing

Improving existing cultivation systems could lead to faster cyanobacterial growth rates. Cell shading reduces growth rates at high cell densities, as light cannot penetrate deeply into the culture. The cells close to the light source absorb

most of the photons but can only use a small fraction of these for photosynthesis. The remainder of the energy is dissipated as heat through mechanisms such as non-photochemical quenching to prevent damage to the cellular photosynthetic machinery (Kirilovsky and Kerfeld, 2016). To address this issue, machine learning has been applied to determine the optimal cell density to minimise cell shading and maximise the growth rate of UTEX 2973 at a given light intensity. This was combined with a semi-continuous cultivation system, which allowed a defined amount of culture to be harvested every 24 hours to maintain the culture at the optimal density. This system resulted in a biomass productivity of 2 g DCW L^{-1} per day, 5-fold higher than the fed-batch culture control (Long *et al.*, 2022). Continuous cultivation of Syn6803 in flat panel photobioreactors has also been reported (Zavřel *et al.*, 2019).

The machine learning approach described above highlights the potential of continuous and semi-continuous cultivation to improve growth rates. However, continuous systems suffer from various drawbacks versus batch cultures. The first is the complexity of balancing diurnal light/dark cycles with the maintenance of cell biomass or specific growth rate at a constant level (Henley, 2019). Another is in the downstream processing of products. In batch systems, the entire culture vessel is emptied, and the solution is passed through sequential unit operations such as centrifugation, filtration, solvent extraction and drying. However, from a process design perspective, continuous systems require more careful planning as the liquid is constantly being drained from the bioreactor. This requires either constant operation of energy-intensive unit operations such as disc stack centrifugation, increasing operational costs, or retaining solutions in holding tanks to be processed downstream in batches. Holding tanks may be useful in some scenarios and could significantly reduce operating costs but extended periods of incubating cells or products in the tank may be detrimental to final product titre or quality. Furthermore, each holding tank and the associated pipes and fittings would further increase capital expenditure. A final disadvantage of continuous systems is the increased contamination risk versus batch cultures. However, all these additional costs could be absorbed if they lead to a significantly higher rate of productivity versus batch cultures. Semi-continuous cultivation has the advantage of facilitating higher biomass production rates than batch cultures whilst reducing

operating costs and process complexity versus continuous cultures. Semi-continuous cultivation at industrial scale has been reported for the microalgae *Tetraselmis* sp. CTP4, which was grown in 100 m^3 outdoor tubular photobioreactors (Pereira *et al.*, 2018). However, cultures were only harvested on 4 occasions across the 60-day cultivation cycle.

With regards to improving downstream processing, one opportunity unique to cyanobacteria is native light-induced cell aggregation mechanisms. While chemical flocculants are often used to help separate cells from media, these can be expensive and require further processing to remove them from the final product (Amaro *et al.*, 2017). Alternatively, blue light could be used to initiate cell aggregation in some cyanobacterial cultures. This would reduce the need for energy-intensive separation technologies such as large-scale centrifugation or filtration. Indeed, blue light has been shown to enhance the flocculation of the cyanobacterium *Limnothrix* sp. strain SK1-2-1 (Sugawara *et al.*, 2018). Furthermore, blue light and low temperature induce cell aggregation of *Thermosynechococcus vulcanus* RKN (Enomoto *et al.*, 2014) (discussed in section 1.2.3). However, it is unclear how widespread blue light-induced cell aggregation is across cyanobacteria species. It remains to be seen whether this phenotype could be genetically engineered or enhanced in industrially relevant species. A combination of chemical flocculants and blue light may increase the flocculation efficiency versus using chemical flocculants alone.

7.5.3 Reducing bioprocessing costs

A TEA discussing cyanobacterial PHB production provides information on the factors contributing the highest costs to a typical cyanobacterial bioprocess. In the base case scenario, total capital expenditure was estimated at $193 million, most of which came from the construction of the open raceway ponds used for cultivation. In contrast, yearly operating costs were estimated at $147 million per year, making up the bulk of the costs over the plant's-lifetime. 60% of these operating costs come from the energy usage required for the dewatering and drying of the biomass (Price *et al.*, 2022). Media costs accounted for only around 9% of operating costs. While the TEA assumed the use of sunlight for photosynthesis, additional electricity would also be required to power lighting in indoor cultivation systems. The cost of wholesale electricity from the UK

national grid has been volatile in recent years (Figure 7-3). However, electricity costs could potentially be decreased through a renewable power purchase agreement. Regardless, electricity cost and usage are critical factors to consider when designing or optimising a cyanobacterial bioprocess.

Figure 7-3. Weekly average of UK wholesale electricity prices 2021-2024. Figure from Ofgem, 2024.

Whilst CO_2 is essentially a free carbon source if bubbling cultures with atmospheric air, cyanobacteria grow significantly faster when using CO_2-enriched air. While the TEA optimistic scenario assumed a free CO_2 supply, the cost of purifying, compressing and transporting CO_2 was estimated at $20 million per year in an alternative scenario (Price *et al.*, 2022). While not as large a fraction of the operating costs as electricity, purchasing CO_2 is still a significant expense. An alternative could be to integrate cyanobacterial cultivation with heavy industry, allowing the use of flue gas captured on-site as a carbon source. While flue gases have varying compositions depending on the source, both natural and genetically engineered cyanobacteria have demonstrated the ability to grow in simulated flue gases (Arata *et al.*, 2012; Chou *et al.*, 2021). This also provides an avenue to utilise this CO_2 as opposed to capturing it and burying it underground.

While electricity and CO_2 feedstock costs were important to the profitability of the proposed PHB production process, the key factor which affected the profitability was the PHB titre (Price *et al.*, 2022). Therefore, the focus for increasing the cost-effectiveness of photosynthetic bioproduction should mostly be on increasing the rate of biomass accumulation and most importantly increasing product titres. Increasing biomass accumulation would therefore not only lead to increased product volumes but also has the potential to reduce operating costs as less energy would be required for drying per unit of product.

7.6 Conclusion

Society can no longer rely on fossil fuel-derived feedstocks to produce our food, clothing, fuel, medicine, and chemicals. Alternative low-carbon or carbon-negative technologies are required to convert atmospheric CO_2 to commercial products and reduce the harmful effects of excessive global warming on our environment. Cyanobacteria have the potential to be part of the solution, demonstrating the ability to convert CO_2 into a wide range of biochemicals. However, the relatively slow growth rates and low product titres achieved by cyanobacteria are the key issues preventing their widespread adoption in the industrial biotechnology sector. While the genetic toolbox for controlling cyanobacterial gene expression is constantly improving, it lags that of common heterotrophic cell chassis. Optogenetics is a technology which has the potential to improve control over gene expression, allowing dynamic metabolic pathway control strategies to be used which can lead to increased product titres.

This work is the first to express and characterise optogenetic tools in the industrially relevant cyanobacterium Syn7002. While the YF1/FixJ system was ineffective for controlling gene expression in Syn7002, the troubleshooting described in this work may inform future attempts at applying this system in both Syn7002 and other cyanobacteria. The CcaS/CcaR system had previously been expressed and utilised for optogenetics in two other cyanobacterial species. However, it had not been applied in more industrially promising species like Syn7002 and an in-depth characterisation of its function under many relevant lighting conditions was missing. This was addressed in the present work by characterising the activity of the system using fluorescence spectrometry and qRT-PCR under different light wavelengths, intensities, multiple ON/OFF cycles,

and light/dark cycles. Targeted genetic modification to the pCpcG2 output promoter resulted in increased system activity under green light and an improved dynamic range. The improved properties of the optimised system were then applied to the production of L-alanine and mannitol, products which have widespread uses in the food and pharmaceutical industries. While production of L-alanine was unsuccessful, this work laid the foundation for future troubleshooting of the process. Finally, the challenging question of how to discover new optogenetic tools was addressed. By growing Syn7002 cultures under blue, orange and white light, genes differentially expressed under specific light wavelengths were identified using RNAseq. The RNAseq dataset provides a basis for further work which may reveal new optogenetic tools by linking photoreceptors to their associated transcriptional regulators and output promoters. Collectively, I believe that this work moves the biotechnology industry one small step closer towards deploying cyanobacteria for low-carbon biochemical production.

8 References

360 Research Reports, 2020. Global L-alanine Sales Market Report 2020. [online] Available at: https://www.360researchreports.com/global L-alanine-industry-research-report-competitive Landscape-market-21412428 [Accessed 16 January 2023].

Abe, K., Miyake, K., Nakamura, M., Kojima, K., Ferri, S., Ikebukuro, K., Sode, K., 2014. Engineering of a green Light inducible gene expression system in Synechocystis sp. PCC6803. Microbial Biotechnology 7, 177–183. https://doi.org/10.1111/1751-7915.12098

Abed, R.M.M., Dobretsov, S., Sudesh, K., 2009. Applications of cyanobacteria in biotechnology. J Appl Microbiol 106, 1–12. https://doi.org/10.1111/j.1365-2672.2008.03918.x

Al-Haj, L., Lui, Y., Abed, R., Gomaa, M., Purton, S., 2016. Cyanobacteria as Chassis for Industrial Biotechnology: Progress and Prospects.-life 6, 42. https://doi.org/10.3390/life6040042

Alvarenga, D.O., Fiore, M.F., Varani, A.M., 2017. A metagenomic approach to cyanobacterial genomics. Frontiers in Microbiology 8, 809. https://doi.org/10.3389/fmicb.2017.00809

Amaro, H.M., Sousa-Pinto, I., Malcata, F.X., Guedes, A.C., 2017. 16 - Microalgal fatty acids—From harvesting until extraction, in: Gonzalez-Fernandez, C., Muñoz, R. (Eds.), Microalgae-Based Biofuels and Bioproducts, Woodhead Publishing Series in Energy. Woodhead Publishing, pp. 369–400. https://doi.org/10.1016/B978-0-08-101023-5.00016-9

Arata, S., Strazza, C., Lodi, A., Del Borghi, A., 2013. Spirulina platensis Culture with Flue Gas Feeding as a Cyanobacteria-Based Carbon Sequestration Option. Chemical Engineering & Technology 36, 91–97. https://doi.org/10.1002/ceat.201100722

Atsumi, S., Higashide, W., Liao, J.C., 2009. Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde. Nature Biotechnology 27, 1177–1180. https://doi.org/10.1038/nbt.1586

Badary, A., Abe, K., Ferri, S., Kojima, K., Sode, K., 2015. The Development and Characterization of an Exogenous Green Light-Regulated Gene Expression System in Marine Cyanobacteria. Marine Biotechnology 17, 245–251. https://doi.org/10.1007/s10126-015-9616-1

Bailey, T.L., 2021. STREME: accurate and versatile sequence motif discovery. Bioinformatics 37, 2834–2840. https://doi.org/10.1093/bioinformatics/btab203

Bailey, T.L., Elkan, C., 1994. Fitting a Mixture Model By Expectation Maximization To Discover Motifs In Biopolymer 9. https://www.aaai.org/Papers/ISMB/1994/ISMB94-004.pdf

Baumschlager, A., Aoki, S.K., Khammash, M., 2017. Dynamic Blue Light-Inducible T7 RNA Polymerases (Opto-T7RNAPs) for Precise Spatiotemporal Gene Expression Control. ACS Synth. Biol. 6, 2157–2167. https://doi.org/10.1021/acssynbio.7b00169

Becker, J., Zelder, O., Häfner, S., Schröder, H., Wittmann, C., 2011. From zero to hero—Design-based systems metabolic engineering of Corynebacterium glutamicum for l Lysine production. Metabolic Engineering 13, 159–168. https://doi.org/10.1016/j.ymben.2011.01.003

Benisch, M., Benzinger, D., Kumar, S., Hu, H., Khammash, M., 2023. Optogenetic closed Loop feedback control of the unfolded protein response optimizes protein production. Metabolic Engineering 77, 32–40. https://doi.org/10.1016/j.ymben.2023.03.001

Bernstein, H.C., McClure, R.S., Hill, E.A., Markillie, L.M., Chrisler, W.B., Romine, M.F., McDermott, J.E., Posewitz, M.C., Bryant, D.A., Konopka, A.E., Fredrickson, J.K., Beliaev, A.S., 2016. Unlocking the Constraints of Cyanobacterial Productivity: Acclimations Enabling Ultrafast Growth. mBio 7. https://doi.org/10.1128/mBio.00949-16

Berntsson, O., Diensthuber, R.P., Panman, M.R., Björling, A., Gustavsson, E., Hoernke, M., Hughes, A.J., Henry, L., Niebling, S., Takala, H., Ihalainen, J.A., Newby, G., Kerruth, S., Heberle, J., Liebi, M., Menzel, A., Henning, R., Kosheleva, I., Möglich, A., Westenhoff, S., 2017. Sequential conformational transitions and α-helical supercoiling regulate a sensor histidine kinase. Nature Communications 8, 284. https://doi.org/10.1038/s41467-017-00300-5

Binder, B.J., Chisholm, S.W., 1995. Cell Cycle Regulation in Marine Synechococcus sp. Strains. Appl. Environ. Microbiol. 61, 708–717. https://doi.org/10.1128/aem.61.2.708-717.1995

Bolay, P., Schlüter, S., Grimm, S., Riediger, M., Hess, W.R., Klähn, S., 2022. The transcriptional regulator RbcR controls ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) genes in the cyanobacterium Synechocystis sp. PCC 6803. New Phytologist 235, 432–445. https://doi.org/10.1111/nph.18139

Bordowitz, J.R., Montgomery, B.L., 2008. Photoregulation of Cellular Morphology during Complementary Chromatic Adaptation Requires Sensor-Kinase-Class Protein RcaE in Fremyella diplosiphon. Journal of Bacteriology 190, 4069–4074. https://doi.org/10.1128/JB.00018-08

Boyden, E.S., Zhang, F., Bamberg, E., Nagel, G., Deisseroth, K., 2005. Millisecond-timescale, genetically targeted optical control of neural activity. Nature Neuroscience 8, 1263–1268. https://doi.org/10.1038/nn1525

Brahamsha, B., Bhaya, D., 2014. Motility in Unicellular and Filamentous Cyanobacteria, in: The Cell Biology of Cyanobacteria.

Browning, D.F., Busby, S.J.W., 2016. Local and global regulation of transcription initiation in bacteria. Nat Rev Microbiol 14, 638–650. https://doi.org/10.1038/nrmicro.2016.103

Burriesci, M., Bhaya, D., 2008. Tracking phototactic responses and modeling motility of Synechocystis sp. strain PCC6803. Journal of Photochemistry and Photobiology B: Biology 91, 77–86. https://doi.org/10.1016/j.jphotobiol.2008.01.012

Bussell, A.N., Kehoe, D.M., 2013. Control of a four-color sensing photoreceptor by a two-color sensing photoreceptor reveals complex light regulation in cyanobacteria. Proceedings of the National Academy of Sciences of the United States of America 110, 12834–12839. https://doi.org/10.1073/pnas.1303371110

Cai, F., Menon, B.B., Cannon, G.C., Curry, K.J., Shively, J.M., Heinhorst, S., 2009. The Pentameric Vertex Proteins Are Necessary for the Icosahedral Carboxysome Shell to Function as a CO2 Leakage Barrier. PLOS ONE 4, e7521. https://doi.org/10.1371/journal.pone.0007521

Campbell, E.L., Hagen, K.D., Chen, R., Risser, D.D., Ferreira, D.P., Meeks, J.C., 2015. Genetic Analysis Reveals the Identity of the Photoreceptor for Phototaxis in Hormogonium Filaments of Nostoc punctiforme. Journal of Bacteriology 197, 782–791. https://doi.org/10.1128/JB.02374-14

Cao, Y.-Q., Li, Q., Xia, P.-F., Wei, L.-J., Guo, N., Li, J.-W., Wang, S.-G., 2017. AraBAD Based Toolkit for Gene Expression and Metabolic Robustness Improvement in Synechococcus elongatus. Scientific Reports 7, 18059. https://doi.org/10.1038/s41598-017-17035-4

Carpine, R., Du, W., Olivieri, G., Pollio, A., Hellingwerf, K.J., Marzocchella, A., Branco dos Santos, F., 2017. Genetic engineering of Synechocystis sp. PCC6803 for poly-ß-hydroxybutyrate overproduction. Algal Research 25, 117–127. https://doi.org/10.1016/j.algal.2017.05.013

Carrasco López, C., García-Echauri, S.A., Kichuk, T., Avalos, J.L., 2020. Optogenetics and biosensors set the stage for metabolic cybergenetics. Current Opinion in Biotechnology 65, 296–309. https://doi.org/10.1016/j.copbio.2020.07.012

Caslake, L.F., Gruber, T.M., Bryant, D.A., 1997. Expression of two alternative sigma factors of Synechococcus sp. strain PCC 7002 is modulated by carbon and nitrogen stress. Microbiology (Reading) 143 (Pt 12), 3807–3818. https://doi.org/10.1099/00221287-143-12-3807

Castenholz, R.W., 2015. General Characteristics of the Cyanobacteria, in: Bergey's Manual of Systematics of Archaea and Bacteria. American Cancer Society, pp. 1–23. https://doi.org/10.1002/9781118960608.cbm00019

Castenholz, R.W., Garcia-Pichel, F., 2012. Cyanobacterial Responses to UV Radiation, in: Whitton, B.A. (Ed.), Ecology of Cyanobacteria II: Their Diversity in Space and Time. Springer Netherlands, Dordrecht, pp. 481–499. https://doi.org/10.1007/978-94-007-3855-3_19

Castillo-Hair, S.M., Baerman, E.A., Fujita, M., Igoshin, O.A., Tabor, J.J., 2019. Optogenetic control of Bacillus subtilis gene expression. Nat Commun 10, 1–11. https://doi.org/10.1038/s41467-019-10906-6

Chan, S., Kanchanatawee, S., Jantama, K., 2012. Production of succinic acid from sucrose and sugarcane molasses by metabolically engineered Escherichia

coli. Bioresource Technology 103, 329–336. https://doi.org/10.1016/j.biortech.2011.09.096

Chen, A.H., Afonso, B., Silver, P.A., Savage, D.F., 2012. Spatial and Temporal Organization of Chromosome Duplication and Segregation in the Cyanobacterium Synechococcus elongatus PCC 7942. PLOS ONE 7, e47837. https://doi.org/10.1371/journal.pone.0047837

Chen, H., Jiang, P., 2019. Metabolic engineering of Escherichia coli for efficient biosynthesis of fluorescent phycobiliprotein. Microbial Cell Factories 18, 58. https://doi.org/10.1186/s12934-019-1100-6

Chen, Y., Taton, A., Go, M., London, R.E., Pieper, L.M., Golden, S.S., Golden, J.W.Y. 2016, n.d. Self-replicating shuttle vectors based on pANS, a small endogenous plasmid of the unicellular cyanobacterium Synechococcus elongatus PCC 7942. Microbiology 162, 2029–2041. https://doi.org/10.1099/mic.0.000377

Chin, T., Okuda, Y., Ikeuchi, M., 2018. Sorbitol production and optimization of photosynthetic supply in the cyanobacterium *Synechocystis* PCC 6803. Journal of Biotechnology 276–277, 25–33. https://doi.org/10.1016/j.jbiotec.2018.04.004

Chou, H.-H., Su, H.-Y., Chow, T.-J., Lee, T.-M., Cheng, W.-H., Chang, J.-S., Chen, H.-J., 2021. Engineering cyanobacteria with enhanced growth in simulated flue gases for high-yield bioethanol production. Biochemical Engineering Journal 165, 107823. https://doi.org/10.1016/j.bej.2020.107823

Conradi, F.D., Zhou, R.-Q., Oeser, S., Schuergers, N., Wilde, A., Mullineaux, C.W., 2019. Factors Controlling Floc Formation and Structure in the Cyanobacterium Synechocystis sp. Strain PCC 6803. Journal of Bacteriology 201, 10.1128/jb.00344-19. https://doi.org/10.1128/jb.00344-19

Cormack, B.P., Valdivia, R.H., Falkow, S., 1996. FACS-optimized mutants of the green fluorescent protein (GFP). Gene, Flourescent Proteins and Applications 173, 33–38. https://doi.org/10.1016/0378-1119(95)00685-0

Cosentino, C., Alberio, L., Gazzarrini, S., Aquila, M., Romano, E., Cermenati, S., Zuccolini, P., Petersen, J., Beltrame, M., Etten, J.L.V., Christie, J.M., Thiel, G., Moroni, A., 2015. Engineering of a light-gated potassium channel. Science 348, 707–710. https://doi.org/10.1126/science.aaa2787

Dasgupta, C.N., 2016. Algae as a Source of Phycocyanin and Other Industrially Important Pigments. Algal Biorefinery: An Integrated Approach 253–276. https://doi.org/10.1007/978-3-319-22813-6_12

David, L., Prado, M., Arteni, A.A., Elmlund, D.A., Blankenship, R.E., Adir, N., 2014. Structural studies show energy transfer within stabilized phycobilisomes independent of the mode of rod–core assembly. Biochimica et Biophysica Acta (BBA) - Bioenergetics 1837, 385–395. https://doi.org/10.1016/j.bbabio.2013.12.014

De Spiegelaere, W., Dern-Wieloch, J., Weigel, R., Schumacher, V., Schorle, H., Nettersheim, D., Bergmann, M., Brehm, R., Kliesch, S., Vandekerckhove, L., Fink, C., 2015. Reference Gene Validation for RT-qPCR, a Note on Different Available Software Packages. PLoS One 10, e0122515. https://doi.org/10.1371/journal.pone.0122515

Deisseroth, K., 2011. Optogenetics. Nature Methods 8, 26–29. https://doi.org/10.1038/nmeth.f.324

Dempo, Y., Ohta, E., Nakayama, Y., Bamba, T., Fukusaki, E., 2014. Molar-Based Targeted Metabolic Profiling of Cyanobacterial Strains with Potential for Biological Production. Metabolites 4, 499–516. https://doi.org/10.3390/metabo4020499

Ding, Q., Ma, D., Liu, G.-Q., Li, Y., Guo, L., Gao, C., Hu, G., Ye, C., Liu, J., Liu, L., Chen, X., 2020. Light-powered Escherichia coli cell division for chemical production. Nat Commun 11, 2262. https://doi.org/10.1038/s41467-020-16154-3

Dong, G., Yang, Q., Wang, Q., Kim, Y.-I., Wood, T.L., Osteryoung, K.W., van Oudenaarden, A., Golden, S.S., 2010. Elevated ATPase Activity of KaiC Applies a Circadian Checkpoint on Cell Division in Synechococcus elongatus. Cell 140, 529–539. https://doi.org/10.1016/j.cell.2009.12.042

Dookeran, Z.A., Nielsen, D.R., 2021. Systematic Engineering of Synechococcus elongatus UTEX 2973 for Photosynthetic Production of l Lysine, Cadaverine, and Glutarate. ACS Synth. Biol. 10, 3561–3575. https://doi.org/10.1021/acssynbio.1c00492

Englund, E., Liang, F., Lindberg, P., 2016. Evaluation of promoters and ribosome binding sites for biotechnological applications in the unicellular cyanobacterium

Synechocystis sp. PCC 6803. Scientific Reports 6, 36640. https://doi.org/10.1038/srep36640

Enomoto, G., Ni-Ni-Win, Narikawa, R., Ikeuchi, M., 2015. Three cyanobacteriochromes work together to form a light color-sensitive input system for c-di-GMP signaling of cell aggregation. PNAS 112, 8082–8087. https://doi.org/10.1073/pnas.1504228112

Enomoto, G., Nomura, R., Shimada, T., Ni-Ni-Win, Narikawa, R., Ikeuchi, M., 2014. Cyanobacteriochrome SesA Is a Diguanylate Cyclase That Induces Cell Aggregation in Thermosynechococcus. Journal of Biological Chemistry 289, 24801–24809. https://doi.org/10.1074/jbc.M114.583674

Expert Market Research, 2024. Global Mannitol Market Report. [online] Available at: https://www.expertmarketresearch.com/reports/mannitol-market [Accessed 3 March 2024].

