

DISSERTATION

DEVELOPMENT OF GENETIC PARTS FOR IMPROVED CONTROL OF TRANSLATION INITIATION IN
SYNECHOCYSTIS SP. PCC 6803 WITH AN APPLICATION IN BIOFUEL PRODUCTION

Submitted by

Jacob Sebesta

Department of Chemical and Biological Engineering

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2021

Doctoral Committee:

Advisor: Christie A.M. Peebles

Graham Peers

Ashok Prasad

Kenneth Reardon

Copyright by Jacob Sebesta 2021

All Rights Reserved

ABSTRACT

DEVELOPMENT OF GENETIC PARTS FOR IMPROVED CONTROL OF TRANSLATION INITIATION IN *SYNECHOCYSTIS SP.* PCC 6803 WITH AN APPLICATION IN BIOFUEL PRODUCTION

Metabolic engineering is developing into a field that can change the way we produce a wide variety of valuable chemicals. Many chemicals are already produced in microbial cultures. Metabolic engineering enables us to modify organisms to produce metabolites they don't usually produce, assuming an enzyme can be identified in another organism that catalyzes the formation of that product (or an enzyme can be designed for that task through protein engineering). The distribution of accumulated metabolites can also be altered. There are some cases where metabolites can be accumulated through cultivation practices. Methods of metabolic engineering to overexpress, knockdown, or knockout native enzymes provide additional tools to alter cellular metabolism and drive accumulation of those products. Precise control over gene expression is central to these efforts.

To avoid competition with human food crops and the resources need to produce them, cyanobacteria may be utilized for production of valuable chemicals. Through photosynthesis, they can utilize carbon dioxide from geological formations or from industrial waste streams. Since most metabolic engineering has been developed in *E. coli* and yeast, it was necessary to first adapt the basic methods for use in cyanobacteria. Along with my co-authors Dr. Allison Werner and Dr. Christie Peebles, we reviewed methods for producing genetically modified *Synechocystis Sp.* PCC6803 (*S.* 6803). To facilitate the generation of strains with many modifications, we covered the method developed in the Peebles Lab for making markerless selections which remove any antibiotic selection markers.

A previous graduate student in the Peebles lab, Stevan Albers, found that strong promoter-ribosome binding site combinations that drove high expression of GFP did not necessarily result in high expression when used to drive expression of a different gene. Therefore, in our work to produce bisabolene in *S. 6803* we tested many ribosome binding sites. In addition, we tested five different codon optimizations of the bisabolene synthase to ensure that expression was not prevented by slow translation elongation. We found that the simple measure of the codon adaptation index (CAI) correlated with expression of the five different codon optimizations. Using a thermodynamic model of translation initiation, we designed ten ribosome binding sites to increase bisabolene synthase expression by 10-fold. Only one of those designs actually approached a 10-fold increase, highlighting the need to continue testing several ribosome binding sites to achieve a desired expression level. Since industrial cultivation of cyanobacteria occurs outdoors, subject to natural light:dark cycles, we tested two of the designed strains in light:dark cycles. The strains reached similar bisabolene titers after being exposed to the same amount of total light period as those previously tested in continuous light. Overall, this work increased the highest bisabolene titer reported in cyanobacteria by approximately 10-fold.

The need to test many ribosome binding sites limits progress in cyanobacterial metabolic engineering. The research of others suggest that ribosome binding sites interact with coding sequences by forming secondary structures with different free energy of folding. The estimation of the free energy of folding may be inaccurate, and, further, the kinetics of such folding may also be important to translation initiation rates. We tested two different designs to limit the impacts that secondary structures that span either side of the start codon may have on translation initiation rates in both *E. coli* and *S. 6803*. Utilization of a 21-nucleotide leader sequence after the start codon to make the sequence context consistent for ribosome binding sites between different coding sequences did not improve the correlation found between the expression of two different reporter genes in either organism. Bicistronic designs use translational coupling between an upstream open reading frame and the gene of interest

with a ribosome binding site contained within the upstream open reading frame to re-initiate translation. This design exploits the helicase activity of ribosomes in elongation mode to actively unfold the secondary structure around the start codon of the gene of interest. We expected this activity to reduce the impacts of secondary structure and improve the correlation in expression between two different reporter genes. Intriguingly, the correlation was much improved in *E. coli*, but not in *S. 6803*.

Together, this dissertation suggests that there are important differences in translation initiation between *E. coli* and *S. 6803*. Improved ribosome binding site design for cyanobacteria would facilitate further increases in terpenoid production both by enabling higher expression of heterologous terpenoid synthases and by reducing the number of strains that must be tested to achieve the desired expression level for each enzyme. Future directions suggested by this work include studies of translation initiation mechanisms in cyanobacteria, development of cell-free expression systems to facilitate rapid testing of many different genetic constructs, and further efforts at pathway engineering to increase terpenoid titer and productivity in cyanobacteria.

ACKNOWLEDGEMENTS

First and foremost, I thank my advisor, Christie Peebles. Her support and patience helped me through the many challenges I faced during my studies. I gradually found confidence in the lab and in communicating my findings with her guidance. I am grateful to her for sending me to many conferences to practice presentations, learn from others, and have fun. Thank you to my committee for helping to improve my ideas and planning of this work and thank you to Dr. Reardon for mentoring me in teaching the labs for CBE101.

I thank Allison Werner for asking great questions, helping me troubleshoot experiments, and building a great photobioreactor that I was lucky to get to use. Yi Ern Cheah provided some initial training on the old GC-FID on his way out. Stevan Albers showed me how to grow *Synechocystis Sp. PCC6803* (S. 6803). Michael Cantrell helped me troubleshoot the Western blots. The work to produce bisabolene in S. 6803 (Chapter 2) was supported by a grant from the National Science Foundation (Award #1332404). The DNA2.0 optimized bisabolene synthase gene was provided as a gift by Dr. Fiona Davies. Dr. Sei Jin Park completed LC-MS/MS analysis of 2.0-10b proteins. Dr. Park also provided helpful suggestions and support for GC-MS troubleshooting and maintenance. Stevan Albers and Yi Ern Cheah contributed to Appendix 1 by developing the protocols for transforming *Synechocystis* used in our lab and teaching Allison and I these methods.

For the work summarized in Chapter 4, I am grateful to Dr. Pauli Kallio for providing the 26 initial plasmids in this study and discussing the work published in Thiel et al. (2018). Jaclyn Adkins and OptiEnz Sensors provided space for growing and storing S. 6803 transformations while the university was shut down due to the SARS-CoV-2 pandemic. This allowed us to restart experiments quickly when the university labs re-opened.

I thank my family for their support, especially my mom, dad, my sister, Darcie, and my stepmom, Alison.

I am grateful to my girlfriend, Caitlin, for her support, constructive feedback, and all the fun we've had together during this time.

Lastly, I must thank the CSU Counseling Center, where the talented counselors helped me to address some of the personal difficulties I faced while a student at CSU.

TABLE OF CONTENTS

ABSTRACT..... ii

ACKNOWLEDGEMENTS..... v

CHAPTER 1: INTRODUCTION..... 1

 Climate change..... 2

 Metabolic engineering importance 2

 Resource requirements for *E. coli* vs cyanobacteria..... 3

 Metabolic engineering in cyanobacteria 4

 The strains..... 4

 Transformation methods 5

 Metabolic engineering methods..... 7

 Metabolic engineering for terpenoid production in cyanobacteria 8

 Cyanobacteria toolbox..... 14

 Previous work in the Peebles Lab 27

 Significance of this work 27

 References 31

CHAPTER 2: IMPROVING EXPRESSION OF A HETEROLOGOUS PROTEIN FOR BIOFUEL PRODUCTION IN
SYNECHOCYSTIS SP. PCC6803 39

 Summary 39

 Background 40

Methods.....	42
Strains and cultivation	42
Bisabolene measurement (GC-MS).....	44
Protein abundance (Western blot)	44
Results.....	46
Codon Optimization	48
Ribosome binding site design	50
Impact on light:dark cycle cultivation on bisabolene titer	54
Bisabolene production in a photobioreactor.....	55
Discussion.....	56
References	62
CHAPTER 3: REVIEW OF TRANSLATION INITIATION IN CYANOBACTERIA.....	65
Motivation.....	65
Challenges in controlling translation initiation for metabolic engineering of cyanobacteria	65
Importance of translation initiation.....	67
The canonical model of bacterial translation initiation.....	68
Overview of the canonical prokaryotic translation initiation	68
Interactions between the ribosome and mRNA	69
The Unified View – Secondary structure and SD-antiSD	70
What is the role of the SD-antiSD interaction?.....	71

Standby-site binding and mRNA unfolding	72
What factors determine the expression level of proteins?	75
Testing of large-scale randomized RBS libraries	76
Alternative mechanisms in bacteria	77
S1-mediated translation initiation	78
70S scanning and leaderless initiation	81
Transcription-Translation Coupling.....	82
Polysomes	83
Operon Re-initiation	83
Translation initiation in cyanobacteria	84
Evidence that things are different in cyanobacteria.....	84
Are there any hints from chloroplasts or cyanophages?	87
Regulation of translation in cyanobacteria	87
Importance for metabolic engineering & synthetic biology.....	88
Conclusions and Future Directions	89
References	91
CHAPTER 4: CONTEXT DEPENDENCE OF RIBOSOME BINDING SITES IN <i>SYNECHOCYSTIS SP.</i> PCC6803 AND <i>E. COLI</i>	97
Summary	97
Introduction	98
Motivation.....	98

Strategies to overcome context dependence of ribosome binding sites	99
Methods.....	100
Strains and cultivation	100
Absorbance and Fluorescence measurements.....	102
Table of sequences	103
Results.....	104
Comparison to Thiel et al. 2018 data.....	106
A consistent N-terminal leader sequence failed to make expression profiles more similar for GFP and YFP.....	107
Bicistronic designs improved the GFP/YFP correlation in <i>E. coli</i> , but not in <i>S. 6803</i>	112
Comparison between species	113
Discussion.....	114
RBSs are not totally modular because their strength depends on the coding sequence.....	114
Bicistronic design translation initiation is modular in <i>E. coli</i> , but not in <i>S. 6803</i>	116
Ribosome binding sites are not modular across species	117
Conclusions	118
Future work.....	120
References	121
CHAPTER 5: DISCUSSION.....	124
Conclusions	124
Future directions.....	126

Directed evolution and high-throughput screening	126
Metabolic valve.....	127
In vitro part characterization	127
High-density cultivation and product extraction	128
Pathway engineering	129
Translation initiation.....	130
References	131
Appendix 1: METHODS FOR GENETICALLY MODIFYING <i>SYNECHOCYSTIS SP.</i> PCC 6803 AND CHARACTERIZING ENGINEERED STRAINS	132
Summary.....	132
Introduction	132
Materials.....	133
Cell Culture.....	133
DNA Design for Homologous Recombination.....	134
Plasmid Construction	134
Natural DNA Uptake and Homologous Recombination.....	135
Cell Culture.....	135
Counterselection for Markerless Transformation	137
Methods.....	137
Plasmid Design and Construction	137
Natural DNA Uptake and Homologous Recombination.....	139

Counterselection for Markerless Recombination.....	141
<i>Synechocystis sp.</i> PCC6803 Electroporation Protocol (version 2).....	143
Western blot protocol for quantifying proteins in <i>Synechocystis sp.</i> PCC6803.....	144
Histag purification of protein from <i>E. coli</i> using spin columns	147
Example: Testing promoter strength in <i>Synechocystis</i> 6803	150
Obtain parent plasmids.....	151
Amplify promoter of interest.....	151
Construct promoter:GFP plasmid	151
Transform <i>Synechocystis</i> 6803 with counterselection plasmid pMK0168	152
<i>Synechocystis</i> 6803 transformation with promoter:GFP plasmid	152
Test promoter by measuring fluorescence.....	152
References	159
Appendix 2: SUPPLEMENTARY INFORMATION FOR CHAPTER 2	166
List of Abbreviations	181

CHAPTER 1: INTRODUCTION

Metabolic engineering is a rapidly growing field in which engineers seek to control the rates of metabolic conversions and the distribution of metabolites present in large-scale cultivation, usually of microbial cultures. Often, the objective is to maximize the production of a single valuable product. To achieve this, it is essential to be able to control gene expression, precisely and accurately. Nielsen and Keasling reviewed the metabolic engineering field recently (Nielsen and Keasling 2016). The brief history of microbial utilization for production of valuable products commenced with fermentations for food products. Genetic modification of microbes to improve their performance started with random mutagenesis and screening, application of which towards penicillin production increased production by 10,000-fold (Thykaer and Nielsen 2003). Dramatic increases in our understanding of cell biology facilitated by full genome sequencing have resulted in impressive improvements in our ability to more precisely alter cell metabolism.

The ability to genetically modify organisms has opened an entirely new field in which engineers have opportunities to address numerous societal problems. Modification of microbial metabolism opens the door to replacing existing processes for producing chemicals that generate pollution. Microbial production of chemicals can be considered renewable and can be completed at lower temperatures and pressures than many petrochemical processes. This is because organisms rely on catalysis of thousands of different enzymes. Production of valuable plant metabolites that are often produced very slowly in minuscule amounts by plants can also be improved by engineering fast growing bacteria or yeast to produce those molecules.

Several companies including Ginkgo Bioworks, Evolva, Inscripta, and Lumen Biosciences have started in the last ten years that provide strain design service. These services promise to deliver strains optimized

for maximum production of a valuable product. Typically, these services rely on automated generation of large numbers of variant strains, followed by product screening, data analysis, and redesign of new strains. This cycle may be repeated until adequate performance is attained. Other companies are being built around novel products that rely on genetic engineering. Bolt Threads has genetically modifying yeast and bacteria to produce spider silk proteins. Butamax and Gevo produce isobutanol as a fuel using *Saccharomyces cerevisiae*. Amyris has engineered *S. cerevisiae* to produce the anti-malarial drug precursor artemisinic acid. This engineering process is expensive and time consuming, however. Nielsen and Keasling (2016) estimated that the development of new strains capable of accumulating products at industrially relevant scales generally takes 6-8 years (or more than 200 worker years) and \$50 million.

Climate change

Providing food and water to a global population that continues to grow presents an enormous challenge to humanity. Disruption of earth systems that agriculture depends on for continued activity by climate change compounds those challenges. Development of biofuels that don't compete with agriculture for arable land and fresh water can be one part of the solution. Biofuels produced by cyanobacteria or algae do not require organic carbon feed like the sugars required for biofuel production by *E. coli* or yeast. In addition, many species of cyanobacteria and algae are salt tolerant and can be grown in brackish or sea water. Utilization of photosynthetic microbes could remove CO₂ from the waste streams from other processes and reduce the carbon impact of liquid fuels.

Metabolic engineering importance

Cyanobacteria and algae do not naturally accumulate large quantities of many metabolites are needed for industrial production of fuels or other products. Technology to manipulate metabolic networks needs to be applied to these organisms to improve performance of this task. Metabolic engineering is a rapidly growing field which applies genetic modifications to organisms to alter cellular function and

metabolic flux, generally to maximize productivity of one or more value products. The principles of metabolic engineering have been used to produce a wide array of molecules in cyanobacteria. Often, genetic elements such as promoters, ribosome binding sites, and protein coding sequences previously characterized for use in *E. coli* are used as a starting point for metabolic engineering projects in cyanobacteria. Reliable tools for manipulating metabolism are needed in order to develop microbial cell factories. Those include tools for modulating expression levels of proteins at the transcriptional level through promoters, and at the translation level through ribosome binding sites. Often, the tools applied towards metabolic engineering in cyanobacteria have been previously developed for use in *E. coli*. As we have found, such tools do not always work the same way in cyanobacteria.

Resource requirements for *E. coli* vs cyanobacteria

Production of biofuels and other valuable products by *E. coli* and yeast requires inputs of organic carbon. This may come in the form of sugars derived from maize or other food crop sources. Because of that requirement, these processes compete with human food crops for agricultural resources such as arable land, fresh water, and fertilizer. Processes that instead rely on cyanobacteria or algae do not require organic carbon inputs. Instead carbon dioxide from deep wells or from the waste streams of combustion processes can be used as inorganic carbon sources that these organisms can fix through photosynthesis.

Many species of cyanobacteria and algae are also salt-tolerant and can be grown in brackish or sea water. This is a major advantage since freshwater resources are limited in many places around the world. Cyanobacteria do require nitrogen and phosphate inputs, however. Researchers are investigating possible sources of these, including from municipal wastewater treatment plants and farmland runoff. In both of these, excess nitrogen and phosphate cause problems, for example, with toxic algae blooms.

Some species of cyanobacteria are also capable of fixing nitrogen, and researchers are also attempting to engineer other species to give them this capability.

To minimize capital costs, cyanobacteria are often grown outdoors in open raceway ponds which are shaped like racetracks with a large paddle wheel mixing the culture. Sunlight is a free source of energy for these cultures, and carbon dioxide-enriched air can be bubbled through them. These conditions can be quite different than what metabolic engineers use to test genetically modified strains. In Chapter 3, on bisabolene production in *S. 6803*, we sought to address at least part of this scale-up problem by growing the best-performing engineered strain in simulated outdoor light conditions. There are many other stresses that engineered strains may face in outdoor ponds that we did not address in this research. For example, the ponds may also be subject to exposure to pathogens and organisms that graze on cyanobacteria, exposure to temperature extremes, and exposure to nutrient starvation. The potential of cyanobacteria and microalgae as sustainable biofactories has been reviewed recently (Benedetti et al. 2018).

Metabolic engineering in cyanobacteria

The strains

There are several strains of cyanobacteria that are commonly used for biological studies as well as for metabolic engineering projects. These include *Synechocystis Sp. PCC 6803 (S. 6803)*, *Synechococcus elongatus sp. PCC 7002 (S. 7002)*, and *Synechococcus elongatus Sp. PCC 7942 (S. 7942)*. These have doubling times of 6.6 hours, 4.1 hours, and 4.9 hours, respectively, when grown with 3% CO₂ supplied (Yao et al. 2016). *S. 6803* and *S. 7942* are freshwater strains, while *S. 7002* tolerates a wider range of salt concentrations and can also tolerate high light intensities. Generally, strains that accumulate glucosylglycerol during osmotic stress (including *S. 7002*) can tolerate salinity up to 200-250% of sea water, while those that instead accumulate sucrose can only tolerate up to 50-100% of seawater. *S.*

6803 and S. 7002 were both found to tolerate double the salinity of seawater (Reed and Stewart 1985). Strategies for engineering other strains to be more salt tolerant have been reviewed recently (Cui et al. 2020).

A faster growing cyanobacteria which has a doubling time of 1.9 hours, *Synechococcus elongatus* UTEX 2973 (S. 2973), had been identified recently and is already being used in metabolic engineering projects. *Synechococcus elongatus* UTEX 2973 was isolated from a mixed culture of *Synechococcus leopoliensis* UTEX 625. That strain had been fast growing but had lost its fast growth ability in the years since it was first described in 1955. (Doubling times above from Yu et al., 2015). S. 2973 and S. 7942 have 99.8% identical genomes, with just 55 SNPs between them (Yu et al., 2015). *Anabaena* sp. PCC 7120 (A. 7120) is one of the nitrogen-fixing cyanobacteria. It has much slower growth rate with a doubling time of 14-15 hours (Callahan and Buikema 2001). Nitrogenase reduces N₂ but is extremely sensitive to oxygen. This is a problem for cyanobacteria which generate oxygen through photosynthesis. Most strains that can fix nitrogen do so in heterocysts, a different cell type in this filamentous species, (A. 7120 is one example) which do not generate oxygen. Other strains have solved this problem by fixing nitrogen only during the night (Stal 2015). Many species of cyanobacteria have fully sequenced genomes, annotation lags behind *E. coli* (and even in *E. coli* so much is unknown) and many or most genes have unknown function and are only annotated as hypothetical proteins.

Transformation methods

Transformation of cyanobacteria is generally a straight-forward process. Plasmids can be constructed that contain some cargo flanked on either side by regions that have an identical sequence to a target region in the chromosome. Some cyanobacteria, like S. 6803, S. 7002, and S. 7942, naturally uptake DNA. With some frequency double stranded breaks may occur and homologous recombination will use the plasmid DNA to repair that break. This can be a rare event, but antibiotic selection markers can be

used to ensure that only the cells that are transformed will grow. Natural uptake and modification of chromosomal DNA in *S. 6803* through homologous recombination was demonstrated in 1988 (Williams 1988). In this method, cells at a high density are simply incubated in the light with a large amount of DNA, and then screened for transformants. Since cyanobacteria are often polyploid, or have multiple copies of their chromosomal DNA (Griese, Lange, and Soppa 2011), it is necessary for mutant screening to check for both the presence of the mutation as well as the absence of the wildtype fragment.

Transformation efficiency was found to vary most strongly with the amount of DNA, and convenient incubation times of 2-6 hours worked well (Williams 1988). Elhai and Wolk (1988) described the method of conjugal transfer for transforming cyanobacteria which allows the plasmid of interest to be transferred from *E. coli* to the cyanobacteria through direct contact with the assistance of an IncP helper plasmid such as RP4 (Elhai and Wolk 1988)

Zang et al. (2007) optimized and compared electroporation, ultrasonic treatment, and natural uptake methods of transformation. The longest homology regions they tested (1,300 base pairs on one side, and 1,700 on the other) resulted in the highest transformation efficiency. They reported that neither electroporation nor ultrasonic treatment improved the transformation efficiency. For natural uptake, plasmid concentrations of 10 µg/ml or greater had much higher transformation efficiency than lower concentrations and cells harvested at an OD of about 0.8 were optimal for transformations. They also found a steep drop-off in transformation efficiency if the plasmid incubation varied from five hours (Zang et al. 2007).

Methylation of foreign DNA using the native methyltransferase of *S. 6803* improves the efficiency of transformation by 11- to 161-fold, presumably by reducing the degradation of the DNA by nucleases. This was demonstrated by cloning *slI0729* and *slr0214* methyltransferases into the integration plasmid which allowed the plasmid to be methylated during cloning in *E. coli* (Wang et al. 2015b).

CRISPR/Cas9-like systems could potentially accelerate transformations. Cas9 nuclease may be toxic to cyanobacteria, and its utilization required transient expression to generate gene knockouts in UTEX2973 (Wendt et al. 2016). As an alternative, an RNA-directed dsDNA nuclease from *Francisella novicida*, Cpf1, has been successfully utilized for generating double-stranded breaks in the fast-growing cyanobacterium, *Synechococcus elongatus* UTEX 2973. Like other CRISPR technology, Cpf1 may be used for facilitating markerless knock-ins, knock-outs, and point mutations. Cpf1 was less toxic than cas9 in UTEX2973, requires shorter sgRNA (single guide RNA) than cas9, and can more easily be applied to multiple targets because it does not require the tracrRNA and crRNA to be fused for correct processing as in the case of cas9 (Ungerer and Pakrasi 2016).

Metabolic engineering methods

A wide variety of metabolic engineering strategies have been developed in the last 30 years (a good recent review: (Chae et al. 2017)). I briefly introduce some of these strategies here and cover some strategies in more detail below in the “Cyanobacteria toolbox” section below if they have been applied in cyanobacteria. Protein engineering is an entire field dedicated to the design and modification of protein functions. Computational models of protein structure can facilitate protein design and compliments strategies such as saturation mutagenesis at specific amino acids followed by screening of the mutants for the desired function. Co-localization of enzymes that operate within a pathway have increased production rates of some metabolites, possibly through substrate-channeling. This has been achieved by fusing the proteins together or using protein scaffolds bearing interaction domains to recruit specific enzymes (for example, (Dueber et al. 2009)), or by designing or repurposing bacterial microcompartments (for example, (Plegaria and Kerfeld 2018)). These strategies can be particularly useful if one or more intermediates in the pathway are toxic because the intermediate is likely to be consumed by the subsequent enzyme in the pathway before it can diffuse away and cause damage. Knockouts of genes in pathways that compete with those that generate the product of interest can

often increase titers if they are not lethal. Genome scale models of metabolism can be used in conjunction with algorithms such as OptKnock to identify genes to knockout to maximize production (Burgard, Pharkya, and Maranas 2003). In cases where a knockout may be lethal, gene repression using antisense RNA or CRISPR interference where the asRNA for dCas9 prevent the RNAP or ribosome from binding the DNA or RNA. Evolutionary engineering generates sequence diversity through error-prone PCR or methods such as MAGE, screens those diverse sequences for function, and repeats the cycle with the best performing strains. Development of a high-throughput screen for the product of interest is required for success in this field, and often biosensors are utilized that can transduce the signal of a specific molecule into something easy to measure such as a fluorescent protein. Gene expression components such as promoters and ribosome binding sites that function well in a variety of organisms and sequence context are essential for many of these strategies.

Metabolic engineering for terpenoid production in cyanobacteria

Cyanobacteria have been engineered to divert metabolic flux to a wide variety of products. As discussed in Chapter 3, we targeted bisabolene as our target molecule. Catalytic hydrogenation of bisabolene produces bisabolane, a saturated fifteen carbon molecule with combustion properties much like Diesel #2 fuel (Peralta-Yahya et al. 2011). Terpenoids are a diverse class of molecules (more than 70,000 (Vickers et al., 2014)) that include commercially valuable photosynthetic pigments. The titers and productivity of terpenoids in engineered cyanobacteria generally lags far behind those for other products that are closer to central metabolism, where metabolite abundances and metabolic fluxes are much higher. Products such as ethanol and 2,3-butandiol have reached the g/L scale, while terpenoids are generally produced at the tens of mg/L scale (Table 1.1, below, summarizes reports of terpenoid production in cyanobacteria).

Cyanobacteria utilize the methylerythritol phosphate (MEP) pathway to convert pyruvate and glyceraldehyde-3-phosphate to isopentenyl pyrophosphate and dimethylallyl pyrophosphate, which form the backbone of terpenoids. In cyanobacteria, a single enzyme, geranylgeranyl pyrophosphate synthase catalyzes the three successive additions of IPP to a DMAPP-derived head-group to form geranyl pyrophosphate, farnesyl pyrophosphate, and geranylgeranyl pyrophosphate. These three metabolites are precursors for diverse terpenoids, including essential photosynthetic pigments including carotenoids and chlorophyll. Eukaryotes, archaea, and some gram-positive bacteria including *Staphylococcus aureus* (Balibar, Shen, and Tao 2009) instead use the mevalonate pathway which draws on acetyl-CoA for terpenoid production.