Fernandez-Rodriguez, J., Moser, F., Song, M., Voigt, C.A., 2017. Engineering RGB color vision into Escherichia coli. Nature Chemical Biology 13, 706–708. https://doi.org/10.1038/nchembio.2390

Ferri, S., Nakamura, M., Ito, A., Nakajima, M., Abe, K., Kojima, K., Sode, K., 2015. Efficient surface-display of autotransporter proteins in cyanobacteria. Algal Research 12, 337–340. https://doi.org/10.1016/j.algal.2015.09.013

Flechtner, Johansen, 2008. The Biological Soil Crusts of the San Nicolas Island: Enigmatic Algae from a Geographically Isolated Ecosystem. https://doi.org/10.3398/1527-0904-68.4.405

Fleming, K.E., O'Shea, E.K., 2018. An RpaA-Dependent Sigma Factor Cascade Sets the Timing of Circadian Transcriptional Rhythms in Synechococcus elongatus. Cell Reports 25, 2937-2945.e3. https://doi.org/10.1016/j.celrep.2018.11.049

Fomicheva, A., Zhou, C., Sun, Q.-Q., Gomelsky, M., 2019. Engineering Adenylate Cyclase Activated by Near-Infrared Window Light for Mammalian Optogenetic Applications. ACS Synth. Biol. 8, 1314–1324. https://doi.org/10.1021/acssynbio.8b00528

Fushimi, K., 2020. Evolution-inspired design of multicolored photoswitches from a single cyanobacteriochrome scaffold. https://doi.org/10.1073/pnas.2004273117

Fushimi, K., Narikawa, R., 2019. Cyanobacteriochromes: photoreceptors covering the entire UV-to-visible spectrum. Current Opinion in Structural Biology, Membranes ● Engineering & Design: Synthetic Signalling 57, 39–46. https://doi.org/10.1016/j.sbi.2019.01.018

Fushimi, K., Rockwell, N.C., Enomoto, G., Ni-Ni-Win, Martin, S.S., Gan, F., Bryant, D.A., Ikeuchi, M., Lagarias, J.C., Narikawa, R., 2016. Cyanobacteriochrome Photoreceptors Lacking the Canonical Cys Residue. Biochemistry 55, 6981–6995. https://doi.org/10.1021/acs.biochem.6b00940

Future Market Insights, 2023. Phycocyanin Market. [online] Available at: https://www.futuremarketinsights.com/reports/phycocyanin-market [Accessed 16 November 2023].

Galperin, M.Y., Nikolskaya, A.N., 2007. Identification of Sensory and Signal-Transducing Domains in Two-Component Signaling Systems. Methods in Enzymology 422, 47–74. https://doi.org/10.1016/S0076-6879(06)22003-2

Garcia-Pichel, F., Belnap, J., Neuer, S., Schanz, F., 2003. Estimates of global cyanobacterial biomass and its distribution. Algological Studies 109, 213–227. https://doi.org/10.1127/1864-1318/2003/0109-0213

Glazer, A.N., 1989. Light guides: Directional energy transfer in a photosynthetic antenna. Journal of Biological Chemistry 264, 1–4. https://doi.org/10.1016/S0021-9258(17)31212-7

Gong, W., Hao, B., Mansy, S.S., Gonzalez, G., Gilles-Gonzalez, M.A., Chan, M.K., 1998. Structure of a biological oxygen sensor: A new mechanism for heme-driven signal transduction. PNAS 95, 15177–15182. https://doi.org/10.1073/pnas.95.26.15177

Gordon, G.C., Pfleger, B.F., 2018. Regulatory Tools for Controlling Gene Expression in Cyanobacteria, in: Zhang, W., Song, X. (Eds.), Synthetic Biology of Cyanobacteria, Advances in Experimental Medicine and Biology. Springer, Singapore, pp. 281–315. https://doi.org/10.1007/978-981-13-0854-3_12

Griese, M., Lange, C., Jörg, S., 2011. Ploidy in cyanobacteria. FEMS Microbiology Letters 323, 124–131. https://doi.org/10.1111/j.1574-6968.2011.02368.x

Gruber, T.M., Gross, C.A., 2003. Multiple Sigma Subunits and the Partitioning of Bacterial Transcription Space. Annual Review of Microbiology 57, 441–466. https://doi.org/10.1146/annurev.micro.57.030502.090913

Guerrero, F., Carbonell, V., Cossu, M., Correddu, D., Jones, P.R., 2012. Ethylene Synthesis and Regulated Expression of Recombinant Protein in Synechocystis sp. PCC 6803. PLoS ONE 7. https://doi.org/10.1371/journal.pone.0050470

Gupta, J.K., Rai, P., Jain, K.K., Srivastava, S., 2020. Overexpression of bicarbonate transporters in the marine cyanobacterium Synechococcus sp. PCC 7002 increases growth rate and glycogen accumulation. Biotechnology for Biofuels 13, 17. https://doi.org/10.1186/s13068-020-1656-8

Gutu, A., O'Shea, E.K., 2013. Two Antagonistic Clock-Regulated Histidine Kinases Time the Activation of Circadian Gene Expression. Molecular Cell 50, 288–294. https://doi.org/10.1016/j.molcel.2013.02.022

Hachicha, Rihab, Elleuch, F., Ben Hlima, H., Dubessay, P., de Baynast, H., Delattre, C., Pierre, G., Hachicha, Ridha, Abdelkafi, S., Michaud, P., Fendri, I., 2022. Biomolecules from Microalgae and Cyanobacteria: Applications and Market Survey. Applied Sciences 12, 1924. https://doi.org/10.3390/app12041924

Henley, W.J., 2019. The past, present and future of algal continuous cultures in basic research and commercial applications. Algal Research 43, 101636. https://doi.org/10.1016/j.algal.2019.101636

Henson, B.J., Hesselbrock, S.M., Watson, L.E., Barnum, S.R., 2004. Molecular phylogeny of the heterocystous cyanobacteria (subsections IV and V) based on nifD. International Journal of Systematic and Evolutionary Microbiology 54, 493–497. https://doi.org/10.1099/ijs.0.02821-0

Heuser, F., Marin, K., Kaup, B., Bringer, S., Sahm, H., 2009. Improving d-mannitol productivity of *Escherichia coli*: Impact of NAD, CO2 and expression of a putative sugar permease from *Leuconostoc pseudomesenteroides*. Metabolic Engineering 11, 178–183. https://doi.org/10.1016/j.ymben.2009.01.006

Hirose, Y., Narikawa, R., Katayama, M., Ikeuchi, M., 2010. Cyanobacteriochrome CcaS regulates phycoerythrin accumulation in Nostoc punctiforme, a group II

chromatic adapter. Proceedings of the National Academy of Sciences 107, 8854–8859. https://doi.org/10.1073/pnas.1000177107

Hirose, Y., Rockwell, N.C., Nishiyama, K., Narikawa, R., Ukaji, Y., Inomata, K., Lagarias, J.C., Ikeuchi, M., 2013. Green/red cyanobacteriochromes regulate complementary chromatic acclimation via a protochromic photocycle. PNAS 110, 4974–4979. https://doi.org/10.1073/pnas.1302909110

Hirose, Y., Shimada, T., Narikawa, R., Katayama, M., Ikeuchi, M., 2008. Cyanobacteriochrome CcaS is the green light receptor that induces the expression of phycobilisome linker protein. Proceedings of the National Academy of Sciences 105, 9528–9533. https://doi.org/10.1073/pnas.0801826105

Holtkamp, W., Kokic, G., Jäger, M., Mittelstaet, J., Komar, A.A., Rodnina, M.V., 2015. Cotranslational protein folding on the ribosome monitored in real time. Science 350, 1104–1107. https://doi.org/10.1126/science.aad0344

Hueso-Gil, A., Nyerges, Á., Pál, C., Calles, B., de Lorenzo, V., 2020. Multiple-Site Diversification of Regulatory Sequences Enables Interspecies Operability of Genetic Devices. ACS Synth. Biol. 9, 104–114. https://doi.org/10.1021/acssynbio.9b00375

Ikeda, M., Takeno, S., 2013. Amino Acid Production by Corynebacterium glutamicum, in: Yukawa, H., Inui, M. (Eds.), Corynebacterium Glutamicum: Biology and Biotechnology, Microbiology Monographs. Springer, Berlin, Heidelberg, pp. 107–147. https://doi.org/10.1007/978-3-642-29857-8_4

Imhoff, J.F., Sahl, H.G., Soliman, G.S.H., Trüper, H.G., 1979. The Wadi Natrun: Chemical composition and microbial mass developments in alkaline brines of Eutrophic Desert Lakes. Geomicrobiology Journal 1, 219–234. https://doi.org/10.1080/01490457909377733

Ishizuka, T., Narikawa, R., Kohchi, T., Katayama, M., Ikeuchi, M., 2007. Cyanobacteriochrome TePixJ of Thermosynechococcus elongatus Harbors Phycoviolobilin as a Chromophore. Plant and Cell Physiology 48, 1385–1390. https://doi.org/10.1093/pcp/pcm106

Iwamoto, K., Shiraiwa, Y., 2005. Salt-Regulated Mannitol Metabolism in Algae. Mar Biotechnol 7, 407–415. https://doi.org/10.1007/s10126-005-0029-4

Jacobsen, J.H., Frigaard, N.-U., 2014. Engineering of photosynthetic mannitol biosynthesis from CO2 in a cyanobacterium. Metabolic Engineering 21, 60–70. https://doi.org/10.1016/j.ymben.2013.11.004

Jacobus, A.P., Gross, J., Evans, J.H., Ceccato-Antonini, S.R., Gombert, A.K., 2021. Saccharomyces cerevisiae strains used industrially for bioethanol production. Essays in Biochemistry 65, 147–161. https://doi.org/10.1042/EBC20200160

Jiang, Z., Swem, L.R., Rushing, B.G., Devanathan, S., Tollin, G., Bauer, C.E., 1999. Bacterial Photoreceptor with Similarity to Photoactive Yellow Protein and Plant Phytochromes. Science 285, 406–409. https://doi.org/10.1126/science.285.5426.406

Jones, M.R., Pinto, E., Torres, M.A., Dörr, F., Mazur-Marzec, H., Szubert, K., Tartaglione, L., Dell'Aversano, C., Miles, C.O., Beach, D.G., McCarron, P., Sivonen, K., Fewer, D.P., Jokela, J., Janssen, E.M. L., 2021. CyanoMetDB, a comprehensive public database of secondary metabolites from cyanobacteria. Water Research 196, 117017. https://doi.org/10.1016/j.watres.2021.117017

Jorissen, H.J.M.M., Quest, B., Remberg, A., Coursin, T., Braslavsky, S.E., Schaffner, K., Tandeau de Marsac, N., Gärtner, W., 2002. Two independent, light-sensing two-component systems in a filamentous cyanobacterium. European Journal of Biochemistry 269, 2662–2671. https://doi.org/10.1046/j.1432-1033.2002.02928.x

Joseph, V., Joseph, A., 2001. Microalgae in Petrochemical Effluent: Growth and Biosorption of Total Dissolved Solids. Bulletin of Environmental Contamination and Toxicology 66, 522–527. https://doi.org/10.1007/s00128-001-0038-6

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2

Jurk, M., Schramm, P., Schmieder, P., 2013. The blue Light receptor YtvA from Bacillus subtilis is permanently incorporated into the stressosome independent of the illumination state. Biochemical and Biophysical Research Communications 432, 499–503. https://doi.org/10.1016/j.bbrc.2013.02.025

Kato, H., Watanabe, S., Nimura-Matsune, K., Chibazakura, T., Tozawa, Y., Yoshikawa, H., 2012. Exploration of a Possible Partnership among Orphan Two-Component System Proteins in Cyanobacterium Synechococcus elongatus PCC 7942. Bioscience, Biotechnology, and Biochemistry 76, 1484–1491. https://doi.org/10.1271/bbb.120172

Katsube, S., Ando, T., Yoneyama, H., 2019. L-Alanine Exporter, AlaE, of Escherichia coli Functions as a Safety Valve to Enhance Survival under Feast Conditions. Int J Mol Sci 20, 4942. https://doi.org/10.3390/ijms20194942

Kerfeld, C.A., Melnicki, M.R., 2016. Assembly, function and evolution of cyanobacterial carboxysomes. Current Opinion in Plant Biology, SI: 31: Physiology and metabolism 2016 31, 66–75. https://doi.org/10.1016/j.pbi.2016.03.009

Kim, P., Porr, B., Mori, T., Kim, Y.-S., Johnson, C.H., Diekman, C.O., Kim, Y.-I., 2020. CikA, an Input Pathway Component, Senses the Oxidized Quinone Signal to Generate Phase Delays in the Cyanobacterial Circadian Clock. J Biol Rhythms 35, 227–234. https://doi.org/10.1177/0748730419900868

Kim, S., Ihara, K., Katsube, S., Ando, T., Isogai, E., Yoneyama, H., 2017. Impact of charged amino acid substitution in the transmembrane domain of l-alanine exporter, AlaE, of Escherichia coli on the l-alanine export. Arch Microbiol 199, 105–114. https://doi.org/10.1007/s00203-016-1279-4

Kirilovsky, D., Kerfeld, C.A., 2016. Cyanobacterial photoprotection by the orange carotenoid protein. Nature Plants 2, 1–7. https://doi.org/10.1038/nplants.2016.180

Klughammer, B., Sültemeyer, D., Badger, M.R., Price, G.D., 1999. The involvement of NAD(P)H dehydrogenase subunits, NdhD3 and NdhF3, in high-affinity CO2 uptake in Synechococcus sp. PCC7002 gives evidence for multiple NDH-1 complexes with specific roles in cyanobacteria. Molecular Microbiology 32, 1305–1315. https://doi.org/10.1046/j.1365-2958.1999.01457.x

Knoop, H., Steuer, R., 2015. A Computational Analysis of Stoichiometric Constraints and Trade-Offs in Cyanobacterial Biofuel Production. Front. Bioeng. Biotechnol. 3. https://doi.org/10.3389/fbioe.2015.00047

Kobayashi, S., Atsumi, S., Ikebukuro, K., Sode, K., Asano, R., 2022. Light-induced production of isobutanol and 3-methyl-1-butanol by metabolically engineered cyanobacteria. Microb Cell Fact 21, 7. https://doi.org/10.1186/s12934-021-01732-x

Kobayashi, S., Nakajima, M., Asano, R., Ferreira, E.A., Abe, K., Tamagnini, P., Atsumi, S., Sode, K., 2019. Application of an engineered chromatic acclimation sensor for red Light-regulated gene expression in cyanobacteria. Algal Research 44. https://doi.org/10.1016/j.algal.2019.101691

Kohler, R., Mooney, R.A., Mills, D.J., Landick, R., Cramer, P., 2017. Architecture of a transcribing-translating expressome. Science 356, 194–197. https://doi.org/10.1126/science.aal3059

Komera, I., Gao, C., Guo, L., Hu, G., Chen, X., Liu, L., 2022. Bifunctional optogenetic switch for improving shikimic acid production in E. coli. Biotechnology for Biofuels and Bioproducts 15, 13. https://doi.org/10.1186/s13068-022-02111-3

Kondo, K., Ochiai, Y., Katayama, M., Ikeuchi, M., 2007. The Membrane-Associated CpcG2-Phycobilisome in Synechocystis: A New Photosystem I Antenna. Plant Physiology 144, 1200–1210. https://doi.org/10.1104/pp.107.099267

Korosh, T.C., Markley, A.L., Clark, R.L., McGinley, L.L., McMahon, K.D., Pfleger, B.F., 2017. Engineering photosynthetic production of L Lysine. Metabolic Engineering 44, 273–283. https://doi.org/10.1016/j.ymben.2017.10.010

Lalwani, M.A., Ip, S.S., Carrasco López, C., Day, C., Zhao, E.M., Kawabe, H., Avalos, J.L., 2021a. Optogenetic control of the lac operon for bacterial chemical and protein production. Nature Chemical Biology 17, 71–79. https://doi.org/10.1038/s41589-020-0639-1

Lalwani, M.A., Kawabe, H., Mays, R.L., Hoffman, S.M., Avalos, J.L., 2021b. Optogenetic Control of Microbial Consortia Populations for Chemical Production. ACS Synth. Biol. 10, 2015–2029. https://doi.org/10.1021/acssynbio.1c00182

Landry, B.P., Stöckel, J., Pakrasi, H.B., 2013. Use of Degradation Tags To Control Protein Levels in the Cyanobacterium Synechocystis sp. Strain PCC 6803. Appl. Environ. Microbiol. 79, 2833–2835. https://doi.org/10.1128/AEM.03741-12

Lau, N.-S., Matsui, M., Abdullah, A.A.-A., 2015. Cyanobacteria: Photoautotrophic Microbial Factories for the Sustainable Synthesis of Industrial Products. BioMed Research International 2015, e754934. https://doi.org/10.1155/2015/754934

Lea-Smith, D.J., Bombelli, P., Vasudevan, R., Howe, C.J., 2016. Photosynthetic, respiratory and extracellular electron transport pathways in cyanobacteria. Biochimica et Biophysica Acta (BBA) - Bioenergetics, Organization and dynamics of bioenergetic systems in bacteria 1857, 247–255. Lee, M., Smith, G.M., Eiteman, M.A., Altman, E., 2004. Aerobic production of alanine by Escherichia coli aceF ldhA mutants expressing the Bacillus sphaericus alaD gene. Appl Microbiol Biotechnol 65, 56–60. https://doi.org/10.1007/s00253-004-1560-3

LEE Filters, 2023. 019 Fire. [online] Available at: https://leefilters.com/colour/019-fire/ [Accessed 2 August 2023].

LEE Filters, 2023. 124 Dark Green. [online] Available at: https://leefilters.com/colour/124-dark-green/ [Accessed 2 August 2023].

Lee, S.-H., Lee, J.E., Kim, Y., Lee, S.-Y., 2016. The Production of High Purity Phycocyanin by Spirulina platensis Using Light-Emitting Diodes Based Two-Stage Cultivation. Appl Biochem Biotechnol 178, 382–395. https://doi.org/10.1007/s12010-015-1879-5

Legris, M., Klose, C., Burgie, E.S., Rojas, C.C.R., Neme, M., Hiltbrunner, A., Wigge, P.A., Schäfer, E., Vierstra, R.D., Casal, J.J., 2016. Phytochrome B integrates light and temperature signals in Arabidopsis. Science 354, 897–900. https://doi.org/10.1126/science.aaf5656

Liew, F.E., Nogle, R., Abdalla, T., Rasor, B.J., Canter, C., Jensen, R.O., Wang, L., Strutz, J., Chirania, P., De Tissera, S., Mueller, A.P., Ruan, Z., Gao, A., Tran, L., Engle, N.L., Bromley, J.C., Daniell, J., Conrado, R., Tschaplinski, T.J., Giannone, R.J., Hettich, R.L., Karim, A.S., Simpson, S.D., Brown, S.D., Leang, C., Jewett, M.C., Köpke, M., 2022. Carbon-negative production of acetone and isopropanol by gas fermentation at industrial pilot scale. Nat Biotechnol 40, 335–344. https://doi.org/10.1038/s41587-021-01195-w

Lindebro, M.C., Poellinger, L., Whitelaw, M.L., 1995. Protein-protein interaction via PAS domains: role of the PAS domain in positive and negative regulation of the bHLH/PAS dioxin receptor-Arnt transcription factor complex. EMBO J 14, 3528–3539.

Lindner, F., Diepold, A., 2022. Optogenetics in bacteria – applications and opportunities. FEMS Microbiology Reviews 46, fuab055. https://doi.org/10.1093/femsre/fuab055

Liu, P., Xu, H., Zhang, X., 2022. Metabolic engineering of microorganisms for L-alanine production. Journal of Industrial Microbiology and Biotechnology 49, kuab057. https://doi.org/10.1093/jimb/kuab057

Liu, T.-S., Wu, K.-F., Jiang, H.-W., Chen, K.-W., Nien, T.-S., Bryant, D.A., Ho, M.-Y., 2023. Identification of a Far-Red Light-Inducible Promoter that Exhibits Light Intensity Dependency and Reversibility in a Cyanobacterium. ACS Synth. Biol. 12, 1320–1330. https://doi.org/10.1021/acssynbio.3c00066

Liu, Y., Tsinoremas, N.F., Johnson, C.H., Lebedeva, N.V., Golden, S.S., Ishiura, M., Kondo, T., 1995. Circadian orchestration of gene expression in cyanobacteria. Genes Dev. 9, 1469–1478. https://doi.org/10.1101/gad.9.12.1469

Lo, C. L., Choudhury, S.R., Irudayaraj, J., Zhou, F.C., 2017. Epigenetic Editing of Ascl1 Gene in Neural Stem Cells by Optogenetics. Scientific Reports 7, 42047. https://doi.org/10.1038/srep42047

Long, B., Fischer, B., Zeng, Y., Amerigian, Z., Li, Q., Bryant, H., Li, M., Dai, S.Y., Yuan, J.S., 2022. Machine learning-informed and synthetic biology-enabled semi-continuous algal cultivation to unleash renewable fuel productivity. Nat Commun 13, 541. https://doi.org/10.1038/s41467-021-27665-y

Long, B.M., Hee, W.Y., Sharwood, R.E., Rae, B.D., Kaines, S., Lim, Y. L., Nguyen, N.D., Massey, B., Bala, S., von Caemmerer, S., Badger, M.R., Price, G.D., 2018. Carboxysome encapsulation of the CO2-fixing enzyme Rubisco in tobacco chloroplasts. Nature Communications 9, 3570. https://doi.org/10.1038/s41467-018-06044-0

Losi, A., Gardner, K.H., Möglich, A., 2018. Blue Light Receptors for Optogenetics. Chem. Rev. 118, 10659–10709. https://doi.org/10.1021/acs.chemrev.8b00163

Luan, G., Qi, Y., Wang, M., Li, Z., Duan, Y., Tan, X., Lu, X., 2015. Combinatory strategy for characterizing and understanding the ethanol synthesis pathway in cyanobacteria cell factories. Biotechnol Biofuels 8, 184. https://doi.org/10.1186/s13068-015-0367-z

Ludwig, M., Bryant, D.A., 2012. Synechococcus sp. Strain PCC 7002 transcriptome: Acclimation to temperature, salinity, oxidative stress, and mixotrophic growth conditions. Frontiers in Microbiology 3, 1–14. https://doi.org/10.3389/fmicb.2012.00354

Luimstra, V.M., Schuurmans, J.M., Hellingwerf, K.J., Matthijs, H.C.P., Huisman, J., 2020. Blue light induces major changes in the gene expression profile of the cyanobacterium Synechocystis sp. PCC 6803. Physiologia Plantarum 170, 10–26. https://doi.org/10.1111/ppl.13086

Luimstra, V.M., Schuurmans, J.M., Verschoor, A.M., Hellingwerf, K.J., Huisman, J., Matthijs, H.C.P., 2018. Blue light reduces photosynthetic efficiency of cyanobacteria through an imbalance between photosystems I and II. Photosynthesis Research 138, 177–189. https://doi.org/10.1007/s11120-018-0561-5

Maccoll, R., 1998. Cyanobacterial Phycobilisomes. Journal of structural biology 124, 311–334. https://doi.org/10.1006/jsbi.1998.4062

Madsen, M.A., Semerdzhiev, S., Amtmann, A., Tonon, T., 2018. Engineering Mannitol Biosynthesis in Escherichia coli and Synechococcus sp. PCC 7002 Using a Green Algal Fusion Protein. ACS Synth. Biol. 7, 2833–2840. https://doi.org/10.1021/acssynbio.8b00238

Maeda, K., Tamura, J., Okuda, Y., Narikawa, R., Midorikawa, T., Ikeuchi, M., 2018. Genetic identification of factors for extracellular cellulose accumulation in the thermophilic cyanobacterium Thermosynechococcus vulcanus: proposal of a novel tripartite secretion system. Molecular Microbiology 109, 121–134. https://doi.org/10.1111/mmi.13977

Mahbub, M., Mullineaux, C.W., 2023. Locations of membrane protein production in a cyanobacterium. Journal of Bacteriology 205, e00209-23. https://doi.org/10.1128/jb.00209-23

Manirafasha, E., Murwanashyaka, T., Ndikubwimana, T., Yue, Q., Zeng, X., Lu, Y., Jing, K., 2017. Ammonium chloride: a novel effective and inexpensive salt solution for phycocyanin extraction from Arthrospira (Spirulina) platensis. J Appl Phycol 29, 1261–1270. https://doi.org/10.1007/s10811-016-0989-y

Markley, A.L., Begemann, M.B., Clarke, R.E., Gordon, G.C., Pfleger, B.F., 2015. Synthetic Biology Toolbox for Controlling Gene Expression in the Cyanobacterium Synechococcus sp. strain PCC 7002. ACS Synth. Biol. 4, 595–603. https://doi.org/10.1021/sb500260k

Martínez-Miranda, J.G., Chairez, I., Durán-Páramo, E., 2022. Mannitol Production by Heterofermentative Lactic Acid Bacteria: a Review. Appl Biochem Biotechnol 194, 2762–2795. https://doi.org/10.1007/s12010-022-03836-5

Mihnev, A., Amtmann, A., 2022. In-silico analysis of cyanobacteriochrome architectures and spectral diversity. https://doi.org/10.1101/2022.09.22.509050

Mihnev, A., Forbes, L., McKenzie, J.D., Cogdell, R.J., Amtmann, A., 2023. Photocycle characterization of a blue-orange cyanobacteriochrome from Synechococcus sp. PCC 7002. https://doi.org/10.1101/2022.09.18.508392

Minas, K., Karunakaran, E., Bond, T., Gandy, C., Honsbein, A., Madsen, M., Amezaga, J., Amtmann, A., Templeton, M.R., Biggs, C.A., Lawton, L., 2015. Biodesalination: an emerging technology for targeted removal of Na+ and Cl− from seawater by cyanobacteria. Desalination and Water Treatment 55, 2647–2668. https://doi.org/10.1080/19443994.2014.940647

Miyake, K., Abe, K., Ferri, S., Nakajima, M., Nakamura, M., Yoshida, W., Kojima, K., Ikebukuro, K., Sode, K., 2014. A green Light inducible lytic system for cyanobacterial cells. Biotechnology for Biofuels 7. https://doi.org/10.1186/1754-6834-7-56

Möglich, A., Ayers, R.A., Moffat, K., 2009. Design and Signaling Mechanism of Light-Regulated Histidine Kinases. Journal of Molecular Biology 385, 1433–1444. https://doi.org/10.1016/j.jmb.2008.12.017

Montgomery, B.L., 2007. Sensing the light: photoreceptive systems and signal transduction in cyanobacteria. Molecular Microbiology 64, 16–27. https://doi.org/10.1111/j.1365-2958.2007.05622.x

Mullineaux, C.W., 2014. Electron transport and light-harvesting switches in cyanobacteria. Front Plant Sci 5. https://doi.org/10.3389/fpls.2014.00007

Mullineaux, C.W., 2005. Function and evolution of grana. Trends in Plant Science 10, 521–525. https://doi.org/10.1016/j.tplants.2005.09.001

Murakami, K.S., 2013. X-ray Crystal Structure of Escherichia coli RNA Polymerase σ70 Holoenzyme *. Journal of Biological Chemistry 288, 9126–9134. https://doi.org/10.1074/jbc.M112.430900

Mustachio, L.M., Aksit, S., Mistry, R.H., Scheffler, R., Yamada, A., Liu, J.M., 2012. The Vibrio cholerae Mannitol Transporter Is Regulated Posttranscriptionally by the MtlS Small Regulatory RNA. Journal of Bacteriology 194, 598–606. https://doi.org/10.1128/jb.06153-11

Nagarajan, A., Pakrasi, H.B., 2016. Membrane-Bound Protein Complexes for Photosynthesis and Respiration in Cyanobacteria, in: eLS. American Cancer Society, pp. 1–8. https://doi.org/10.1002/9780470015902.a0001670.pub2

Nagata, Y., 1999. d-Amino Acids in Nature, in: Pályi, G., Zucchi, C., Caglioti, L. (Eds.), Advances in BioChirality. Elsevier Science B.V., Amsterdam, pp. 271–283. https://doi.org/10.1016/B978-008043404-9/50018-8

Nakajima, M., Ferri, S., Rögner, M., Sode, K., 2016. Construction of a Miniaturized Chromatic Acclimation Sensor from Cyanobacteria with Reversed Response to a Light Signal. Scientific Reports 6. https://doi.org/10.1038/srep37595

Narikawa, R., Kohchi, T., Ikeuchi, M., 2008. Characterization of the photoactive GAF domain of the CikA homolog (SyCikA, Slr1969) of the cyanobacterium Synechocystis sp. PCC 6803. Photochem Photobiol Sci 7, 1253–1259. https://doi.org/10.1039/b811214b

Narikawa, R., Nakajima, T., Aono, Y., Fushimi, K., Enomoto, G., Ni-Ni-Win, Itoh, S., Sato, M., Ikeuchi, M., 2015. A biliverdin-binding cyanobacteriochrome from the chlorophyll d –bearing cyanobacterium Acaryochloris marina. Scientific Reports 5, 7950. https://doi.org/10.1038/srep07950

Narikawa, R., Zikihara, K., Okajima, K., Ochiai, Y., Katayama, M., Shichida, Y., Tokutomi, S., Ikeuchi, M., 2006. Three Putative Photosensory Light, Oxygen or Voltage (LOV) Domains with Distinct Biochemical Properties from the

Filamentous Cyanobacterium Anabaena sp. PCC 7120. Photochemistry and Photobiology 82, 1627–1633. https://doi.org/10.1111/j.1751-1097.2006.tb09822.x

Niopek, D., Benzinger, D., Roensch, J., Draebing, T., Wehler, P., Eils, R., Di Ventura, B., 2014. Engineering light-inducible nuclear localization signals for precise spatiotemporal control of protein dynamics in living cells. Nature Communications 5, 4404. https://doi.org/10.1038/ncomms5404

Nixon, P.J., Michoux, F., Yu, J., Boehm, M., Komenda, J., 2010. Recent advances in understanding the assembly and repair of photosystem II. Annals of Botany 106, 1–16. https://doi.org/10.1093/aob/mcq059

Nomura, C.T., Sakamoto, T., Bryant, D.A., 2006. Roles for heme-copper oxidases in extreme high Light and oxidative stress response in the cyanobacterium Synechococcus sp. PCC 7002. Archives of Microbiology 185, 471–479. https://doi.org/10.1007/s00203-006-0107-7

Ohlendorf, R., Schumacher, C.H., Richter, F., Möglich, A., 2016. Library-Aided Probing of Linker Determinants in Hybrid Photoreceptors. ACS Synth. Biol. 5, 1117–1126. https://doi.org/10.1021/acssynbio.6b00028

Ohlendorf, R., Vidavski, R.R., Eldar, A., Moffat, K., Möglich, A., 2012. From Dusk till Dawn: One-Plasmid Systems for Light-Regulated Gene Expression. Journal of Molecular Biology 416, 534–542. https://doi.org/10.1016/j.jmb.2012.01.001

Olson, Evan J, Tzouanas, C.N., Tabor, J.J., 2017. A photoconversion model for full spectral programming and multiplexing of optogenetic systems. Molecular Systems Biology 13, 926. https://doi.org/10.15252/msb.20167456

Olson, J.M., 2006. Photosynthesis in the Archean era. Photosynthesis Research 88, 109–117. https://doi.org/10.1007/s11120-006-9040-5

Olson, E.J., Hartsough, L.A., Landry, B.P., Shroff, R., Tabor, J.J., 2014. Characterizing bacterial gene circuit dynamics with optically programmed gene expression signals. Nat Methods 11, 449–455. https://doi.org/10.1038/nmeth.2884

Ong, N.T., Tabor, J.J., 2018. A Miniaturized Escherichia coli Green Light Sensor with High Dynamic Range. ChemBioChem 19, 1255–1258. https://doi.org/10.1002/cbic.201800007

Opačić, M., Vos, E.P.P., Hesp, B.H., Broos, J., 2010. Localization of the Substrate-binding Site in the Homodimeric Mannitol Transporter, EIImtl, of Escherichia coli*. Journal of Biological Chemistry 285, 25324–25331. https://doi.org/10.1074/jbc.M110.122523

Parmar, A., Singh, N.K., Kaushal, A., Sonawala, S., Madamwar, D., 2011. Purification, characterization and comparison of phycoerythrins from three different marine cyanobacterial cultures. Bioresource Technology 102, 1795–1802. https://doi.org/10.1016/j.biortech.2010.09.025

Pattanaik, B., Whitaker, M.J., Montgomery, B.L., 2012. Light Quantity Affects the Regulation of Cell Shape in Fremyella diplosiphon. Front. Microbiol. 3. https://doi.org/10.3389/fmicb.2012.00170

Pereira, H., Páramo, J., Silva, J., Marques, A., Barros, A., Maurício, D., Santos, T., Schulze, P., Barros, R., Gouveia, L., Barreira, L., Varela, J., 2018. Scale-up and large-scale production of Tetraselmis sp. CTP4 (Chlorophyta) for CO2 mitigation: from an agar plate to 100-m3 industrial photobioreactors. Sci Rep 8, 5112. https://doi.org/10.1038/s41598-018-23340-3

Pereira, S.B., Sousa, A., Santos, M., Araújo, M., Serôdio, F., Granja, P., Tamagnini, P., 2019. Strategies to Obtain Designer Polymers Based on Cyanobacterial Extracellular Polymeric Substances (EPS). International Journal of Molecular Sciences 20, 5693. https://doi.org/10.3390/ijms20225693

Pope, M.A., Hodge, J.A., Nixon, P.J., 2020. An Improved Natural Transformation Protocol for the Cyanobacterium Synechocystis sp. PCC 6803. Frontiers in Plant Science 11, 372. https://doi.org/10.3389/fpls.2020.00372

Pouzet, S., Banderas, A., Le Bec, M., Lautier, T., Truan, G., Hersen, P., 2020. The Promise of Optogenetics for Bioproduction: Dynamic Control Strategies and Scale-Up Instruments. Bioengineering 7, 151. https://doi.org/10.3390/bioengineering7040151

Price, S., Kuzhiumparambil, U., Pernice, M., Ralph, P., 2022. Techno-economic analysis of cyanobacterial PHB bioplastic production. Journal of Environmental Chemical Engineering 10, 107502. https://doi.org/10.1016/j.jece.2022.107502

Pritam, P., Sarnaik, A.P., Wangikar, P.P., 2023. Metabolic engineering of Synechococcus elongatus for photoautotrophic production of mannitol.

Biotechnology and Bioengineering 120, 2363–2370. https://doi.org/10.1002/bit.28479

Pu, L., Yang, S., Xia, A., Jin, F., 2018. Optogenetics Manipulation Enables Prevention of Biofilm Formation of Engineered Pseudomonas aeruginosa on Surfaces. ACS Synth. Biol. 7, 200–208. https://doi.org/10.1021/acssynbio.7b00273

Ranzani, A.T., Wehrmann, M., Kaiser, J., Juraschitz, M., Weber, A.M., Pietruschka, G., Gerken, U., Mayer, G., Möglich, A., 2022. Light-Dependent Control of Bacterial Expression at the mRNA Level. ACS Synth. Biol. 11, 3482–3492. https://doi.org/10.1021/acssynbio.2c00365

Rast, A., Heinz, S., Nickelsen, J., 2015. Biogenesis of thylakoid membranes. Biochimica et Biophysica Acta (BBA) - Bioenergetics, SI: Chloroplast Biogenesis 1847, 821–830. https://doi.org/10.1016/j.bbabio.2015.01.007

Raven, J.A., Beardall, J., Quigg, A., 2020. Light-Driven Oxygen Consumption in the Water-Water Cycles and Photorespiration, and Light Stimulated Mitochondrial Respiration, in: Larkum, A.W.D., Grossman, A.R., Raven, J.A. (Eds.), Photosynthesis in Algae: Biochemical and Physiological Mechanisms, Advances in Photosynthesis and Respiration. Springer International Publishing, Cham, pp. 161–178. https://doi.org/10.1007/978-3-030-33397-3_8

Reis, A.C., Salis, H.M., 2020. An Automated Model Test System for Systematic Development and Improvement of Gene Expression Models. ACS Synth. Biol. 9, 3145–3156. https://doi.org/10.1021/acssynbio.0c00394

Revyakin, A., Liu, C., Ebright, R.H., Strick, T.R., 2006. Abortive Initiation and Productive Initiation by RNA Polymerase Involve DNA Scrunching. Science 314, 1139–1143. https://doi.org/10.1126/science.1131398

Riaz-Bradley, A., 2019. Transcription in cyanobacteria: a distinctive machinery and putative mechanisms. Biochemical Society Transactions 47, 679–689. https://doi.org/10.1042/BST20180508

Rippka, R., Deruelles, J., Waterbury, J.B., Herdman, M., Stanier, R., 1979. Generic Assignments, Strain Histories and Properties of Pure Cultures of Cyanobacteria. Microbiology 111, 1–61. https://doi.org/10.1099/00221287-111-1-1

Rippka, R., Waterbury, J., Cohen-Bazire, G., 1974. A cyanobacterium which lacks thylakoids. Arch. Microbiol. 100, 419–436. https://doi.org/10.1007/BF00446333

Rockwell, N.C., Lagarias, J.C., 2017. Phytochrome diversification in cyanobacteria and eukaryotic algae. Current Opinion in Plant Biology, 37 Physiology and metabolism 2017 37, 87–93. https://doi.org/10.1016/j.pbi.2017.04.003

Rockwell, N.C., Martin, S.S., Lagarias, J.C., 2012. Red/Green Cyanobacteriochromes: Sensors of Color and Power. Biochemistry 51, 9667–9677. https://doi.org/10.1021/bi3013565

Rohnke, B.A., Singh, S.P., Pattanaik, B., Montgomery, B.L., 2018. RcaE-Dependent Regulation of Carboxysome Structural Proteins Has a Central Role in Environmental Determination of Carboxysome Morphology and Abundance in Fremyella diplosiphon. mSphere 3. https://doi.org/10.1128/mSphere.00617-17

Roy, S., Llewellyn, C.A., Egeland, E.S., Johnsen, G., 2011. Phytoplankton Pigments: Characterization, Chemotaxonomy and Applications in Oceanography. Cambridge University Press.

Ruff, E.F., Record, M.T., Artsimovitch, I., 2015. Initial Events in Bacterial Transcription Initiation. Biomolecules 5, 1035–1062. https://doi.org/10.3390/biom5021035

Ruffing, A.M., Jensen, T.J., Strickland, L.M., 2016. Genetic tools for advancement of Synechococcus sp. PCC 7002 as a cyanobacterial chassis. Microb Cell Fact 15, 190. https://doi.org/10.1186/s12934-016-0584-6

Saha, B.C., 2006. Production of mannitol from inulin by simultaneous enzymatic saccharification and fermentation with Lactobacillus intermedius NRRL B-3693. Enzyme and Microbial Technology 39, 991–995. https://doi.org/10.1016/j.enzmictec.2006.02.001

Sanfilippo, J.E., Garczarek, L., Partensky, F., Kehoe, D.M., 2019. Chromatic Acclimation in Cyanobacteria: A Diverse and Widespread Process for Optimizing Photosynthesis. Annual Review of Microbiology 73, 407–433. https://doi.org/10.1146/annurev-micro-020518-115738

Santos-Merino, M., Singh, A.K., Ducat, D.C., 2019. New applications of synthetic biology tools for cyanobacterial metabolic engineering. Frontiers in Bioengineering and Biotechnology 7. https://doi.org/10.3389/fbioe.2019.00033

Sato, T., Takamatsu, S., Yamamoto, K., Umemura, I., Tosa, T., Chibata, I., 1982. Production of L-Alanine from Ammonium Fumarate Using Two Types of Immobilized Microbial Cells, in: Chibata, Ichiro, Fukui, S., Wingard, L.B. (Eds.), Enzyme Engineering. Springer US, Boston, MA, pp. 271–272. https://doi.org/10.1007/978-1-4615-9290-7_60

Schirrmeister, B.E., Gugger, M., Donoghue, P.C.J., 2015. Cyanobacteria and the Great Oxidation Event: evidence from genes and fossils. Palaeontology 58, 769–785. https://doi.org/10.1111/pala.12178

Schmidl, S.R., Ekness, F., Sofjan, K., Daeffler, K.N.-M., Brink, K.R., Landry, B.P., Gerhardt, K.P., Dyulgyarov, N., Sheth, R.U., Tabor, J.J., 2019. Rewiring bacterial two-component systems by modular DNA-binding domain swapping. Nat Chem Biol 15, 690–698. https://doi.org/10.1038/s41589-019-0286-6

Schmidl, S.R., Sheth, R.U., Wu, A., Tabor, J.J., 2014. Refactoring and Optimization of Light-Switchable Escherichia coli Two-Component Systems. ACS Synth. Biol. 3, 820–831. https://doi.org/10.1021/sb500273n

Schneider, G.J., Hasekorn, R., 1988. RNA polymerase subunit homology among cyanobacteria, other eubacteria and archaebacteria. Journal of Bacteriology 170, 4136–4140. https://doi.org/10.1128/jb.170.9.4136-4140.1988

Schneider, G.J., Tumer, N.E., Richaud, C., Borbely, G., Haselkorn, R., 1987. Purification and characterization of RNA polymerase from the cyanobacterium Anabaena 7120. Journal of Biological Chemistry 262, 14633–14639. https://doi.org/10.1016/S0021-9258(18)47843-X

Sclavi, B., Zaychikov, E., Rogozina, A., Walther, F., Buckle, M., Heumann, H., 2005. Real-time characterization of intermediates in the pathway to open complex formation by Escherichia coli RNA polymerase at the T7A1 promoter. Proceedings of the National Academy of Sciences 102, 4706–4711. https://doi.org/10.1073/pnas.0408218102

Sedoud, A., López-Igual, R., Rehman, A. ur, Wilson, A., Perreau, F., Boulay, C., Vass, I., Krieger Liszkay, A., Kirilovsky, D., 2014. The Cyanobacterial Photoactive

Orange Carotenoid Protein Is an Excellent Singlet Oxygen Quencher. The Plant Cell 26, 1781–1791. https://doi.org/10.1105/tpc.114.123802

Shin, Y., Berry, J., Pannucci, N., Haataja, M.P., Toettcher, J.E., Brangwynne, C.P., 2017. Spatiotemporal Control of Intracellular Phase Transitions Using Light-Activated optoDroplets. Cell 168, 159-171.e14. https://doi.org/10.1016/j.cell.2016.11.054

Shono, C., Ariyanti, D., Abe, K., Sakai, Y., Sakamoto, I., Tsukakoshi, K., Sode, K., Ikebukuro, K., 2020. A Green Light-Regulated T7 RNA Polymerase Gene Expression System for Cyanobacteria. Mar Biotechnol. https://doi.org/10.1007/s10126-020-09997-w

Singh, A., Sharma, P., Saran, A.K., Singh, N., Bishnoi, N.R., 2013. Comparative study on ethanol production from pretreated sugarcane bagasse using immobilized Saccharomyces cerevisiae on various matrices. Renewable Energy 50, 488–493. https://doi.org/10.1016/j.renene.2012.07.003