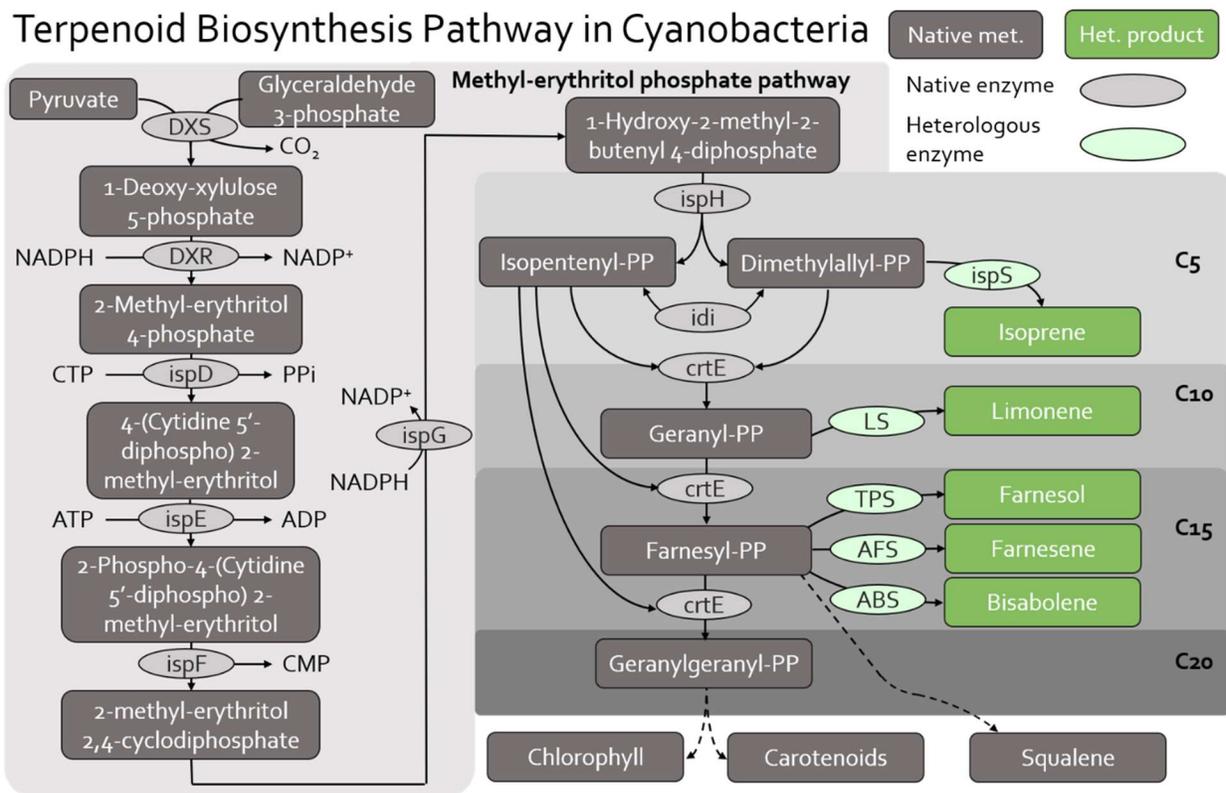


Figure 1.1: The terpenoid biosynthesis pathway in cyanobacteria. The five-carbon molecules that form all terpenoid backbones, isopentenyl pyrophosphate and dimethylallyl pyrophosphate, are produced from pyruvate and glyceraldehyde-3-phosphate via the methyl-erythritol pathway. The enzyme crtE then produces ten-, fifteen-, and twenty-carbon backbones for this diverse class of molecules. Examples of heterologous products that have been made in cyanobacteria are shown on the right side.

Examples of efforts to engineer cyanobacteria to produce other terpenoids such as isoprene, limonene, farnesene, bisabolene, and bisabolol are discussed in this section. A common thread observed is that titers of terpenoid products is greatly enhanced by continuous removal of the product. Isoprene is volatile enough that it can be collected from the gas phase. This presents challenges for capturing the product. However, as Gao et al. (2016) showed, continuous removal of the headspace vapor of the bioreactors increased the total isoprene produced over what was measured when a sealed container was used. For the less volatile molecules like limonene and bisabolene, cultures are grown with an oil overlay of dodecane where most of the product is harvested. It is unclear how these large, hydrophobic molecules are transported from inside cells to the overlay. A molecular dynamics model made clear that 10-15 carbon terpenoids can traverse bacterial cell membranes. Specifically, insertion of these molecules into the cell membrane incurs an energetic penalty which is partly compensated for by proximity of cells to the dodecane phase (Vermaas et al. 2018). Decreased growth rates of *Dunaliella salina* with increasing rates of dodecane sparging, along with light microscopy showed that dodecane can lyse cells, releasing their contents (Kleinegris, van Es, et al. 2011; Kleinegris, Janssen, et al. 2011). Continuous removal of the product may reduce the negative impact that product inhibition can have on enzyme catalysis.

Among the terpenoid product titers discussed here, a vast difference is found between the maximum isoprene titer achieved and the titers achieved for the others. This could be due to differences in the transport rates and the related impact of product inhibition. It may also be due to differences in precursor abundance. In cyanobacteria, a single enzyme (crtE) catalyzes three successive reactions which each add another five carbons from isopentenyl pyrophosphate to a dimethylallyl pyrophosphate head group. Given the hydrophobicity of the terpenoid backbone being constructed in this process, it may be that crtE usually binds the backbone until the full 20-carbon geranylgeranyl pyrophosphate (GGPP) molecule is completed. This would limit the abundance of the intermediates, geranyl

pyrophosphate and farnesyl pyrophosphate which are precursors for heterologous products limonene and bisabolene, respectively. It was recently demonstrated that the GGPP synthase in *S. 7002* (*crtE*) mainly produces GGPP (Feng et al., 2020).

In an early prototyping experiment to produce heterologous terpenoid synthases in cyanobacteria, a single construct for bisabolene synthase (from *Abies grandis*) or limonene (from *Mentha spicata*) expression demonstrated that these terpenoids could be produced in *S. 7002*. Titters of each were higher when a dodecane overlay was applied, reaching 4 mg/L for limonene and 0.6 mg/L for bisabolene. Glycogen synthesis knockout failed to improve those titters. Instead, those strains exhibited over-accumulation of several organic acids (Davies et al. 2014).

Expression of all seven genes in the mevalonate (MVA) pathway from bacterial sources increased the isoprene production by 2.5-fold over a strain of *S. 7002* that only expressed isoprene synthase (Bentley, Zurbriggen, and Melis 2014). This modest increase may be due to the relatively small pool sizes of the precursor metabolites for the MVA pathway. Gao et al. (2016) observed that acetyl-CoA was 20-fold less abundant in *S. 7942* than it was in *E. coli*, while glyceraldehyde-3-phosphate and pyruvate were 5- and 21-fold more abundant in *S. 7942* than they were in *E. coli*, respectively.

Chaves et al. (2016) increased isoprene production by only co-expressing the isopentenyl diphosphate isomerase from *Pseudomonas pneumoniae* with isoprene synthase. Interestingly, knockout of the *cpc* operon that codes for phycocyanin could also increase isoprene production. Knockout of phycocyanin increased production of carotenoids, which would likely necessitate increased flux through the MEP pathway that isoprene synthase also relies on (Chaves et al. 2016).

A strain of *Synechococcus elongatus* engineered by Gao et al. (2016) achieved 1.26 g/L isoprene after 21 days, equivalent to 40% of the fixed carbon. They first compared expression levels of codon-optimized and non-optimized isoprene synthase gene from six plants, as well as the catalytic activity of the six

different genes. The best performing strain combined the best of those strains with IDI overexpression to increase the pool of available dimethylallyl pyrophosphate. In addition, to channel DMAPP substrate to isoprene synthase, IDI was fused to isoprene synthase. The proximity of the two enzymes was expected to help isoprene synthase compete with native enzymes for DMAPP (Gao et al. 2016).

Codon optimized *ispS* from kudzu (*Pueraria montana*) was expressed from different promoters, *PrbcL*, *Ptac*, *PpsaA**, with and without co-expression of *dxs*, a purported bottleneck in the precursor pathway (methyl-erythritol phosphate, or MEP, pathway). *PrbcL* driven *ispS* alone resulted in the highest expression of *ispS*, and the highest isoprene specific productivity of 1.2 ng/mL/hr/OD (equivalent to 93 ug/g DCW) in closed cultures. Online detection of headspace isoprene, in which isoprene was continuously removed, increased this to 4.2 ng/mL/hr/OD (336 ug/g DCW), suggesting negative feedback on the MEP pathway. Performance was also tested in high salt concentrations to simulate growth in seawater. Isoprene production lagged in 600mM NaCl due to an extended lag phase and, likely, to a diversion of metabolic flux towards osmoprotectants such as sucrose and glucosylglycerol (Pade et al. 2016).

Dienst 2020 utilized high density cultivation to increase terpenoid titers in *S. 6803*. The *petE* promoter and the strongest bicistronic design (BCD) from Mutalik et al. (2013) (BCD2) were used to express bisabolene, bisabolol, or patchoulol synthases (protein abundance not quantified). The titers reached 179 mg/L, 96.3 mg/L, and 17.3 mg/L, respectively for these products.

Following a strategy demonstrated earlier in *E. coli* (for example, (Martin et al. 2003; Pitera et al. 2007)), Bentley et al. (2014) expressed the six mevalonate pathway enzymes from *Enterococcus faecalis* and *Streptococcus pneumoniae* in *S. 6803* to increase the production of isoprene (a five-carbon terpenoid). Although expression of the mevalonate pathway enzymes were not optimized, this strategy did increase isoprene production by about 2.5-fold (Bentley, Zurbriggen, and Melis 2014).

Table 1.1: A summary of reports of terpenoid production in cyanobacteria, arranged first by the product precursor (a terpenoid backbone with five carbons (IPP), ten carbons (GPP), fifteen carbons (FPP), or twenty carbons (GGPP)), then by year of the report (adapted from Lin and Pakrasi (2019)).

Product Precursor	Species	Product	Titer (mg/L)	Specific titer	Productivity	Specific productivity	TPS Source Org.	Construct	First Author	Year	RBS design	Codon opt.	Promoter
IPP	S. 6803	soprene				0.05 mg/g DCW/d	<i>Pueraria montana</i>	PsbA2-ips	Lindberg	2010	PsbA2 RBS	DNA2.0	PsbA2
	S. 6803	soprene	0.31	0.250 mg/g DCW		0.03 mg/g DCW/d	<i>Pueraria montana</i>	PsbA2-ips, MVA pathway	Bentley	2014	PsbA2 RBS	DNA2.0	PsbA2
	S. 7942	soprene	1,360		60 mg/L/d		<i>Eucalyptus globulus</i>	Ptc-idi-ips-dxs, ips from <i>Eucalyptus</i>	Gao	2016	trc	Yes, but unclear	trc
	S. 6803	soprene		0.336 mg/g DCW		0.101 mg/L/OD/d	<i>Pueraria montana</i>	Ptbl-ips + Ptbl-ips	Paide	2016	Ptbl, tac	removed rare codons	Ptbl, Ptbl
	S. 6803	soprene	2.5		0.63 mg/L/d		<i>Pueraria montana</i>	CpcB-ips fusion protein	Chaves	2017	CpcB	Yes, but unclear	CpcB
	S. 6803	soprene		2.8 mg/g DCW		2.8 mg/g DCW/d	<i>Eucalyptus globulus</i>	Ptbl-ips	Englund	2018			
	S. 6803	soprene		2.8 mg/g DCW		2.8 mg/g DCW/d	<i>Eucalyptus globulus</i>	Ptbl-ips, Ptbl-idi-dxs	Englund	2018	BCD or Ribof	Same Designer	trc
	S. 7002	limonene	4		1 mg/L/d		<i>Wentha spicata</i>	Ptbl-ims, delete gfs	Davies	2014	CpcB	DNA2.0	CpcB
	S. 6803	limonene	1		0.03 mg/L/d		<i>Schizonepeta tenuifolia</i>	Ptbl-ims, Ptbl-dxs-idi-crte	Kiyota	2014	trc	Seescript	trc
	S. 7942	limonene	2.5		0.63 mg/L/d		<i>Wentha spicata</i>	Ptbl-ims	Wang	2016	RBS calculator (1)	Yes, but unclear	PsbA from sea plant
GPP	S. 6803	limonene	6.7		0.96 mg/L/d		<i>Wentha spicata</i> (gpps from <i>A. grandis</i>)	Ptbl-ims, Ptbl-dxs-gpps, Ptbl-rpi-rpe	Lin	2017	trc	IDT	trc
	A. 7120	limonene	0.52		0.043 mg/L/d		<i>Sitka spruce</i>	Ptbl-Ptbl-ims-dxs-idi-gpps	Halfmann	2014b	PsbA1	No, cloned from plant	PsbA1
	S. 6803	phellandrene		3.2 mg/g DCW		1.6 mg/g DCW/d	<i>Lavandula angustifolia</i>	CpcB-PHL fusion protein	Formighieri	2015	CpcB	DNA2.0	CpcB
	S. 6803	phellandrene		10 mg/g DCW		5 mg/g DCW/d	<i>Lavandula angustifolia</i>	MVA pathway, CpcB-PHL fusion protein	Formighieri	2016	CpcB	DNA2.0	CpcB
	S. 6803	phellandrene		5.95 mg/g DCW		2.95 mg/g DCW/d	<i>Lavandula angustifolia</i> (gpps from <i>Picea abies</i>)	CpcB-PHL, NptI-gpps fusion proteins	Betterle	2018	CpcB	previous Mellis	CpcB, Ptc
	S. 7942	amorphadiene	19.8		1.98 mg/L/d		<i>Artemisia annua</i>	Ptbl-dxs-idi-ispA, Ptbl-ads	Choi	2016	trc	DNA2.0	trc
	S. 7002	bisabolene	0.6		0.15 µg/L/d		<i>Abies grandis</i>	CpcB-bis5, delete gfs	Davies	2014	CpcB	DNA2.0	CpcB
	S. 6803	bisabolene	179.4		1.45 mg/OD/L		<i>Abies grandis</i>	High density cultivation, PpetE + BCD2	Dienst	2020	BCD2	DNA2.0	PpetE
	S. 6803	bisabolene	7.9		2.6 mg/OD		<i>Abies grandis</i>	Varied RBS, and codon usage	Seberta	2020	RBS Calculator	Several	Ptczop
	S. 6803	bisabolol	96.3		0.688 mg/OD/L		<i>Warricaria recutita</i>	High density cultivation, PpetE + BCD2	Dienst	2020	BCD2	DNA2.0	PpetE
FPP	S. 7942	arnesene	4.6		0.66 mg/L/d		<i>Malus domestica Borkh.</i> (apple seed)	Ptbl-dxs-idi-ispA, Ptbl-afs	Lee	2017	trc	DNA2.0	trc
	A. 7120	arnesene	0.3		0.02 mg/L/d		<i>Picea abies</i>	Ptbl-Ptbl-afs	Halfmann	2014a	PsbA1	DNA2.0	PsbA1
	S. 6803	pachulol	17.3		0.17 mg/OD/L		<i>Pogostemon cablin</i>	High density cultivation, PpetE + BCD2	Dienst	2020	BCD2	DNA2.0	PpetE
	S. 6803	squalene		4.98 mg/L/OD			<i>Saccharomyces cerevisiae</i>	Ptbl-dxs-idi-ispA, Ptbl-sqs	Choi	2016	trc	DNA2.0	trc
	S. 7942	squalene		79.2 mg/g DCW		6.09 mg/g DCW/d	<i>Saccharomyces cerevisiae</i>	dxs, idi, ispA, cpcB-SQS fusion, or PPS-SQS fusion	Choi	2017	trc	DNA2.0	trc
	S. 6803	geranylinalool		0.360 mg/g DCW		0.18 mg/g DCW/d	<i>Nicotiana attenuata</i>	CpcB-TPS fusion	Formighieri	2017	CpcB	No, cloned from plant	PsbA2, Ptbl
	S. 6803	manoyl oxide		0.45 mg/g DCW		0.11 mg/g DCW/d	<i>Coleus forskohlii</i>	Ptbl-CTPS-dxs, Ptbl-B-DXS-GGPPS-TP52-TP53	Englund	2015	RBS*	No, cloned from plant	PsbA2, Ptbl

Cyanobacteria toolbox

Cloning

Modular cloning systems are important for standardizing assembly of genetic parts in synthetic biology. Precise definition of parts is essential for facilitating combinatorial assembly of plasmids that may include variable origins of replication, homologous regions for chromosomal integration, antibiotic selection markers, promoters, ribosome binding sites, genes of interest, terminators, and secondary proteins that may regulate expression of the gene of interest. The syntax of these assembly/design languages consists of standardized sequences flanking each genetic part including the restriction enzyme recognition sequences and the overlapping sticky end sequences that are generated on digestion. Typically, each part interface has its own unique sticky ends which ensures that the parts are assembled in the correct order (e.g., promoter-coding sequence-terminator). The modularity of these cloning systems also facilitate the design and documentation of plasmids, and automated cloning in the biofoundries that are beginning to generate thousands of strains every day. Two such systems have recently been developed: SyneBrick (Kim et al., 2017) and CyanoGate (Vasudevan et al., 2019).

SyneBrick vectors were developed for gene expression in *S. 7942* for chromosomal integration. They are based on Bg1Brick formatting and contain the pUC origin of replication, one of three flanking pairs of neutral site homologous regions, a selection marker, an inducer gene (or none), a promoter, and a gene of interest. The ribosome binding sites utilized are inherited from promoter sequences. CyanoGate is a similar modular cloning system for cyanobacteria protein expression based on the Plant MoClo syntax which relies on Type IIS restriction enzymes BsaI and BbsI. Unlike other restriction enzymes, Type IIS enzymes cut outside of their recognition sequence (and recognize asymmetric sequences so that they only cut on one side of the recognition sequence), allowing for the generation of unique single stranded “sticky ends” which guide Golden Gate assembly. This system was used to adapt pPMQAK1 with an

RSF1010 origin of replication and pSEVA421 with an RK2 origin to this syntax. Copy numbers in S.6803 were estimated at 31 and 10 for these plasmids, respectively (Vasudevan et al. 2019).

Neutral sites for homologous recombination

Heterologous genes are often chromosomally integrated into neutral sites which are not expected to result in other phenotypic changes to the organism. In S. 6803, a commonly used neutral site is the slr0168 hypothetical protein gene. Common practice is to use the two halves of the gene as the homologous regions flanking the synthetic construct to be introduced. Transcription of the slr0168 gene could result in unwanted and unexpected interference with the expression of the gene of interest.

Transcriptomic data has been used to identify new neutral sites for S. 6803 which have no background transcription, in both the chromosome and an endogenous plasmid. Those sites with the longest stretch of un-transcribed DNA were chosen for study. Characterization by YFP expression in two chromosomal sites were similar to each other, while expression was around 10x higher for the endogenous plasmid integration. Expression from the pCC5.2 plasmid neutral site was also much higher than from the pPMQAK1 plasmid with RSF1010 origin of replication. Expression of YFP from this site using the strong promoter Pcp560 resulted in YFP accumulating to an estimated 20% of total cellular protein (Ng, Berla, and Pakrasi 2015).

Pinto et al. (2015) tested 13 chromosomal integration sites in S. 6803 for the potential as neutral sites. These sites were identified by selecting small open reading frames annotated as hypothetical or unknown proteins with no transmembrane domains, no reported interaction with other known proteins, and minimal similarity to other known proteins. Of those sites several were shown to have low transcription levels (potentially, this low transcription level could easily be unique to the growth conditions they tested). Of the five sites identified, only two (slr0271 and slr0397) were shown to grow at the same rate as WT in each of five different representative growth conditions. Examination of the

proteome of GFP-expressing mutants from these five sites showed that only one native protein was differentially expressed. None of the sites from Ng et al. (2015) were considered here because they didn't meet the screening requirements listed above. At the time of this study, 1532 of 3264 ORFs were still unknown or hypothetical. A full set of barcoded single-gene knockout strains of *S. 6803* has recently been constructed with the expectation that the function of more of these unknown proteins will be identified (Gale et al. 2019).

A frequently used neutral site corresponds to the gene *Slr0168* which has unknown function, but its deletion doesn't result in a phenotype. Transformation site originally from Williams 1988 that has been used many times since then for genomic integration that interrupt the native gene without an observed phenotype being generated by the disruption (Williams 1988). It has been found that protein corresponding to *slr0168* is secreted (Sergeyenko and Los 2000). The *psbA2* gene is also sometimes used because there are two copies of the gene. Finally, a site between *slr2030* and *2031* has also been used (Sato et al. 2001; Englund et al. 2015).

Replicative plasmids

Several plasmid backbones have been developed for expressing heterologous genes in cyanobacteria. Expression from plasmids avoids potential interference from (or of) the genomic DNA. It is also slightly more convenient for transformations because it not necessary to allow time for homologous recombination and it is also not necessary to ensure that each copy of the chromosome contains the correct modification. Most of the replicative plasmids used in cyanobacteria contain the broad-host range RSF1010 replicon which can replicate in *E. coli* as well as in the cyanobacteria. pPMQAK1 uses the RSF1010 replicon and contains a BioBrick standard cloning site (H. H. Huang et al. 2010b). The pDF plasmids contain the same replicon, and include the spectinomycin resistance gene, the *lacI^q* gene for repression of promoters containing lac operators, and follows the cutsite-part format of SpeI-Promoter-

KpnI-ORF-PstI (Guerrero et al. 2012). Several other replicative plasmids for use in cyanobacteria were summarized by Heidorn et al. (2011).

More recently, Ferreira et al. (2018) have introduced the pSEVA plasmid series built on the Standard European Vector Architecture (SEVA) as another option for use in cyanobacteria. The SEVA architecture is formed by cargo, replication origin and selection marker variable regions separated by the oriT conjugation origin and the T0 and T1 terminators. Those used by Ferreira use the RSF1010 replicon (Ferreira et al. 2018). Liu and Pakrasi (2018) have also adapted two of the endogenous plasmids of *S. 6803* for use in biotechnology by combining them with the pUC118 plasmid. This facilitates cloning by enabling the plasmids to replicate in *E. coli* with a high copy number. Specifically, pCA-UC118 and pCB-SC101 were constructed by combining pUC118 (high copy number in *E. coli*) with the *S. 6803* endogenous plasmid pCA2.4 (former), and pSC101 (low copy number) with the endogenous plasmid pCB2.4. Both of the resulting plasmids were capable of generating about 50% higher expression of GFP than the same construct in an RSF1010 backbone plasmid, suggesting that both had higher copy numbers in *S. 6803* than the RSF1010 plasmid (Liu and Pakrasi 2018).

Knockdowns with CRISPR interference or antisense RNA

Gene knockouts have been an important tool for reducing metabolic flux through pathways that compete with the product for precursors. In many cases, genes that would be helpful to delete are essential to growth. Therefore, new tools have been developed to reduce expression of such essential genes, often at a certain point in time when the culture can be optimally switched from prioritizing growth to prioritizing production. The concept of a metabolic valve, where flux through competing pathways can be reduced to the level that maximizes productivity, has been developed in the lab of Dr. Kristala Prather (for example, (Gupta et al. 2017)). This concept has been applied to production of valuable chemicals in cyanobacteria. Often, the CRISPR/dCas9 system is used to this end where the deactivated Cas9 (mutated to eliminate its nuclease activity) binds the DNA either within the promoter

to inhibit transcription initiation, or within the target gene, inhibiting transcription elongation. When the promoter is targeted, this action functions much like transcription factors that repress transcription initiation except that single guide RNA can be designed to direct the dCas9 to any target. An important limitation has been identified by a few papers that have found that expression of multiple guide RNA targeting multiple sites can quickly overwhelm the dCas9 resulting in reduced efficiency in repression.

Yao et al. knocked down expression of GFP all the way to background level of wild-type *S. 6803* when *P_{trc}* was used to express the sgRNA (single guide RNA) which targeted the 5' region of the coding sequence and dCas9. Weaker promoters for this purpose resulted in less effective repression, and expressing two or more different sgRNAs resulted in less effective repression (Yao et al. 2016). The same lab of Dr. Elton Hudson used the same approach to reduce expression of up to six native genes to increase production of fatty alcohols. Those six genes were selected as those that utilized the C16 and C18 fatty acid ACP which would compete with the heterologous fatty acid-ACP reductase expressed. The best performing strain targeted all six genes even though the repression of each individual gene was less efficient in this strain than those that targeted fewer genes for repression (Kaczmarzyk et al. 2018).

Gordon et al. (2016) also observed that balancing expression of sgRNAs and dCas9 was necessary for optimal repression in *S. 7002*. The RBS Calculator was also used to design RBSs to tune expression of dCas9. However, the predicted RBS strength did not correlate with repression of the reporter gene being repressed. The utility of this system was further demonstrated by successfully repressing important genes that would be difficult to completely knockout without incurring growth defects, including carboxysome shell proteins, the phycobilisome operon, and glutamine synthetase I (Gordon et al. 2016).

CRISPRi using dCas9 was also successfully applied in *S. 7942*. The authors observed that repression was slightly more effective when the sgRNA targeted the start of the coding sequence than when it targeted the -35 region of the promoter (C.-H. Huang et al. 2016). Liu, Johnson, and Pakrasi recently developed a

similar CRISPR interference tool by deactivating the nuclease activity of the Cpf (analog of Cas9) (Liu, Johnson, and Pakrasi 2020). Expression of an antisense RNA may be used to for a similar purpose. Zess, Begemann, and Pflieger expressed a small antisense RNA under control of the tet repressor that, when induced, could repress expression of GFP by more than 50% by binding the 5'-UTR and reducing the translation initiation rate (Zess, Begemann, and Pflieger 2016).

Promoters

Promoters are the DNA sequences that control the rate of transcription initiation. As one of the main controls on protein expression, many studies have been carried out to identify promoters of varying strengths, and inducible promoters that can be switched on. As with many other synthetic biology tools, many of the promoters used in cyanobacteria have been imported from use in *E. coli*. However, there are many native promoters from cyanobacteria that have been characterized, especially those that are inducible by light. Since cyanobacteria will be grown in the light, it is important to note that the commonly used inducer for promoters containing the tet operator, anhydrotetracycline (aTc), is light sensitive. Reduced gene expression in an aTc-induced strain has been observed after 48 hours due to this light sensitivity (Kim et al. 2017).

A wide variety of promoters inducible by certain types of light have been identified in cyanobacteria including those inducible by green light (Abe et al. 2014), UV-B light (Máté et al. 1998), high light intensity (Mazouni et al. 1998; Eriksson et al. 2000; Salem and van Waasbergen 2004; Muramatsu and Hihara 2007), or no light (dark inducible) (Imamura, Asayama, and Shirai 2004). The Peebles Lab has also contributed to the identification and characterization of promoters in cyanobacteria as discussed below. To identify light-inducible promoters, Werner et al. (2018) selected promoters from genes that were shown to oscillate in light:dark cycle by microarray data (Beck et al. 2014) and defined as the 500nt upstream from the start codons of genes except that sequences were truncated to remove other open reading frames. Of the nineteen promoters tested, four conferred light:dark cycle correlated expression

of bacterial luciferase. PhlIC, Prbp1, Pslr0006, PsigA were strongly light entrained in 12:12 light:dark cycles. Generally, transcripts of luxAB for these four promoters rose sharply after 1 hour in light, and fell back near the dark-phase level by three hours after the onset of light. Expression of luxAB in these strains ceased oscillation when shifted to continuous light. Expression from PhlIC conferred a light-intensity response such that increasing the light level from ~200 μE to 500 μE resulted in a nearly 10-fold higher expression one hour after light onset (Werner et al. 2018).

Albers et al. (2015) characterized expression levels in *S. 6803* of promoters commonly used for heterologous expression in *E. coli* and adapted Ptac to be inducible by IPTG through addition of lac operators. PsigA, Ptac (Psc4-0, removed lac operators), Ptrc (Psc5-0, removed lac operators), Ptic (Psc6-0, removed lac operators). Ptac had higher expression in *S. 6803* than it did in *E. coli*, and Ptic had lower expression in *S. 6803*. Ptrc had similar expression in both organisms. Psc1-2, Psc2-2, Psc3-2: Ptic with 3, 2, 1 nt removed between -35 and -10: Interestingly, Psc3-2 had slightly higher expression than Psc6-2 (Albers, Gallegos, and Peebles 2015b).

Camsund et al. altered the trc promoter to improve the repression by lac operators which had previously been found to poorly repress transcription initiation. They tested 8 BioBrick promoters (BBa_J23### series) which were originally designed based on the consensus *E. coli* promoter sequence, plus the native promoters PnirA, PpetE, and PrnpB. The BioBrick promoters spanned a wide range of expression levels both higher and lower than the native promoters. Ribosome binding sites are often inherited from promoters, meaning the entire region from the start codon to the 5'-end of the promoter region is treated as "the promoter." Somewhat uniquely, a single ribosome binding site, RBS*, was used consistently throughout this study. These promoters are also known as the Anderson collection from the iGEM registry and were suggested as a useful set of orthogonal promoters for use in cyanobacteria based on the results found. A series of Ptrc variants were also tested in order to improve the Ptrc repression by changing the spacing between the 5' lac operator and the -35 box. However, this effort

failed to improve the fold change in expression (from uninduced to induced) in S. 6803 (Camsund, Heidorn, and Lindblad 2014).

Surprisingly, Huang et al. (2010) found that the BioBrick promoters BBa_R0040 (tet), BBa_R0051 (lambda phage PR), BBa_R0010 (Plac) had minimal expression or below the detection limit. Seven variants of rbcl promoter between -277 and the TSS generated a range of expression levels, but all were much weaker than Ptrc1O and Ptrc2O (one lac operator and two lac operator versions of Ptrc) (H. H. Huang et al. 2010b).

Huang and Lindblad (2013) designed and characterized a set of trc promoter variants in S. 6803. For example, promoter L12 was generated by changing the -10 element of BBa_R0040 to TATAAT, the consensus sequence in S. 6803. This change had the unexpected result of decreasing, rather than increasing, the expression level. Varied the sequence between the -10 element and the transcription start site of the L12 promoter to obtain a wider range of expression levels. L03 swapped the AG at -6 and -5 from the TSS to GC and increased the expression level by about 450-fold. An improved induction ratio of 83 was also achieved by this promoter. Ptrc1O still had the highest maximum expression. However, the uninduced expression was nearly the same as the induced level. (H.-H. Huang and Lindblad 2013)

PnirA was shown to be responsive to nitrate concentration over the range 10-100 μM . However, expression reached a maximum (between 4 and 8 hours after exposure) and then receded (Ivanikova, McKay, and Bullerjahn 2005). Briggs et al. observed that copper ions at 3 μM could induce expression of the plastocyanin gene petE, but that accumulation of the corresponding transcript was independent of the copper concentration (Briggs, Pecoraro, and McIntosh 1990). Subsequent studies, such as that by Englund et al., (2016) suggested that the amount of copper present in BG11 is enough to induce expression of PpetE. PcoaT and PnrsB were characterized using bacterial luciferase. PcoaT-luxAB showed

a detection range of 0.3-6 μM for Co^{2+} and 1-3 μM for Zn^{2+} induction. PnrsB-luxAB showed a detection range of 0.2-6 μM of Ni^{2+} (Peca et al. 2008). Guerrero et al. (2012) used ethylene production as a reporter for promoter activity in *S. 6803*. Ptrc and PA1lacO-1 had high activity, PpetE and PcoaT had medium production, and Psmt and PluxRI (a quorum sensing promoter) had low activity.

Markley et al. (2015) tested two constitutive promoter libraries in *S. 7002*. One was generated by error-prone PCR of PcpCB from *S. 6803*, truncated to 89 nucleotides. The truncation reduced expression of YFP by about half, but one of the 29 library members had a similar expression level to the full PcpCB sequence. All others had reduced expression level. The second promoter set was a selection of 13 the BioBrick BBa_J23119-derived series. Addition of the lac operator to make the truncated PcpCB promoter inducible, reduced expression by 80%, co-expression of lacI further reduced expression, which was relieved by addition of 1mM IPTG. Interestingly, they found only a weak correlation in expression between *E. coli* and *S. 7002* (Markley et al. 2015). Generally, it has been difficult to improve function with random mutations. It is necessary to generate and screen a much larger library to sample more of the potential expression level space.

In another promoter library, Sengupta et al. (2019) Generated a library of 48 promoter-RBS fragments (300 bp 5' of the start codons) using error-prone PCR starting with PrbcL and PcpCB. Although developed in *S. 7942*, the promoters had activity in *S. elongatus* PCC11801 and PCC11802, though the expression profiles were somewhat different. High carbon (1% CO_2) still generally reduced expression from the PrbcL mutant promoters as it does in the native sequence. Also, high light generally reduced expression from the PcpCB mutants as it does in the wild-type promoter. Six of the mutants of PrbcL abolished measurable activity. Some of the remaining mutants increased activity while others decreased it. The variants of both PrbcL and PcbCB were also tested in *E. coli*. Expression was poorly correlated between *S. 7942* and *E. coli* for both sets. (Sengupta et al. 2019)

Englund et al. (2016) tested characterized several native promoters in *S. 6803*, with a focus on those which were expected to be inducible by metal ions (PnrsB, PnrsD, PnrsS, PcoaT, PziaA, PpetE, PrbcL1A, PpsaA, PpsbA2, PrnpB). PnrsB was tightly regulated with low expression the absence of nickel or cobalt and expression similar to that of PpsbA2 when induced by those metal ions (for a 39-fold induction rate). Expression from induced PnrsB was significantly stronger than from PrnpB or PrbcL1A. The expression from the PpsbA2 was strongly dependent on the length of the promoter sequence used, mainly because the shortest variant did not include a binding site for an anti-sense RNA that regulates expression of psbA. They also observed that expression from the PpsbA2L promoter increased at higher metal ion concentrations. PnrsB had a significantly wider dynamic range in response to metal ion concentration however, with lower metal concentrations resulting in lower expression from PnrsB (0.5-10uM). Of the *S. 6803* promoters tested, only PpsbA2 had measurable activity in *E. coli* (Englund, Liang, and Lindberg 2016).

Ferreira et al. (2018) tested PrnpB, three T7 promoter variants, psbA2, tac, four trc variants, native ggpS promoter, and five BioBrick promoters. The pSEVA251, pSEVA351, pSEVA451 backbones are variant plasmids built on SEVA architecture. All used the RSF1010 origin of replication and kanamycin-, chloramphenicol, or spectinomycin/streptomycin- resistances (respectively). In stark contrast to Huang et al. (2010), the BBa_R0051 (lambda phage P_R) promoter far exceeded the strength of the PrnpB (by ~20-fold) and the other BioBrick promoters (by ~120-fold). As a reporter, they used GFP generator BBa_E0240 which relies on the RBS BBa_B0032 for translation initiation. *S.6803* constitutively expressing the T7 RNA polymerase were also used to test four T7 promoter variants. Each of these generated expression between one quarter and one half that of the PrnpB promoter. De-repression of promoters coning lac or tet operators only resulted in about a two-fold increase in expression, continuing the challenges in developing inducible promoters in *S. 6803* with a wide dynamic range. (Ferreira et al. 2018)

Jin et al. also examined T7 promoters. T7 RNAP and T7 promoters do offer the ability to generate transcripts at least somewhat orthogonally to the native transcription machinery. Attempts to increase the expression from the T7 promoter by increasing the expression of T7 RNAP resulted in toxicity from the high level of T7 RNAP expression. The nickel-inducible *nrsB* promoter was successfully utilized to drive expression of the T7 RNAP. Mutation of just two nucleotides adjacent to the transcription start site of the T7 promoter resulted in a more than 2-fold increase in the reporter expression level (Jin, Lindblad, and Bhaya 2019).

Liu and Pakrasi (2018) tested twelve promoters copied from the *S. 6803* genome and the *P_{trc10}* promoter resulting in an 8000-fold range of expression of eYFP. The *cpcB* promoter resulted in the highest expression followed by *P_{trc10}*. The promoters tested have different lengths and the authors did not explain how they selected the 5' end of the promoters. Each one did contain the RBS inherited from the promoter (Liu and Pakrasi 2018). Potential interactions between the RBS and the coding sequence in the mRNA may impact translation initiation rates. Therefore, if another gene of interest is expressed using this promoter set, a different expression profile may be observed.

Wang et al. (2018) discovered one of the strongest cyanobacteria promoter, which happened to be a combination of *P_{petE}* from *S. 6803* and *P_{psbA}* from *Amaranthus hybridus*, and was designated *P_{psbA*}*. They measured 7-fold higher ethylene production from a strain that used this promoter to express ethylene forming enzyme (*efe*) than a strain that used the strong promoter *P_{cpc560}*. Unlike most other studies of the relative expression of different promoters, this one utilized the same RBS (“RBSv4”) for each promoter. This should be expected to better isolate transcription initiation as the rate being measured (Wang et al. 2018).

The other “super strong” promoter for expression in *S. 6803* is *P_{cpc560}* found by Zhou et al. (2014). This promoter consists of the 560 nucleotides upstream of the start codon of the native *cpcB* gene.

Expression of two heterologous genes (separately) using this promoter reached approximately 15% of total soluble protein. The strength of this promoter is due to the fourteen transcription factor binding sites, and possibly also due to it containing two separate transcription start sites (Zhou et al. 2014).

Ribosome binding sites

Heidorn et al. (2011) compared expression of GFPmut3B in *E. coli* and *S. 6803* using BioBrick RBSs B0030, B0032, and B0034 as well as RBS* which contains the Shine-Dalgarno sequence with the *S. 6803* consensus spacing from the start codon. RBS* resulted in the highest expression in *S. 6803* – about 2-fold higher than B0030 and about 4-fold higher than B0032 and B0034. In *E. coli*, fluorescence measurements were generally an order of magnitude higher, and B0034 and B0030 had the highest expression, followed by RBS* and, at about 1/6 the expression of B0034, B0032. In other words, the expression profiles for the four RBSs were significantly different between the two organisms. (Heidorn et al. 2011)

Markley et al. (2015) generated 11 RBS variants in *S. 7002*, focusing on mutations within the Shine-Dalgarno sequence. The set of RBSs generated about a 30-fold range of expression, and measured expression had a weak correlation to the predictions made by the RBS Calculator. Generally, the transcript abundance for RBSs had low variability except for one which had about three times the abundance of the others. This RBS, with SD sequence 'AGGAGA' also resulted in the highest expression of the YFP reporter gene (Markley et al. 2015). The transcript abundance may have been higher due to a higher translation initiation rate which caused higher ribosome occupancy of the mRNA which blocks degradation by RNases.

Englund et al. (2016) tested 11 RBS sequences in *S. 6803* including eight BioBrick parts, the RBS* previously tested by the Lindblad Lab, and the native RBSs from the *psbA2* and *rbcL* genes. In order to assess the context dependence of the measured translation initiation rates of the RBSs, both EYFP with

the PpetE promoter and mTagBFP with the PpsbA2S promoter were combined with each of the RBSs. The expression profiles of the two reporters for this RBS set were somewhat different. Six of the eleven had similar relative expression for each reporter, while the other five showed different expression levels. The mTagBFP set of constructs was also tested in *E. coli*. This resulted in a similar expression profile to mTagBFP tested in *S. 6803*, except that the fluorescence divided by the optical density was generally about an order of magnitude higher in *E. coli* (Englund, Liang, and Lindberg 2016).

Thiel et al. (2018) also examined the impact that the reporter gene coding sequence may have on the expression profile of a set of ribosome binding sites. They measured the GFP and YFP expression when driven by 13 different ribosome binding sites including RBS*, the native RBSs for *cpcB*, *psbA2* and *rbcL*, the RBS from *S. 7002 cpcB*, the RBS inherited from the *lac* promoter, and six selected from those designed by the RBS Calculator (Salis, Mirsky, and Voigt 2009). The expression profile was similar but varied significantly for a few RBSs, and the overall Spearman rank correlation coefficient was just 0.543. This is remarkable because the 5' region of the GFP and YFP coding sequences used only differ by codon usage, and have similar nucleotide sequences (Thiel et al. 2018).

Liu and Pakrasi (2018) swapped the 22 nucleotides 5' of twenty native start codons in *S. 6803* for the RBS in between *P_{trc10}* and *eYFP*. Only the constructs with RBS sequences from *ndhJ* and *psaF* exceeded the original RBS contained within the *P_{trc10}* promoter. Nine of the RBS sequences resulted in expression below the limit of detection. (Liu and Pakrasi 2018)

Wang et al. (2018) found that an *EcoRV* cloning scar between the *PpsbA2** promoter and RBS dramatically decreased ethylene production compared with a strain that lacked the scar. They completed rational design of the RBS sequence to vary from RBSv4 to generate 14 total RBSs. Two of the RBS designs increased expression by about 2.5-fold. The highest expressing strain (v33) reached 12.6%

of total soluble protein (Wang et al. 2018). This is too small of a set of sequences to draw many overarching conclusions about how to design RBSs.

Several tools are available to study cyanobacteria at the system level. Widely used methods to measure the abundance of all mRNA, proteins, and metabolites as well as the metabolic fluxes through many reactions have also been applied to cyanobacteria. In addition to such measurements, several metabolic models that describe the flux through all reactions in cells have been developed. These system-wide measurements and predictions promise to improve understanding of cell function and the adjustments that cells make in response to perturbations such as gene knockouts, knockdowns, and overexpression or expression of heterologous enzymes.

Previous work in the Peebles Lab

Two observations from previous graduate students in the lab suggested that our projects to improve biofuel production from *S. 6803* were being stymied by poor control over the expression of heterologous enzymes. Yi Ern Cheah attempted to express the *TesA* gene from *E. coli* in *S. 6803*, but was unable to detect the protein. He did, however, detect the mRNA for that gene. This suggested that there was a problem with translation of *TesA*. Secondly, our initial attempt to produce bisabolene in *S. 6803* (by Stevan Albers) failed to produce measurable bisabolene. The plasmid used for this relied on the *Ptic2op* promoter and the bisabolene synthase gene from the grand fir tree which was codon optimized for expression in *S. 6803*. This prototype strain failed and suggested to us that the enzyme was not being expressed at a sufficient level to generate bisabolene.

Significance of this work

This dissertation addresses several aspects of metabolic engineering of cyanobacteria. Out of the many problems that need to be addressed in this field, I focused on the precise control over the expression of proteins in *S. 6803*. Although many projects have succeeded in expressing heterologous proteins in

cyanobacteria, the expression level for a given promoter-ribosome binding site-coding sequence combination has been unpredictable, and often very low, hindering the function of engineered strains. Usually, this means a low titer of the desired product. This thesis focuses specifically on improving translation initiation control.

First, I present methods for metabolic engineering in cyanobacteria including cultivation, plasmid design, transformation, western blotting, and selection of fluorescent proteins for reporter genes. These methods are fundamental to engineering cyanobacteria. Much of this chapter was co-written with Dr. Allison Werner and Dr. Christie Peebles. It was written to help new researchers join the cyanobacteria synthetic biology and metabolic engineering community. An example genetic modification workflow is provided, taking the reader from gene cloning, through transformation of *S. 6803*, finishing with measurements of GFP.

Second, I report on the use of codon optimization and ribosome binding site design for improving expression of heterologous genes. Specifically, bisabolene synthase was expressed in 19 different constructs, one of which increased the titer of bisabolene achieved by nearly 10-fold over a previously report. We concluded that many ribosome binding sites need to be tested to achieve a desired expression level. The simple codon adaptation index (CAI) correlated well with the bisabolene titer and bisabolene synthase expression level of the five different codon optimizations, and we determined that the codon optimizations completed by large commercial gene synthesis companies generally worked reasonably well. To measure the impact of light-dark cycles, one important aspect of scale-up from bench-scale to industrial-scale, production of bisabolene was measured from cultures grown in simulated outdoor light conditions.

Third, I review the literature which examines translation initiation in cyanobacteria. This was motivated by our observations that genetic parts that work well in *E. coli* don't always work well in cyanobacteria,

leading us to question how well translation initiation mechanisms are conserved between them. For example, what explains why, in *S. 6803*, only 26% of genes have a Shine-Dalgarno sequence in the 5'-untranslated region? Several mechanisms independent of the Shine-Dalgarno sequence interaction between mRNA and 16S rRNA are found in cyanobacteria. However, further research is needed to determine the significance of these alternative mechanisms including how they impact translation initiation rates and how ribosome binding sites should be designed in this context.

Fourth, we observed that coding sequences can interact with ribosome binding sites to impact translation initiation rates. We replicated the work of Thiel et al. (2018), measuring the expression profiles of either yellow fluorescent protein (YFP) or green fluorescent protein (GFP) from 13 ribosome binding sites including native *S.6803* sequences and sequences designed by the RBS Calculator. Significant differences in the expression profiles suggest that ribosome binding sites are not modular as currently design. This represents a significant problem in synthetic biology and metabolic engineering, because the uncertainty in expression levels for a given RBS-coding sequence combination necessitates the testing of several combinations to achieve a desired expression level. Two methods for reducing the impact of the coding sequence on the translation initiation rates were tested in both *E. coli* and *S. 6803*. Insertion of 21 nucleotide leader sequence was intended to simply keep the coding sequence context constant for different genes. However, this failed to improve the correlation between GFP and YFP in either organism. Bicistronic designs dramatically improved the correlation in *E. coli*, but did not in *S. 6803*.

Finally, the findings reported are summarized in the Discussion. I synthesize overarching conclusions drawn from those findings. Several different lines of future research are suggested by this dissertation. Perhaps the highest impact of these would be the development of a cell-free transcription-translation system using *S. 6803* lysate. Such a system would dramatically accelerate testing genetic parts for use in

cyanobacteria and facilitate furthering our understanding of translation initiation and other mechanisms in cyanobacteria.

References

- Abe, Koichi, Kotone Miyake, Mayumi Nakamura, Katsuhiko Kojima, Stefano Ferri, Kazunori Ikebukuro, and Koji Sode. 2014. "Engineering of a Green-Light Inducible Gene Expression System in *Synechocystis* Sp. PCC6803: Engineered Green Light Gene Expression System." *Microbial Biotechnology* 7 (2): 177–83. <https://doi.org/10.1111/1751-7915.12098>.
- Albers, Stevan C, Victor A Gallegos, and Christie A M Peebles. 2015a. "Engineering of Genetic Control Tools in *Synechocystis* Sp. PCC 6803 Using Rational Design Techniques." *Journal of Biotechnology* 216 (December): 36–46. <http://dx.doi.org/10.1016/j.jbiotec.2015.09.042>.
- Albers, Stevan C., Victor A. Gallegos, and Christie A.M. Peebles. 2015b. "Engineering of Genetic Control Tools in *Synechocystis* Sp. PCC 6803 Using Rational Design Techniques." *Journal of Biotechnology* 216 (December): 36–46. <https://doi.org/10.1016/j.jbiotec.2015.09.042>.
- Albers, Stevan C, and Christie A M Peebles. 2016. "Evaluating Light-Induced Promoters for the Control of Heterologous Gene Expression in *Synechocystis* Sp. PCC 6803." *Biotechnology Progress*, November, n/a-n/a. <https://doi.org/10.1002/btpr.2396>.
- Angermayr, S Andreas, Alex Gorchs Rovira, and Klaas J Hellingwerf. 2015. "Metabolic Engineering of Cyanobacteria for the Synthesis of Commodity Products." *Trends in Biotechnology* 33 (6): 352–61. <http://dx.doi.org/10.1016/j.tibtech.2015.03.009>.
- Balibar, Carl J., Xiaoyu Shen, and Jianshi Tao. 2009. "The Mevalonate Pathway of *Staphylococcus Aureus*." *Journal of Bacteriology* 191 (3): 851–61. <https://doi.org/10.1128/JB.01357-08>.
- Beck, Christian, Stefanie Hertel, Anne Rediger, Robert Lehmann, Anika Wiegand, Adrian Kölsch, Beate Heilmann, Jens Georg, Wolfgang R. Hess, and Ilka M. Axmann. 2014. "Daily Expression Pattern of Protein-Encoding Genes and Small Noncoding RNAs in *Synechocystis* Sp. Strain PCC 6803." Edited by H. L. Drake. *Applied and Environmental Microbiology* 80 (17): 5195–5206. <https://doi.org/10.1128/AEM.01086-14>.
- Benedetti, Manuel, Valeria Vecchi, Simone Barera, and Luca Dall'Osto. 2018. "Biomass from Microalgae: The Potential of Domestication towards Sustainable Biofactories." *Microbial Cell Factories* 17 (1). <https://doi.org/10.1186/s12934-018-1019-3>.
- Bentley, Fiona K., Andreas Zurbruggen, and Anastasios Melis. 2014. "Heterologous Expression of the Mevalonic Acid Pathway in Cyanobacteria Enhances Endogenous Carbon Partitioning to Isoprene." *Molecular Plant* 7 (1): 71–86. <https://doi.org/10.1093/mp/sst134>.
- Betterle, Nico, and Anastasios Melis. 2018. "Heterologous Leader Sequences in Fusion Constructs Enhance Expression of Geranyl Diphosphate Synthase and Yield of β -Phellandrene Production in Cyanobacteria (*Synechocystis*)." *ACS Synthetic Biology* 7 (3): 912–21. <https://doi.org/10.1021/acssynbio.7b00431>.
- Briggs, Linda M., Vincent L. Pecoraro, and Lee McIntosh. 1990. "Copper-Induced Expression, Cloning, and Regulatory Studies of the Plastocyanin Gene from the Cyanobacterium *Synechocystis* Sp. PCC 6803." *Plant Molecular Biology* 15 (4): 633–42. <https://doi.org/10.1007/BF00017837>.
- Burgard, Anthony P., Priti Pharkya, and Costas D. Maranas. 2003. "Optknock: A Bilevel Programming Framework for Identifying Gene Knockout Strategies for Microbial Strain Optimization." *Biotechnology and Bioengineering* 84 (6): 647–57. <https://doi.org/10.1002/bit.10803>.
- Callahan, S. M., and W. J. Buikema. 2001. "The Role of HetN in Maintenance of the Heterocyst Pattern in *Anabaena* Sp. PCC 7120." *Molecular Microbiology* 40 (4): 941–50. <https://doi.org/10.1046/j.1365-2958.2001.02437.x>.
- Camsund, Daniel, Thorsten Heidorn, and Peter Lindblad. 2014. "Design and Analysis of LacI-Repressed Promoters and DNA-Looping in a Cyanobacterium." *Journal of Biological Engineering* 8 (1): 4. <https://doi.org/10.1186/1754-1611-8-4>.