Singh, N.K., Sonani, R.R., Rastogi, R.P., Madamwar, D., 2015. The phycobilisomes: an early requisite for efficient photosynthesis in cyanobacteria. EXCLI J 14, 268–289. https://doi.org/10.17179/excli2014-723

Song, J.Y., Cho, H.S., Cho, J.I., Jeon, J.S., Lagarias, J.C., Park, Y.I., 2011. Near-UV cyanobacteriochrome signaling system elicits negative phototaxis in the cyanobacterium Synechocystis sp. PCC 6803. Proceedings of the National Academy of Sciences of the United States of America 108, 10780–10785. https://doi.org/10.1073/pnas.1104242108

Sozer, O., Komenda, J., Ughy, B., Domonkos, I., Laczkó-Dobos, H., Malec, P., Gombos, Z., Kis, M., 2010. Involvement of Carotenoids in the Synthesis and Assembly of Protein Subunits of Photosynthetic Reaction Centers of Synechocystis sp. PCC 6803. Plant and Cell Physiology 51, 823–835. https://doi.org/10.1093/pcp/pcq031

Stadnichuk, I.N., Krasilnikov, P.M., Zlenko, D.V., 2015. Cyanobacterial phycobilisomes and phycobiliproteins. Microbiology 84, 101–111. https://doi.org/10.1134/S0026261715020150

Stamatakis, K., Tsimilli-Michael, M., Papageorgiou, G.C., 2014. On the question of the light-harvesting role of β-carotene in photosystem II and photosystem I

core complexes. Plant Physiology and Biochemistry, Photosynthesis Research for Sustainability 81, 121–127. https://doi.org/10.1016/j.plaphy.2014.01.014

Stevens, S.E., Patterson, C.O.P., Myers, J., 1973. The Production of Hydrogen Peroxide by Blue-Green Algae: A Survey. Journal of Phycology 9, 427–430. https://doi.org/10.1111/j.1529-8817.1973.tb04116.x

Sugawara, T., Chinzei, M., Numano, S., Kitazaki, C., Asayama, M., 2018. Flocculation and pentadecane production of a novel filamentous cyanobacterium Limnothrix sp. strain SK1-2-1. Biotechnol Lett 40, 829–836. https://doi.org/10.1007/s10529-018-2525-4

Swint-Kruse, L., Matthews, K.S., 2009. Allostery in the LacI/GalR family: variations on a theme. Current Opinion in Microbiology, Cell Regulation 12, 129–137. https://doi.org/10.1016/j.mib.2009.01.009

Szekeres, E., Sicora, C., Dragoş, N., Drugă, B., 2014. Selection of proper reference genes for the cyanobacterium Synechococcus PCC 7002 using real-time quantitative PCR. FEMS Microbiology Letters 359, 102–109. https://doi.org/10.1111/1574-6968.12574

Takors, R., Kopf, M., Mampel, J., Bluemke, W., Blombach, B., Eikmanns, B., Bengelsdorf, F.R., Weuster-Botz, D., Dürre, P., 2018. Using gas mixtures of CO, CO2 and H2 as microbial substrates: the do's and don'ts of successful technology transfer from laboratory to production scale. Microbial Biotechnology 11, 606–625. https://doi.org/10.1111/1751-7915.13270

Taton, A., Erikson, C., Yang, Y., Rubin, B.E., Rifkin, S.A., Golden, J.W., Golden, S.S., 2020. The circadian clock and darkness control natural competence in cyanobacteria. Nature Communications 11, 1688. https://doi.org/10.1038/s41467-020-15384-9

Tonon, T., Li, Y., McQueen-Mason, S., 2017. Mannitol biosynthesis in algae: more widespread and diverse than previously thought. New Phytologist 213, 1573–1579. https://doi.org/10.1111/nph.14358

Tschirhart, T., Shukla, V., Kelly, E.E., Schultzhaus, Z., NewRingeisen, E., Erickson, J.S., Wang, Z., Garcia, W., Curl, E., Egbert, R.G., Yeung, E., Vora, G.J., 2019. Synthetic Biology Tools for the Fast-Growing Marine Bacterium Vibrio

natriegens. ACS Synth. Biol. 8, 2069–2079. https://doi.org/10.1021/acssynbio.9b00176

Turner, S., Pryer, K.M., Miao, V.P.W., Palmer, J.D., 1999. Investigating Deep Phylogenetic Relationships among Cyanobacteria and Plastids by Small Subunit rRNA Sequence Analysis1. Journal of Eukaryotic Microbiology 46, 327–338. https://doi.org/10.1111/j.1550-7408.1999.tb04612.x

Uhlenbusch, I., Sahm, H., Sprenger, G.A., 1991. Expression of an L-alanine dehydrogenase gene in Zymomonas mobilis and excretion of L-alanine. Applied and Environmental Microbiology 57, 1360–1366. https://doi.org/10.1128/aem.57.5.1360-1366.1991

Umena, Y., Kawakami, K., Shen, J.-R., Kamiya, N., 2011. Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 Å. Nature 473, 55–60. https://doi.org/10.1038/nature09913

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., Rozen, S.G., 2012. Primer3—new capabilities and interfaces. Nucleic Acids Res 40, e115. https://doi.org/10.1093/nar/gks596

Usherenko, S., Stibbe, H., Muscó, M., Essen, L.-O., Kostina, E.A., Taxis, C., 2014. Photo-sensitive degron variants for tuning protein stability by light. BMC Syst Biol 8, 128. https://doi.org/10.1186/s12918-014-0128-9

Uyeda, J.C., Harmon, L.J., Blank, C.E., 2016. A Comprehensive Study of Cyanobacterial Morphological and Ecological Evolutionary Dynamics through Deep Geologic Time. PLOS ONE 11, e0162539. https://doi.org/10.1371/journal.pone.0162539

Vermaas, W.F., 2001. Photosynthesis and Respiration in Cyanobacteria, in: eLS. American Cancer Society. https://doi.org/10.1038/npg.els.0001670

Vijayan, V., Jain, I.H., O'Shea, E.K., 2011. A high resolution map of a cyanobacterial transcriptome. Genome Biol 12, R47. https://doi.org/10.1186/gb-2011-12-5-r47

Wardani, A.K., Nagahisa, K., Shimizu, H., Shioya, S., 2007. Reduction of lactate production in Lactococcus lactis, a combined strategy: metabolic engineering by introducing foreign alanine dehydrogenase gene and hemin addition. World J Microbiol Biotechnol 23, 947–953. https://doi.org/10.1007/s11274-006-9319-x

Whitton, B.A., 2012. Ecology of Cyanobacteria II: Their Diversity in Space and Time. Springer Science & Business Media. https://doi.org/10.1007/0-306-46855-7

Wholesale Market Indicators, Ofgem. [WWW Document] URL https://www.ofgem.gov.uk/energy-data-and-research/data-portal/wholesale-market-indicators (accessed 12.1.24).

Wilson, A., Punginelli, C., Gall, A., Bonetti, C., Alexandre, M., Routaboul, J.-M., Kerfeld, C.A., Grondelle, R. van, Robert, B., Kennis, J.T.M., Kirilovsky, D., 2008. A photoactive carotenoid protein acting as light intensity sensor. PNAS 105, 12075–12080. https://doi.org/10.1073/pnas.0804636105

Wiltbank, L.B., Kehoe, D.M., 2019. Diverse light responses of cyanobacteria mediated by phytochrome superfamily photoreceptors. Nature Reviews Microbiology 17, 37–50. https://doi.org/10.1038/s41579-018-0110-4

Wiltbank, L.B., Kehoe, D.M., 2016. Two Cyanobacterial Photoreceptors Regulate Photosynthetic Light Harvesting by Sensing Teal, Green, Yellow, and Red Light. mBio 7, 10.1128/mbio.02130-15. https://doi.org/10.1128/mbio.02130-15

Włodarczyk, A., Selão, T.T., Norling, B., Nixon, P.J., 2020. Newly discovered Synechococcus sp. PCC 11901 is a robust cyanobacterial strain for high biomass production. Commun Biol 3, 1–14. https://doi.org/10.1038/s42003-020-0910-8

Wu, W., Du, W., Gallego, R.P., Hellingwerf, K.J., van der Woude, A.D., Branco dos Santos, F., 2020. Using osmotic stress to stabilize mannitol production in Synechocystis sp. PCC6803. Biotechnol Biofuels 13, 117. https://doi.org/10.1186/s13068-020-01755-3

Wu, Y.I., Frey, D., Lungu, O.I., Jaehrig, A., Schlichting, I., Kuhlman, B., Hahn, K.M., 2009. A genetically encoded photoactivatable Rac controls the motility of living cells. Nature 461, 104–108. https://doi.org/10.1038/nature08241

Xie, F., Wang, J., Zhang, B., 2023. RefFinder: a web-based tool for comprehensively analyzing and identifying reference genes. Funct Integr Genomics 23, 125. https://doi.org/10.1007/s10142-023-01055-7

Xiong, Q., Feng, J., Li, S., Zhang, G., Qiao, Z., Chen, Z., Wu, Y., Lin, Y., Li, T., Ge, F., Zhao, J., 2015. Integrated Transcriptomic and Proteomic Analysis of the

Global Response of Synechococcus to High Light Stress. & Cellular Proteomics 14, 1038–1053. https://doi.org/10.1074/mcp.M114.046003

Xu, Y., Alvey, R.M., Byrne, P.O., Graham, J.E., Shen, G., Bryant, D.A., 2011. Expression of Genes in Cyanobacteria: Adaptation of Endogenous Plasmids as Platforms for High Level Gene Expression in Synechococcus sp. PCC 7002, in: Carpentier, R. (Ed.), Photosynthesis Research Protocols, Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 273–293. https://doi.org/10.1007/978-1-60761-925-3_21

Xu, Y., Tiago Guerra, L., Li, Z., Ludwig, M., Charles Dismukes, G., Bryant, D.A., 2013. Altered carbohydrate metabolism in glycogen synthase mutants of *Synechococcus* sp. strain PCC 7002: Cell factories for soluble sugars. Metabolic Engineering 16, 56–67. https://doi.org/10.1016/j.ymben.2012.12.002

Yang, Y., Lam, V., Adomako, M., Simkovsky, R., Jakob, A., Rockwell, N.C., Cohen, S.E., Taton, A., Wang, J., Lagarias, J.C., Wilde, A., Nobles, D.R., Brand, J.J., Golden, S.S., 2018. Phototaxis in a wild isolate of the cyanobacterium Synechococcus elongatus. Proceedings of the National Academy of Sciences 115, E12378–E12387. https://doi.org/10.1073/pnas.1812871115

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., Madden, T.L., 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 13, 134. https://doi.org/10.1186/1471-2105-13-134

Yuan, S.-F., Yi, X., Johnston, T.G., Alper, H.S., 2020. De novo resveratrol production through modular engineering of an Escherichia coli–Saccharomyces cerevisiae co-culture. Microbial Cell Factories 19, 143. https://doi.org/10.1186/s12934-020-01401-5

Zavřel, T., Faizi, M., Loureiro, C., Poschmann, G., Stühler, K., Sinetova, M., Zorina, A., Steuer, R., Červený, J., 2019. Quantitative insights into the cyanobacterial cell economy. eLife 8, e42508. https://doi.org/10.7554/eLife.42508

Zayner, J.P., Antoniou, C., French, A.R., Hause, R.J., Sosnick, T.R., 2013. Investigating Models of Protein Function and Allostery With a Widespread Mutational Analysis of a Light-Activated Protein. Biophys J 105, 1027–1036. https://doi.org/10.1016/j.bpj.2013.07.010

Zehr, J.P., 2011. Nitrogen fixation by marine cyanobacteria. Trends in Microbiology 19, 162–173. https://doi.org/10.1016/j.tim.2010.12.004

Zerulla, K., Chimileski, S., Näther, D., Gophna, U., Papke, R.T., Soppa, J., 2014. DNA as a Phosphate Storage Polymer and the Alternative Advantages of Polyploidy for Growth or Survival. PLOS ONE 9, e94819. https://doi.org/10.1371/journal.pone.0094819

Zerulla, K., Ludt, K., Soppa, J., 2016. The ploidy level of Synechocystis sp. PCC 6803 is highly variable and is influenced by growth phase and by chemical and physical external parameters. Microbiology 162, 730–739. https://doi.org/10.1099/mic.0.000264

Zhang, L., Bryan, S.J., Selão, T.T., 2022. Sustainable citric acid production from CO2 in an engineered cyanobacterium. Frontiers in Microbiology 13.

Zhang, M., Gu, L., Cheng, C., Zhu, J., Wu, H., Ma, J., Dong, W., Kong, X., Jiang, M., Ouyang, P., 2017. High-yield production of mannitol by Leuconostoc pseudomesenteroides CTCC G123 from chicory-derived inulin hydrolysate. Journal of Industrial Microbiology and Biotechnology 44, 1237–1244. https://doi.org/10.1007/s10295-017-1953-9

Zhang, P., Yang, J., Cho, E., Lu, Y., 2020. Bringing Light into Cell-Free Expression. ACS Synth. Biol. 9, 2144–2153. https://doi.org/10.1021/acssynbio.0c00211

Zhang, X., Betterle, N., Hidalgo Martinez, D., Melis, A., 2021. Recombinant Protein Stability in Cyanobacteria. ACS Synth. Biol. 10, 810–825. https://doi.org/10.1021/acssynbio.0c00610

Zhang, X., Jantama, K., Moore, J.C., Shanmugam, K.T., Ingram, L.O., 2007. Production of l-alanine by metabolically engineered Escherichia coli. Appl Microbiol Biotechnol 77, 355–366. https://doi.org/10.1007/s00253-007-1170-y

Zhao, C., Li, Z., Li, T., Zhang, Y., Bryant, D.A., Zhao, J., 2015. High-yield production of extracellular type-I cellulose by the cyanobacterium Synechococcus sp. PCC 7002. Cell Discov 1, 1–12. https://doi.org/10.1038/celldisc.2015.4

Zhao, E.M., Lalwani, M.A., Chen, J.-M., Orillac, P., Toettcher, J.E., Avalos, J.L., 2021. Optogenetic Amplification Circuits for Light-Induced Metabolic Control. ACS Synth. Biol. 10, 1143–1154. https://doi.org/10.1021/acssynbio.0c00642

Zhao, E.M., Suek, N., Wilson, M.Z., Dine, E., Pannucci, N.L., Gitai, Z., Avalos, J.L., Toettcher, J.E., 2019. Light-based control of metabolic flux through assembly of synthetic organelles. Nat Chem Biol 15, 589–597. https://doi.org/10.1038/s41589-019-0284-8

Zhao, E.M., Zhang, Y., Mehl, J., Park, H., Lalwani, M.A., Toettcher, J.E., Avalos, J.L., 2018. Optogenetic regulation of engineered cellular metabolism for microbial chemical production. Nature 555, 683–687. https://doi.org/10.1038/nature26141

Zheng, L., Zheng, Z., Li, X., Wang, G., Zhang, K., Wei, P., Zhao, J., Gao, N., 2021. Structural insight into the mechanism of energy transfer in cyanobacterial phycobilisomes. Nat Commun 12, 5497. https://doi.org/10.1038/s41467-021-25813-y

Zheng, X., O'Shea, E.K., 2017. Cyanobacteria Maintain Constant Protein Concentration despite Genome Copy-Number Variation. Cell Reports 19, 497–504. https://doi.org/10.1016/j.celrep.2017.03.067

Zhou, L., Deng, C., Cui, W.-J., Liu, Z.-M., Zhou, Z.-M., 2016. Efficient L-Alanine Production by a Thermo-Regulated Switch in Escherichia coli. Appl Biochem Biotechnol 178, 324–337. https://doi.org/10.1007/s12010-015-1874-x

Zhou, J., Zhang, H., Meng, H., Zhu, Y., Bao, G., Zhang, Y., Li, Y., Ma, Y., 2014. Discovery of a super-strong promoter enables efficient production of heterologous proteins in cyanobacteria. Sci Rep 4, 4500. https://doi.org/10.1038/srep04500

Zhu, Y., Graham, J.E., Ludwig, M., Xiong, W., Alvey, R.M., Shen, G., Bryant, D.A., 2010. Roles of xanthophyll carotenoids in protection against photoinhibition and oxidative stress in the cyanobacterium Synechococcus sp. strain PCC 7002. Archives of Biochemistry and Biophysics, Carotenoids 504, 86–99. https://doi.org/10.1016/j.abb.2010.07.007

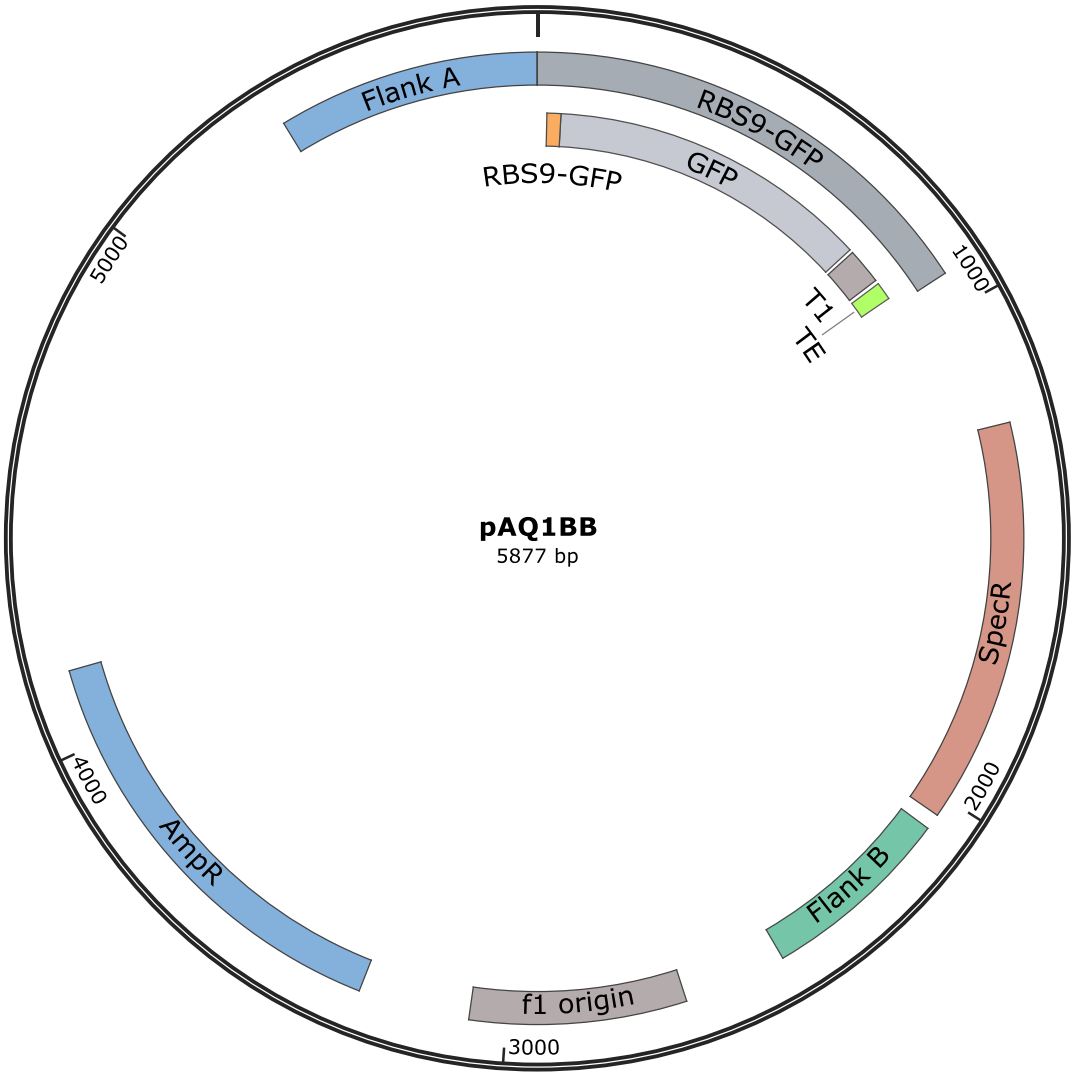
Zinicovscaia, I., Cepoi, L., Povar, I., Chiriac, T., Rodlovskaya, E., Culicov, O.A., 2018. Metal Uptake from Complex Industrial Effluent by Cyanobacteria

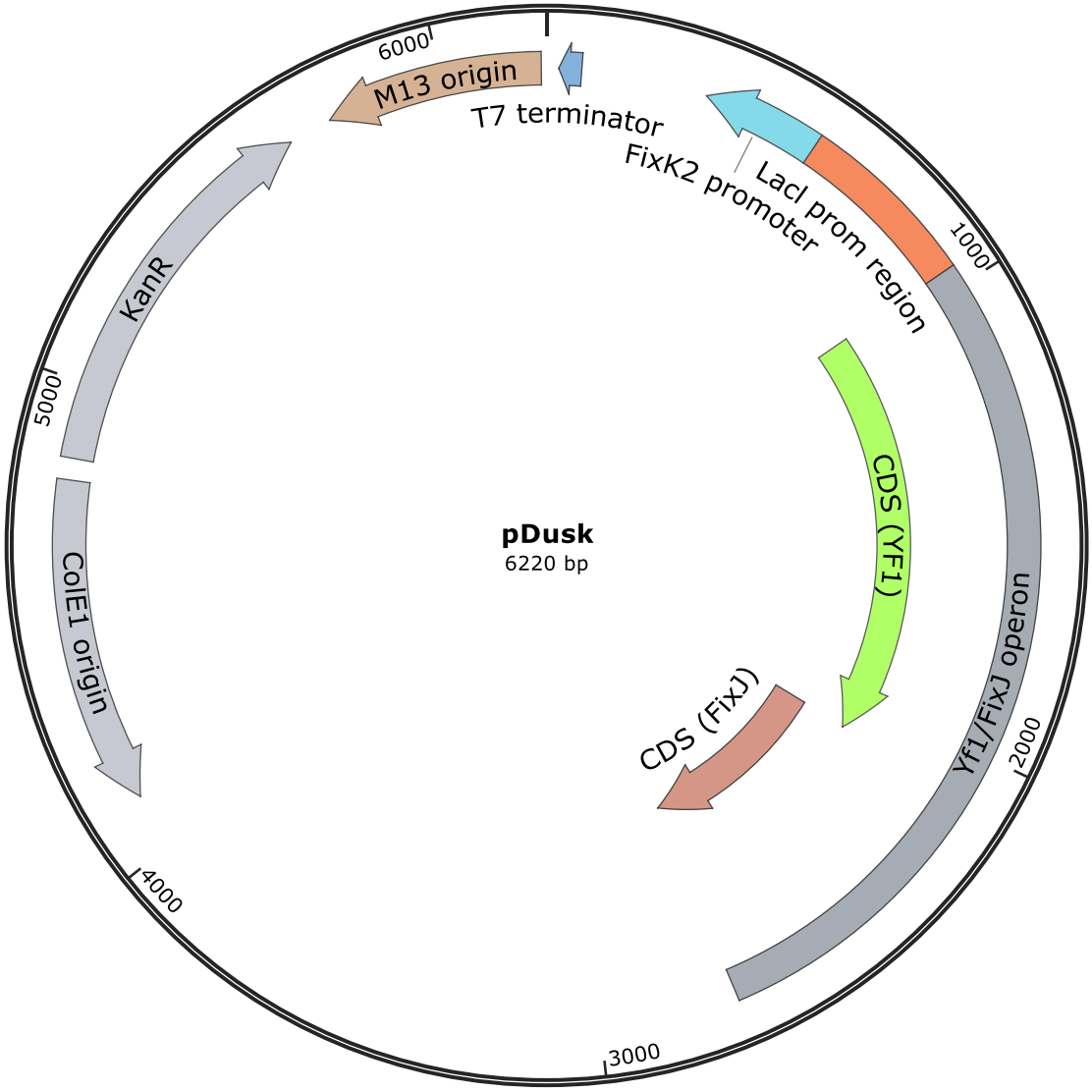
Arthrospira platensis. Water Air Soil Pollut 229, 220. https://doi.org/10.1007/s11270-018-3873-3