- Camsund, Daniel, and Peter Lindblad. 2014. "Engineered Transcriptional Systems for Cyanobacterial Biotechnology." *Frontiers in Bioengineering and Biotechnology*.
- Chae, Tong Un, So Young Choi, Je Woong Kim, Yoo-Sung Ko, and Sang Yup Lee. 2017. "Recent Advances in Systems Metabolic Engineering Tools and Strategies." *Current Opinion in Biotechnology* 47 (October): 67–82. <https://doi.org/10.1016/j.copbio.2017.06.007>.
- Chaves, Julie E., Paloma Rueda Romero, Henning Kirst, and Anastasios Melis. 2016. "Role of Isopentenyl-Diphosphate Isomerase in Heterologous Cyanobacterial (*Synechocystis*) Isoprene Production." *Photosynthesis Research*, July. <https://doi.org/10.1007/s11120-016-0293-3>.
- Cheah, Yi Ern, Stevan C Albers, and Christie A M Peebles. 2013. "A Novel Counter-Selection Method for Markerless Genetic Modification in *Synechocystis* Sp. PCC 6803." *Biotechnology Progress* 29 (1): 23–30. <https://doi.org/10.1002/btpr.1661>.
- Choi, Sun Young, Hyun Jeong Lee, Jaeyeon Choi, Jiye Kim, Sang Jun Sim, Youngsoo Um, Yunje Kim, Taek Soon Lee, Jay D. Keasling, and Han Min Woo. 2016. "Photosynthetic Conversion of CO₂ to Farnesyl Diphosphate-Derived Phytochemicals (Amorpha-4,11-Diene and Squalene) by Engineered Cyanobacteria." *Biotechnology for Biofuels* 9 (1). <https://doi.org/10.1186/s13068-016-0617-8>.
- Choi, Sun Young, Jin-Young Wang, Ho Seok Kwak, Sun-Mi Lee, Youngsoo Um, Yunje Kim, Sang Jun Sim, Jong-il Choi, and Han Min Woo. 2017. "Improvement of Squalene Production from CO₂ in *Synechococcus Elongatus* PCC 7942 by Metabolic Engineering and Scalable Production in a Photobioreactor." *ACS Synthetic Biology* 6 (7): 1289–95. <https://doi.org/10.1021/acssynbio.7b00083>.
- Cui, Jinyu, Tao Sun, Lei Chen, and Weiwen Zhang. 2020. "Engineering Salt Tolerance of Photosynthetic Cyanobacteria for Seawater Utilization." *Biotechnology Advances* 43 (November): 107578. <https://doi.org/10.1016/j.biotechadv.2020.107578>.
- Davies, Fiona K., Victoria H. Work, Alexander S. Beliaev, and Matthew C. Posewitz. 2014. "Engineering Limonene and Bisabolene Production in Wild Type and a Glycogen-Deficient Mutant of *Synechococcus* Sp. PCC 7002." *Frontiers in Bioengineering and Biotechnology* 2 (June). <https://doi.org/10.3389/fbioe.2014.00021>.
- Dienst, Dennis, Julian Wichmann, Oliver Mantovani, João S. Rodrigues, and Pia Lindberg. 2020. "High Density Cultivation for Efficient Sesquiterpenoid Biosynthesis in *Synechocystis* Sp. PCC 6803." *Scientific Reports* 10 (1). <https://doi.org/10.1038/s41598-020-62681-w>.
- Dueber, John E, Gabriel C Wu, G Reza Malmirchegini, Tae Seok Moon, Christopher J Petzold, Adeeti V Ullal, Kristala L J Prather, and Jay D Keasling. 2009. "Synthetic Protein Scaffolds Provide Modular Control over Metabolic Flux." *Nature Biotechnology* 27 (8): 753–59. <https://doi.org/10.1038/nbt.1557>.
- Elhai, Jeff, and C.Peter Wolk. 1988. "[83] Conjugal Transfer of DNA to Cyanobacteria." In *Methods in Enzymology*, 167:747–54. Elsevier. [https://doi.org/10.1016/0076-6879\(88\)67086-8](https://doi.org/10.1016/0076-6879(88)67086-8).
- Engler, Carola, and Sylvestre Marillonnet. 2014. "Golden Gate Cloning." In *DNA Cloning and Assembly Methods*, edited by Svein Valla and Rahmi Lale, 119–31. Totowa, NJ: Humana Press. https://doi.org/10.1007/978-1-62703-764-8_9.
- Englund, Elias, Johan Andersen-Ranberg, Rui Miao, Björn Hamberger, and Pia Lindberg. 2015. "Metabolic Engineering of *Synechocystis* Sp. PCC 6803 for Production of the Plant Diterpenoid Manoyl Oxide." *ACS Synthetic Biology* 4 (12): 1270–78. <https://doi.org/10.1021/acssynbio.5b00070>.
- Englund, Elias, Feiyan Liang, and Pia Lindberg. 2016. "Evaluation of Promoters and Ribosome Binding Sites for Biotechnological Applications in the Unicellular Cyanobacterium *Synechocystis* Sp. PCC 6803." *Scientific Reports* 6 (November): 36640. <https://doi.org/10.1038/srep36640>.
- Englund, Elias, Kiyan Shabestary, Elton P. Hudson, and Pia Lindberg. 2018. "Systematic Overexpression Study to Find Target Enzymes Enhancing Production of Terpenes in *Synechocystis* PCC 6803,

- Using Isoprene as a Model Compound." *Metabolic Engineering* 49 (September): 164–77. <https://doi.org/10.1016/j.ymben.2018.07.004>.
- Eriksson, Jan, Gaza F. Salih, Haile Ghebramedhin, and Christer Jansson. 2000. "Deletion Mutagenesis of the 5' PsbA2 Region in *Synechocystis* 6803: Identification of a Putative Cis Element Involved in Photoregulation." *Molecular Cell Biology Research Communications* 3 (5): 292–98. <https://doi.org/10.1006/mcbr.2000.0227>.
- Feng, Yuchi, R. Marc L. Morgan, Paul D. Fraser, Klaus Hellgardt, and Peter J. Nixon. 2020. "Crystal Structure of Geranylgeranyl Pyrophosphate Synthase (CrtE) Involved in Cyanobacterial Terpenoid Biosynthesis." *Frontiers in Plant Science* 11: 589. <https://doi.org/10.3389/fpls.2020.00589>.
- Ferreira, Eunice A, Catarina C Pacheco, Filipe Pinto, José Pereira, Pedro Lamosa, Paulo Oliveira, Boris Kirov, Alfonso Jaramillo, and Paula Tamagnini. 2018. "Expanding the Toolbox for *Synechocystis* Sp. PCC 6803: Validation of Replicative Vectors and Characterization of a Novel Set of Promoters." *Synthetic Biology* 3 (1). <https://doi.org/10.1093/synbio/ysy014>.
- Formighieri, Cinzia, and Anastasios Melis. 2015. "A Phycocyanin·phellandrene Synthase Fusion Enhances Recombinant Protein Expression and β -Phellandrene (Monoterpene) Hydrocarbons Production in *Synechocystis* (Cyanobacteria)." *Metabolic Engineering* 32 (November): 116–24. <https://doi.org/10.1016/j.ymben.2015.09.010>.
- Formighieri, Cinzia, and Anastasios Melis. 2016. "Sustainable Heterologous Production of Terpene Hydrocarbons in Cyanobacteria." *Photosynthesis Research* 130 (1–3): 123–35. <https://doi.org/10.1007/s11120-016-0233-2>.
- Formighieri, Cinzia, and Anastasios Melis. 2017. "Heterologous Synthesis of Geranylinalool, a Diterpenol Plant Product, in the Cyanobacterium *Synechocystis*." *Applied Microbiology and Biotechnology* 101 (7): 2791–2800. <https://doi.org/10.1007/s00253-016-8081-8>.
- Gale, Grant A. R., Alejandra A. Schiavon Osorio, Lauren A. Mills, Baojun Wang, David J. Lea-Smith, and Alistair J. McCormick. 2019. "Emerging Species and Genome Editing Tools: Future Prospects in Cyanobacterial Synthetic Biology." *Microorganisms* 7 (10): 409. <https://doi.org/10.3390/microorganisms7100409>.
- Gao, Xiang, Fang Gao, Deng Liu, Hao Zhang, Xiaoqun Nie, and Chen Yang. 2016. "Engineering the Methylerythritol Phosphate Pathway in Cyanobacteria for Photosynthetic Isoprene Production from CO₂." *Energy & Environmental Science* 9 (4): 1400–1411. <https://doi.org/10.1039/C5EE03102H>.
- Gibson, Daniel G, Lei Young, Ray-Yuan Chuang, J Craig Venter, Clyde A Hutchison, and Hamilton O Smith. 2009. "Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases." *Nat Meth* 6 (5): 343–45.
- Gordon, Gina C., Travis C. Korosh, Jeffrey C. Cameron, Andrew L. Markley, Matthew B. Begemann, and Brian F. Pfeleger. 2016. "CRISPR Interference as a Titratable, Trans-Acting Regulatory Tool for Metabolic Engineering in the Cyanobacterium *Synechococcus* Sp. Strain PCC 7002." *Metabolic Engineering* 38 (November): 170–79. <https://doi.org/10.1016/j.ymben.2016.07.007>.
- Griese, Marco, Christian Lange, and Jörg Soppa. 2011. "Ploidy in Cyanobacteria." *FEMS Microbiology Letters* 323 (2): 124–31. <https://doi.org/10.1111/j.1574-6968.2011.02368.x>.
- Guerrero, Fernando, Verónica Carbonell, Matteo Cossu, Danilo Correddu, and Patrik R. Jones. 2012. "Ethylene Synthesis and Regulated Expression of Recombinant Protein in *Synechocystis* Sp. PCC 6803." Edited by Brett Neilan. *PLoS ONE* 7 (11): e50470. <https://doi.org/10.1371/journal.pone.0050470>.
- Gupta, Apoorv, Irene M Brockman Reizman, Christopher R Reisch, and Kristala L J Prather. 2017. "Dynamic Regulation of Metabolic Flux in Engineered Bacteria Using a Pathway-Independent

- Quorum-Sensing Circuit." *Nature Biotechnology* 35 (3): 273–79.
<https://doi.org/10.1038/nbt.3796>.
- Halfmann, Charles, Liping Gu, William Gibbons, and Ruanbao Zhou. 2014a. "Genetically Engineering Cyanobacteria to Convert CO₂, Water, and Light into the Long-Chain Hydrocarbon Farnesene." *Applied Microbiology and Biotechnology* 98 (23): 9869–77. <https://doi.org/10.1007/s00253-014-6118-4>.
- Halfmann, Charles, Liping Gu, and Ruanbao Zhou. 2014b. "Engineering Cyanobacteria for the Production of a Cyclic Hydrocarbon Fuel from CO₂ and H₂O." *Green Chem.* 16 (6): 3175–85.
<https://doi.org/10.1039/C3GC42591F>.
- Heidorn, Thorsten, Daniel Camsund, Hsin-Ho Huang, Pia Lindberg, Paulo Oliveira, Karin Stensjö, and Peter Lindblad. 2011. "Synthetic Biology in Cyanobacteria." In *Methods in Enzymology*, 497:539–79. Elsevier. <http://linkinghub.elsevier.com/retrieve/pii/B978012385075100024X>.
- Huang, Chun-Hung, Claire R. Shen, Hung Li, Li-Yu Sung, Meng-Ying Wu, and Yu-Chen Hu. 2016. "CRISPR Interference (CRISPRi) for Gene Regulation and Succinate Production in Cyanobacterium *S. Elongatus* PCC 7942." *Microbial Cell Factories* 15 (1). <https://doi.org/10.1186/s12934-016-0595-3>.
- Huang, H H, D Camsund, P Lindblad, and T Heidorn. 2010a. "Design and Characterization of Molecular Tools for a Synthetic Biology Approach towards Developing Cyanobacterial Biotechnology." *Nucleic Acids Res* 38. <https://doi.org/10.1093/nar/gkq164>.
- Huang, H. H., D. Camsund, P. Lindblad, and T. Heidorn. 2010b. "Design and Characterization of Molecular Tools for a Synthetic Biology Approach towards Developing Cyanobacterial Biotechnology." *Nucleic Acids Research* 38 (8): 2577–93. <https://doi.org/10.1093/nar/gkq164>.
- Huang, Hsin-Ho, and Peter Lindblad. 2013. "Wide-Dynamic-Range Promoters Engineered for Cyanobacteria." *Journal of Biological Engineering* 7 (1): 10. <https://doi.org/10.1186/1754-1611-7-10>.
- Imamura, Sousuke, Munehiko Asayama, and Makoto Shirai. 2004. "In Vitro Transcription Analysis by Reconstituted Cyanobacterial RNA Polymerase: Roles of Group 1 and 2 Sigma Factors and a Core Subunit, RpoC2: Roles of Cyanobacterial Sigma Factors & RpoC2." *Genes to Cells* 9 (12): 1175–87. <https://doi.org/10.1111/j.1365-2443.2004.00808.x>.
- Ivanikova, Natalia V., R. Michael L. McKay, and George S. Bullerjahn. 2005. "Construction and Characterization of a Cyanobacterial Bioreporter Capable of Assessing Nitrate Assimilatory Capacity in Freshwaters: Cyanobacterial Nitrate Bioreporter." *Limnology and Oceanography: Methods* 3 (2): 86–93. <https://doi.org/10.4319/lom.2005.3.86>.
- Jin, Haojie, Peter Lindblad, and Devaki Bhaya. 2019. "Building an Inducible T7 RNA Polymerase/T7 Promoter Circuit in *Synechocystis* Sp. PCC6803." *ACS Synthetic Biology* 8 (4): 655–60. <https://doi.org/10.1021/acssynbio.8b00515>.
- Kaczmarzyk, Danuta, Ivana Cengic, Lun Yao, and Elton P. Hudson. 2018. "Diversion of the Long-Chain Acyl-ACP Pool in *Synechocystis* to Fatty Alcohols through CRISPRi Repression of the Essential Phosphate Acyltransferase PlsX." *Metabolic Engineering* 45 (January): 59–66. <https://doi.org/10.1016/j.ymben.2017.11.014>.
- Kaneko, Takakazu, Shusei Sato, Hirokazu Kotani, Ayako Tanaka, Erika Asamizu, Yasukazu Nakamura, Nobuyuki Miyajima, et al. 1996. "Sequence Analysis of the Genome of the Unicellular Cyanobacterium *Synechocystis* Sp. Strain PCC6803. II. Sequence Determination of the Entire Genome and Assignment of Potential Protein-Coding Regions." *DNA Research* 3 (3): 109–36. <https://doi.org/10.1093/dnares/3.3.109>.
- Kim, Wook Jin, Sun-Mi Lee, Youngsoon Um, Sang Jun Sim, and Han Min Woo. 2017. "Development of SyneBrick Vectors As a Synthetic Biology Platform for Gene Expression in *Synechococcus Elongatus* PCC 7942." *Frontiers in Plant Science* 8: 293. <https://doi.org/10.3389/fpls.2017.00293>.

- Kiyota, Hiroshi, Yukiko Okuda, Michiho Ito, Masami Yokota Hirai, and Masahiko Ikeuchi. 2014. "Engineering of Cyanobacteria for the Photosynthetic Production of Limonene from CO₂." *Journal of Biotechnology* 185 (September): 1–7. <https://doi.org/10.1016/j.jbiotec.2014.05.025>.
- Kleinegris, Dorinde M. M., Marjon A. van Es, Marcel Janssen, Willem A. Brandenburg, and René H. Wijffels. 2011. "Phase Toxicity of Dodecane on the Microalga *Dunaliella Salina*." *Journal of Applied Phycology* 23 (6): 949–58. <https://doi.org/10.1007/s10811-010-9615-6>.
- Kleinegris, Dorinde M.M., Marcel Janssen, Willem A. Brandenburg, and René H. Wijffels. 2011. "Two-Phase Systems: Potential for in Situ Extraction of Microalgal Products." *Biotechnology Advances* 29 (5): 502–7. <https://doi.org/10.1016/j.biotechadv.2011.05.018>.
- Kok, Stefan de, Leslie H Stanton, Todd Slaby, Maxime Durot, Victor F Holmes, Kedar G Patel, Darren Platt, et al. 2014. "Rapid and Reliable DNA Assembly via Ligase Cycling Reaction." *ACS Synthetic Biology* 3 (2): 97–106. <https://doi.org/10.1021/sb4001992>.
- Lee, Hyun Jeong, Jiwon Lee, Sun-Mi Lee, Youngsoon Um, Yunje Kim, Sang Jun Sim, Jong-Il Choi, and Han Min Woo. 2017. "Direct Conversion of CO₂ to α -Farnesene Using Metabolically Engineered *Synechococcus Elongatus* PCC 7942." *Journal of Agricultural and Food Chemistry* 65 (48): 10424–28. <https://doi.org/10.1021/acs.jafc.7b03625>.
- Lin, Po-Cheng, Rajib Saha, Fuzhong Zhang, and Himadri B. Pakrasi. 2017. "Metabolic Engineering of the Pentose Phosphate Pathway for Enhanced Limonene Production in the Cyanobacterium *Synechocystis* Sp. PCC 6803." *Scientific Reports* 7 (1). <https://doi.org/10.1038/s41598-017-17831-y>.
- Lin, Po-Cheng, and Himadri B. Pakrasi. 2019. "Engineering Cyanobacteria for Production of Terpenoids." *Planta* 249 (1): 145–54. <https://doi.org/10.1007/s00425-018-3047-y>.
- Lindberg, Pia, Sungsoon Park, and Anastasios Melis. 2010. "Engineering a Platform for Photosynthetic Isoprene Production in Cyanobacteria, Using *Synechocystis* as the Model Organism." *Metabolic Engineering* 12 (1): 70–79. <https://doi.org/10.1016/j.ymben.2009.10.001>.
- Liu, Deng, Virginia M. Johnson, and Himadri B. Pakrasi. 2020. "A Reversibly Induced CRISPRi System Targeting Photosystem II in the Cyanobacterium *Synechocystis* Sp. PCC 6803." *ACS Synthetic Biology* 9 (6): 1441–49. <https://doi.org/10.1021/acssynbio.0c00106>.
- Liu, Deng, and Himadri B. Pakrasi. 2018. "Exploring Native Genetic Elements as Plug-in Tools for Synthetic Biology in the Cyanobacterium *Synechocystis* Sp. PCC 6803." *Microbial Cell Factories* 17 (1). <https://doi.org/10.1186/s12934-018-0897-8>.
- Markley, Andrew L., Matthew B. Begemann, Ryan E. Clarke, Gina C. Gordon, and Brian F. Pfleger. 2015. "Synthetic Biology Toolbox for Controlling Gene Expression in the Cyanobacterium *Synechococcus* Sp. Strain PCC 7002." *ACS Synthetic Biology* 4 (5): 595–603. <https://doi.org/10.1021/sb500260k>.
- Martin, Vincent J J, Douglas J Pitera, Sydnor T Withers, Jack D Newman, and Jay D Keasling. 2003. "Engineering a Mevalonate Pathway in *Escherichia Coli* for Production of Terpenoids." *Nature Biotechnology* 21 (7): 796–802. <https://doi.org/10.1038/nbt833>.
- Máté, Z., L. Sass, M. Szekeres, I. Vass, and F. Nagy. 1998. "UV-B-Induced Differential Transcription of *PsbA* Genes Encoding the D1 Protein of Photosystem II in the Cyanobacterium *Synechocystis* 6803." *The Journal of Biological Chemistry* 273 (28): 17439–44. <https://doi.org/10.1074/jbc.273.28.17439>.
- Mazouni, K., S. Bulteau, C. Cassier-Chauvat, and F. Chauvat. 1998. "Promoter Element Spacing Controls Basal Expression and Light Inducibility of the Cyanobacterial *SecA* Gene." *Molecular Microbiology* 30 (5): 1113–22. <https://doi.org/10.1046/j.1365-2958.1998.01145.x>.
- Muramatsu, M., and Y. Hihara. 2007. "Coordinated High-Light Response of Genes Encoding Subunits of Photosystem I Is Achieved by AT-Rich Upstream Sequences in the Cyanobacterium *Synechocystis*

- Sp. Strain PCC 6803." *Journal of Bacteriology* 189 (7): 2750–58.
<https://doi.org/10.1128/JB.01903-06>.
- Nakao, Mitsuteru, Shinobu Okamoto, Mitsuyo Kohara, Tsunakazu Fujishiro, Takatomo Fujisawa, Shusei Sato, Satoshi Tabata, Takakazu Kaneko, and Yasukazu Nakamura. 2010. "CyanoBase: The Cyanobacteria Genome Database Update 2010." *Nucleic Acids Research* 38 (Database issue): D379–81. <https://doi.org/10.1093/nar/gkp915>.
- Ng, Andrew H., Bertram M. Berla, and Himadri B. Pakrasi. 2015. "Fine-Tuning of Photoautotrophic Protein Production by Combining Promoters and Neutral Sites in the Cyanobacterium *Synechocystis* Sp. Strain PCC 6803." Edited by K. E. Wommack. *Applied and Environmental Microbiology* 81 (19): 6857–63. <https://doi.org/10.1128/AEM.01349-15>.
- Nielsen, Jens, and Jay D. Keasling. 2016. "Engineering Cellular Metabolism." *Cell* 164 (6): 1185–97. <https://doi.org/10.1016/j.cell.2016.02.004>.
- Pade, Nadin, Sabrina Erdmann, Heike Enke, Frederik Dethloff, Ulf Dühning, Jens Georg, Juliane Wambutt, et al. 2016. "Insights into Isoprene Production Using the Cyanobacterium *Synechocystis* Sp. PCC 6803." *Biotechnology for Biofuels* 9 (1). <https://doi.org/10.1186/s13068-016-0503-4>.
- Peca, Loredana, Péter B. Kás, Zoltan Máté, Andrea Farsang, and Imre Vass. 2008. "Construction of Bioluminescent Cyanobacterial Reporter Strains for Detection of Nickel, Cobalt and Zinc." *FEMS Microbiology Letters* 289 (2): 258–64. <https://doi.org/10.1111/j.1574-6968.2008.01393.x>.
- Peralta-Yahya, Pamela P., Mario Ouellet, Rossana Chan, Aindrila Mukhopadhyay, Jay D. Keasling, and Taek Soon Lee. 2011. "Identification and Microbial Production of a Terpene-Based Advanced Biofuel." *Nature Communications* 2 (September): 483. <https://doi.org/10.1038/ncomms1494>.
- Pitera, Douglas J., Chris J. Paddon, Jack D. Newman, and Jay D. Keasling. 2007. "Balancing a Heterologous Mevalonate Pathway for Improved Isoprenoid Production in *Escherichia Coli*." *Metabolic Engineering* 9 (2): 193–207. <https://doi.org/10.1016/j.ymben.2006.11.002>.
- Plegaria, Jefferson S, and Cheryl A Kerfeld. 2018. "Engineering Nanoreactors Using Bacterial Microcompartment Architectures." *Current Opinion in Biotechnology* 51 (June): 1–7. <https://doi.org/10.1016/j.copbio.2017.09.005>.
- Quax, Tessa E.F., Nico J. Claassens, Dieter Söll, and John van der Oost. 2015. "Codon Bias as a Means to Fine-Tune Gene Expression." *Molecular Cell* 59 (2): 149–61. <https://doi.org/10.1016/j.molcel.2015.05.035>.
- Reed, R. H., and W. D. P. Stewart. 1985. "Osmotic Adjustment and Organic Solute Accumulation in Unicellular Cyanobacteria from Freshwater and Marine Habitats." *Marine Biology* 88 (1): 1–9. <https://doi.org/10.1007/BF00393037>.
- Salem, K., and L. G. van Waasbergen. 2004. "Light Control of HliA Transcription and Transcript Stability in the Cyanobacterium *Synechococcus Elongatus* Strain PCC 7942." *Journal of Bacteriology* 186 (6): 1729–36. <https://doi.org/10.1128/JB.186.6.1729-1736.2004>.
- Salis, Howard M. 2011. "Chapter Two - The Ribosome Binding Site Calculator." In *Synthetic Biology, Part B Computer Aided Design and DNA Assembly*, edited by Christopher Voigt B T - Methods in Enzymology, Volume 498:19–42. Academic Press. <http://dx.doi.org/10.1016/B978-0-12-385120-8.00002-4>.
- Salis, Howard M, Ethan A Mirsky, and Christopher A Voigt. 2009. "Automated Design of Synthetic Ribosome Binding Sites to Control Protein Expression." *Nature Biotechnology* 27 (10): 946–50. <https://doi.org/10.1038/nbt.1568>.
- Satoh, Soichirou, Masahiko Ikeuchi, Mamoru Mimuro, and Ayumi Tanaka. 2001. "Chlorophyll *b* Expressed in Cyanobacteria Functions as a Light-Harvesting Antenna in Photosystem I through Flexibility of the Proteins." *Journal of Biological Chemistry* 276 (6): 4293–97. <https://doi.org/10.1074/jbc.M008238200>.

- Sebesta, Jacob, and Christie AM. Peebles. 2020. "Improving Heterologous Protein Expression in *Synechocystis* Sp. PCC 6803 for Alpha-Bisabolene Production." *Metabolic Engineering Communications* 10 (June): e00117. <https://doi.org/10.1016/j.mec.2019.e00117>.
- Sengupta, Annesha, Avinash Vellore Sunder, Sujata V. Sohoni, and Pramod P. Wangikar. 2019. "Fine-Tuning Native Promoters of *Synechococcus Elongatus* PCC 7942 To Develop a Synthetic Toolbox for Heterologous Protein Expression." *ACS Synthetic Biology* 8 (5): 1219–23. <https://doi.org/10.1021/acssynbio.9b00066>.
- Sergeyenko, T. V., and D. A. Los. 2000. "Identification of Secreted Proteins of the Cyanobacterium *Synechocystis* Sp. Strain PCC 6803." *FEMS Microbiology Letters* 193 (2): 213–16. <https://doi.org/10.1111/j.1574-6968.2000.tb09426.x>.
- Stal, Lucas J. 2015. "Nitrogen Fixation in Cyanobacteria." In *ELS*, edited by John Wiley & Sons Ltd, 1–9. Chichester, UK: John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470015902.a0021159.pub2>.
- Thiel, Kati, Edita Mulaku, Hariharan Dandapani, Csaba Nagy, Eva-Mari Aro, and Pauli Kallio. 2018. "Translation Efficiency of Heterologous Proteins Is Significantly Affected by the Genetic Context of RBS Sequences in Engineered Cyanobacterium *Synechocystis* Sp. PCC 6803." *Microbial Cell Factories* 17 (1). <https://doi.org/10.1186/s12934-018-0882-2>.
- Thykaer, Jette, and Jens Nielsen. 2003. "Metabolic Engineering of Beta-Lactam Production." *Metabolic Engineering* 5 (1): 56–69. [https://doi.org/10.1016/s1096-7176\(03\)00003-x](https://doi.org/10.1016/s1096-7176(03)00003-x).
- Ungerer, Justin, and Himadri B. Pakrasi. 2016. "Cpf1 Is A Versatile Tool for CRISPR Genome Editing Across Diverse Species of Cyanobacteria." *Scientific Reports* 6 (1). <https://doi.org/10.1038/srep39681>.
- Vasudevan, Ravendran, Grant A. R. Gale, Alejandra A. Schiavon, Anton Puzorjov, John Malin, Michael D. Gillespie, Konstantinos Vavitsas, et al. 2019. "CyanoGate: A Modular Cloning Suite for Engineering Cyanobacteria Based on the Plant MoClo Syntax." *Plant Physiology* 180 (1): 39–55. <https://doi.org/10.1104/pp.18.01401>.
- Vermaas, Josh V., Gayle J. Bentley, Gregg T. Beckham, and Michael F. Crowley. 2018. "Membrane Permeability of Terpenoids Explored with Molecular Simulation." *The Journal of Physical Chemistry B* 122 (45): 10349–61. <https://doi.org/10.1021/acs.jpcc.8b08688>.
- Vickers, Claudia E., Mareike Bongers, Qing Liu, Thierry Delatte, and Harro Bouwmeester. 2014. "Metabolic Engineering of Volatile Isoprenoids in Plants and Microbes: Exploiting Volatile Isoprenoids for Biotechnology." *Plant, Cell & Environment* 37 (8): 1753–75. <https://doi.org/10.1111/pce.12316>.
- Wang, Bo, Carrie Eckert, Pin-Ching Maness, and Jianping Yu. 2018. "A Genetic Toolbox for Modulating the Expression of Heterologous Genes in the Cyanobacterium *Synechocystis* Sp. PCC 6803." *ACS Synthetic Biology* 7 (1): 276–86. <https://doi.org/10.1021/acssynbio.7b00297>.
- Wang, Bo, Jianping Yu, Weiwen Zhang, and Deirdre R. Meldrum. 2015a. "Premethylation of Foreign DNA Improves Integrative Transformation Efficiency in *Synechocystis* Sp. Strain PCC 6803." *Applied and Environmental Microbiology*, October, AEM.02575-15. <https://doi.org/10.1128/AEM.02575-15>.
- . 2015b. "Premethylation of Foreign DNA Improves Integrative Transformation Efficiency in *Synechocystis* Sp. Strain PCC 6803." Edited by M. J. Pettinari. *Applied and Environmental Microbiology* 81 (24): 8500–8506. <https://doi.org/10.1128/AEM.02575-15>.
- Wang, Xin, Wei Liu, Changpeng Xin, Yi Zheng, Yanbing Cheng, Su Sun, Runze Li, et al. 2016. "Enhanced Limonene Production in Cyanobacteria Reveals Photosynthesis Limitations." *Proceedings of the National Academy of Sciences* 113 (50): 14225–30. <https://doi.org/10.1073/pnas.1613340113>.
- Wendt, Kristen E., Justin Ungerer, Ryan E. Cobb, Huimin Zhao, and Himadri B. Pakrasi. 2016. "CRISPR/Cas9 Mediated Targeted Mutagenesis of the Fast Growing Cyanobacterium

- Synechococcus Elongatus UTEX 2973." *Microbial Cell Factories* 15 (1).
<https://doi.org/10.1186/s12934-016-0514-7>.
- Werner, Allison, Katelyn Oliver, Alexander Dylan Miller, Jacob Sebesta, and Christie A.M. Peebles. 2018. "Discovery and Characterization of Synechocystis Sp. PCC 6803 Light-Entrained Promoters in Diurnal Light:Dark Cycles." *Algal Research* 30 (March): 121–27.
<https://doi.org/10.1016/j.algal.2017.12.012>.
- Williams, John G.K. 1988. "[85] Construction of Specific Mutations in Photosystem II Photosynthetic Reaction Center by Genetic Engineering Methods in Synechocystis 6803." In *Methods in Enzymology*, 167:766–78. Elsevier.
<http://linkinghub.elsevier.com/retrieve/pii/0076687988670881>.
- Yao, Lun, Ivana Cengic, Josefine Anfelt, and Elton P. Hudson. 2016. "Multiple Gene Repression in Cyanobacteria Using CRISPRi." *ACS Synthetic Biology* 5 (3): 207–12.
<https://doi.org/10.1021/acssynbio.5b00264>.
- Yoshihara, Shizue, XiaoXing Geng, Shinobu Okamoto, Kei Yura, Takashi Murata, Mitiko Go, Masayuki Ohmori, and Masahiko Ikeuchi. 2001. "Mutational Analysis of Genes Involved in Pilus Structure, Motility and Transformation Competency in the Unicellular Motile Cyanobacterium Synechocystis Sp. PCC6803." *Plant and Cell Physiology* 42 (1): 63–73. <https://doi.org/10.1093/pcp/pce007>.
- Zang, Xiaonan, Bin Liu, Shunmei Liu, K. K. I. U. Arunakumara, and Xuecheng Zhang. 2007. "Optimum Conditions for Transformation of Synechocystis Sp. PCC 6803." *Journal of Microbiology (Seoul, Korea)* 45 (3): 241–45.
- Zess, Erin K., Matthew B. Begemann, and Brian F. Pflieger. 2016. "Construction of New Synthetic Biology Tools for the Control of Gene Expression in the Cyanobacterium *Synechococcus* Sp. Strain PCC 7002: Construction of New Synthetic Biology Tools." *Biotechnology and Bioengineering* 113 (2): 424–32. <https://doi.org/10.1002/bit.25713>.
- Zhou, Jie, Haifeng Zhang, Hengkai Meng, Yan Zhu, Guanhui Bao, Yanping Zhang, Yin Li, and Yanhe Ma. 2014. "Discovery of a Super-Strong Promoter Enables Efficient Production of Heterologous Proteins in Cyanobacteria." *Scientific Reports* 4 (March). <https://doi.org/10.1038/srep04500>.

CHAPTER 2: IMPROVING EXPRESSION OF A HETEROLOGOUS PROTEIN FOR BIOFUEL PRODUCTION IN *SYNECHOCYSTIS SP.* PCC6803¹

Summary

Cyanobacterial biofuels have the potential to reduce the cost and climate impacts of biofuel production because primary carbon fixation and conversion to fuel are completed together in the cultivation of the cyanobacteria. Cyanobacterial biofuels, therefore, do not rely on costly organic carbon feedstocks that heterotrophs require, which reduces competition for agricultural resources such as arable land and freshwater. However, the published product titer achieved for most molecules of interest using cyanobacteria lag behind what has been achieved using yeast and *Escherichia coli* (*E. coli*) cultures. In *Synechocystis* sp. PCC 6803 (*S. 6803*), we attempted to increase the product titer of the sesquiterpene, bisabolene, which may be converted to bisabolane, a possible diesel replacement. We tested 19 strains of genetically modified *S. 6803* with five different codon usage sequences of the bisabolene synthase from the grand fir tree (*Abies grandis*). At least three ribosome binding sites (most were designed using the RBS Calculator) were tested for each codon usage sequence. We also tested strains with and without the farnesyl pyrophosphate synthase gene from *E. coli*. Bisabolene titers after five days of growth in continuous light ranged from un-detected to 7.8 mg/L. Bisabolene synthase abundance was measured

¹ This chapter was published as:

Sebesta, Jacob, and Christie AM. Peebles. 2020. "Improving Heterologous Protein Expression in *Synechocystis* Sp. PCC 6803 for Alpha-Bisabolene Production." *Metabolic Engineering Communications* 10 (June): e00117. <https://doi.org/10.1016/j.mec.2019.e00117>.

and found to be well correlated with titer. Select strains were also tested in 12:12 light:dark cycles, where similar titers were reached after the same amount of light exposure time. One engineered strain was also tested in photobioreactors exposed to a simulated outdoor light pattern with maximum light intensity of $1,600 \mu\text{mol photons m}^{-2} \text{s}^{-1}$. Here, the bisabolene titer reached 22.2 mg/L after 36 days of growth. Dramatic improvements in our ability to control gene expression in cyanobacteria such as *S. 6803*, and the co-utilization of additional metabolic engineering methods, are needed in order for these titers to improve to the levels reported for engineered *E. coli*.

Background

Development of biofuel production technology offers one potential path for generating renewable energy that could reduce the concentrations of atmospheric greenhouse gases generated from human activity. Metabolic engineering of photoautotrophs such as cyanobacteria and microalgae offers the potential to design processes that directly convert sunlight into biofuel products or precursors. This avoids the necessity of growing crops for carbohydrate feedstocks required for heterotrophic cultivation as is currently implemented in the microbial conversion of maize to biofuels. Cyanobacteria and microalgae have higher areal biomass productivities than land crops and do not require arable land (Dismukes et al., 2008). A comparison of microalgal biodiesel to soybean biodiesel has shown the net energy ratio (energy consumed by all processing steps divided by the energy produced) to be more favorable in microalgal biodiesel (Batan et al., 2010).

Though cyanobacteria generally also grow more slowly than heterotrophs, some species such as *Synechococcus* UTEX 2973 approach the growth rate of *Saccharomyces cerevisiae*, with a doubling time of 1.9 hr (Yu et al., 2015). Many species of cyanobacteria are salt tolerant, allowing their growth in seawater (Pade and Hagemann, 2014), and municipal wastewater streams have also been proposed as media that provide nitrogen and phosphorous (Gonçalves et al., 2016; Hughes et al., 2018). The

advantages of these species may be eventually realized by translating what is learned from our current model cyanobacteria for use with those species. *Synechocystis* sp. PCC 6803 (*S. 6803*) is a model cyanobacteria and has been widely studied from the perspectives of metabolic engineering as well as from the perspective of photoautotroph biology. The chromosome of *S. 6803* can be easily modified using the organism's native homologous recombination mechanisms. In addition, several replicative plasmids have been used to modify *S. 6803* without modifying the chromosome (Ferreira et al., 2018; Huang et al., 2010; Jin et al., 2018; Liu and Pakrasi, 2018). The genome of *S. 6803* was sequenced in 1996 (Kaneko et al., 1996), and other genome projects listed on the CyanoBase website (<http://genome.microbedb.jp/CyanoBase>) have reached 376 cyanobacterial species.

A robust research community is engaged in developing and testing diverse genetic parts and studying the biology of cyanobacteria. Many genes from different organisms have been expressed heterologously in cyanobacteria. The genetic elements necessary for expressing these proteins, including promoters and ribosome binding sites (RBSs), have been directly adapted from use in *Escherichia coli* (*E. coli*) or have been elements copied from the cyanobacteria species itself (Huang and Lindblad, 2013; Wang et al., 2018). The RBS Calculator has also been applied for the development of those genetic parts in cyanobacteria (Markley et al., 2015). A recent review (Carroll et al., 2018) covers these topics in detail, including the advancements achieved in metabolic engineering of cyanobacteria in terms of the titers achieved for many products.

One class of molecules, terpenoids, have been targeted for production in cyanobacteria which may be utilized in industries ranging from pharmaceuticals, to commodity chemicals and fuels. One successful example of metabolic engineering in cyanobacteria is provided by Gao *et al.*, who achieved a product titer of 1.26 g/L of the five-carbon terpenoid, isoprene in *Synechococcus elongatus* by implementing many common metabolic engineering strategies in combination (Gao et al., 2016).

Similar product titers have not been achieved for more complex terpenoids. For example, the C10 monoterpene, limonene, has been produced at titers of 1 mg/L after 30 days of cultivation (Kiyota et al., 2014), and 6.7 mg/L after 7 days (Lin et al., 2017). The C15 sesquiterpene, caryophyllene, was produced at a titer of 46 µg/L after seven days (Reinsvold et al., 2011). Pattanaik and Lindberg have provided a review of terpenoid production in cyanobacteria (Pattanaik and Lindberg, 2015).

Davies *et al.* previously engineered *Synechococcus* sp. PCC 7002 to produce 0.6 mg/L bisabolene by expressing a codon optimized sequence bisabolene synthase from *Abies grandis* (*A. grandis*), using the strong, constitutive *cpcBA* promoter from *S. 6803* (Davies et al., 2014). In this work we increased bisabolene production in *S. 6803* by varying codon usage and RBS sequences to control expression of bisabolene synthase. We utilized a counterselection method (Cheah et al., 2013) and inducible promoter (Albers et al., 2015) previously developed in our lab. Five codon optimizations of the bisabolene synthase gene from *A. grandis* were compared and, for each codon optimization, three or four RBS sequences designed by the RBS Calculator (Salis et al., 2009) were utilized. The co-expression of farnesyl pyrophosphate synthase from *E. coli* was also hypothesized to increase the supply of the substrate molecule for bisabolene synthase and therefore increase the bisabolene titer. Here, we present the impact these variations in genetic sequences had on bisabolene synthase expression and on bisabolene production.

Methods

Strains and cultivation

S. 6803 seed cultures were started from freezer stocks (5 % DMSO, stored at -80 °C) and generally grown in shake flasks containing 45 mL Bg-11 media (Stanier et al., 1979) with phosphate increased to 1 mM, buffered with 10mM TES-NaOH, pH 8.0 . All strains were grown at 30 °C under fluorescent light at approximately 200 +/- 20 µmol photons/m²/s with shaking at 200 rpm. For continuous light

experiments, 0.9 mL of 1 M sodium bicarbonate was added every twelve hours of growth starting at 0 hours. For 12:12 light:dark cycles, in which the lights were turned off while shaking continued, 0.9 mL of 1 M sodium bicarbonate was added every 24 hours, at the beginning of the light period. Bisabolene synthase (and when the gene was also present, farnesyl pyrophosphate synthase) was induced by addition of IPTG to 1 mM at 12 hours after inoculation, and 9 mL of dodecane (Sigma-Aldrich) was added directly to flasks at the time of induction.

Photobioreactor experiments were completed using 3 L photobioreactors (Allen Scientific Glass, Boulder, CO, USA) containing 1.5 L BG-11 (with 1 mM phosphate) buffered with 10 mM TES-NaOH, pH 8.0. An air stream enriched to 5% CO₂ was supplied to the bottom of the reactor at a rate of 200 mL min⁻¹, following the reactor setup and operation described by Werner et al. (Werner et al., 2019). However, in this case, the cultures were not entrained to the light:dark cycle before commencing the experiment. We simulated outdoor sunlight exposure for the photobioreactors with two 4000k white LED panels (Reliance Laboratories, Port Townsend, WA, USA) programmed with a sinusoidal light intensity curve peaking at a light intensity of approximately 1600 μmol photons m⁻² s⁻¹, with 12 hours of darkness. The reactors were inoculated from shake flask cultures to an initial OD_{730 nm} of 2.0, which we found was necessary for the cultures to survive the high light intensity. A 150 mL dodecane overlay and IPTG to 1mM were added to this at the time of inoculation of the PBRs.

E. coli (DH5α) was used for plasmid construction and was grown at 37°C in lysogeny broth. Initial plasmids used for generating a markerless bisabolene synthase expressing strain were assembled using ligase cycling reactions (Kok et al., 2014) (PCR of parts using New England Biolabs Phusion DNA polymerase, ligase cycling using Ampligase from Lucigen). Bisabolene synthase codon usage variants were synthesized by GenScript ('GS') and IDT ('IDT', 'EuH', and 'HCR') or PCR amplified from a plasmid ('2.0'), and inserted into plasmids using Gibson Assembly (Gibson et al., 2009). RBS variant plasmids were generated using ligase cycling. Correct colonies were screened *via* colony PCR, and then sequenced

(GeneWiz or QuintaraBio). Transformation of WT S. 6803 was completed as previously described by Cheah *et al.* (2013). Colonies were screened by colony PCR and sequenced confirmed by one Sanger sequencing reaction originating in the upstream homologous region, and covering the promoter, RBS, and 5' region of the bisabolene synthase gene.

Bisabolene measurement (GC-MS)

Samples of dodecane were taken directly from dodecane layers of shake flask and photobioreactor cultures to measure the bisabolene concentration. These samples were analyzed on an Agilent 6893 GC equipped with an Agilent 5973N mass spectrometer. A 50 m Agilent HP-5 column was used to separate hydrophobic molecules. Oven temperature was ramped from 180 °C to 260 °C at a rate of 15 °C/min. Bisabolene quantities were calculated based on a standard curve generated using serial dilutions of bisabolene (mixture of isomers, AlfaAesar). The software, Automated Mass spectral Deconvolution and Identification System (AMDIS), was used to analyze the spectral data and identify bisabolene (chemdata.nist.gov).

Protein abundance (Western blot)

Cultures of each strain were grown for 48 hours, starting at OD₇₃₀ = 0.05, with induction by 1 mM IPTG added at 12 hrs, and no dodecane added. Cells were harvested by centrifugation, frozen in liquid nitrogen, and stored at -80 °C for less than three weeks. Frozen cell pellets were thawed on ice and resuspended with 0.5 mL ice cold PBS, pH 7.4. These were centrifuged and the supernatant fully removed before being resuspended in 500 µL lysis buffer consisting of 1x PBS, pH 7.4, 0.1% Triton X-100, 1 mM DTT, and 1x HALT protease inhibitor (ThermoScientific). Cells were lysed by sonication (Misonix model S-4000, with a microtip at 45% power, 3s on/3s off for 2 minutes processing time), and the cell debris removed by centrifugation (12,000 x g for 10 minutes). The total protein concentration in the lysate was measured using a Pierce BCA Protein Assay kit and a BMG FLUOstar Omega microplate

reader. For each bioreplicate, 25 µg of total protein was loaded into a lane of a BioRad MiniPROTEAN TGX pre-cast gel. A relative standard curve was included on each gel which consisted of the pooled lysate from three 50 mL flasks of the 2.0-10xB strain grown using the same methods as used to obtain protein samples from the other strains. Four different amounts of this lysate were loaded in four lanes of each gel to generate a relative response curve. Proteins were transferred from the gel to PVDF membrane at 100 V for 100 minutes on ice in 1x TRIS/Glycine/SDS, and 20% methanol. Membranes were blocked overnight at 4 °C in PBS, 0.05% Tween20, 5 mM EDTA, and 5% nonfat milk. Mouse anti-histag antibody (ThermoFisher Scientific) was diluted 1:2,500 in 1x PBS, 0.05% Tween20, 0.5% nonfat milk, and 5 mM EDTA, and the membrane was incubated in the primary antibody solution with gentle shaking for 2 hours at room temperature. The membranes were then washed in PBS with 0.05% Tween20 once for 15 minutes and then twice for 5 minutes each before being incubated in a 1:25,000 dilution of goat anti-mouse antibody conjugated to horseradish peroxidase (ThermoFisher Scientific) (in 1x PBS, 0.05% Tween20, and 0.5% nonfat milk) for 2 hours with gentle shaking at room temperature. Again, the membranes were then washed in PBS 0.05% Tween20 once for 15 minutes and then twice for 5 minutes each before being incubated at room temperature for 5 minutes with 500 µL of SuperSignal West Femto Maximum Sensitivity ECL substrate (Thermo Scientific) between sheets of plastic wrap. Membranes were imaged using a UVP BioChemi gel imager and band intensity was quantified using ImageJ software (Schneider et al., 2012).

To quantify the fraction of total protein in the engineered strains represented by bisabolene synthase, immobilized metal affinity chromatography was utilized to purify bisabolene synthase. *E. coli* (DE3 strain lemo21, New England Biolabs) transformed with the plasmid expressing the GS100-op construct was grown to OD600 of 1.2, at which point IPTG was added to 1 mM and the culture incubated with shaking at 22 °C for 24 hours. Cells were centrifuged and the pellets frozen at -80 °C. Cells were resuspended in 1x PBS, pH 7.4, 0.1% Triton X-100, 1 mM DTT, and 1x HALT protease inhibitor, and sonicated to lyse the

cells. BioRad Profinity IMAC resin was used to purify the histagged bisabolene synthase. The binding, wash, and elution buffers suggested by the Profinity manual were used.

Results

We initially designed a strain of *S. 6803* to express the bisabolene synthase from *A. grandis* (Figure 2.1), with the sequence codon optimized by GenScript for expression in *S. 6803*. This strain used the IPTG-inducible *P_{tic2op}* promoter and the RBS inherited from that promoter (Albers et al., 2015). After five days of growth in shake flasks under continuous light, this strain produced 1.1 ± 0.08 mg/L bisabolene. According to the metabolic network of *S. 6803* available in KEGG, the native terpenoid synthase pathway uses one enzyme to catalyze the production of geranyl pyrophosphate, farnesyl pyrophosphate and geranylgeranyl pyrophosphate. We hypothesized that this may result in low concentrations of farnesyl pyrophosphate in the cell, and that co-expression of a farnesyl pyrophosphate synthase may increase the production of bisabolene.

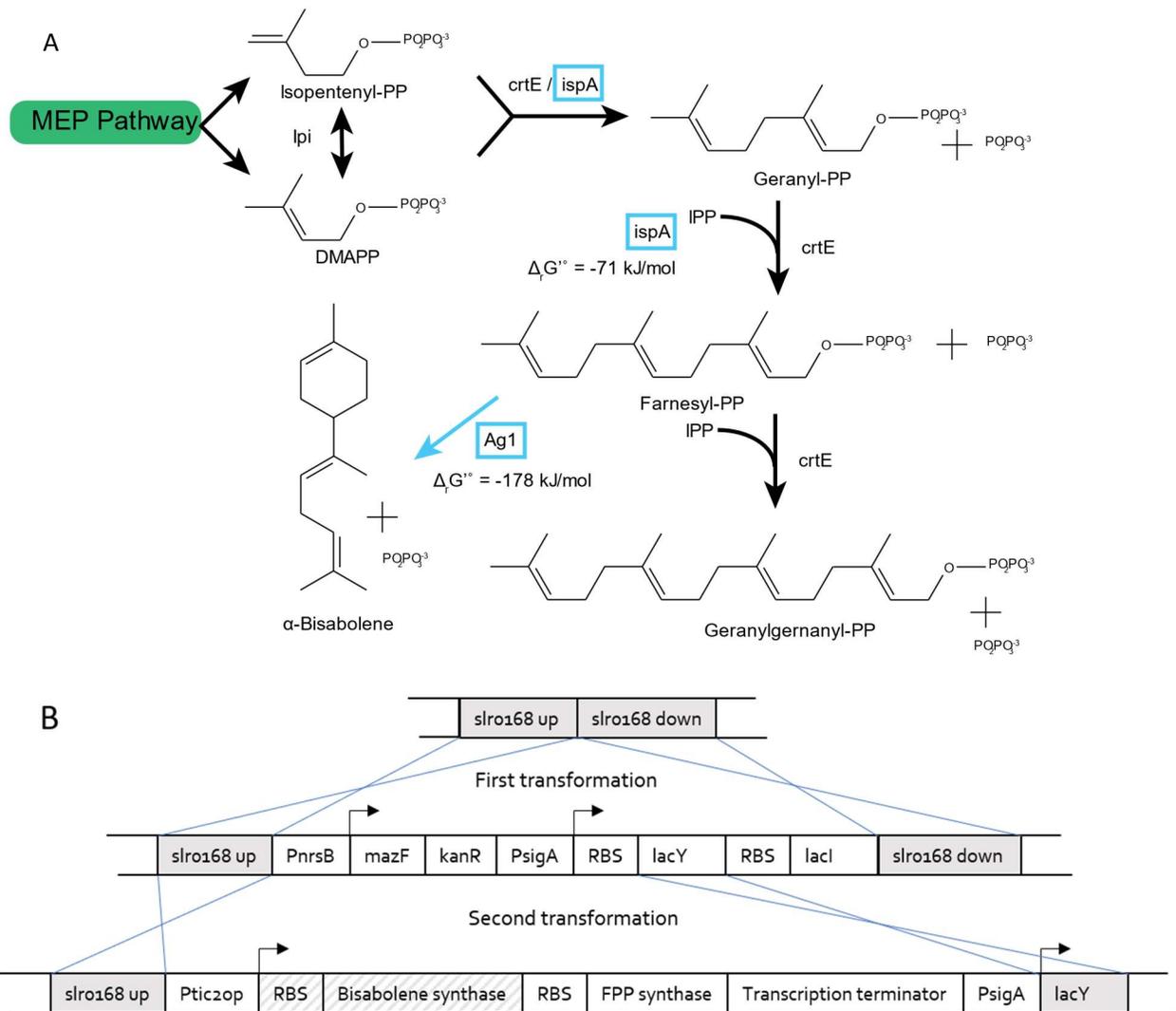


Figure 2.1: Terpene synthesis in *S. 6803* (A). Blue boxes indicate the heterologous genes used *ispA* gene from *E. coli* and the bisabolene synthase gene from *Abies grandis*. The two-step selection/counterselection transformation allows the generation of strains without selection markers (B).

We cloned the farnesyl pyrophosphate synthase from *E. coli*, codon optimized by GenScript, into an operon downstream of the bisabolene synthase gene. Denoted 'GS-1x', this engineered strain produced bisabolene at 1.6 ± 0.2 mg/L when grown in 45 mL BG-11 in shake flasks for five days. Since the strain that expressed both genes had a significantly higher bisabolene titer, we included farnesyl pyrophosphate synthase in all future strains. To further increase the product titer of bisabolene, we attempted to increase the expression of bisabolene synthase in *S. 6803* by testing different codon optimized sequences and RBS sequences. In these designs, Ptic2op initiated transcription (when induced

by addition of IPTG) of a bicistronic operon including a GenScript codon optimized sequences for *A. grandis* bisabolene synthase and farnesyl pyrophosphate synthase from *E. coli*. This sequence, along with constitutively expressed *lacY* and *lacI* for Ptic2op induction/repression, was integrated into the slr0168 neutral site containing the *mazF* counterselection cassette (Cheah et al., 2013). This cassette replaced the nickel-inducible counter-selection marker (*mazF*) and the kanamycin resistance gene, such that no selection marker was present in the final strain. No growth defects were observed due to either the presence of a dodecane layer or due to the expression of bisabolene synthase and farnesyl pyrophosphate synthase (see Figure A2.S1). During the exponential growth phase, the specific growth rates were all between 0.09 and 0.11 hr⁻¹, corresponding to a doubling time between 7.3 hours and 6.1 hours.

Codon Optimization

Five codon optimization strategies were tested (the sequences are given in Table S2). The GenScript optimization of bisabolene synthase was described above, denoted 'GS.' IDT (Integrated DNA Technologies, Skokie, IL USA) codon optimization tool was utilized to optimize the bisabolene synthase gene, denoted 'IDT,' for *S. 6803*. A bisabolene synthase gene optimized by DNA2.0 for expression in *Synechococcus p.* PCC 7002 was provided as a gift from Dr. Fiona Davies (Davies et al., 2014). Strains using this sequence are denoted '2.0'. The free software, EuGene, was utilized to design two additional codon optimizations (Gaspar et al., 2012). One utilized only the harmonization algorithm, denoted 'EuH', which considers the codon usage frequency of the native host organism and attempts to match that frequency in the design for the target organism. This function was limited by a lack of published gene sequences from *A. grandis*. Only 16 gene sequences were available from the Dendrome Project (now available at treegenesdb.org) to calculate codon usage for the native organism. Another sequence, denoted 'EuHCR', utilized the harmonization, codon context, and remove repeats rules.