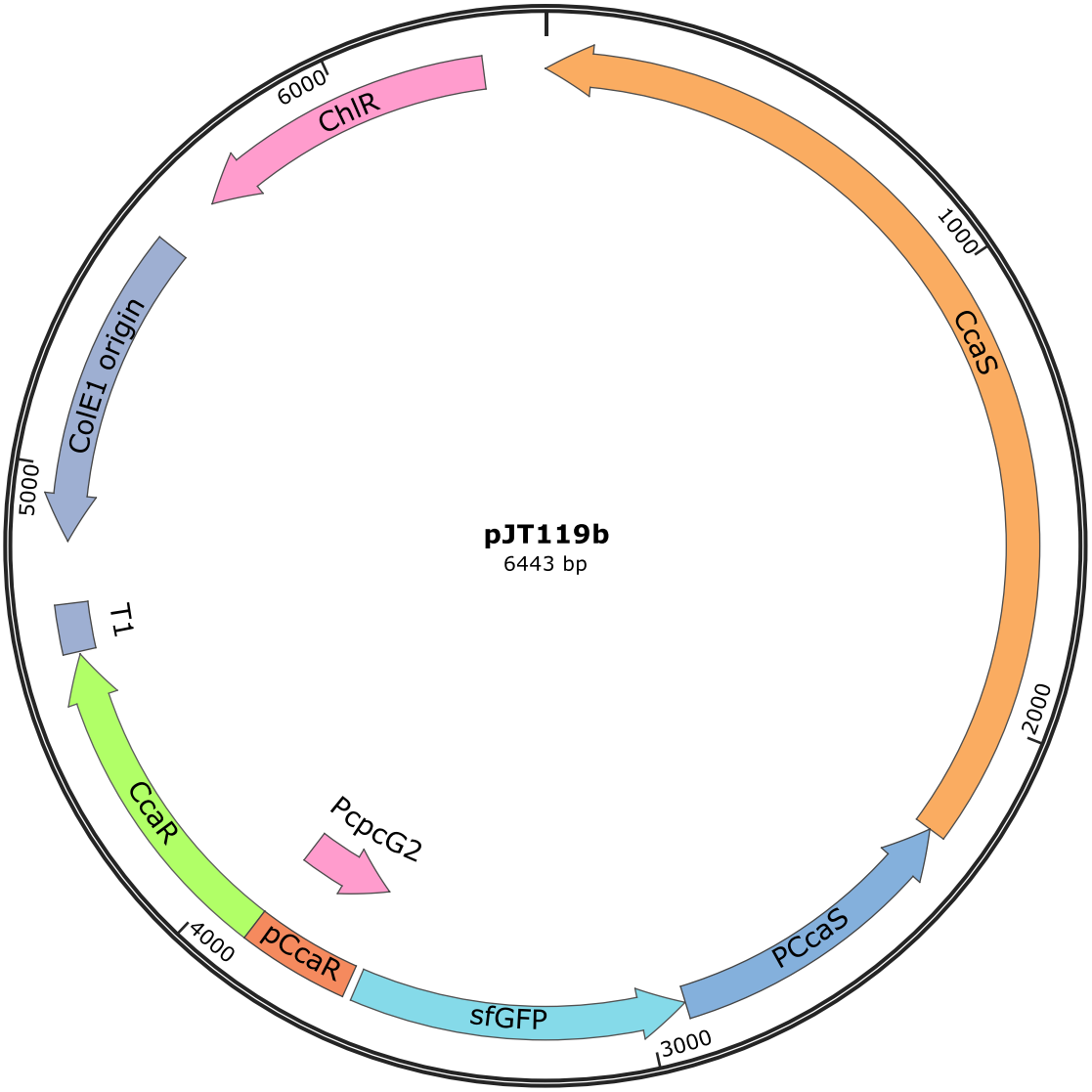
Zoltowski, B.D., Nash, A.I., Gardner, K.H., 2011. Variations in Protein–Flavin Hydrogen Bonding in a Light, Oxygen, Voltage Domain Produce Non-Arrhenius Kinetics of Adduct Decay. Biochemistry 50, 8771–8779. https://doi.org/10.1021/bi200976a

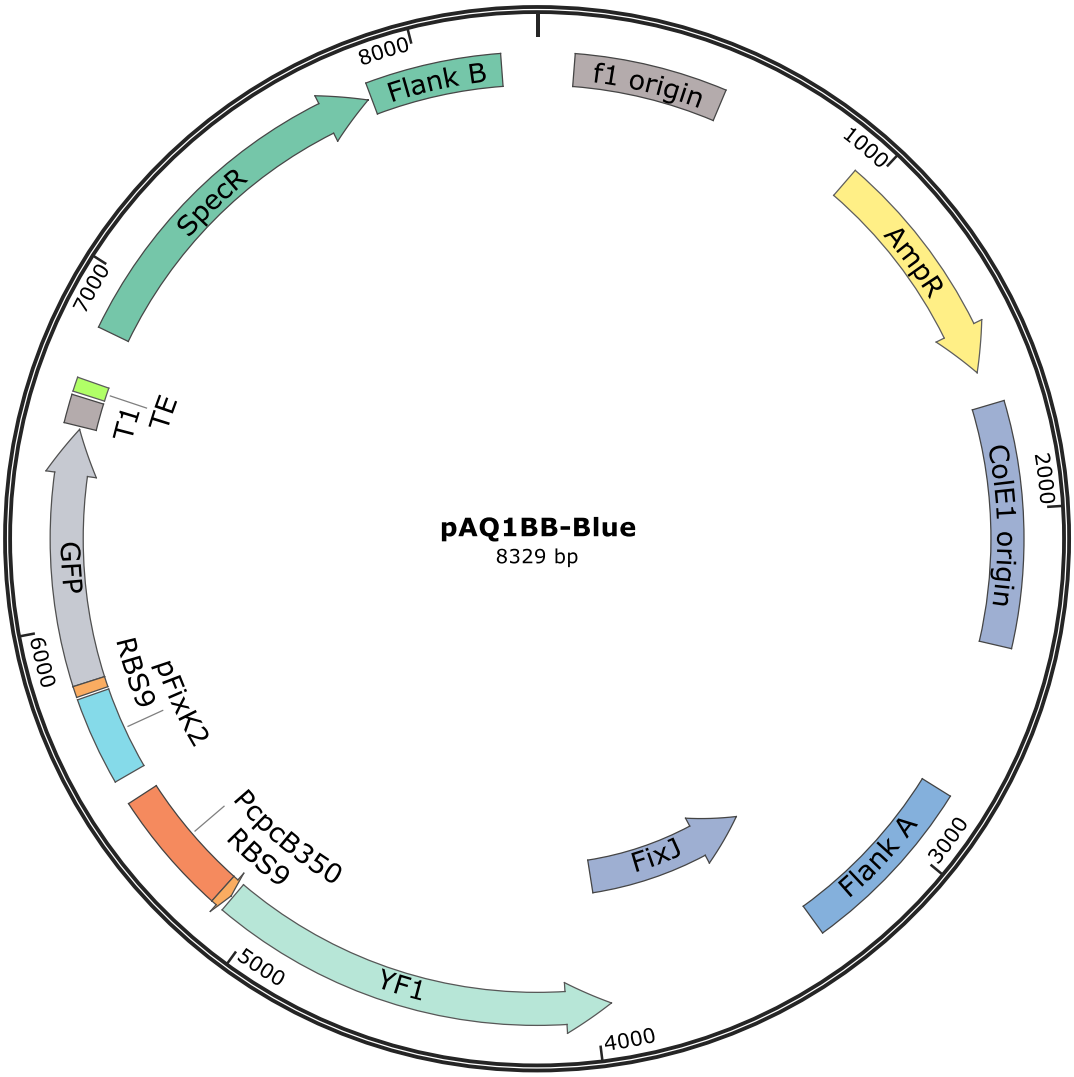
Appendix I

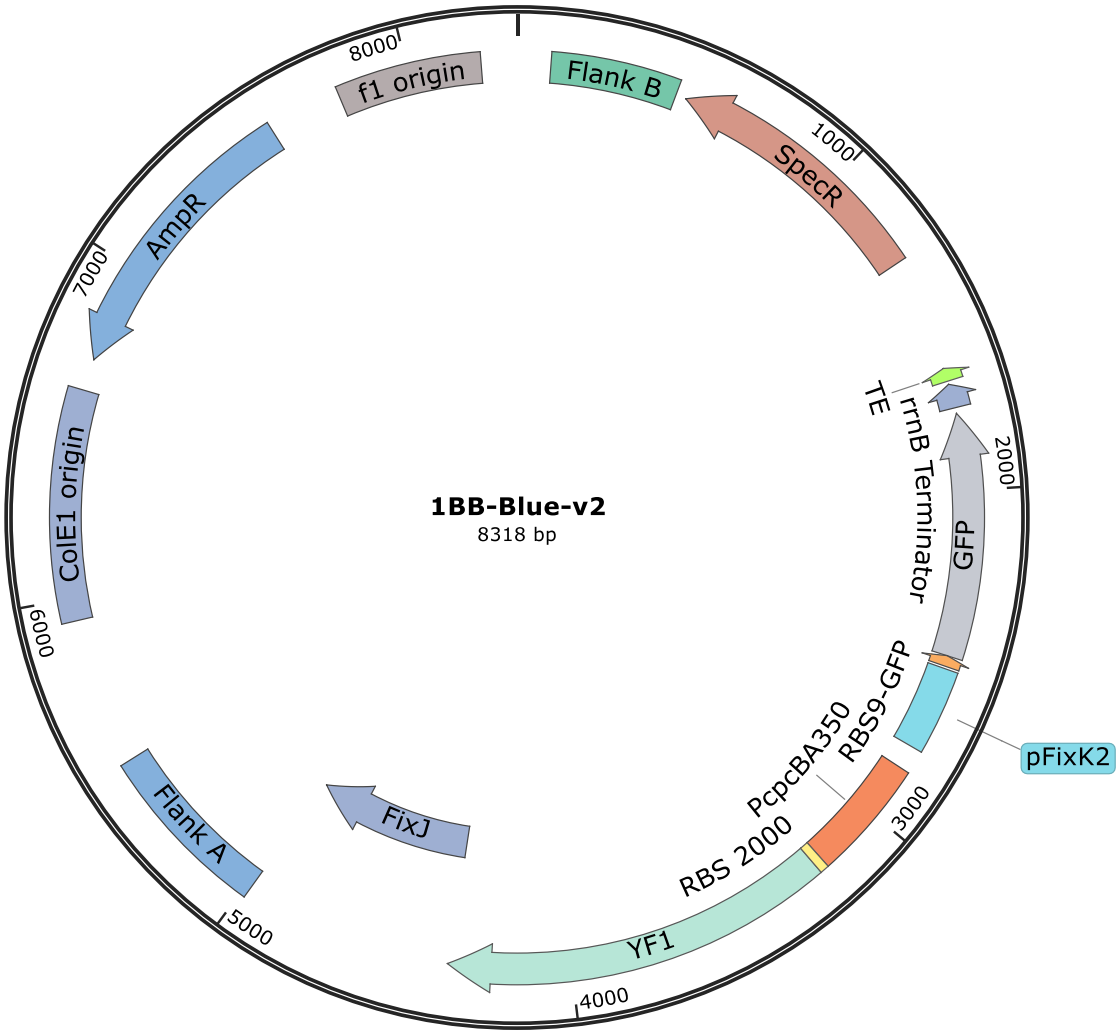
Plasmid maps of all plasmids used in this project.

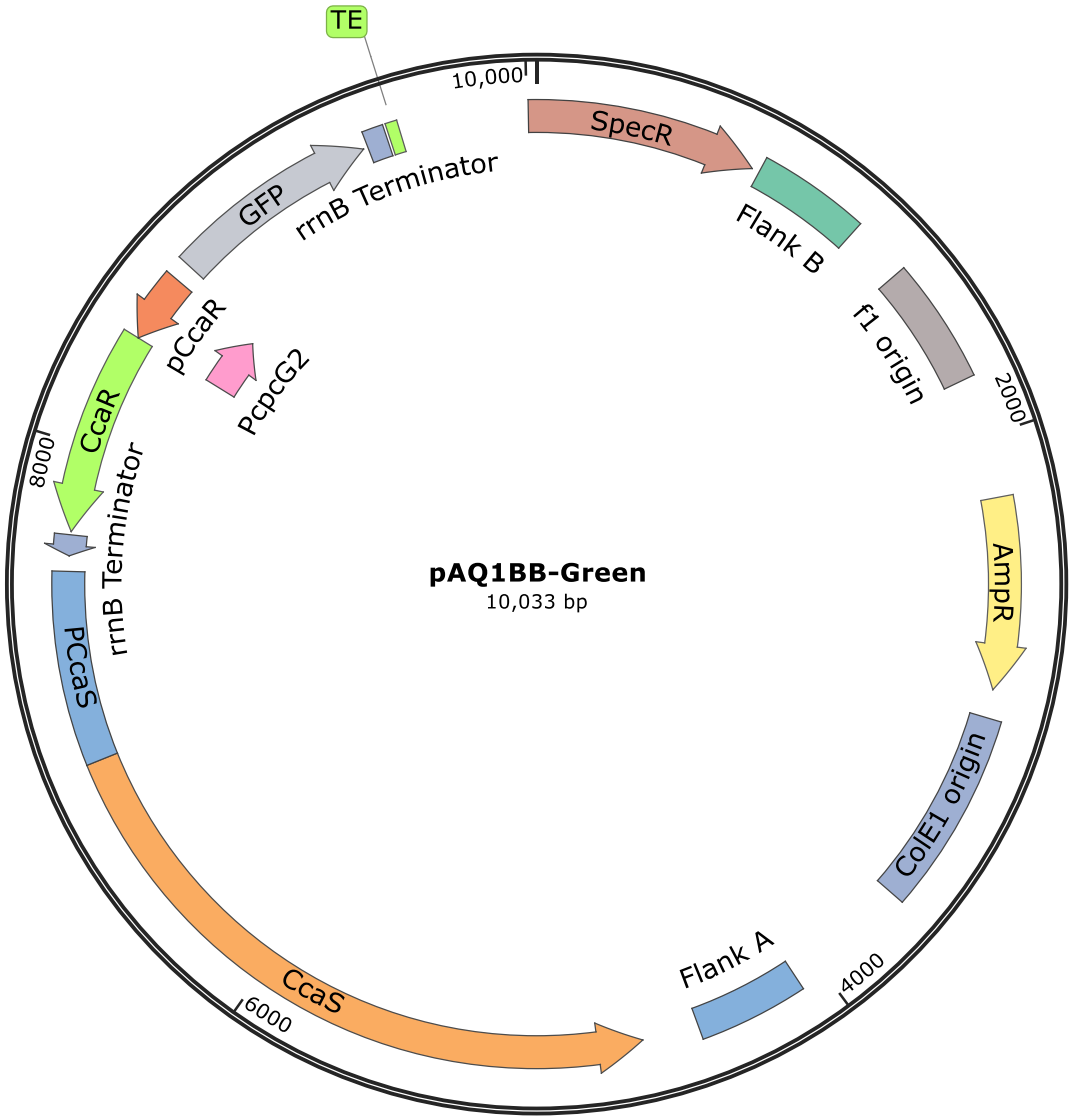


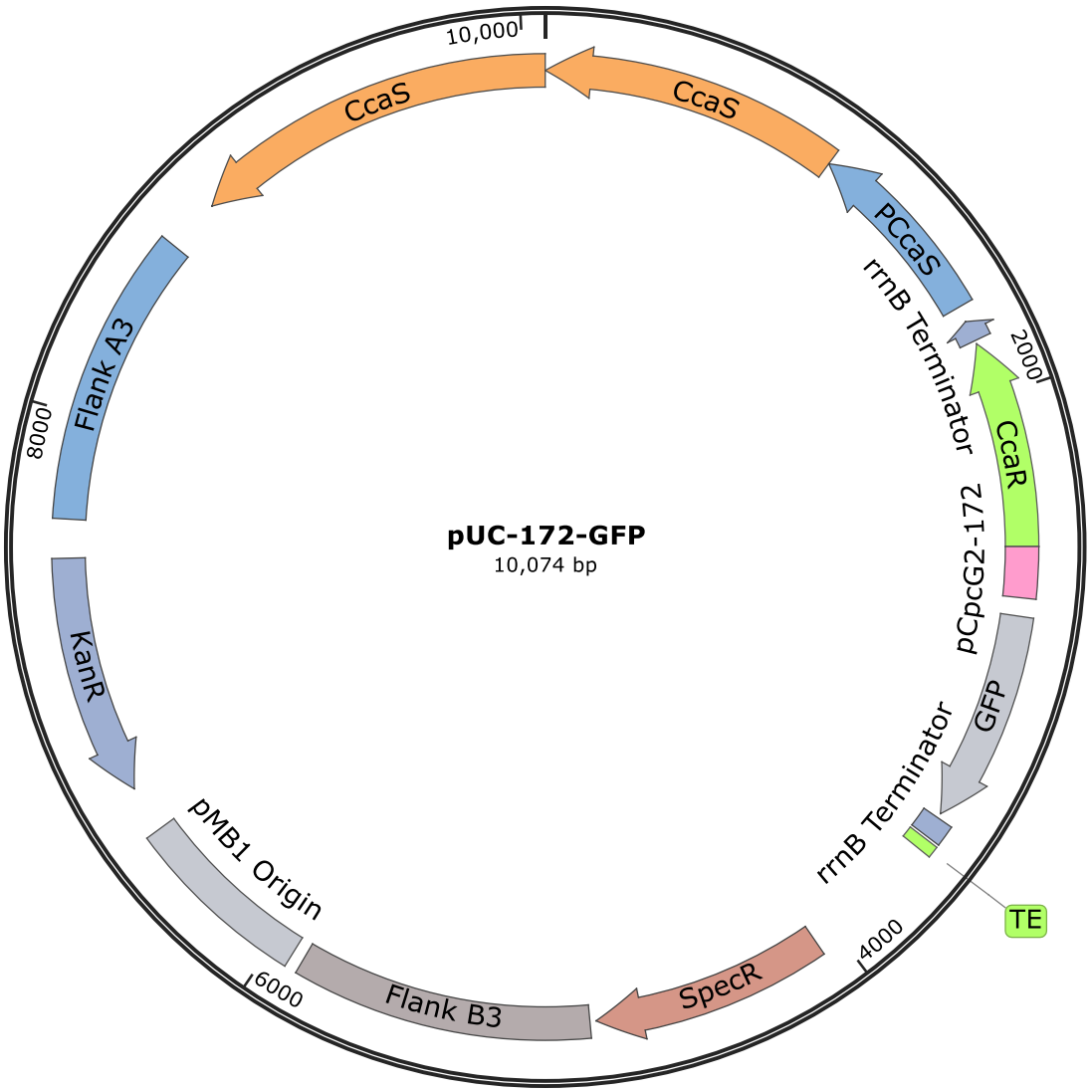

pDusk
6220 bp
6000
1000
2000
3000
4000
5000
M13 origin
T7 terminator
FixK2 promoter
LacI prom region
CDS (YF1)
CDS (FixJ)
Yf1/FixJ operon
KanR
ColE1 origin