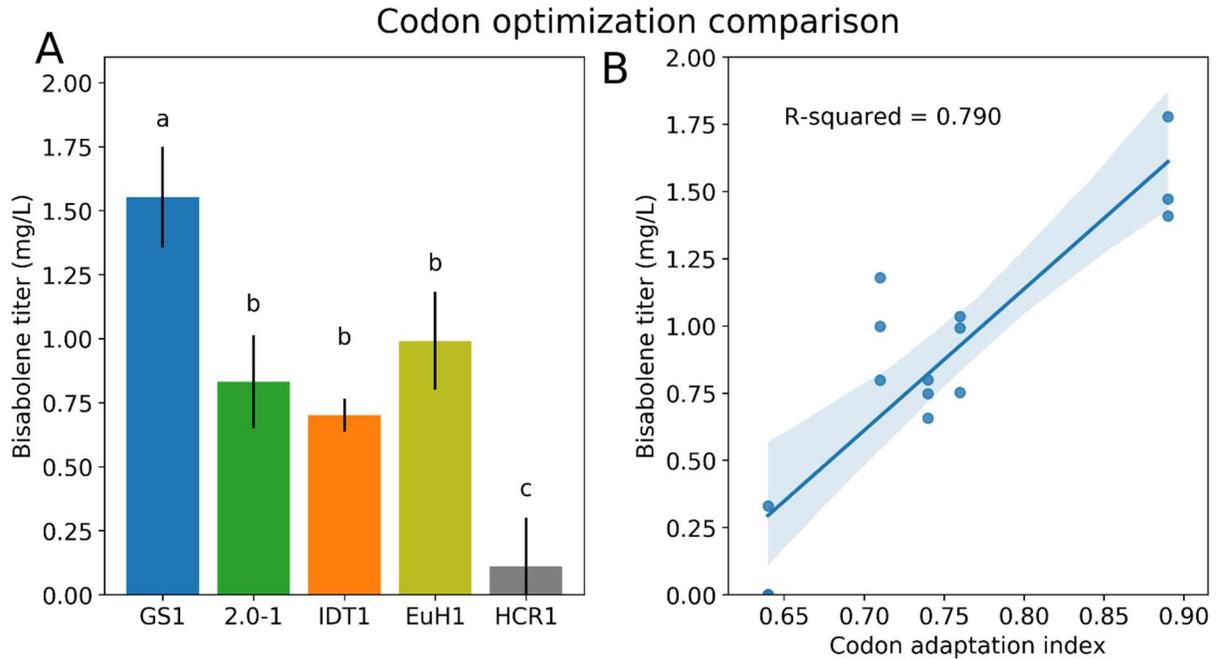


Figure 2.2: Comparison of bisabolene titer for five different codon optimizations of bisabolene synthase (A). (B) shows the linear regression between the codon adaptation index for each of these strains versus bisabolene titer. Shading represents the 95% confidence interval for the fit. Error bars indicate the standard deviation of the titer measured from three biological replicates. An ANOVA determined that there are significant differences between the codon optimizations, and a Tukey test indicated similarities as shown by the compact letter designation annotations (paired t-test with $p < 0.05$).

The five sequences have between 73% and 80% sequence identity. The codon adaptation index (CAI) compares the codon usage of a given gene sequence to the codon usage frequency of the organism (Sharp and Li, 1987). The GenScript optimization had the highest CAI at 0.89, while the DNA2.0, IDT, and EuH each had CAI between 0.76 and 0.71. The EuHCR-optimized gene had a CAI of 0.64. We measured the bisabolene titer of five strains of *S. 6803* expressing each of the codon optimizations using identical promoters (Ptic2op) and the RBS inherent to the that promoter. After five days of growth in continuous light, the bisabolene titer of these strains varied from 0.1 ± 0.2 mg/L for the EuHCR strain to 1.6 ± 0.2 mg/L for the GenScript-optimized strain (see Figure 2.2). The EuH, DNA2.0, and IDT optimized strains, which had similar CAI, also had similar bisabolene titers of 1.0 ± 0.2 , 0.8 ± 0.2 , and 0.7 ± 0.1 mg/L, respectively. Bisabolene titer and CAI were well correlated with an R2 value of 0.85 determined for the linear regression.

Ribosome binding site design

RBSs were designed using the RBS Calculator (Espah Borujeni et al., 2014; Espah Borujeni and Salis, 2016; Espah Borujeni et al., 2017; Salis et al., 2009) forward engineering tool (free energy version v1.1) using the pre-sequence 'GATAACAATT' corresponding to the sequence of the Ptic2op promoter just 5' of the RBS and transcription start site. The first 30 nucleotides of codon optimized bisabolene synthase sequence were entered as the gene sequence (see Table 2.1). Two RBSs were designed for each codon optimized gene sequence with a target expression level of 2000 (arbitrary units), or about ten times the expression level predicted by the RBS Calculator for the GS-1x strain. After we constructed strains based on these designs, the RBS Calculator was revised (free energy version v2.1), and we recalculated predicted expression levels using the updated calculator. The translation initiation rates predicted by both versions of the calculator for each RBS used in this study are provided in Table A2.S1.

Table 2.1: RBS sequences used in this work. The preceding sequence and the first 30 nucleotides of the gene used to design the RBSs in RBS Calculator are also shown.

Strain	RBS sequence	Gene sequence (first 30nt)
Pre-sequence: AATTGTGAGCGGATAACAATT		
GS1x	TCACACAGGAAACAGAATCAT	ATGGCTGGAGTGTCTGCCGTGAGCAAAGTG
GS10xA	AAACCTACGTAAACCCCTTTTTAAGGTAAAAG	ATGGCTGGAGTGTCTGCCGTGAGCAAAGTG
GS10xB	ACCAACACCTTTTAGAAGGGGTAAATTATA	ATGGCTGGAGTGTCTGCCGTGAGCAAAGTG
GS100x	ATCCCCCAAACCAAGGGAGGTTAAGA	ATGGCTGGAGTGTCTGCCGTGAGCAAAGTG
2.0-1x	TCACACAGGAAACAGAATCAT	ATGGCCGGTGTGAGCGCAGTGAGTAAAGTG
2.0-10xA	GAGGAGACGGACCCTTTCGAAGACGTTTAGGTAAG	ATGGCCGGTGTGAGCGCAGTGAGTAAAGTG
2.0-10xB	TTATTCTAAAATCTAACTATTATAGGAAGAGATT	ATGGCCGGTGTGAGCGCAGTGAGTAAAGTG
IDT1x	TCACACAGGAAACAGAATCAT	ATGGCTGGAGTCTCCGCGGTGAGTAAAGTT
IDT10xA	CAATAGCATCTATATAAAACATATCGGTAAAA	ATGGCTGGAGTCTCCGCGGTGAGTAAAGTT
IDT10xB	TCGGTAGCCGAAAAAAATCCAAGTAGGTATCGAA	ATGGCTGGAGTCTCCGCGGTGAGTAAAGTT
EuH1x	TCACACAGGAAACAGAATCAT	ATGGCCGGAGTGAGTGCCGTCAGCAAGGTG
EuH10xA	AACAGGAATATACTATTTAGAGGTACGGTAAACAT	ATGGCCGGAGTGAGTGCCGTCAGCAAGGTG
EuH10xB	CACACAGAAAGGAGAAGTCAGAAAACAA	ATGGCCGGAGTGAGTGCCGTCAGCAAGGTG
HCR1x	TCACACAGGAAACAGAATCAT	ATGGCCGGGGTATCGGCGGTTTCCAAGGTT

HCR10x A	GCGCAGCACATCGCAACAATAAAAGGGCTAT	ATGGCGGGGGTATCGGCGGTTTCCAAGGTT
HCR10x B	TTCACAAAATTCTTTTAGTTTAGGGCTCAAC	ATGGCGGGGGTATCGGCGGTTTCCAAGGTT
Pre-sequence: CACCATCATCACCATTAAATAA		
FPPS	CGAGGAAAACCAT	ATGGATTTTCCCCAACAACTGGAAGCCTGC

Bisabolene titer for strains designed for higher translation initiation rates

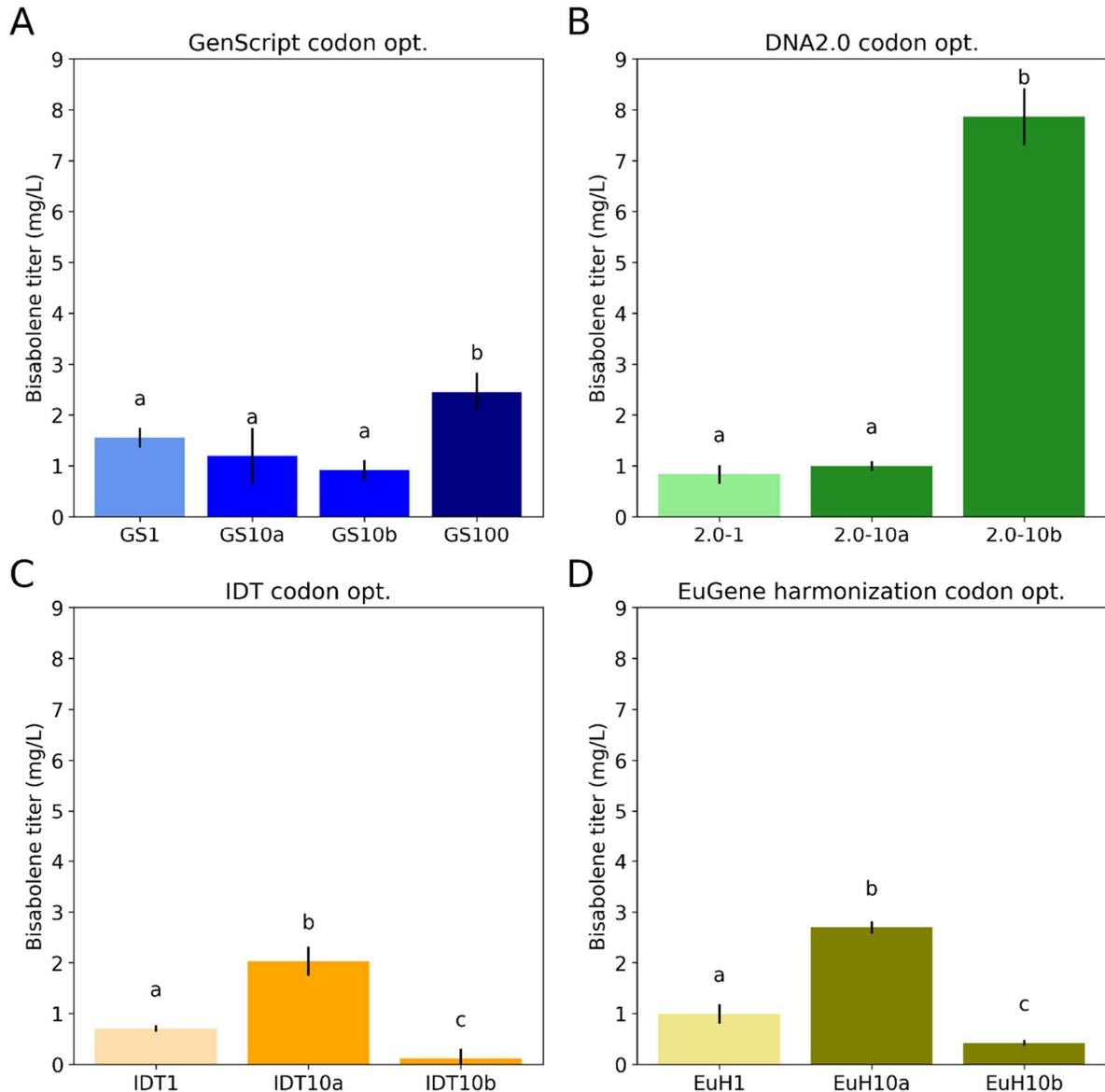


Figure 2.3: Bisabolene titer for *S. 6803* strains with RBS sequences designed for higher translation initiation rates for codon optimized gene sequence designed by GenScript (A), DNA2.0 for expression in *Synechococcus sp.* PCC7002 (B), IDT (C), and EuGene (D). Error bars indicate the standard deviation of the titer measured from three biological replicates. ANOVA tests indicate differences between RBS designs within each codon optimization group ($\alpha = 0.05$) except the HCR codon optimization. Tukey tests

for each group indicated statistically significant differences as shown by the compact letter designation annotations (paired t-test with $p < 0.05$).

Utilizing different RBSs for each codon optimization allowed us to reach a bisabolene titer of 7.9 ± 0.6 mg/L for one strain (2.0-10xB), though most strains generated titers between 0.4 and 2.7 mg/L, and the bisabolene concentration was below the detection limit in two cases (HCR-10xA and HCR-10xB) (see Figure 2.3). The predicted expression levels from the RBS Calculator correlated poorly with bisabolene titers. Within each set of strains that used the same codon optimized bisabolene synthase, RBS sequences designed to have higher translation initiation rates either increased or decreased the bisabolene titer from the base case titer for that codon optimization. For example, one RBS sequence designed to have 10-fold higher expression than the base case GenScript codon optimization strain had a statistically similar titer, while the other actually had a lower titer of 0.9 ± 0.2 mg/L compared to 1.6 ± 0.2 mg/L. The strain with an RBS designed to have expression 100-fold higher than that for the same base case strain did achieve a higher bisabolene titer of 2.5 ± 0.4 mg/L. The two designed RBS for the EuHCR codon optimization resulted in strains that failed to produce detectable levels of bisabolene (corresponding to a titer of 0.01 mg/L). The RBS Calculator-predicted expression level only poorly trended with bisabolene titer (Figure 2.4).

Bisabolene synthase concentration correlations

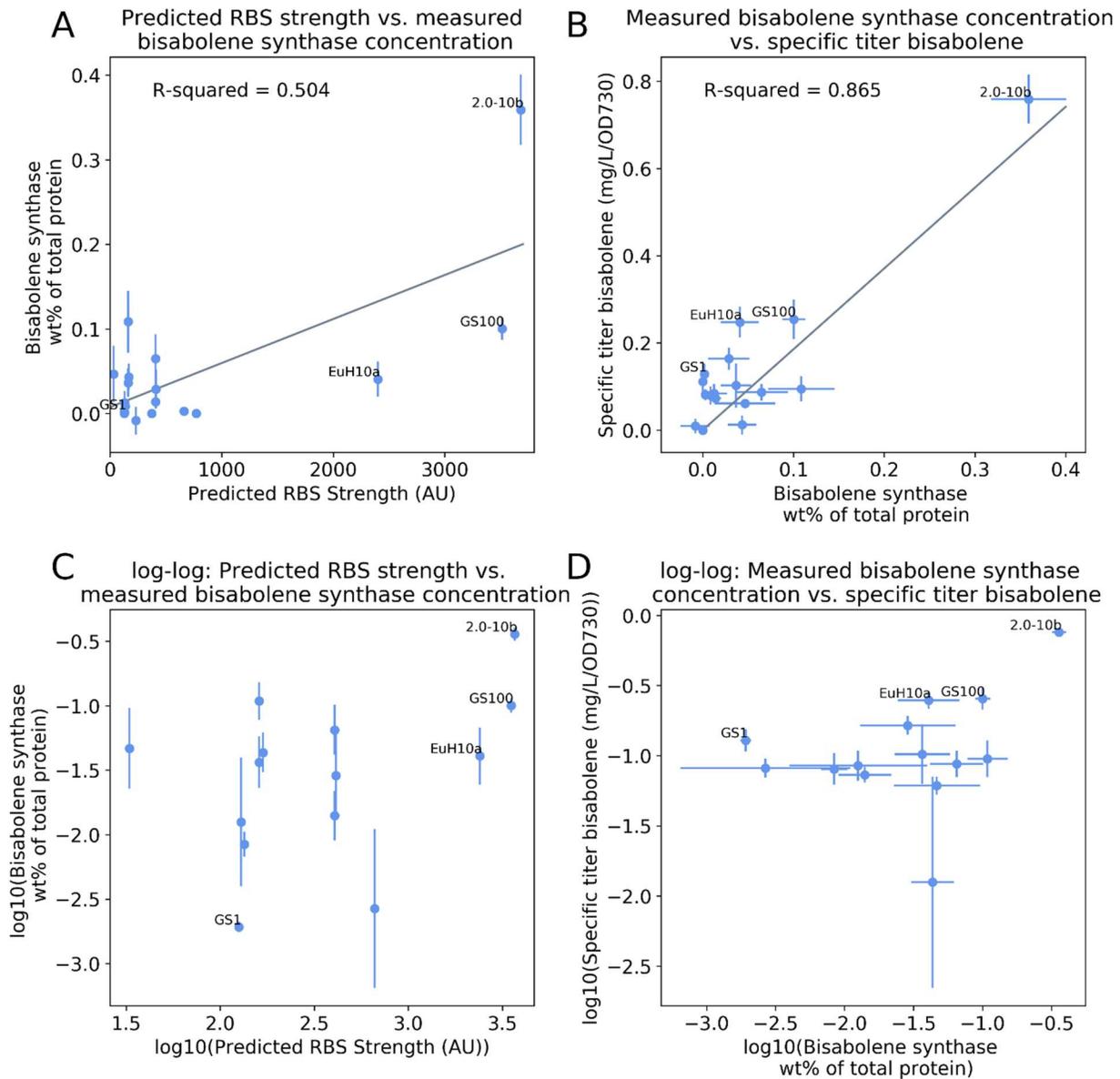


Figure 2.4: Bisabolene synthase relative protein abundance measured by Western blot versus bisabolene specific titer (A) and relative measured bisabolene synthase abundance versus the RBS Calculator v2.1 predicted translation initiation rate (B). Error bars represent the standard deviation of the Western blot signals or bisabolene specific titer from three biological replicates. 2.4(C) shows the log-log plot of 2.4(A), above it, and (D) shows the log-log plot of 2.4(B).

We compared the measured bisabolene titer to the relative bisabolene concentration for these strains using Western blotting to validate the assumption that bisabolene titer could be used as a proxy for gene expression. Bisabolene titer and relative bisabolene synthase concentration were correlated with an R^2 -value of 0.87, though the correlation is strongly influenced by the measurements for 2.0-10xB,

which was in outlier on both axes (see Figure 2.4). The correlation between relative bisabolene synthase concentration and the RBS Calculator v2.1-predicted translation initiation rate was weaker, with an R^2 -value of 0.50. A final Western blot containing a serial dilution of purified bisabolene synthase was used to estimate the fraction of total protein represented by bisabolene synthase in these samples. For the highest expressing strain, 2.0-10xB, bisabolene synthase was only 0.4% of the total protein. The presence of bisabolene synthase and farnesyl pyrophosphate synthase in the 2.0-10xB strain was also verified using LC-MS/MS (data not shown).

Impact on light:dark cycle cultivation on bisabolene titer

Cheah et al. (2013) previously explored an important, but often overlooked, aspect of cyanobacterial metabolic engineering. Engineered strains are often tested in shake flasks under continuous light, though industrial cultivation is likely to occur outside and be subject to diurnal light fluctuations. They found that strains successfully engineered to produce higher titers of free fatty acids in continuous light lost all advantage over wild type cells with regards to titer when grown in 12:12 light-dark cycles. Since bisabolene is not produced by wild type *S. 6803*, we could only explore the impact of light-dark cycles on the mutant strains. Two strains, GS-1x and 2.0-10xB, were grown in 12:12 light dark cycles for either 5 days or 10 days. When grown for 5 days in light-dark cycles, the titer of each was less than half of what it had been when grown for the same time period in continuous light (0.6 vs. 1.6 mg/L for GS-1x and 2.6 vs 7.8mg/L for 2.0-10xB, see Figure 2.5). Growth in light-dark cycles for 10 days exposed the cultures to the same total amount of light as the 5-day continuous light cultures, and these cultures ended up with cell densities closer to the 5-day continuous light cultures than the 5-day light-dark cycle cultures. Both strains achieved higher titers of bisabolene (2.7mg/L for GS-1x and 9.7 mg/L for 2.0-10xB) when grown in 10-day light-dark cycles. The statistical significance of these comparisons is unclear when comparing on a specific titer basis. The specific titer of GS-1x grown for 10-days in a light:dark cycle was higher than the 5-day continuous light culture, whereas, for 2.0-10xB, the 10-day light:dark cycle cultures had a

similar titer to the cultures grown for 5-days in continuous light. Only for 2.0-10xB was the specific titer when grown in a 5-day light-dark cycle significantly different than when grown in a 5-day continuous light condition.

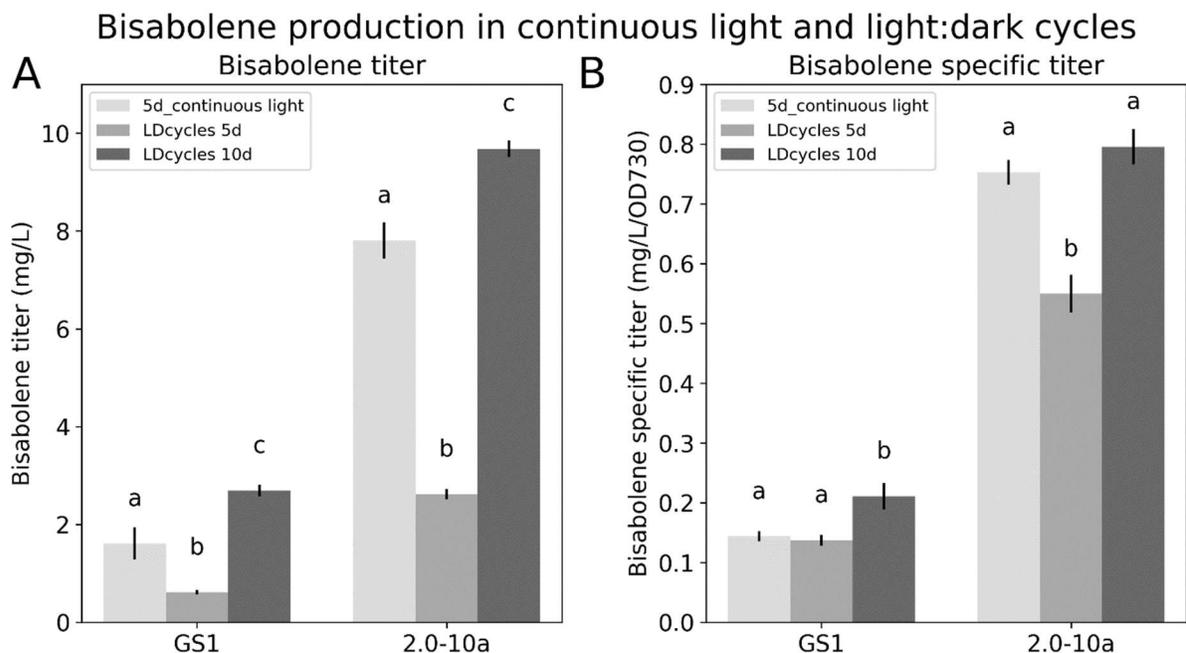


Figure 2.5: Comparison of bisabolene titer and productivity for two strains grown either in continuous light (white bars) or in 12:12 light-dark cycles. Error bars indicate the standard deviation of the titer measured from three biological replicates. ANOVA tests indicated differences in titer and specific titer depending on the cultivation conditions ($\alpha = 0.05$). Tukey tests of the titers for each strain (A) indicate significant differences all paired cultivation conditions. Tukey tests on the specific titer (B) indicated similarities as shown by the compact letter designation annotations (paired t-test with $p < 0.05$).

Bisabolene production in a photobioreactor

Finally, the best-performing strain tested in shake flasks, 2.0-10xB, was grown in a 3 L photobioreactor containing 1.5 L BG-11, aerated with 200 mL/min. 5% CO₂-enriched air, following the reactor setup and operation described by Werner et al. (Werner et al., 2019) including simulated outdoor sunlight exposure for the photobioreactors. After six days, the titer reached 7.4 ± 0.8 mg bisabolene/L and plateaued (Figure 2.6).

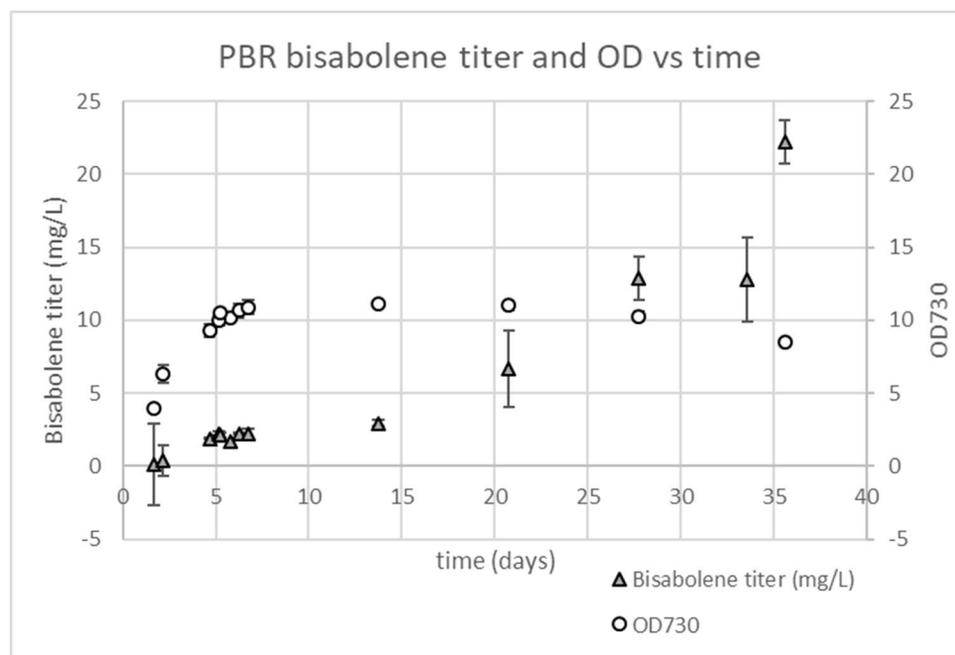


Figure 2.6: Bisabolene titer and growth over time in photobioreactors grown with sinusoidal light:dark cycles. Error bars indicate the standard deviation of the measurements from three biological replicates. Interestingly, the titer increased during stationary phase and reached 22.2 ± 1.0 mg bisabolene/L at 36 days. There was no significant stripping bisabolene from the dodecane layer due to the bubbling of gas through the reactor (data not shown).