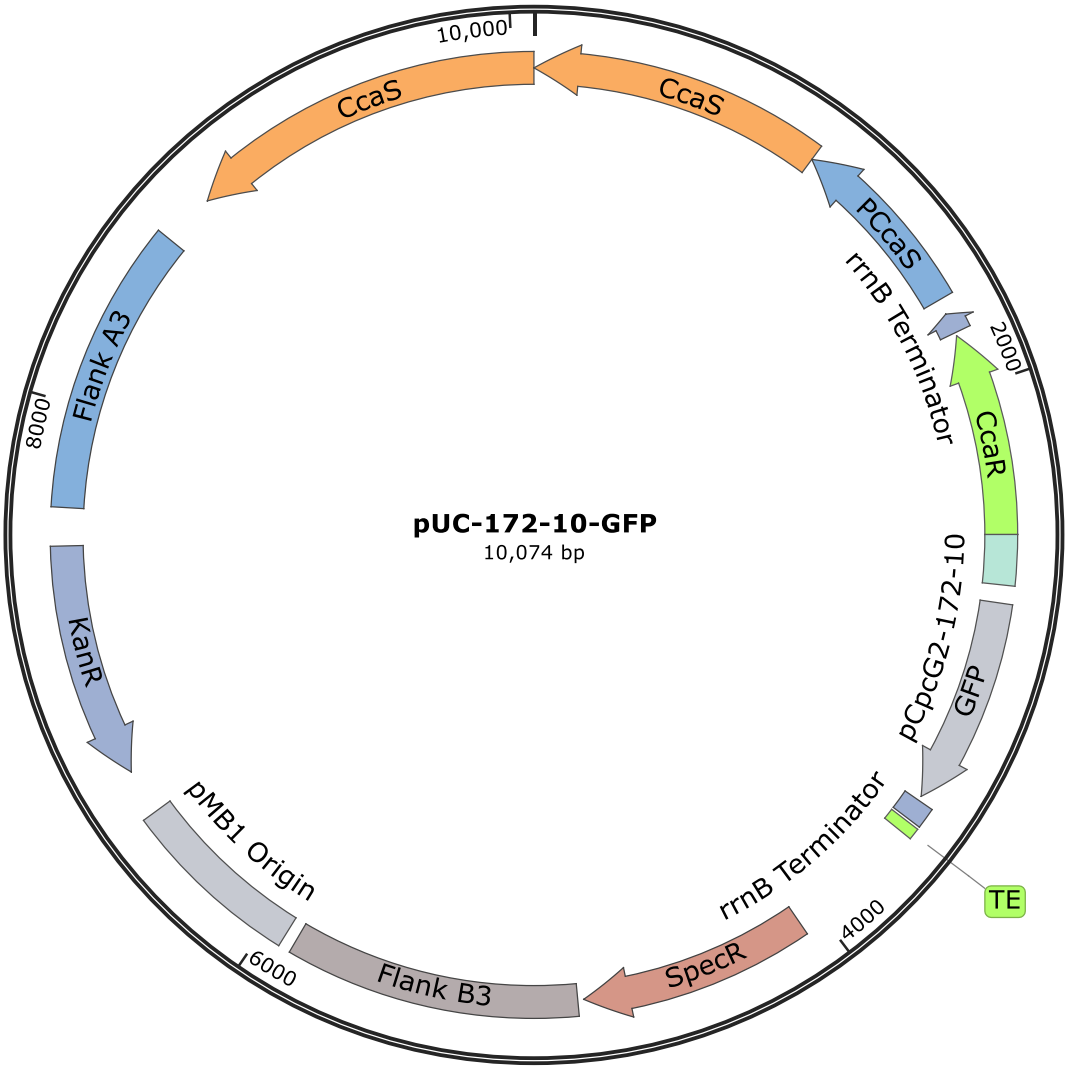

pJT119b
6443 bp
1000
2000
3000
4000
5000
6000
CcaS
PCcaS
sfGFP
pCcaR
CcaR
PcpcG2
T1
ColE1 origin
ChlR

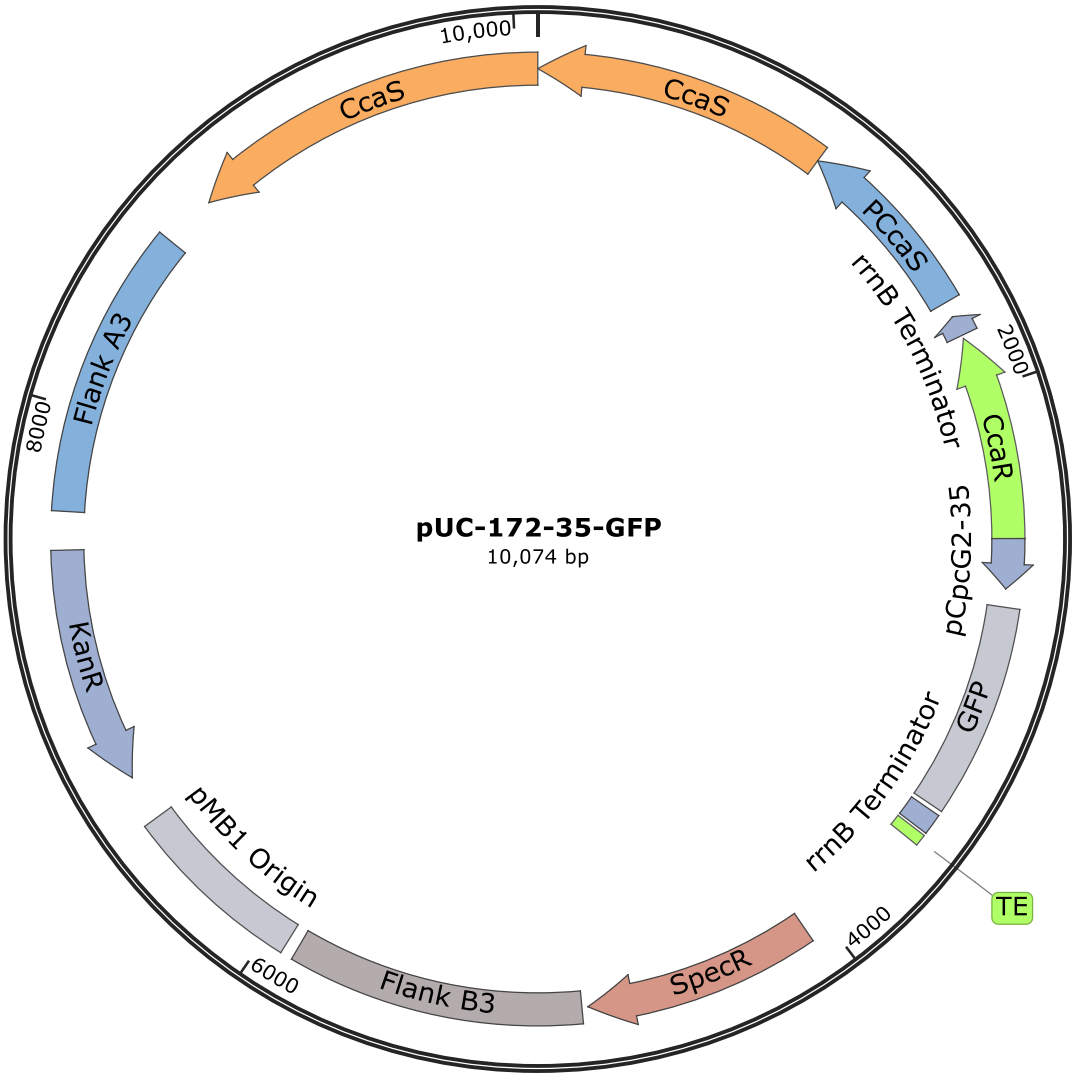

pAQ1BB-Blue
8329 bp
8000
1000
2000
3000
4000
5000
6000
7000
Flank B
f1 origin
AmpR
ColE1 origin
Flank A
FixJ
YF1
PcpcB350
RBS9
pFixK2
RBS9
GFP
T1
TE
SpecR

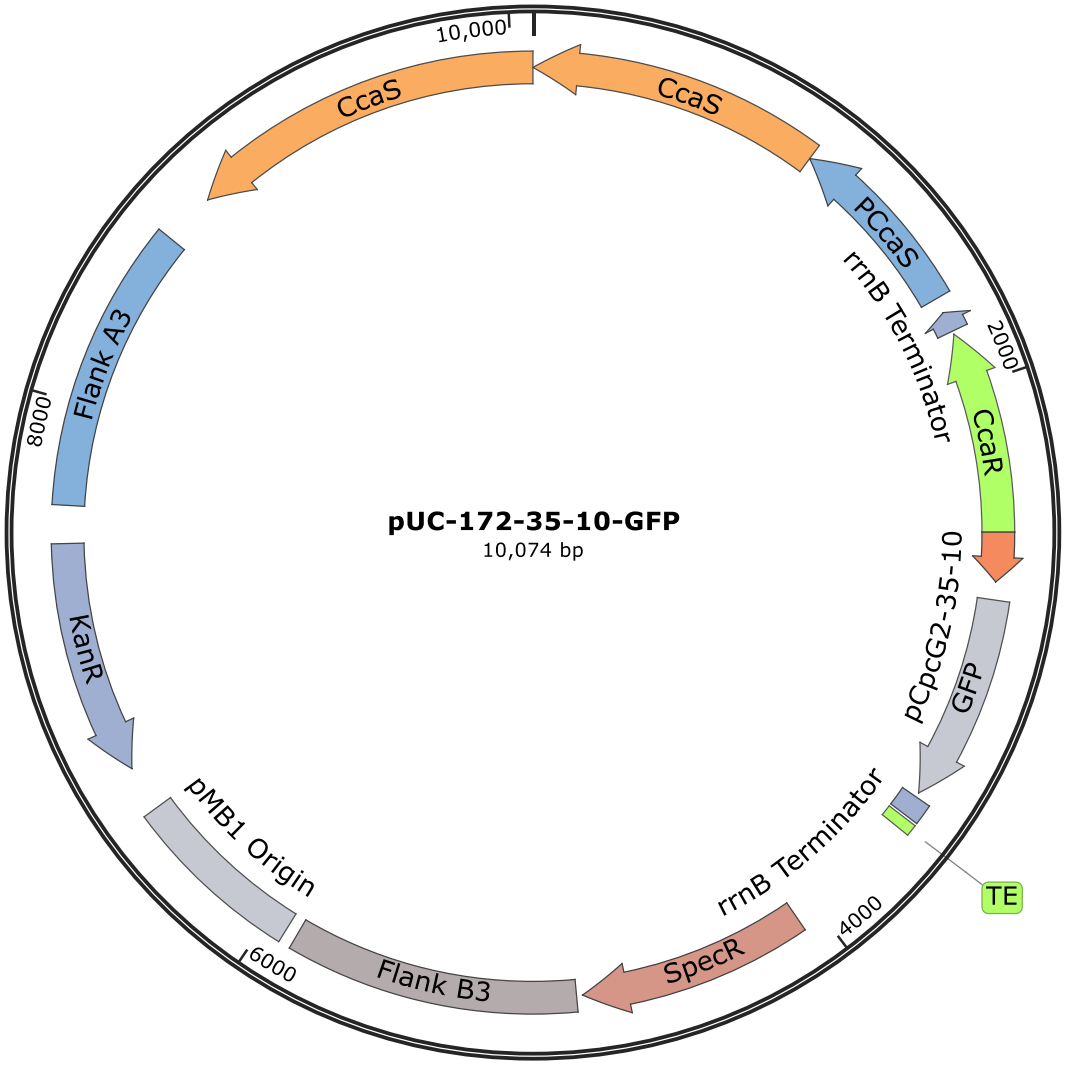

1BB-Blue-v2
8318 bp
8000
f1 origin
Flank B
1000
SpecR
7000
AmpR
ColE1 origin
6000
Flank A
5000
FixJ
4000
YF1
RBS 2000
PcpcBA350
3000
RBS9-GFP
pFixK2
GFP
rrnB Terminator
TE
2000

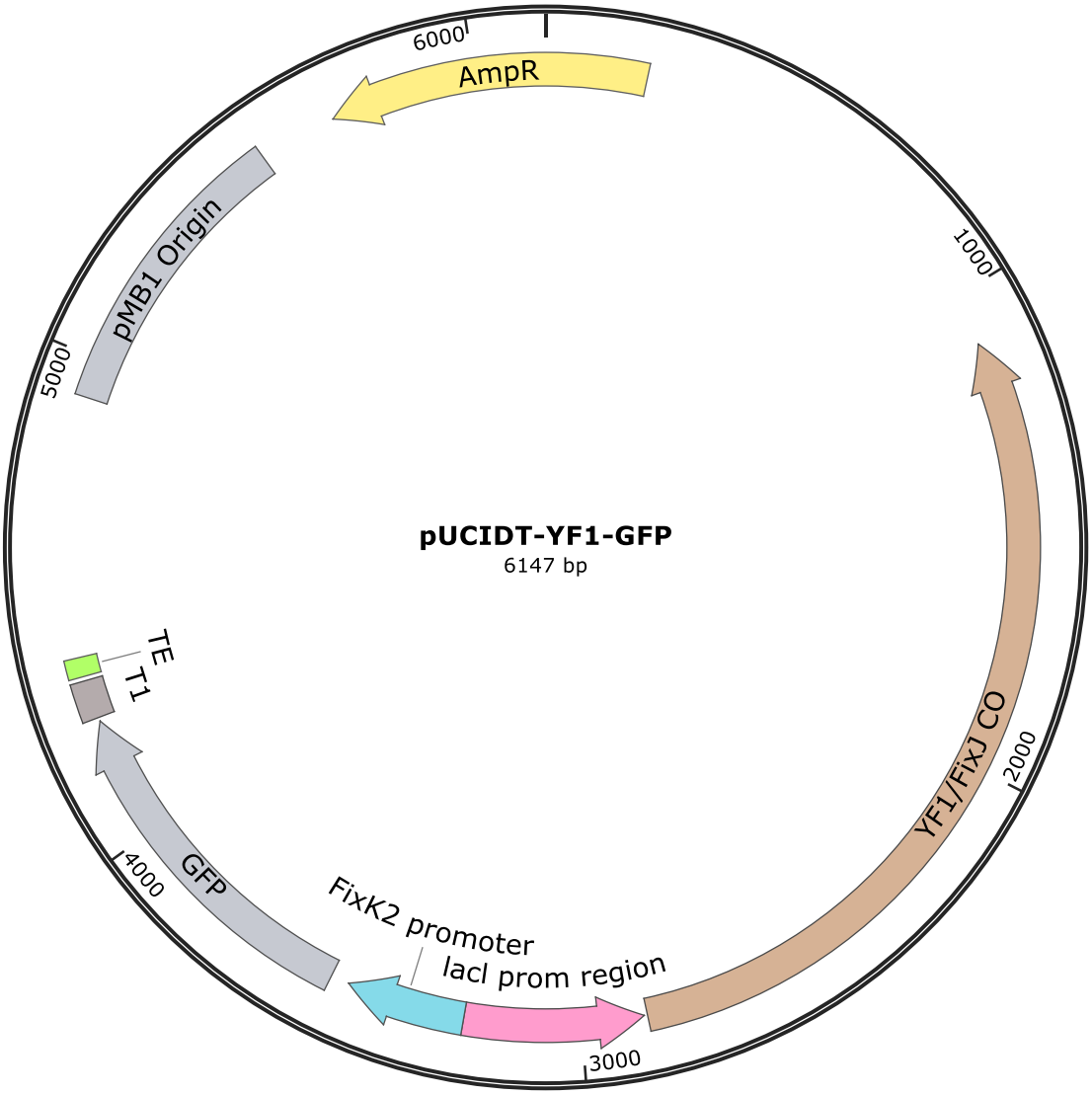

TE
10,000
SpecR
Flank B
f1 origin
2000
AmpR
ColE1 origin
4000
Flank A
CcaS
6000
PCcaS
rrnB Terminator
8000
CcaR
pCcaR
PcpcG2
GFP
rrnB Terminator
pAQ1BB-Green
10,033 bp

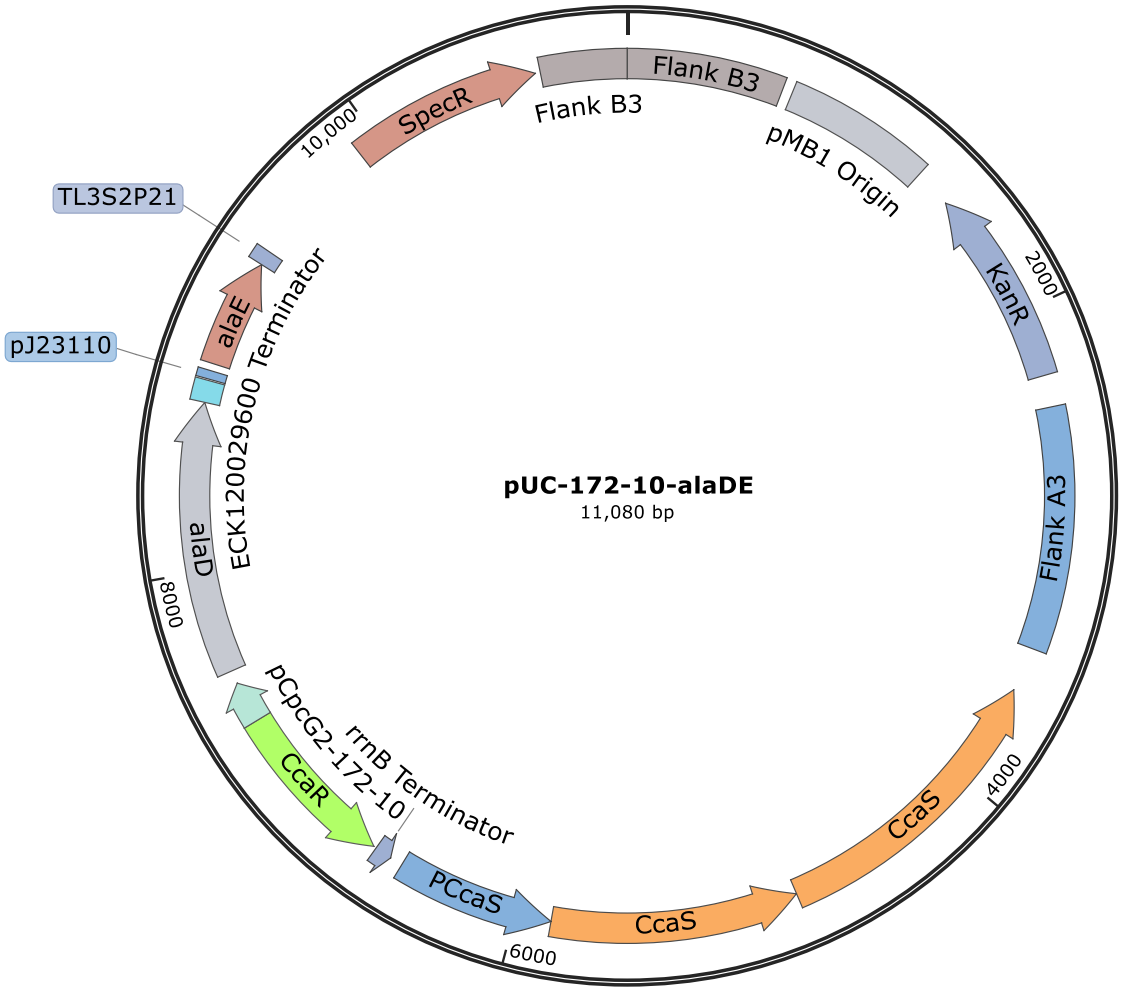

pUC-172-GFP
10,074 bp
10,000
2000
4000
6000
8000
CcaS
CcaS
PCcaS
rrnB Terminator
CcaR
pCpcG2-172
GFP
rrnB Terminator
TE
SpecR
Flank B3
pMB1 Origin
KanR
Flank A3

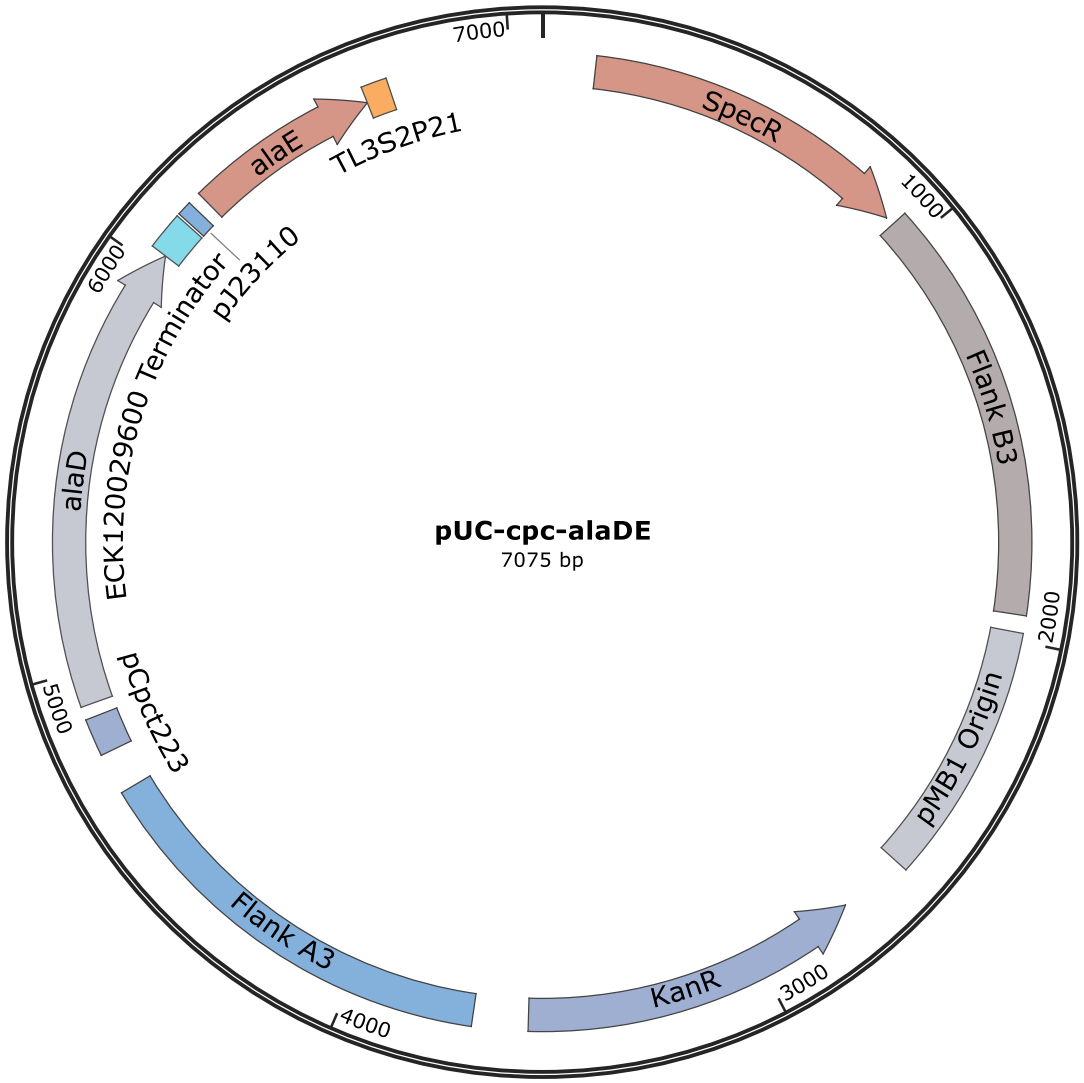

pUC-172-10-GFP
10,074 bp
10,000
CcaS
CcaS
PCcaS
rrnB Terminator
2000
CcaR
pCpcG2-172-10
GFP
TE
4000
rrnB Terminator
SpecR
Flank B3
6000
pMB1 Origin
KanR
8000
Flank A3

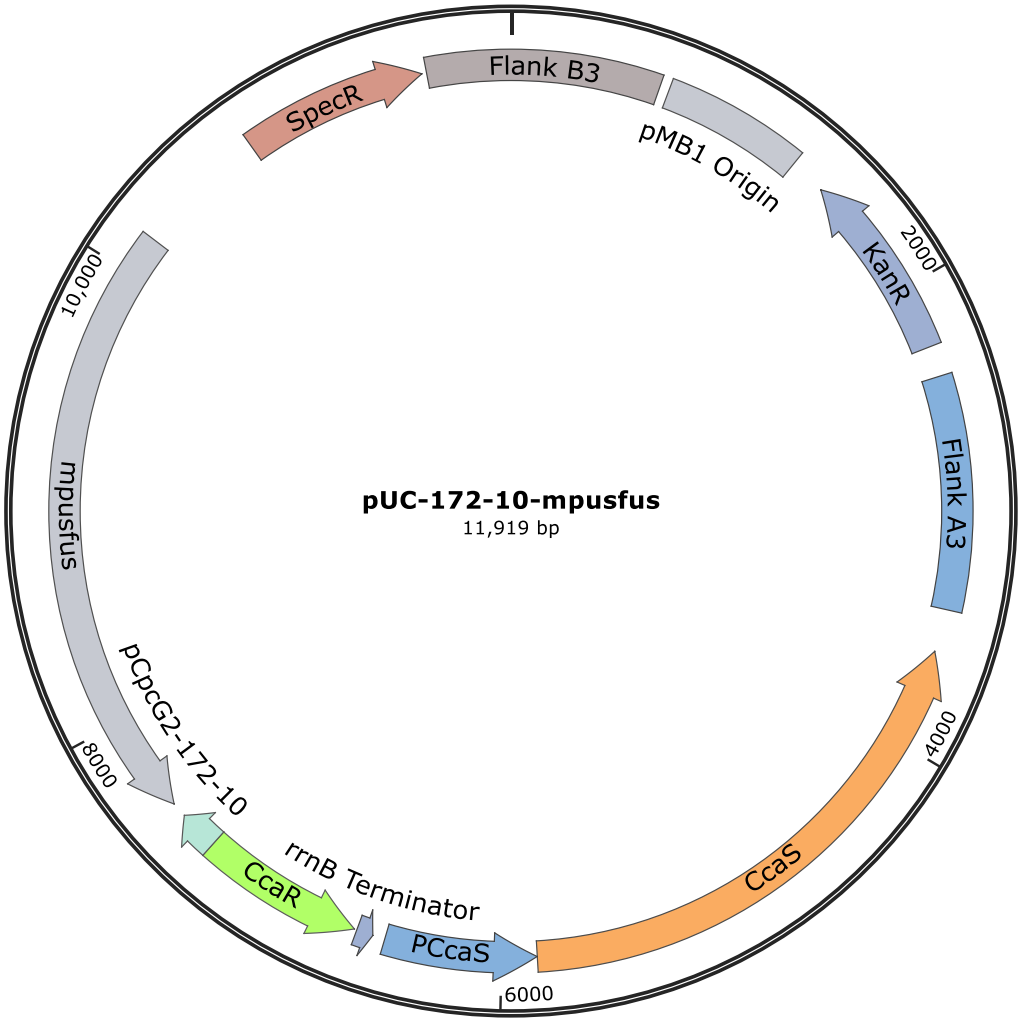

pUC-172-35-GFP
10,074 bp
10,000
2000
4000
6000
8000
CcaS
CcaS
PCcaS
rrnB Terminator
CcaR
pCpcG2-35
GFP
rrnB Terminator
TE
SpecR
Flank B3
pMB1 Origin
KanR
Flank A3

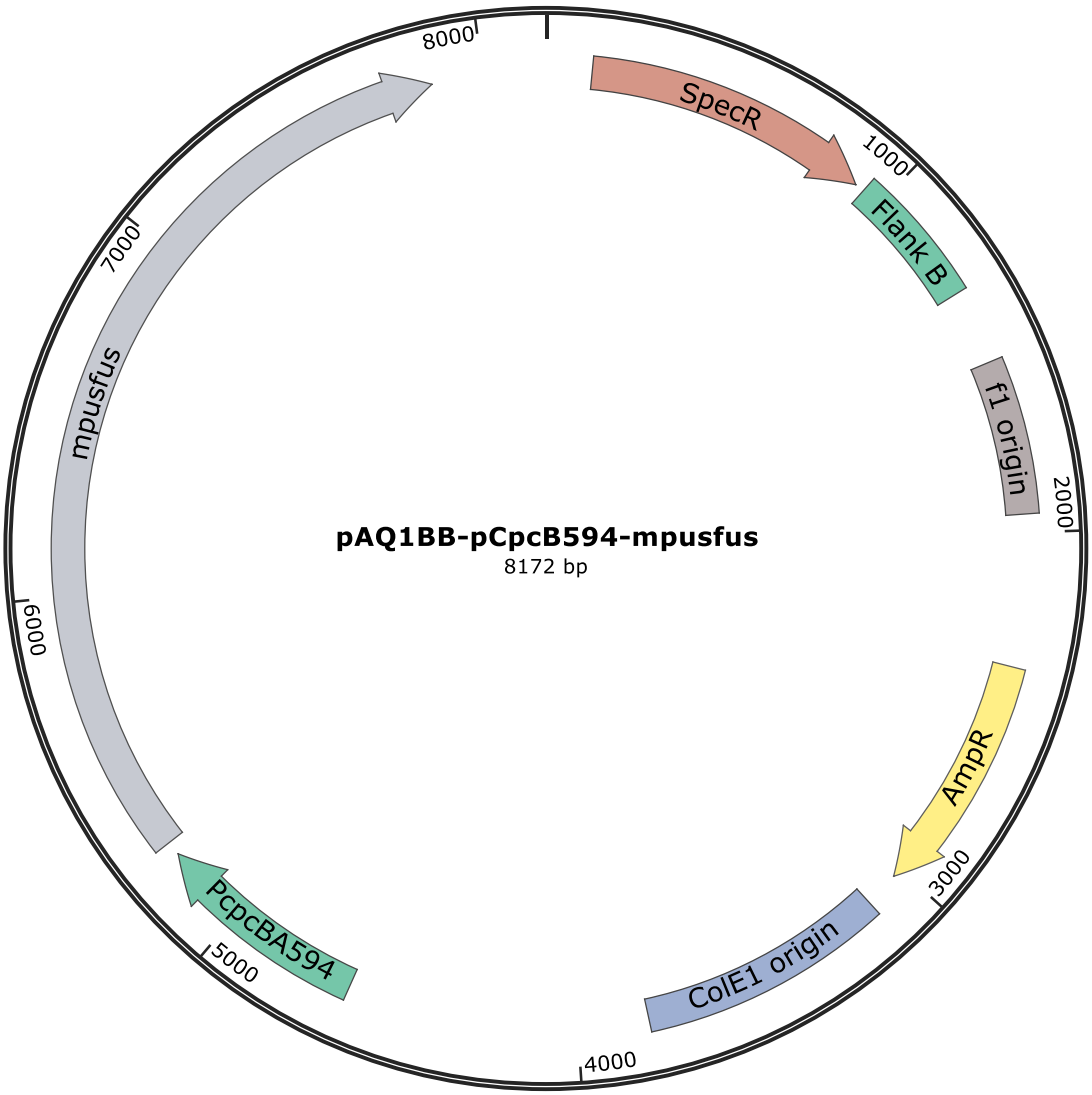

10,000
CcaS
CcaS
PCcaS
rrnB Terminator
2000
CcaR
pCpcG2-35-10
GFP
TE
4000
rrnB Terminator
SpecR
Flank B3
6000
pMB1 Origin
KanR
8000
Flank A3
pUC-172-35-10-GFP
10,074 bp


6000
AmpR
1000
pMB1 Origin
5000
pUCIDT-YF1-GFP
6147 bp
TE
T1
YF1/FixJ CO
2000
GFP
4000
FixK2 promoter
lacI prom region
3000


pUC-172-10-alaDE
11,080 bp
10,000
SpecR
Flank B3
Flank B3
pMB1 Origin
2000
KanR
Flank A3
4000
CcaS
CcaS
6000
PCcaS
CcaR
PCpcG2-172-10
rrnB Terminator
8000
alaD
ECK120029600 Terminator
alaE
pJ23110
TL3S2P21


pUC-cpc-alaDE
7075 bp
7000
1000
2000
3000
4000
5000
6000
SpecR
Flank B3
pMB1 Origin
KanR
Flank A3
pCpct223
alaD
ECK120029600 Terminator
pJ23110
alaE
TL3S2P21


pUC-172-10-mpusfus
11,919 bp
SpecR
Flank B3
pMB1 Origin
KanR
Flank A3
CcaS
PCcaS
rrnB Terminator
CcaR
pCpcG2-172-10
mpusfus
2000
4000
6000
8000
10,000


pAQ1BB-pCpcB594-mpusfus
8172 bp
8000
1000
2000
3000
4000
5000
6000
7000
SpecR
Flank B
f1 origin
AmpR
ColE1 origin
PcpcBA594
mpusfus

Appendix II

'Blue-specific genes' identified as being differentially expressed in blue light by RNAseq.

Blue only DEGs from Venny	Cyanobase Annotation	gene name	Basemean
SYNPCC7002_A0016	glyoxalase family protein superfamily	NA	344.61
SYNPCC7002_A0042	conserved hypothetical protein	NA	1090.83
SYNPCC7002_A0046	two-component response regulator	NA	1933.73
SYNPCC7002_A0107	conserved hypothetical protein	NA	527.90
SYNPCC7002_A0157	photosystem II D1 subunit PsbA-II (Qb protein)	psbA-II	132407.40
SYNPCC7002_A0184	conserved hypothetical protein	NA	1035.95
SYNPCC7002_A0205	hydrogenase assembly chaperone HypC/HupF	hypC	22.44
SYNPCC7002_A0211	conserved hypothetical protein	NA	168.23
SYNPCC7002_A0255	glycosyl transferase, WecB/TagA/CpsF family	NA	388.55
SYNPCC7002_A0261	conserved hypothetical membrane protein	NA	624.34
SYNPCC7002_A0278	thiamine biosynthesis protein ThiS	thiS	28.55
SYNPCC7002_A0286	threonine synthase-II	thrC-II	3860.20
SYNPCC7002_A0295	two-component hybrid sensor and regulator	NA	2209.13
SYNPCC7002_A0416	conserved hypothetical protein	NA	141.92
SYNPCC7002_A0469	conserved hypothetical protein	NA	819.05
SYNPCC7002_A0476	DnaK-type molecular chaperone	dnaK	227.22
SYNPCC7002_A0536	hypothetical protein	NA	43.43
SYNPCC7002_A0586	probable dioxygenase of extradiol dioxygenase family	NA	240.32
SYNPCC7002_A0658	conserved hypothetical membrane protein	NA	72.54
SYNPCC7002_A0664	nitrilase	NA	173.16
SYNPCC7002_A0685	possible ribosomal protein S6 modification protein	rimK	99.71
SYNPCC7002_A0698	hypothetical protein	NA	13.95
SYNPCC7002_A0704	hydrolase, alpha/beta fold family domain protein	NA	75.50

SYNPCC7002_A0762	conserved hypothetical protein	NA	158.57
SYNPCC7002_A0788	ribonuclease E	rne	2627.17
SYNPCC7002_A0793	AhpC/TSA family protein	NA	3618.43
SYNPCC7002_A0822	conserved hypothetical protein	NA	272.33
SYNPCC7002_A0854	NADH dehydrogenase subunit 5	ndhF	3834.52
SYNPCC7002_A0855	SipA; possible regulator of histidine kinase NblS	sipA	103.46
SYNPCC7002_A0865	sensory transduction histidine kinase	gidA	113.82
SYNPCC7002_A0901	conserved hypothetical protein	NA	2027.75
SYNPCC7002_A0920	circadian clock protein KaiB	kaiB	76.69
SYNPCC7002_A1017	conserved hypothetical protein containing Piwi domain	NA	156.23
SYNPCC7002_A1031	preprotein translocase, SecE subunit	secE	433.44
SYNPCC7002_A1070	ammonium transporter, putative	amt	309.80
SYNPCC7002_A1073	conserved hypothetical protein	NA	76.26
SYNPCC7002_A1175	conserved hypothetical protein TIGR00278	NA	287.88
SYNPCC7002_A1205	conserved hypothetical membrane protein	NA	1339.86
SYNPCC7002_A1215	Hypothetical protein	NA	327.06
SYNPCC7002_A1239	ferredoxin Like protein	NA	474.08
SYNPCC7002_A1285	Nitrogen regulatory protein P-II	glnB	3763.15
SYNPCC7002_A1291	flavodoxin	NA	14.21
SYNPCC7002_A1352	fructose-bisphosphate aldolase, class II, Calvin cycle subtype	fba	32118.91
SYNPCC7002_A1360	hypothetical protein	NA	2485.58
SYNPCC7002_A1387	conserved hypothetical membrane protein	NA	760.26
SYNPCC7002_A1399	5S ribosomal RNA	NA	60.43
SYNPCC7002_A1404	SCP Like extracellular protein subfamily, putative	NA	518.47
SYNPCC7002_A1418	photosystem q(b) protein	psbA	304817.10
SYNPCC7002_A1431	conserved hypothetical protein	NA	745.49
SYNPCC7002_A1432	Na+-transporting NADH:ubiquinone oxidoreductase subunit 2	NA	293.08
SYNPCC7002_A1443	pyruvate:ferredoxin (flavodoxin) oxidoreductase	nifJ	4541.91

SYNPCC7002_A1469	Secreted and surface protein containing fasciclin Like repeats	NA	2462.97
SYNPCC7002_A1471	hypothetical protein	NA	260.62
SYNPCC7002_A1540	conserved hypothetical protein	NA	579.11
SYNPCC7002_A1553	conserved hypothetical membrane protein	NA	37.12
SYNPCC7002_A1587	Universal stress protein (Usp)	NA	636.93
SYNPCC7002_A1604	conserved hypothetical protein	NA	1272.44
SYNPCC7002_A1614	thioredoxin	trx	6844.88
SYNPCC7002_A1652	light-independent protochlorophyllide reductase, B subunit	chlB	12037.60
SYNPCC7002_A1715	UDP-glucose 4-epimerase	galE	346.16
SYNPCC7002_A1721	ABC-2 type transporter protein	NA	288.28
SYNPCC7002_A1732	yrdC domain protein	NA	1230.21
SYNPCC7002_A1741	ribosomal protein S21	rpsU	1487.55
SYNPCC7002_A1752	LCCL domain containing protein	NA	1632.08
SYNPCC7002_A1832	sigma factor	rpoD	114.58
SYNPCC7002_A1845	conserved hypothetical protein (DUF370)	NA	1933.93
SYNPCC7002_A1882	HTH containing protein	NA	96.22
SYNPCC7002_A1893	ribosomal protein L32	rpmF	1438.07
SYNPCC7002_A2032	gamma-carotene 1' hydroxylase; myxoxanthophyll biosynthesis protein	cruF	275.66
SYNPCC7002_A2068	glutaredoxin	grxC	22.12
SYNPCC7002_A2086	conserved hypothetical protein	NA	1994.70
SYNPCC7002_A2094	conserved hypothetical protein	NA	193.65
SYNPCC7002_A2102	N-acetyl-gamma-glutamyl-phosphate reductase	argC	2075.57
SYNPCC7002_A2103	UvrD/REP helicase domain protein	NA	755.37
SYNPCC7002_A2183	conserved hypothetical protein	NA	119.07
SYNPCC7002_A2199	photosystem II D2 protein	psbD	43362.43
SYNPCC7002_A2259	hypothetical protein	NA	21.65
SYNPCC7002_A2377	conserved hypothetical protein	NA	14.61
SYNPCC7002_A2396	1-acyl-sn-glycerol-3-phosphate acyltransferase	plsC	85.51
SYNPCC7002_A2448	conserved hypothetical protein	NA	3759.45

SYNPCC7002_A2481	conserved hypothetical membrane protein	NA	184.88
SYNPCC7002_A2507	iron transport protein	sufA	595.79
SYNPCC7002_A2524	tRNA-Thr	NA	11.13
SYNPCC7002_A2526	conserved hypothetical protein	NA	2941.16
SYNPCC7002_A2567	Protein of unknown function (DUF751)	NA	25.53
SYNPCC7002_A2579	hypothetical protein	NA	57756.90
SYNPCC7002_A2600	CheW Like domain containing protein	NA	174.57
SYNPCC7002_A2602	two-component response regulator	NA	868.25
SYNPCC7002_A2606	prohibitin	NA	479.10
SYNPCC7002_A2635	conserved hypothetical protein	NA	2570.02
SYNPCC7002_A2754	ammonium transporter	amt	2029.33
SYNPCC7002_A2833	fatty acid desaturase	desE	997.30
SYNPCC7002_A2837	conserved hypothetical protein	NA	35.55
SYNPCC7002_C0017	site-specific recombinase for integration and excision	NA	10.85
SYNPCC7002_D0005	HupE / UreJ family protein	NA	457.12
SYNPCC7002_D0008	nitrile hydratase beta subunit	nthB	344.00
SYNPCC7002_E0036	conserved hypothetical protein	NA	15.62
SYNPCC7002_G0043	hypothetical protein	NA	12.48
SYNPCC7002_G0046	conserved hypothetical protein	NA	32.40
SYNPCC7002_G0115	ABC transporter, ATP-binding protein	NA	37.26
SYNPCC7002_G0117	conserved hypothetical membrane protein	NA	2037.66
SYNPCC7002_G0141	phosphonate ABC transporter, ATP-binding protein	phnC	20.25
SYNPCC7002_G0145	ATP synthase epsilon subunit	atpC-II	24.82
SYNPCC7002_G0147	conserved hypothetical protein	NA	84.05
SYNPCC7002_G0148	ATP synthase F0, A subunit	atpB-II	312.32
SYNPCC7002_G0149	ATP synthase c subunit	atpH-II	116.37
SYNPCC7002_G0150	ATP synthase b subunit	atpF-II	542.89
SYNPCC7002_G0152	ATP synthase F1, gamma subunit	atpG-II	128.29

Appendix III

Table 9.2. 'Orange-specific genes' identified as being differentially expressed in orange light by RNAseq.

Orange DEGs from Venny	Cyanobase Annotation	gene name	Basemean
SYNPCC7002_A0011	conserved hypothetical protein, DUF820	NA	371.37
SYNPCC7002_A0013	hypothetical protein	NA	2083.24
SYNPCC7002_A0048	methyl-accepting chemotaxis protein	NA	42066.77
SYNPCC7002_A0049	two-component hybrid sensor and regulator	NA	14404.41
SYNPCC7002_A0050	conserved hypothetical protein	NA	1469.24
SYNPCC7002_A0056	conserved hypothetical protein	NA	47.41
SYNPCC7002_A0073	2-phosphopyruvate hydratase (enolase)	eno	591.40
SYNPCC7002_A0078	conserved hypothetical protein	NA	7762.24
SYNPCC7002_A0095	glucose-1-phosphate adenylyltransferase	glgC	4125.20
SYNPCC7002_A0124	conserved hypothetical protein with saccharopine dehydrogenase/reductase domain	NA	189.31
SYNPCC7002_A0127	Protein of unknown function (DUF636) family	NA	71.21
SYNPCC7002_A0130	2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldolase	eda	155.50
SYNPCC7002_A0131	transcriptional regulator, MerR family	NA	539.12
SYNPCC7002_A0132	conserved hypothetical protein	NA	278.21
SYNPCC7002_A0136	6,7-dimethyl-8-ribityllumazine synthase (Riboflavin synthase beta chain)	ribH	372.80
SYNPCC7002_A0158	soluble lytic transglycosylase	slt	357.45
SYNPCC7002_A0171	transcription regulator of rubisco operon, RbcR (LysR family)	rbcR	227.60
SYNPCC7002_A0172	NADH2 dehydrogenase (plastoquinone) chain 5	ndhF-III	383.40

SYNPCC7002_A0173	proton-translocating NADH-quinone oxidoreductase (NADH dehydrogenase subunit 4)	ndhD-III	292.37
SYNPCC7002_A0187	conserved hypothetical protein	NA	3617.83
SYNPCC7002_A0213	carboxyl-terminal protease (periplasmic)	prc	10611.18
SYNPCC7002_A0224	Peptidase family M48	NA	2069.43
SYNPCC7002_A0225	TPR-repeat protein	NA	428.53
SYNPCC7002_A0263	light-repressed protein LrtA (ribosomal subunit interface protein)	lrtA	43849.33
SYNPCC7002_A0267	conserved hypothetical protein	NA	3370.33
SYNPCC7002_A0282	Glutaredoxin, GrxC family	grxC	420.14
SYNPCC7002_A0307	dihydroneopterin aldolase	folB	115.37
SYNPCC7002_A0312	conserved hypothetical protein	NA	1307.34
SYNPCC7002_A0344	possible rare lipoprotein A	rlpA	735.52
SYNPCC7002_A0349	ATP-dependent metalloprotease, FtsH family	NA	11642.89
SYNPCC7002_A0359	conserved hypothetical protein	NA	375.55
SYNPCC7002_A0366	nicotinate-nucleotide pyrophosphorylase	nadC	2053.19
SYNPCC7002_A0375	cytochrome cM	cytM	61.12
SYNPCC7002_A0376	conserved hypothetical protein	NA	49.17
SYNPCC7002_A0381	conserved hypothetical protein	NA	1081.17
SYNPCC7002_A0392	conserved hypothetical protein	NA	171.40
SYNPCC7002_A0403	glycosyl transferase, group 1	NA	1289.73
SYNPCC7002_A0422	aspartyl-tRNA synthetase	aspS	282.88
SYNPCC7002_A0425	conserved hypothetical protein, TPR domain	NA	398.49
SYNPCC7002_A0433	Chromosome partitioning protein, ParB family	NA	322.09
SYNPCC7002_A0434	Peptidase family M3	prlC	2463.53
SYNPCC7002_A0449	Protein of unknown function (DUF820) family	NA	819.62
SYNPCC7002_A0456	3-isopropylmalate dehydratase, small subunit	leuD	258.44
SYNPCC7002_A0457	6-pyruvoyl tetrahydropterin synthase	NA	46.30