Discussion

Following other work to generate bisabolene from microbial cultures (Davies et al., 2014; Peralta-Yahya et al., 2011), we sought first to show that bisabolene could be produced in genetically modified *S. 6803*. The bisabolene synthase from *A. grandis* was selected as it had shown the highest activity in *E. coli* among four codon-optimized genes from trees (Peralta-Yahya et al., 2011). We anticipated that *S. 6803* may have a small concentration of the precursor metabolite, farnesyl pyrophosphate (FPP), because it appears to lack a dedicated farnesyl pyrophosphate synthase. Instead, the production of FPP in *S. 6803* likely depends on geranylgeranyl pyrophosphate synthase sometimes releasing FPP before another isopentenyl pyrophosphate reacts with FPP in the active site. Therefore, we co-expressed FPP synthase from *E. coli*, codon optimized by GenScript for expression in *S. 6803*, in an operon structure with

bisabolene synthase. Transcription of this bicistronic mRNA was initiated by the tic promoter. This promoter has relatively strong expression in *S. 6803* (Albers et al., 2015), and the presence of two lac operators maintains low expression until induced by IPTG. This was expected to be useful in the event that the expression of the two genes was toxic to *S. 6803*. We did not, however, find there to be significant impacts on growth from the expression of bisabolene synthase. The initial strain produced 1.6 ± 0.2 mg bisabolene/L after 5 days or 0.31 ± 0.04 mg/L/day, a significantly higher titer than previously reported 0.6 mg/L bisabolene titer in cyanobacteria after 4 days (Davies et al., 2014).

Building on an initial proof of concept that bisabolene synthase from *A. grandis* could be expressed and functional in *S. 6803*, we sought to increase the expression of this gene and increase the bisabolene titer achieved. Lacking a high-throughput screen for bisabolene production, we constructed a set of *S. 6803* strains with varied codon usage (five variants) and varied RBS sequence (three or four RBS sequences for each codon usage variant).

The simple measure of CAI correlated with bisabolene titer when compared using the same RBS. All three commercially designed gene sequences could be used to generate functional bisabolene synthase. When RBS sequences were varied, the range of bisabolene titers achieved using each of these three codon optimizations were significantly overlapping. Kudla et al. (2009) measured the fluorescence of 154 different coding sequences of GFP expressed in *E. coli* and found that the stability of mRNA secondary structure near the RBS accounted for about half of all variation in the expression level, while the CAI was a poor predictor of fluorescence (Kudla et al., 2009). This suggests that it may also be possible to improve heterologous gene expression without changing much of the sequence, and thereby avoid the requirement to synthesize the full gene sequences which may or may not be expressed at a higher level than the wild-type sequence.

The RBS sequences tested in this work show that the sequence of the 5' untranslated region clearly impacts gene expression, but at this time we can't make accurate predictions about the expression level of a gene in *S. 6803* based only on this sequence. The RBS Calculator poorly predicted the relative expression of bisabolene synthase in *S. 6803*. Similar results were found by Wang et al. (2017) when expressing ethylene forming enzyme in *S. 6803*. This tool was originally validated using fluorescent protein expression in *E. coli*, and subsequently in *Pseudomonas fluorescens*, *Salmonella typhimurium* LT2, and *Corynebacterium glutamicum* (Farasat et al., 2014; Salis et al., 2009). Although we do expect translation to be a similar process across species of prokaryotes, the only concession the RBS Calculator makes to differences between organisms appears to be the 16s rRNA sequence which interacts with the RBS in the mRNA. *S. 6803* and *E. coli* have nearly identical sequences in the last 14 nucleotides at the 3' end of their 16S rRNA and, therefore, The RBS Calculator uses the same anti-Shine-Dalgarno sequence for *S.* as it does for *E. coli* (only the final nucleotide of the 16s rRNA differs between these organisms). The translation initiation rates predicted using the RBS Calculator for the sequences used in study are generally the same or similar between the two organisms.

The ability to accurately predict relative gene expression rates would facilitate the development of more complex genetic circuits in *S. 6803* for applications in industrial biotechnology. Genetic studies provide some evidence that translation may be different between *E. coli* and *S. 6803*. For example, a smaller proportion of genes in *S. 6803* than in *E. coli* appear to be initiated by Shine-Dalgarno (SD) sequences (26% for *S. 6803* versus 57% for *E. coli*) (Ma et al., 2002). It is not clear whether SD sequences should be expected to increase translation initiation rates in *S. 6803* as much as they do in *E. coli*. SD-antiSD hybridization is thought to reduce the impact of mRNA secondary structure on translation initiation (de Smit and van Duin, 1994). However, cyanobacterial mRNA that lack SD sequences are generally predicted to have weaker secondary structure adjacent to the start codon on either side than the mRNAs of γ -proteobacteria that lack a SD sequence (Scharff et al., 2011). Optimal spacing between the

start codon and SD sequence can also influence translation initiation rates. RBS Calculator penalizes deviation in spacing from five nucleotides between the SD and start codon (Salis et al., 2009). Ma et al. found the most frequent spacing for between SD and the start codon to be five for *E. coli*, and seven for *S. 6803* (Ma et al., 2002).

We were also surprised that cultivation of the best performing strain in photobioreactors supplemented with CO₂ did not significantly increase the product titer beyond what was measured from shake flask cultivation. We expected these cultures to reach higher optical densities in the CO₂-rich environment and also to have higher precursor availability for bisabolene synthase. However, both the maximum optical density and the titer after five days growth were similar to what was measured from shake flask cultures. It is likely that, in both cases, a nutrient other than carbon becomes limiting and prevents further growth and bisabolene production. For example, Clark et al. (2018) increased the concentration of nitrate by more than 10-fold, the concentration of phosphate by nearly 80-fold and the concentration of iron by more than 35-fold in Media A to reach higher cell densities of *Synechococcus* sp. PCC 7002 (Clark et al., 2018).

In both shake flask and PBR cultures subjected to light:dark cycles, bisabolene titer continued to increase after the cultures reached stationary phase. The bisabolene titer in PBR cultures increased significantly after they reached stationary phase and the specific titer increased from 0.26 ± 0.03 mg/L/OD after 14 days of growth to 2.6 ± 0.19 mg/L/OD after 36 days of growth. This result suggests that the MEP pathway may remain active during stationary phase, and continue to make IPP, DMAPP, and FPP available for bisabolene synthesis. Carotenoid production is also reliant on the MEP pathway and may increase during the stationary phase as the cells experience low light in the high-density culture. Carotenoids are also an important component of photoprotection, and their increased production may also be necessary for the periods of time that the cell are exposed to the high light intensity of the surface of the culture.

These results are similar to the findings of others. In terms of the product titer, Davies et al. (2014) expressed the same DNA2.0 codon optimized bisabolene synthase sequence in *Synechococcus* sp. PCC 7002 using a *cpcBA* promoter from *S. 6803*. Their tests resulted in a bisabolene titer of 0.6 mg/L after 96 hours of growth in continuous light (6 $\mu\text{g/L/hr}$). Wichmann et al. (2018) engineered strains of the algae *Chlamydomonas reinhardtii* to produce titers of 3.9 mg/L of bisabolene in phototrophic conditions and 11.0 mg/L in mixotrophic conditions after seven days of growth (Wichmann et al., 2018). Wang et al. (2017) tested a small set of RBS sequences driving expression of ethylene forming enzyme in *S. 6803*, achieving a 2.5-fold increase above the base strain.

We utilized a solvent layer of dodecane to collect produced bisabolene, following the work of others (Davies et al., 2014; Peralta-Yahya et al., 2011). The dodecane layer formed an emulsion after a few days of growth in both the shake flasks and the photobioreactors. Occasionally, samples of this emulsion required brief centrifugation in order to collect a completely organic sample. We did not find any growth inhibition due to the addition of dodecane in shake flasks and bisabolene was not detected following extraction (Bligh and Dyer, 1959) of cells harvested from cultures grown with dodecane (data not shown). Molecular dynamics models have supported the hypothesis that bisabolene may diffuse through the cell membrane into aqueous media, or, more rapidly, into dodecane in contact with the cell (Vermaas et al., 2018). Conversely, it has also been reported, in studies using *Dunaliella salina*, that a stagnant dodecane did not result in cell death while dodecane sparged through a culture did result in cell death (Kleinegris et al., 2011). The collection of excreted, hydrophobic, and volatile products from cyanobacterial or microalgal cultures which are grown outside to utilize sunlight is an area that requires further research into both the mechanism by which they are excreted (as this may limit the production rate), as well as into cost effective means of capturing the excreted product.

Our findings reinforce that it continues to be necessary to test combinations of genetic components in *S. 6803* to obtain a desired outcome because present tools for predicting the effects of the components

are inadequate. Combinations of RBSs and codon optimizations resulted in 3.3-fold increase in the concentration of bisabolene synthase over our original strain. However, even the highest expressing strain, bisabolene production likely remains limited by the expression of this enzyme. In this strain, 2.0-10xB, bisabolene synthase was estimated at just 0.4% of the total protein. Continued progress in metabolic engineering of cyanobacteria would benefit from improved understanding of translation initiation mechanisms in cyanobacteria and how they may differ from our understanding of translation initiation mechanisms in *E. coli*. Further improvements to genetic component design rules are needed to reduce the impacts of context dependence on the function of the components.

References

- Albers, S.C., Gallegos, V.A., Peebles, C.A.M., 2015. Engineering of genetic control tools in *Synechocystis* sp. PCC 6803 using rational design techniques. *J. Biotechnol.* 216, 36–46.
<https://doi.org/10.1016/j.jbiotec.2015.09.042>
- Batan, L., Quinn, J., Willson, B., Bradley, T., 2010. Net Energy and Greenhouse Gas Emission Evaluation of Biodiesel Derived from Microalgae. *Environ. Sci. Technol.* 44, 7975–7980.
<https://doi.org/10.1021/es102052y>
- Bligh, E.G., Dyer, W.J., 1959. A RAPID METHOD OF TOTAL LIPID EXTRACTION AND PURIFICATION. *Can. J. Biochem. Physiol.* 37, 911–917. <https://doi.org/10.1139/o59-099>
- Carroll, A.L., Case, A.E., Zhang, A., Atsumi, S., 2018. Metabolic engineering tools in model cyanobacteria. *Metab. Eng.* <https://doi.org/10.1016/j.ymben.2018.03.014>
- Cheah, Y.E., Albers, S.C., Peebles, C.A.M., 2013. A novel counter-selection method for markerless genetic modification in *Synechocystis* sp. PCC 6803. *Biotechnol. Prog.* 29, 23–30.
<https://doi.org/10.1002/btpr.1661>
- Clark, R.L., McGinley, L.L., Purdy, H.M., Korosh, T.C., Reed, J.L., Root, T.W., Pflieger, B.F., 2018. Light-optimized growth of cyanobacterial cultures: Growth phases and productivity of biomass and secreted molecules in light-limited batch growth. *Metab. Eng.* 47, 230–242.
<https://doi.org/10.1016/j.ymben.2018.03.017>
- Davies, F.K., Work, V.H., Beliaev, A.S., Posewitz, M.C., 2014. Engineering Limonene and Bisabolene Production in Wild Type and a Glycogen-Deficient Mutant of *Synechococcus* sp. PCC 7002. *Front. Bioeng. Biotechnol.* 2. <https://doi.org/10.3389/fbioe.2014.00021>
- de Smit, M.H., van Duin, J., 1994. Translational initiation on structured messengers. Another role for the Shine-Dalgarno interaction. *J. Mol. Biol.* 235, 173–184.
- Dismukes, G.C., Carrieri, D., Bennette, N., Ananyev, G.M., Posewitz, M.C., 2008. Aquatic phototrophs: efficient alternatives to land-based crops for biofuels. *Curr. Opin. Biotechnol.* 19, 235–240.
<https://doi.org/10.1016/j.copbio.2008.05.007>
- Espah Borujeni, A., Channarasappa, A.S., Salis, H.M., 2014. Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Res.* 42, 2646–2659. <https://doi.org/10.1093/nar/gkt1139>
- Espah Borujeni, A., Salis, H.M., 2016. Translation Initiation is Controlled by RNA Folding Kinetics via a Ribosome Drafting Mechanism. *J. Am. Chem. Soc.* <https://doi.org/10.1021/jacs.6b01453>
- Espah Borujeni, A., Cetnar, D., Farasat, I., Smith, A., Lundgren, N., Salis, H.M., 2017. Precise quantification of translation inhibition by mRNA structures that overlap with the ribosomal footprint in N-terminal coding sequences. *Nucleic Acids Res.* 45, 5437–5448.
<https://doi.org/10.1093/nar/gkx061>
- Farasat, I., Kushwaha, M., Collens, J., Easterbrook, M., Guido, M., Salis, H.M., 2014. Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria. *Mol. Syst. Biol.* 10, 731–731. <https://doi.org/10.15252/msb.20134955>
- Ferreira, E.A., Pacheco, C.C., Pinto, F., Pereira, J., Lamosa, P., Oliveira, P., Kirov, B., Jaramillo, A., Tamagnini, P., 2018. Expanding the toolbox for *Synechocystis* sp. PCC 6803: validation of replicative vectors and characterization of a novel set of promoters. *Synth. Biol.* 3. <https://doi.org/10.1093/synbio/ysy014>
- Gao, X., Gao, F., Liu, D., Zhang, H., Nie, X., Yang, C., 2016. Engineering the methylerythritol phosphate pathway in cyanobacteria for photosynthetic isoprene production from CO₂. *Energy Environ. Sci.* 9, 1400–1411. <https://doi.org/10.1039/C5EE03102H>

- Gaspar, P., Oliveira, J.L., Frommlet, J., Santos, M.A.S., Moura, G., 2012. EuGene: maximizing synthetic gene design for heterologous expression. *Bioinformatics* 28, 2683–2684. <https://doi.org/10.1093/bioinformatics/bts465>
- Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., Smith, H.O., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345. <https://doi.org/10.1038/nmeth.1318>
- Gonçalves, A.L., Rodrigues, C.M., Pires, J.C.M., Simões, M., 2016. The effect of increasing CO₂ concentrations on its capture, biomass production and wastewater bioremediation by microalgae and cyanobacteria. *Algal Res.* 14, 127–136. <https://doi.org/10.1016/j.algal.2016.01.008>
- Huang, H.H., Camsund, D., Lindblad, P., Heidorn, T., 2010. Design and characterization of molecular tools for a Synthetic Biology approach towards developing cyanobacterial biotechnology. *Nucleic Acids Res.* 38, 2577–2593. <https://doi.org/10.1093/nar/gkq164>
- Huang, H.-H., Lindblad, P., 2013. Wide-dynamic-range promoters engineered for cyanobacteria. *J. Biol. Eng.* 7, 10. <https://doi.org/10.1186/1754-1611-7-10>
- Hughes, A.R., Sulesky, A., Andersson, B., Peers, G., 2018. Sulfate amendment improves the growth and bioremediation capacity of a cyanobacteria cultured on municipal wastewater centrate. *Algal Res.* 32, 30–37. <https://doi.org/10.1016/j.algal.2018.03.004>
- Jin, H., Wang, Y., Idoine, A., Bhaya, D., 2018. Construction of a Shuttle Vector Using an Endogenous Plasmid From the Cyanobacterium *Synechocystis* sp. PCC6803. *Front. Microbiol.* 9. <https://doi.org/10.3389/fmicb.2018.01662>
- Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M., Tabata, S., 1996. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res. Int. J. Rapid Publ. Rep. Genes Genomes* 3, 109–136.
- Kiyota, H., Okuda, Y., Ito, M., Hirai, M.Y., Ikeuchi, M., 2014. Engineering of cyanobacteria for the photosynthetic production of limonene from CO₂. *J. Biotechnol.* 185, 1–7. <https://doi.org/10.1016/j.jbiotec.2014.05.025>
- Kleinegris, D.M.M., van Es, M.A., Janssen, M., Brandenburg, W.A., Wijffels, R.H., 2011. Phase toxicity of dodecane on the microalga *Dunaliella salina*. *J. Appl. Phycol.* 23, 949–958. <https://doi.org/10.1007/s10811-010-9615-6>
- Kok, S. de, Stanton, L.H., Slaby, T., Durot, M., Holmes, V.F., Patel, K.G., Platt, D., Shapland, E.B., Serber, Z., Dean, J., Newman, J.D., Chandran, S.S., 2014. Rapid and Reliable DNA Assembly *via* Ligase Cycling Reaction. *ACS Synth. Biol.* 3, 97–106. <https://doi.org/10.1021/sb4001992>
- Kudla, G., Murray, A.W., Tollervey, D., Plotkin, J.B., 2009. Coding-Sequence Determinants of Gene Expression in *Escherichia coli*. *Science* 324, 255–258. <https://doi.org/10.1126/science.1170160>
- Lin, P.-C., Saha, R., Zhang, F., Pakrasi, H.B., 2017. Metabolic engineering of the pentose phosphate pathway for enhanced limonene production in the cyanobacterium *Synechocystis* sp. PCC 6803. *Sci. Rep.* 7. <https://doi.org/10.1038/s41598-017-17831-y>
- Liu, D., Pakrasi, H.B., 2018. Exploring native genetic elements as plug-in tools for synthetic biology in the cyanobacterium *Synechocystis* sp. PCC 6803. *Microb. Cell Factories* 17. <https://doi.org/10.1186/s12934-018-0897-8>
- Ma, J., Campbell, A., Karlin, S., 2002. Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures. *J. Bacteriol.* 184, 5733–5745.

- Markley, A.L., Begemann, M.B., Clarke, R.E., Gordon, G.C., Pflieger, 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>
- Pade, N., Hagemann, M., 2014. Salt Acclimation of Cyanobacteria and Their Application in Biotechnology. *Life* 5, 25–49. <https://doi.org/10.3390/life5010025>
- Pattanaik, B., Lindberg, P., 2015. Terpenoids and Their Biosynthesis in Cyanobacteria. *Life* 5, 269–293. <https://doi.org/10.3390/life5010269>
- Peralta-Yahya, P.P., Ouellet, M., Chan, R., Mukhopadhyay, A., Keasling, J.D., Lee, T.S., 2011. Identification and microbial production of a terpene-based advanced biofuel. *Nat. Commun.* 2, 483. <https://doi.org/10.1038/ncomms1494>
- Reinsvold, R.E., Jinkerson, R.E., Radakovits, R., Posewitz, M.C., Basu, C., 2011. The production of the sesquiterpene β -caryophyllene in a transgenic strain of the cyanobacterium *Synechocystis*. *J. Plant Physiol.* 168, 848–852. <https://doi.org/10.1016/j.jplph.2010.11.006>
- Salis, H.M., Mirsky, E.A., Voigt, C.A., 2009. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* 27, 946–950. <https://doi.org/10.1038/nbt.1568>
- Scharff, L.B., Childs, L., Walther, D., Bock, R., 2011. Local Absence of Secondary Structure Permits Translation of mRNAs that Lack Ribosome-Binding Sites. *PLoS Genet.* 7, e1002155. <https://doi.org/10.1371/journal.pgen.1002155>
- Schneider, C.A., Rasband, W.S., Eliceiri, K.W., 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675. <https://doi.org/10.1038/nmeth.2089>
- Sharp, P.M., Li, W.-H., 1987. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 15, 1281–1295. <https://doi.org/10.1093/nar/15.3.1281>
- Stanier, R.Y., Deruelles, J., Rippka, R., Herdman, M., Waterbury, J.B., 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>
- Vermaas, J.V., Bentley, G.J., Beckham, G.T., Crowley, M.F., 2018. Membrane Permeability of Terpenoids Explored with Molecular Simulation. *J. Phys. Chem. B* 122, 10349–10361. <https://doi.org/10.1021/acs.jpcc.8b08688>
- Wang, B., Eckert, C., Maness, P.-C., Yu, J., 2018. A Genetic Toolbox for Modulating the Expression of Heterologous Genes in the Cyanobacterium *Synechocystis* sp. PCC 6803. *ACS Synth. Biol.* 7, 276–286. <https://doi.org/10.1021/acssynbio.7b00297>
- Werner, A., Broeckling, C.D., Prasad, A., Peebles, C.A.M., 2019. A comprehensive time-course metabolite profiling of the model cyanobacterium *Synechocystis* sp. PCC 6803 under diurnal light:dark cycles. *Plant J.* <https://doi.org/10.1111/tpj.14320>
- Wichmann, J., Baier, T., Wentnagel, E., Lauersen, K.J., Kruse, O., 2018. Tailored carbon partitioning for phototrophic production of (E)- α -bisabolene from the green microalga *Chlamydomonas reinhardtii*. *Metab. Eng.* 45, 211–222. <https://doi.org/10.1016/j.ymben.2017.12.010>
- Yu, J., Liberton, M., Cliften, P.F., Head, R.D., Jacobs, J.M., Smith, R.D., Koppelaar, D.W., Brand, J.J., Pakrasi, H.B., 2015. *Synechococcus elongatus* UTEX 2973, a fast growing cyanobacterial chassis for biosynthesis using light and CO₂. *Sci. Rep.* 5. <https://doi.org/10.1038/srep08132>

CHAPTER 3: REVIEW OF TRANSLATION INITIATION IN CYANOBACTERIA

Motivation

Ribosomes are the manufacturing facilities of cells which produce the machines which cell uses to do chemistry (metabolism) and build structures. RNA polymerase reads the genetic code and generates mRNA transcripts that include the sequences that code for proteins. The process of producing transcripts is highly regulated in terms of when and how many copies of the code to produce.

Translation by the ribosomes is also subject to regulation, and the rate of translation initiation is thought to be a rate-limiting step in translation (Kozak 2005).

Protein production overall is a resource intensive process. Up to 50% of energy use in rapidly growing bacteria may be devoted to translation (Russell and Cook 1995). In *E. coli*, 55% of the cell dry weight is composed of proteins (Neidhardt and Curtiss 1996). Given the high resource requirements of protein synthesis, it is essential for cells to be able to accurately produce proteins, and to produce the right proteins at the right time. Both transcription and translation initiation are regulated for these ends.

In this review, we focus on the mechanisms used by cyanobacterial cells to regulate translation initiation and to identify correct start codons within transcripts. As most research on prokaryotic translation initiation has occurred in *E. coli*, that perspective is presented before reviewing the research that has, so far, suggested that the mechanisms utilized by cyanobacteria may be somewhat different than they are in *E. coli*.

Challenges in controlling translation initiation for metabolic engineering of cyanobacteria

Beyond the interest in understanding how translation is initiated, metabolic engineers need to be able to reliably control translation initiation to have success in manipulating cellular metabolism. Despite all that we do know about the mechanisms of translation initiation, it is generally necessary to test many

ribosome binding site sequences in order to achieve the desired expression level. Many studies have characterized the relative “strength” of ribosome binding sites using reporter genes. Unfortunately, the relative strengths are dependent on the coding sequence as well and not in a predictable manner. Something is missing from our model of translation initiation.

Those of us performing metabolic engineering of cyanobacteria have tended to adopt the designs and tactics previously developed in *E. coli* metabolic engineering projects. To manipulate translation initiation in cyanobacteria, we have either utilized ribosome binding sites characterized in *E. coli*, copied the 20 or so nucleotides upstream of highly expressed native genes as RBSs as had been done in *E. coli*, or have used the RBS Calculator to design RBSs and predict their strength.

Heidorn et al. (2011) characterized three iGEM BioBrick RBSs (developed for use in *E. coli*) in *Synechocystis sp.* PCC6803 (S. 6803) as well as one sequence simply designed for strong SD-antiSD binding designated ‘RBS*.’ Markley et al. (2015) varied the SD sequence within an RBS and measured fluorescent protein expression in *Synechococcus sp.* 7002. Englund, Liang, and Lindberg (2016) tested thirteen RBSs, including eight iGEM biobrick sequences, three native sequences, and RBS* and tested each RBS for initiating translation of YFP and BFP. Thiel et al. (2018) also tested 13 RBSs in S. 6803, including some native sequences and several sequences designed by the RBS Calculator and previously tested in *E. coli*. Thiel et al. also demonstrated the impact that the downstream coding sequence can have on translation initiation rates by comparing the expression of GFP and YFP driven by the 13 RBS. Despite some sequence similarity in the coding sequences, the expression profiles between GFP and YFP were significantly different. Wang et al. (2018) made select mutations to an RBS to generate twelve other sequences and measured ethylene production when the RBSs were used to initiate translation of ethylene forming enzyme. Predictions from the RBS Calculator were compared against the measurements. The RBS Calculator performed particularly poorly in this case, with the predictions and measurements seemingly anti-correlated. Liu and Pakrasi tested 20 RBSs derived from native 5’-

untranslated regions of native genes. Surprisingly, they measured fluorescence similar to the background for 9 of the RBS when driving YFP expression. These included the sequences preceding highly expressed genes *cpcB*, *rbcL*, and *psbA2*.

Importance of translation initiation

Many start codons may exist in a given transcript. In order to avoid producing nonsense amino acid sequences without function, ribosomes must accurately identify the correct start codons. This may be achieved *via* specific binding between the 16S ribosomal RNA and the Shine-Dalgarno (SD) sequence often found in the 5'-untranslated region of transcripts at a defined distance from the start codon. However, many genes have internal SD-like sequences, and significant portions of many prokaryotic genomes lack such a sequence near the start codon. An alternative mechanism likely exists where the correct start codon is selected by virtue of having minimized secondary structure of the adjacent mRNA making it uniquely accessible to the ribosome.