SYNPCC7002_A0470	sodium-dependent bicarbonate transporter	NA	392.18
SYNPCC7002_A0487	conserved hypothetical protein	NA	11308.07
SYNPCC7002_A0490	ribosomal protein L21	rplU	660.28
SYNPCC7002_A0503	conserved hypothetical protein	NA	1330.11
SYNPCC7002_A0505	conserved hypothetical protein	NA	353.44
SYNPCC7002_A0518	Protein of unknown function (DUF564) family (Ycf21)	NA	199.91
SYNPCC7002_A0530	conserved hypothetical protein	NA	835.98
SYNPCC7002_A0534	transporter, major facilitator superfamily	NA	89.77
SYNPCC7002_A0550	amidophosphoribosyltransferase	purF	6120.32
SYNPCC7002_A0580	Solanesyl diphosphate synthase	sdsA	1187.12
SYNPCC7002_A0590	arsenical resistance operon repressor, ArsR family	NA	23.84
SYNPCC7002_A0591	efflux transporter, RND family, MFP subunit	NA	193.32
SYNPCC7002_A0598	conserved hypothetical protein	NA	71.54
SYNPCC7002_A0618	ATP-dependent RNA helicase DeaD	deaD	10291.80
SYNPCC7002_A0644	conserved hypothetical protein TIGR00701	NA	859.14
SYNPCC7002_A0679	conserved hypothetical protein	NA	241.21
SYNPCC7002_A0682	photosystem I subunit II	psaD	12987.33
SYNPCC7002_A0691	probable ribosome-associated GTPase A	NA	299.48
SYNPCC7002_A0692	conserved hypothetical protein	NA	68.61
SYNPCC7002_A0693	chaperone protein DnaJ	dnaJ	944.65
SYNPCC7002_A0707	magnesium-protoporphyrin IX monomethyl ester aerobic oxidative cyclase (oxygen-dependent)	acsF	7317.07
SYNPCC7002_A0724	dehydrogenase subunit Like protein	betA	294.05
SYNPCC7002_A0757	DNA repair protein RecO	recO	362.02
SYNPCC7002_A0772	conserved hypothetical membrane protein	NA	37.22
SYNPCC7002_A0784	tRNA-Pro	NA	30.15
SYNPCC7002_A0806	conserved hypothetical protein	NA	927.35
SYNPCC7002_A0809	Photosystem II reaction center N protein.-related protein	NA	278.83

SYNPCC7002_A0823	uroporphyrinogen decarboxylase	hemE	1148.28
SYNPCC7002_A0831	thiamine biosynthesis protein ThiC	thiC	4069.06
SYNPCC7002_A0843	carboxyl-terminal protease	ctpA	678.34
SYNPCC7002_A0853	ferredoxin-NADP reductase	petH	2993.40
SYNPCC7002_A0882	adenylosuccinate synthetase (IMP--aspartate ligase)	purA	1553.04
SYNPCC7002_A0894	radical SAM domain protein; Fe-S oxidoreductase family 2	NA	795.41
SYNPCC7002_A0900	two-component hybrid sensory kinase; GAF domain	NA	723.72
SYNPCC7002_A0924	NADH dehydrogenase subunit G	ndhG	1240.30
SYNPCC7002_A0925	NADH-plastoquinone oxidoreductase, I subunit	ndhI	1594.62
SYNPCC7002_A0944	putative GTPase	NA	827.59
SYNPCC7002_A0955	30S ribosomal protein S1	rpsA	4697.16
SYNPCC7002_A0957	Photosystem II reaction center, PsbT protein	psbT	6258.73
SYNPCC7002_A0960	glutathione-dependent formaldehyde-activating enzyme	gfa	159.83
SYNPCC7002_A0966	conserved hypothetical protein (DUF29)	NA	145.68
SYNPCC7002_A0985	nicotinamide nucleotide transhydrogenase, chain alpha, part 2	pntC	358.38
SYNPCC7002_A0991	inner membrane protein YidC/OxaA	yidC	3692.39
SYNPCC7002_A0992	Predicted RNA-binding protein	NA	838.25
SYNPCC7002_A1000	magnesium-chelatase, subunit H	chlH	1276.02
SYNPCC7002_A1015	FHA domain protein	NA	1305.00
SYNPCC7002_A1023	3-oxoacyl-[acyl-carrier-protein] synthase II	fabF	1388.96
SYNPCC7002_A1029	ribosomal protein L11	rplK	869.85
SYNPCC7002_A1034	conserved hypothetical protein	NA	4112.54
SYNPCC7002_A1043	ribosomal protein S11	rpsK	2879.51
SYNPCC7002_A1063	ribosomal protein L23	rplW	1026.07
SYNPCC7002_A1065	ribosomal protein L3	rplC	3512.05
SYNPCC7002_A1075	conserved hypothetical protein	NA	403.61
SYNPCC7002_A1097	cytochrome b6/f complex, alternative iron-sulfur subunit	NA	100.28
SYNPCC7002_A1142	sodium-coupled permease	NA	761.01

SYNPCC7002_A1174	hydrolase, alpha/beta fold family domain protein	NA	478.50
SYNPCC7002_A1176	cell division protein	ftsH	2945.54
SYNPCC7002_A1179	lipoprotein, putative	NA	4025.19
SYNPCC7002_A1202	sigma-70 factor	sigB	5351.26
SYNPCC7002_A1203	aspartate aminotransferase	NA	2156.87
SYNPCC7002_A1208	tRNA-Met	NA	94.57
SYNPCC7002_A1213	conserved hypothetical protein	NA	11765.06
SYNPCC7002_A1228	inorganic pyrophosphatase	ppa	2039.37
SYNPCC7002_A1240	conserved hypothetical protein	NA	152.99
SYNPCC7002_A1246	D-3-phosphoglycerate dehydrogenase	serA	831.92
SYNPCC7002_A1247	ribosomal protein L11 methyltransferase	prmA	277.56
SYNPCC7002_A1261	two-component hybrid sensor and regulator	NA	17171.88
SYNPCC7002_A1278	ketol-acid reductoisomerase	ilvC	628.83
SYNPCC7002_A1280	DNA-directed DNA polymerase I	polA	879.43
SYNPCC7002_A1302	glutamyl-tRNA reductase	hemA	384.88
SYNPCC7002_A1316	thymidylate kinase	tmk	57.13
SYNPCC7002_A1327	Hypothetical protein	NA	273.67
SYNPCC7002_A1332	soluble hydrogenase 42 kD subunit DHSS	NA	1048.35
SYNPCC7002_A1333	conserved hypothetical protein	NA	1281.62
SYNPCC7002_A1335	conserved hypothetical protein	NA	17.81
SYNPCC7002_A1344	sulfate adenylyltransferase	sat	368.04
SYNPCC7002_A1353	3-isopropylmalate dehydratase, large subunit	leuC	2409.88
SYNPCC7002_A1357	Protein of unknown function (PF01936)	NA	37.15
SYNPCC7002_A1369	conserved hypothetical protein	NA	146.12
SYNPCC7002_A1395	16S ribosomal RNA	NA	149841.70
SYNPCC7002_A1398	23S ribosomal RNA	NA	7276946.00
SYNPCC7002_A1425	lignostilbene-alpha,beta-dioxygenase	NA	254.93
SYNPCC7002_A1436	ABC-type branched-chain amino acid transport system, ATPase component	NA	86.05

SYNPCC7002_A1460	transaldolase	tal	6558.51
SYNPCC7002_A1474	aldo/keto reductase family	NA	461.03
SYNPCC7002_A1477	conserved hypothetical protein	NA	104.28
SYNPCC7002_A1481	lipoprotein release ABC-type transport system permease component	NA	996.76
SYNPCC7002_A1482	ABC transporter, ATP-binding protein	NA	80.48
SYNPCC7002_A1487	conserved hypothetical protein	NA	254.92
SYNPCC7002_A1488	hypothetical protein	NA	870.47
SYNPCC7002_A1490	hypothetical protein	NA	15525.12
SYNPCC7002_A1494	adenylylsulfate kinase	cysC	146.48
SYNPCC7002_A1505	ABC-type multidrug/protein/lipid transport system, ATPase component	NA	1055.85
SYNPCC7002_A1514	putative glycosyltransferase, group 1 family protein	NA	674.49
SYNPCC7002_A1515	conserved hypothetical membrane protein	NA	1565.26
SYNPCC7002_A1516	glycosyl transferase, group 1 family protein	NA	994.70
SYNPCC7002_A1518	conserved hypothetical protein	NA	141.74
SYNPCC7002_A1525	cell shape determining protein, MreB/Mrl family subfamily	NA	554.58
SYNPCC7002_A1528	haloacid dehalogenase Like hydrolase domain protein	NA	1181.66
SYNPCC7002_A1531	2-succinyl-6-hydroxy-2, 4-cyclohexadiene-1-carboxylic acid synthase/2-oxoglutarate decarboxylase	menD	346.19
SYNPCC7002_A1532	glycogen synthase 1	glgA1	626.82
SYNPCC7002_A1539	conserved hypothetical protein	NA	380.64
SYNPCC7002_A1551	conserved hypothetical membrane protein	NA	12.43
SYNPCC7002_A1582	conserved hypothetical protein	NA	2766.86
SYNPCC7002_A1591	conserved hypothetical protein	NA	5971.99
SYNPCC7002_A1602	conserved hypothetical protein	NA	2098.30
SYNPCC7002_A1603	conserved hypothetical protein	NA	5930.72
SYNPCC7002_A1636	imidazoleglycerol phosphate synthase, cyclase subunit	hisF	11.03

SYNPCC7002_A1649	ferric uptake regulator	fur	448.73
SYNPCC7002_A1668	conserved hypothetical protein	NA	28.67
SYNPCC7002_A1675	global nitrogen regulator	NA	1978.31
SYNPCC7002_A1679	conserved hypothetical protein (DUF820)	NA	444.30
SYNPCC7002_A1692	Pyrimidine operon attenuation protein/uracil phosphoribosyltransferase	pyrR	32.12
SYNPCC7002_A1693	conserved hypothetical protein	NA	17835.55
SYNPCC7002_A1695	conserved hypothetical protein (DUF1232)	NA	141.80
SYNPCC7002_A1699	TPR-repeat containing protein	NA	547.34
SYNPCC7002_A1714	S-adenosylmethionine synthetase	metK	1865.94
SYNPCC7002_A1740	hypothetical protein	NA	2168.25
SYNPCC7002_A1790	conserved hypothetical membrane protein	NA	131.86
SYNPCC7002_A1805	NADH dehydrogenase subunit F4	ndhF4	2757.73
SYNPCC7002_A1806	NADH dehydrogenase subunit D4	ndhD 4	1441.97
SYNPCC7002_A1807	CO_2 hydration protein	NA	1233.91
SYNPCC7002_A1821	putative phycobilisome degradation protein	nblA	589.54
SYNPCC7002_A1828	coproporphyrinogen III oxidase, aerobic	hemF	1506.93
SYNPCC7002_A1836	transcriptional regulator, Fur family	NA	1081.47
SYNPCC7002_A1837	branched-chain amino acid ABC transporter periplasmic amino acid-binding protein	NA	514.44
SYNPCC7002_A1849	LexA repressor	lexA	1151.49
SYNPCC7002_A1872	conserved hypothetical protein	NA	11.96
SYNPCC7002_A1881	hypothetical protein	NA	16.85
SYNPCC7002_A1884	hypothetical protein	NA	197.01
SYNPCC7002_A1902	heat shock protein	htpG	767.69
SYNPCC7002_A1910	apocytochrome f precursor	petA	4044.77
SYNPCC7002_A1918	Protein of unknown function (DUF552)	NA	12494.34
SYNPCC7002_A1933	NblA-related protein	NA	575.22
SYNPCC7002_A1939	hypothetical protein	NA	140.50

SYNPCC7002_A1949	Ubiquinone biosynthesis hydroxylase, UbiH/UbiF/VisC/COQ6 family	ubiH	738.93
SYNPCC7002_A1958	translation elongation factor	NA	1375.01
SYNPCC7002_A1965	DNA-binding protein HU	NA	8823.14
SYNPCC7002_A1976	glycosyl transferase, group 1 family protein	NA	60.89
SYNPCC7002_A1977	AMP-binding enzyme	NA	610.16
SYNPCC7002_A1978	oxidoreductase, short chain dehydrogenase/reductase family	NA	124.74
SYNPCC7002_A1985	imidazoleglycerol-phosphate dehydratase	hisB	146.55
SYNPCC7002_A1986	conserved hypothetical protein	NA	325.70
SYNPCC7002_A1991	Heme oxygenase 2	ho2	92.55
SYNPCC7002_A1992	putative magnesium-protoporphyrin IX monomethyl ester aerobic oxidative cyclase	acsF	495.44
SYNPCC7002_A2000	NADH dehydrogenase subunit D1	ndhD 1	4620.54
SYNPCC7002_A2001	hypothetical protein	NA	157.98
SYNPCC7002_A2022	putative sucrose phosphorylase	NA	235.84
SYNPCC7002_A2027	conserved hypothetical protein	NA	2523.55
SYNPCC7002_A2028	Cell envelope-related transcriptional attenuator domain containing protein	NA	2826.89
SYNPCC7002_A2036	ABC transporter, ATPase subunit	NA	225.92
SYNPCC7002_A2045	DNA-directed RNA polymerase, beta subunit	rpoB	9540.44
SYNPCC7002_A2051	conserved hypothetical protein	NA	87.42
SYNPCC7002_A2063	ribosomal protein S7	rpsG	1418.20
SYNPCC7002_A2064	30S ribosomal protein S12	rpsL	5377.55
SYNPCC7002_A2095	conserved hypothetical protein (DUF1001)	NA	595.80
SYNPCC7002_A2107	transposase	NA	1852.90
SYNPCC7002_A2117	conserved hypothetical protein	NA	279.37
SYNPCC7002_A2118	cellulose synthase catalytic subunit	NA	471.55
SYNPCC7002_A2119	STAS domain protein	NA	120.45
SYNPCC7002_A2122	transposase	NA	6280.17
SYNPCC7002_A2123	urease accessory protein, UreD	ureD	219.36

SYNPCC7002_A2128	hypothetical protein	NA	90.66
SYNPCC7002_A2139	alpha-glucan phosphorylase	NA	3403.91
SYNPCC7002_A2141	transposase	NA	6353.85
SYNPCC7002_A2151	photosystem II reaction centre M protein	psbM	2869.18
SYNPCC7002_A2152	putative ferredoxin (2Fe-2S)	NA	230.11
SYNPCC7002_A2163	putative small-conductance mechanosensitive channel	NA	586.78
SYNPCC7002_A2170	hypothetical protein	NA	207.46
SYNPCC7002_A2171	conserved hypothetical protein	NA	13265.79
SYNPCC7002_A2175	conserved repeat domain protein	NA	6032.26
SYNPCC7002_A2177	hypothetical protein	NA	1753.00
SYNPCC7002_A2182	ABC-2 type transporter superfamily	NA	458.99
SYNPCC7002_A2190	putative zinc-binding oxidoreductase	NA	293.85
SYNPCC7002_A2201	ABC transporter amino acid-binding protein	NA	1235.58
SYNPCC7002_A2202	secreted pentapeptide repeats protein	NA	425.48
SYNPCC7002_A2214	phycocyanin alpha subunit phycocyanobilin lyase, CpcF	cpcF	64.38
SYNPCC7002_A2216	transposase	NA	300.31
SYNPCC7002_A2234	preprotein translocase, SecG subunit	secG	206.04
SYNPCC7002_A2242	ABC-2 type transporter superfamily protein	NA	56.33
SYNPCC7002_A2261	conserved hypothetical protein	NA	1255.96
SYNPCC7002_A2271	hybrid sensory kinase	NA	1489.15
SYNPCC7002_A2280	hydrolase, alpha/beta fold family	NA	215.71
SYNPCC7002_A2286	phosphate ABC transporter, permease protein	pstA	197.26
SYNPCC7002_A2295	conserved hypothetical protein (DUF805)	NA	717.73
SYNPCC7002_A2300	penicillin-binding protein	NA	2131.15
SYNPCC7002_A2311	type II NADH dehydrogenase A	ndbA	1581.99
SYNPCC7002_A2314	DNA modification methyltransferase	NA	250.18
SYNPCC7002_A2319	Type 3 multicopper oxidase	NA	517.90
SYNPCC7002_A2320	conserved hypothetical protein	NA	271.97

SYNPCC7002_A2321	Integral membrane protein superfamily	NA	82.44
SYNPCC7002_A2345	light-independent protochlorophyllide reductase, N subunit	chlN	3278.71
SYNPCC7002_A2346	conserved hypothetical protein	NA	3689.38
SYNPCC7002_A2347	light-independent protochlorophyllide reductase, iron-sulfur ATP-binding protein	chlL	8144.48
SYNPCC7002_A2348	peptide deformylase	def	4270.59
SYNPCC7002_A2355	FAD dependent oxidoreductase, putative	NA	570.69
SYNPCC7002_A2367	3-oxoacyl-(acyl-carrier-protein) reductase	fabG	313.41
SYNPCC7002_A2371	bicarbonate transporter, BicA	bicA	434.50
SYNPCC7002_A2376	conserved hypothetical protein	NA	26.97
SYNPCC7002_A2384	DNA polymerase III alpha subunit	dnaE	1315.13
SYNPCC7002_A2386	hypothetical protein	NA	496.41
SYNPCC7002_A2387	conserved hypothetical protein	NA	5470.93
SYNPCC7002_A2395	conserved hypothetical protein	NA	428.26
SYNPCC7002_A2400	conserved hypothetical protein	NA	119.36
SYNPCC7002_A2403	conserved hypothetical protein	NA	152.73
SYNPCC7002_A2425	polyphosphate kinase	ppk	671.85
SYNPCC7002_A2427	permease, YjgP/YjgQ family protein	NA	195.28
SYNPCC7002_A2457	chaperonin, 10 kDa protein	groES	112.02
SYNPCC7002_A2458	chaperonin, 60 kDa protein	groEL	1680.99
SYNPCC7002_A2474	glycosyl transferase, group 1 family protein	rfaG	320.32
SYNPCC7002_A2509	S layer domain protein	NA	278.27
SYNPCC7002_A2531	conserved hypothetical protein	NA	14360.43
SYNPCC7002_A2532	conserved hypothetical protein	NA	88.87
SYNPCC7002_A2538	signal peptide peptidase SppA, 36K type	sppA	179.27
SYNPCC7002_A2548	ferredoxin PetF2	petF2	124.36
SYNPCC7002_A2570	4Fe-4S binding domain protein	NA	1406.35
SYNPCC7002_A2571	ATPase like protein	NA	2328.04
SYNPCC7002_A2588	hypothetical protein	NA	20.23

SYNPCC7002_A2590	short-chain alcohol dehydrogenase family	NA	194.05
SYNPCC7002_A2597	conserved hypothetical protein	NA	3003.62
SYNPCC7002_A2670	hydrolase, alpha/beta fold family	NA	2470.07
SYNPCC7002_A2689	Predicted ATPase of ABC class	NA	380.47
SYNPCC7002_A2690	universal stress protein	NA	9565.57
SYNPCC7002_A2696	erthyrocyte band 7 integral membrane protein	NA	3231.11
SYNPCC7002_A2709	tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase	trmU	166.38
SYNPCC7002_A2716	hypothetical protein	NA	1794.16
SYNPCC7002_A2731	hypothetical protein	NA	10.10
SYNPCC7002_A2732	conserved hypothetical protein	NA	2815.31
SYNPCC7002_A2738	mechanosensitive ion channel family protein	NA	523.32
SYNPCC7002_A2753	conserved hypothetical protein	NA	395.79
SYNPCC7002_A2760	conserved hypothetical protein	NA	144.29
SYNPCC7002_A2772	conserved hypothetical protein	NA	343.18
SYNPCC7002_A2785	23S ribosomal RNA	NA	6638172.00
SYNPCC7002_A2788	16S ribosomal RNA	NA	151768.80
SYNPCC7002_A2802	conserved hypothetical protein	NA	244.93
SYNPCC7002_A2810	carotenoid binding protein	NA	3900.61
SYNPCC7002_A2839	ATP-binding protein of ABC transporter	NA	171.06
SYNPCC7002_A2841	glucosylglycerol 3-phosphatase	stpA	388.67
SYNPCC7002_A2848	sphingomyelinase-c (Sphingomyelin phosphodiesterase)	sph	339.90
SYNPCC7002_A2850	hypothetical protein	NA	5002.93
SYNPCC7002_A2863	sensory transduction system regulatory protein (diguanylate cyclase (GGDEF) domain)	NA	515.45
SYNPCC7002_A2864	transposase	NA	6300.48
SYNPCC7002_A2869	sensory box/GGDEF family protein	NA	5173.79
SYNPCC7002_C0004	type 1 restriction-modification system specificity subunit	NA	106.62
SYNPCC7002_C0008	conserved hypothetical protein	NA	3463.77
SYNPCC7002_D0010	CobW/P47K family protein	NA	669.01

SYNPCC7002_D0011	hypothetical protein	NA	142.98
SYNPCC7002_E0004	hypothetical protein	NA	21.46
SYNPCC7002_E0038	DNA methylase	NA	1677.78
SYNPCC7002_E0039	Conserved hypothetical protein, with zinc-binding domain	NA	638.64
SYNPCC7002_F0007	conserved hypothetical protein	NA	591.77
SYNPCC7002_F0008	ATP-dependent nuclease subunit A	addA	487.85
SYNPCC7002_F0011	hypothetical protein	NA	11.00
SYNPCC7002_F0021	conserved hypothetical protein (DUF559)	NA	94.06
SYNPCC7002_F0028	conserved hypothetical protein, PIN domain	NA	81.11
SYNPCC7002_F0065	conserved hypothetical protein; TPR-repeat domain	NA	1599.30
SYNPCC7002_F0083	conserved hypothetical protein	NA	21.34
SYNPCC7002_F0084	conserved hypothetical protein, nucleotidyltransferase domain	NA	67.71
SYNPCC7002_F0088	conserved hypothetical protein	NA	53.14
SYNPCC7002_F0102	PIN domain protein	NA	404.40
SYNPCC7002_G0010	hypothetical protein	NA	200.43
SYNPCC7002_G0026	conserved hypothetical protein	NA	136.49
SYNPCC7002_G0056	hypothetical protein	NA	14.73
SYNPCC7002_G0070	HlyD family secreted protein	NA	712.07
SYNPCC7002_G0077	transposase	NA	6167.80
SYNPCC7002_G0088	FecCD transport (permease) family	fecC	110.40
SYNPCC7002_G0107	ABC transporter, transmembrane and ATP-binding domains	NA	609.08
SYNPCC7002_G0108	hypothetical protein	NA	467.83
SYNPCC7002_G0110	hypothetical protein	NA	932.66
SYNPCC7002_G0113	Bacterial extracellular solute-binding lipoprotein	NA	147.72
SYNPCC7002_G0130	hypothetical protein	NA	4309.36
SYNPCC7002_G0161	conserved hypothetical protein	NA	26.98