In the cyanobacterium, *Synechocystis sp.* PCC6803 (*S.* 6803), only about 26% of genes have a SD sequence (Ma, Campbell, and Karlin 2002). Several research groups have undertaken projects to precisely control translation initiation rates in this species for the expression of heterologous proteins which may have applications in biofuels or the production of other valuable chemicals. Generally, there has been less success in controlling translation initiation rates in cyanobacteria than in *E. coli* including in projects which have utilized the expression level prediction software, RBS Calculator. It is therefore necessary that we take a closer look at translation initiation in cyanobacteria. Is the translation initiation mechanism as widely conserved across prokaryotes as we biological engineers have assumed? If we want to control translation initiation and design/engineer this process, what might be missing from our engineering perspective?

The canonical model of bacterial translation initiation

The Shine-Dalgarno sequence (SD) has often been presumed to be the defining element of the ribosome binding site. The SD, however, is only one component of the ribosome-mRNA interactions that help to initiate translation at the correct start codon in bacteria. In fact, many mRNA lack any SD sequence.

Most genes in *Synechocystis Sp.* PCC6803 lack a Shine-Dalgarno sequence. As will be discussed below, several other elements such as the standby site, S1 ribosomal protein binding, mRNA secondary structure, and the base-pairing between start codon and the anticodon may be used to identify the start codon. These elements also contribute to controlling the rate of translation initiation from the mRNA.

Overview of the canonical prokaryotic translation initiation

The prokaryotic ribosome consists of 30S and 50S subunits which together contain approximately 55 proteins and three ribosomal RNAs of 120 (5S), 1,500 (16S), and 2,900 (23S) nucleotides with a total molecular mass of about $2.6\text{-}2.9 \times 10^6$ Daltons (K.E. van Holde and W.E. Hill 1974). Laursen et al. (2005) provide a useful review of prokaryotic translation initiation. Translation initiation is the rate-limiting step in protein synthesis. It is therefore of interest to metabolic engineers as an important mechanism to modulate for expressing heterologous enzymes. As Gualerzi and Pon (2015) observed, the central role of the Shine-Dalgarno sequence in translation initiation has been “taken as dogma.” Shine-Dalgarno sequence base-pairing with the 3'-end of 16S rRNA does play an important role in the efficiency and accuracy of translation initiation. However, it cannot sufficiently explain how bacteria control translation initiation given that many genes lack a Shine-Dalgarno sequence.

The process of formation of the 70S initiation complex is briefly described here. Initially, IF2-GFP and IF3 bind the 30S subunit, followed by IF1. The position of IF1 allows it to block the P-site which may ensure that the initiator tRNA binds to the A-site first. The initiator tRNA, fMet-tRNA, and the mRNA then bind, in either order. Any secondary structure that blocks access to the start codon must then unfold, forming

the 30S pre-initiation complex. Complex interactions between the initiation factors, the mRNA, and the tRNA are responsible for moving the tRNA to the P-site. Conformational changes in the 30S pre-initiation complex facilitate the binding of fMet-tRNA anti-codon to the start codon, forming the more stable 30S initiation complex. This “locking” process is the rate-limiting step in the formation of the 30S initiation complex. The 50S ribosomal subunit may then bind and IF1 and IF3 then dissociate. GTP bound to IF2 is hydrolyzed and IF2 is released. The 70S initiation complex may then commence translation by catalyzing the first peptidyltransferase reaction (Laursen et al. 2005). An in-depth review of bacterial translation initiation which is summarized here has been provided by Gualerzi and Pon (Gualerzi and Pon 2015).

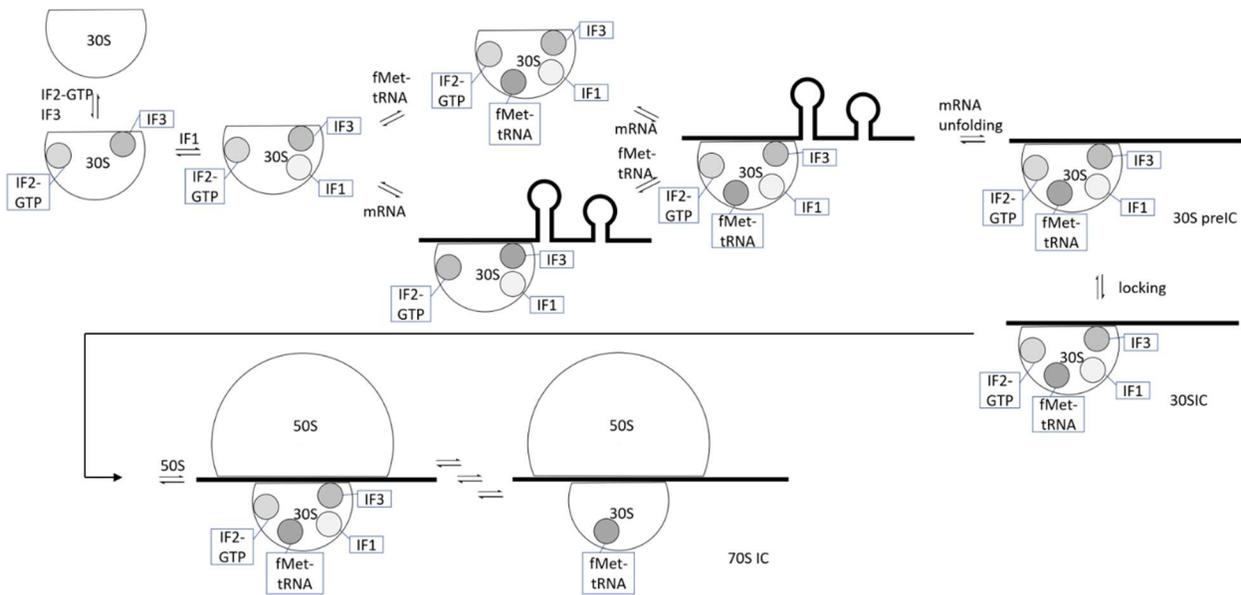


Figure 3.1: The canonical model of bacterial translation initiation (adapted from Gualerzi and Pon, 2015). The 30S ribosomal subunit first binds initiation factors, IF2-GTP and IF3, followed by IF1. The initiator tRNA (fMet-tRNA) and mRNA are then bound in either order. At this point, the mRNA must unfold for the 30S pre-initiation complex to form. This complex can then transition to the more stable 30S initiation complex. The 50S subunit can then bind and, following several steps in which the initiation factors are released, the 70S initiation complex is formed.

Interactions between the ribosome and mRNA

Looking more closely at the interactions between the ribosome and the mRNA, footprinting assays have provided evidence for extensive interaction between the 30S ribosome and mRNA. Hydroxyl radical footprinting indicated that the region between -35 and +19 from the start codon interfaced with the

ribosome (Hüttenhofer and Noller 1994). This region of the mRNA adjacent to the start codon is called the ribosome binding site. Specific interactions between the ribosome and the mRNA are accomplished by base pairing between the 16S rRNA and the Shine-Dalgarno sequence of the 5'-untranslated region upstream of the start codon and between the tRNA anticodon and the start codon. These interactions facilitate the identification of the correct start codon which will define the start of the coding sequence and the reading frame of the protein coded for. In other words, these interactions are important regulators for accuracy in protein production.

The Unified View – Secondary structure and SD-antiSD

Nakagawa found that highly expressed genes with significant mRNA folding energy tended to have Shine-Dalgarno sequences and suggested that the SD-antiSD interaction may be used to facilitate translation initiation in the presence of this thermodynamic barrier of unfolding the mRNA (S. Nakagawa et al. 2010). This interpretation was echoed by Wei and Xia (2019) in their study of translation initiation in cyanobacteria. These interpretations follow the idea that was summarized by Tokumasa Nakamoto (Nakamoto 2006). His hypothesis was that start codons were uniquely accessible to the ribosome, allowing translation to initiate specifically. Nakamoto observed that mRNA is significantly structured, with 60-70% of nucleotides being base-paired, and he pointed to several publications from the 1960's which suggest that ribosomes bind RNA quite promiscuously in the absence of secondary structures. Nakamoto and Vogel found that cell-free transcription and translation systems that accurately produced phage proteins also could initiate translation from random positions within poly(U), poly(CU), poly(ACU), and poly(GU) RNAs (Nakamoto and Vogl 1978). In this framework, the SD-antiSD is not considered an important component of identifying the correct start codon. Instead, the SD sequence is used to overcome strong secondary structures in the RBS and initiate translation and, thus, is not needed when the RBS has weak secondary structure.

Marzi et al. determined the structures of several states of the pre-initiation complex that forms on the rpsO mRNA. These revealed that the ribosomal protein S15 stabilizes mRNA secondary structure, which prevents the mRNA from entering the mRNA channel of the ribosome and inhibits translation initiation. Inhibition can be relieved by competitive binding of S15 to the 16D rRNA (in other words, by an increase in concentration of 16S rRNA) (Marzi et al. 2007).

Saito et al. (2020) recently used orthogonal ribosomes with mutated aSD sequences to test whether SD-antiSD base-pairing was necessary for identification of correct start codons. They found that this base pairing was not necessary for correct identification, but that both SD-antiSD base pairing and A-rich regions can enhance initiation rates (Saito, Green, and Buskirk 2020). Komarova et al. found that longer SD sequences decreased the concentration of the reporter gene beta-galactosidase (Komarova et al. 2002).

What is the role of the SD-antiSD interaction?

mRNA may base pair with themselves to form secondary structures consisting of stems and un-paired loops. Such structures within the ribosome binding site prevent interaction with the ribosome which can only bind single-stranded mRNA. Many studies have confirmed the hypothesized “unified view” which states that translation initiation generally occurs at start codons with low secondary structure and that SD sequences can help ribosomes overcome secondary structures in the region.

De Smit and van Duin observed a strong negative correlation between the stability of a hairpin containing a ribosome binding site and protein expression (M. H. de Smit and van Duin 1990). Olsthoorn et al. (1994 and 1995) examined the translation initiation of the coat protein mRNA of bacteriophage MS2 with mutations and monitored the evolution of the phage. Specific mutations were made to disrupt secondary structure in the 5'-UTR or the SD sequence. Mutations that weakened or strengthened the secondary structure were compensated for in subsequent generations by other mutations in the

sequence to revert the stability to its wild-type strength, while none of the mutations that arose disrupted the SD sequence (R. C. Olsthoorn, Licis, and van Duin 1994). Similarly, mutations were made to shorten or extend the SD sequence in this gene's translation initiation region. Future generations compensated for extended SD-antiSD complementarity with mutations that increased secondary structure strength. Mutations that decreased structure strength arose following shortening of the SD sequence (R. C. L. Olsthoorn, Zoog, and Duin 1995). These results supported the hypothesis that the SD sequence can be used to overcome secondary structure barriers to translation initiation and that the SD sequences is unnecessary in mRNA with weak secondary structure. The balance of SD-antiSD hybridization and secondary structure free energy changes was suggested to be controlling the rate of translation initiation. In a separate study Olsthoorn and van Duin observed that stems of the hairpins in the MS2 A-protein gene could be extended without affecting the phage fitness, but extended loops in the same hairpin decreased fitness. This led the authors to conclude that there is selection pressure against single-stranded RNA (R. C. Olsthoorn and van Duin 1996). This may be due to the vulnerability of the single-stranded RNA to RNases.

Examination of evolutionary changes in start codons and SD sequences in *E. coli* mRNA suggested that high protein expression levels could be maintained in mRNA for which the start codon mutated from the preferred AUG by contemporaneous mutations that increased SD-antiSD base-pairing (Belinky, Rogozin, and Koonin 2017). Complementarity between the 16S rRNA and a region 3'- of the start codon of some mRNAs suggested another specific interaction (downstream box) (Sprengart, Fatscher, and Fuchs 1990), but the interaction does not appear to have a role in translation initiation (Moll et al. 2001).

Standby-site binding and mRNA unfolding

De Smit and van Duin followed up their previous studies on the influence of mutations that impacted mRNA secondary structures on translation initiation by considering the mRNA folding kinetics, rather

than just the folding free energy. They found that mRNA hairpins typically fold too quickly for the small subunit of the ribosome to bind and reasoned that an unstructured region of mRNA must be present to increase local concentrations of ribosomes which can traverse the mRNA to the RBS in the brief periods when it is unstructured. This unstructured region has been referred to as the standby site (Maarten H. de Smit and van Duin 2003).

If transcription and translation are not closely coupled, the mRNA may hybridize with itself. This hybridization can present a thermodynamic barrier to translation initiation if the translation initiation region is not single-stranded and available to the 30S ribosome for binding. Two models have emerged for how the ribosome can overcome this barrier. Either the 30S ribosome can only bind the mRNA when it exists in the unfolded state or it may bind first to an upstream site, called the “standby site”, and may travel to the start codon as the mRNA unfolds. Studer and Joseph (2006) measured the binding kinetics of 30S ribosomes of mRNA designed to have different secondary structures. mRNA with weaker secondary structure bound 30S ribosome with higher affinity, and those that lacked SD sequences but had weak secondary structure were also capable of binding the 30S ribosome. The length of the SD sequence did not have a significant effect on the binding kinetics (Studer and Joseph 2006).

The presence of initiation factors (IF1, IF2, and IF3) or the initiator tRNA did not impact the binding kinetics. However, the presence of both initiation factors and the initiator tRNA did stabilize the 30S binding of mRNA, especially in structured mRNA. Further experiments demonstrated that the SD sequence was necessary for the unfolding of strong secondary structure, along with both the IFs and initiator tRNA. Unfolding was also dependent on GTP being bound to IF2, but hydrolysis of GTP was not necessary suggesting that the unfolding of the mRNA was a passive process rather than one driven by some helicase activity (Studer and Joseph 2006). Reviewing translational repression mechanisms, Schlx and Worhunsky concluded that kinetic barriers to translation were common and needed to be considered in addition to thermodynamic models of translation initiation (Schlx and Worhunsky 2003).

This work has been followed by numerous studies that support the finding that the SD sequence may be necessary for structured translation initiation regions, but not for those with weak structures. Kudla et al. (2009) observed that strong secondary structure in the translation initiation region inhibited translation initiation in synthetic ribosome binding sites in *E. coli*. Scharff et al. (2011) determined that mRNA that lacked both a SD sequence and had low secondary structure could be translated.

What is the role of S1-binding?

Takyar et al. (2015) demonstrated that the ribosome of *E. coli* has helicase activity in elongation mode, and that the S3 and S4 proteins are likely the specific proteins having this activity (Takyar, Hickerson, and Noller 2005). This helicase activity is the basis for the engineered translation initiation systems called “bicistronic designs” where an small upstream open reading frame puts the ribosome in elongation mode to take advantage of helicase activity to unfold structures around a reinitiation site which consists of an RBS and a start codon of the gene of interest which overlaps with the stop codon of the upstream reading frame (Mutalik et al. 2013). How does the ribosome overcome secondary structure before it enters elongation mode?

The ribosomal S1 protein also interacts with AU-rich region 5' of where the Shine-Dalgarno sequence is typically found. This S1 protein was shown to be essential for translation of the *ssb* mRNA in *E. coli* (Boni et al. 1991). Komarova et al. found that translation initiation rates could be increased by addition of A/U-rich S1 binding sites. Following a previous finding of others that S1 could bind BoxA, they altered an RBS to include BoxA to determine if it might be used to enhance translation initiation rates. BoxA is known better as an anti-terminator in the *E. coli rrrB* operon which prevents transcription termination at a hairpin and allows the full operon to be transcribed. It has the sequence 5'-GCUCUUUAACA-3', in an mRNA that is not translated. Translation was strongly enhanced by addition of BoxA to the RBS, 5' of the Shine-Dalgarno sequence. Their work suggested that S1 binding in translation initiation regions should also be considered when evaluating ribosome binding (Komarova et al. 2002).

Duval et al. (2013) found that the *E. coli* S1 protein helps the 30S ribosome bind structured mRNAs. Their work also demonstrated that S1 helps to destabilize secondary structures in natural mRNA, but characterized the activity of S1 as an RNA chaperone that promotes binding rather than a helicase that actively unfolds the RNA. The S1 protein acts as a ratchet that prevents secondary structure from reforming after it unfolds. In addition, they found that mRNA containing a SD sequence could be translated efficiently without the S1 protein present, but that S1 was needed for those that lacked an SD sequence (Duval et al. 2013).

The *S. 6803* ribosomal S1 was not associated with the rest of ribosomal proteins in gradient sedimentation (Riediger et al. 2020). Baers et al. (2019) found S1 in the cytosol fraction, and S3 in the thylakoid membrane fraction in their proteome mapping of *S. 6803* (Baers et al. 2019). The authors suggested that these proteins may have a role in protein localization.

What factors determine the expression level of proteins?

Gingold and Pilpel (2011) provide a concise review of the factors which impact translation initiation and elongation rates. They include a helpful discussion around whether translation rates are limited by a single factor or by a combination of factors. This is especially difficult to delineate given that each sequence change can impact multiple factors such as the mRNA degradation rate, codon usage, secondary structure of the mRNA, the ribosome elongation rate, and proper folding of the protein (Gingold and Pilpel 2011).

Kudla et al. (2010) found that expression levels of GFP genes with synonymous codon variations was better correlated with the folding free energy than with the codon adaptation index. However, Tuller et al. (2010) found that, in natural genes in *E. coli*, both codon usage and folding free energy were determinants of translation rates (Tamir Tuller et al. 2010). Tuller and Zur later reviewed the many, often overlapping, translation regulating signals found on either side of the start codon. These included

the prevalence of weak mRNA secondary structure around the start codon, and sequence-related degradation rates (T. Tuller and Zur 2015). These factors are difficult to separate because changing one nucleotide in the sequence changes multiple parameters. In addition, those parameters sometimes have multiple explanations. For example, A/U-rich sequences in the 5'-UTR may increase protein expression by reducing folding energy of the mRNA. Such sequences can also be bound by the ribosomal S1 protein in *E. coli* which is an RNA chaperone which helps position the ribosome for single stranded mRNA binding as it unfolds.

Testing of large-scale randomized RBS libraries

A large library of 18 nucleotide RBS sequences was subjected to three rounds of selection based on approximate translation rates. The library was enriched only slightly in SD sequences. It was also enriched in cytosine in many positions which the authors suggested also may base-pair with the 16S rRNA outside of the aSD sequence (Barendt et al. 2012). A follow-up study using shorter 5'-UTR leaders resulted in an RBS library in which 54% were non-SD sequences. The sequences had a position-independent G/U bias which often could be explained as additional nucleotides which may base-pair with the 16S rRNA outside the aSD. For RBSs which did contain an SD sequence, A/U-rich sequences were often also present, suggesting that these two motifs work together to produce high expression rates. The authors suggested that the longer 5'UTR used in the previous report allowed for more interactions between the ribosome and mRNA, giving rise to different motifs in the RBS (specifically, C-rich sequences with the longer leader, and G/U-rich sequences with the shorter leader, though both types were expected to contribute to base-pairing between the 16S rRNA and the mRNA) (Barendt et al. 2013).

Evfratov et al. (2017) tested a large library of completely randomized RBS sequences in *E. coli*. The sequences were sorted by expression level (cerulean fluorescent protein) using Flowseq. High

expressing fractions were enriched in A and depleted of C. A/G enrichment in the -7 to -13 region relative to the start codon corresponding to the SD sequence was also observed. Estimated secondary structure folding energy positively correlated with translation efficiency. Machine learning was applied to model the impacts of several parameters on expression level including the length and location of SD sequence, start codon used, secondary structures, and A/U-rich sequences. They found a potential benefit of multiple SD sequences and start codons as well as AG repeats for high expression, but also concluded that precise prediction of translation efficiency remains difficult.

Alternative mechanisms in bacteria

A recent commentary provides a useful primer on the emergence of ideas around alternative mechanisms of translation initiation (Babitzke and O'Connor 2017). Why are there different mechanisms across species? Hockenberry et al. attempted to answer this question. They found covariance of SD sequence usage and factors such as high growth rates indicating the necessity of high efficiency translation (Hockenberry et al. 2018). It could be that SD-antiSD sequences facilitate higher rates of translation initiation. However, the discovery of alternative mechanisms of translation initiation in bacteria further our understanding of how translation is controlled by the cell and provide opportunities for designing new expression systems in metabolic engineering. Such discoveries may also suggest why current designs are proving unpredictable in the resulting expression level.

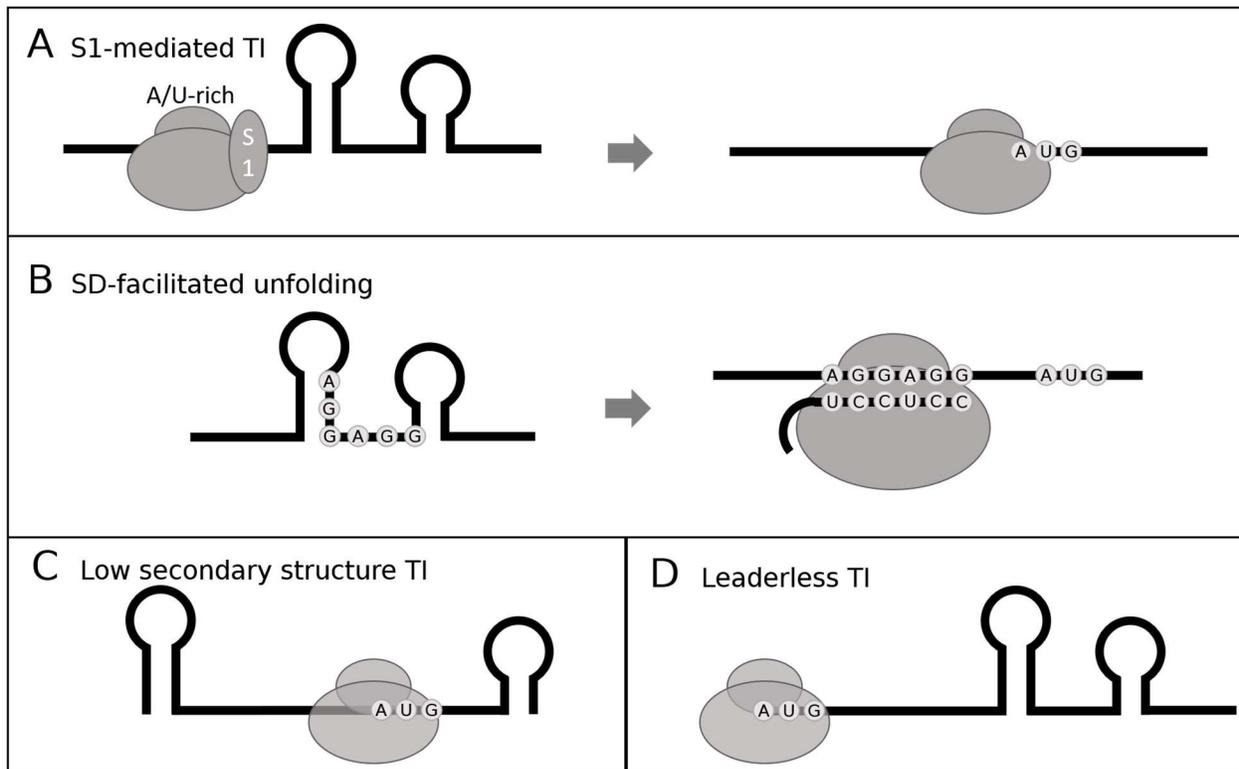


Figure 3.2: Translation initiation may occur *via* several different mechanisms. A) The 30S ribosome can bind an A/U-rich region of the mRNA with low secondary structure. The ribosomal S1 protein can act as an RNA chaperone, gradually reducing the secondary structure that prevents binding the start codon. B) and C) Under the “Unified View,” SD-antiSD base-pairing can compensate for high energetic penalty of unfolding strong mRNA structures (B), while translation initiation regions with low secondary structure strength do not require such compensation (C). D) 70S ribosomes can also initiation translation of mRNA that lack a 5'-UTR.

S1-mediated translation initiation

Several studies have implicated the ribosomal S1 protein in translation initiation. S1 is generally thought to bind A/U-rich regions of mRNA and act as an RNA chaperone that ratchets open secondary structure. This facilitates 30S subunit binding of the RBS and subsequent translation initiation.

The S1 protein was shown to bind U-enriched sequences of phage RNA, as well as the U-rich region of *E. coli* *ssb* mRNA that precedes the SD sequence (Boni et al. 1991). S1 is essential for translation initiation in *E. coli*. Sorensen, Fricke, and Pedersen deleted the *rpsA* gene that encodes the ribosomal protein S1 in *E. coli*. Growth was only possible when this knockout strain was induced to express a copy of the gene from a plasmid. Their studies also revealed that reduced S1 expression resulted in depletion of amino

acids, and that ribosomes depleted of S1 failed to elongate peptide chains (Sørensen, Fricke, and Pedersen 1998). Previously, mRNA with strong interaction between the SD and antiSD was reported to obviate the need for S1 in translation initiation (Farwell, Roberts, and Rabinowitz 1992).

Mutsuda and Sugiura (2006) studied the translation initiation of wild-type and mutant mRNA for the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) in *Synechococcus elongatus* PCC7942 and PCC6301 using an in vitro translation system. The translation initiation region of the *rbcS* gene contains a SD sequence of AGGAU. Mutation of this sequence to ACCAU should reduce SD-antiSD interaction. This mutation actually led to an increase in the translation of *rbcS*. In contrast, deletions of the pyrimidine-rich sequence between -46 and -14 from the start codon (though still 5' of the SD sequence) dramatically lowered translation of the gene. The authors confirmed by immunoprecipitation that the ribosomal S1 protein binds the -46 to -14 region of the *rbcS* mRNA and that deleting nucleotides in that region reduced or abolished S1 binding. The S1-mRNA interaction appears to be the most important sequence element for translation initiation of *rbcS*. Unlike the *E. coli* S1 protein which binds an AC-rich region, the *Synechococcus* S1 binds a CU-rich region (Mutsuda and Sugiura 2006).

There are two analogs to the *E. coli* S1 protein which are 31 and 38 kDa compared with 61 kDa for the *E. coli* S1. *S. 6803* also has two sequence homologs of similar size to those in *S. 6301*, *rbcS1A* and *rbcS1B*. These proteins have not been characterized at this time. Mutsuda and Sugiura did not provide direct evidence contrary to the Salah et al. (2009) results that suggest that cyanobacteria S1 protein may not bind the ribosome. Due to these contradictory results, further experimentation is needed to confirm the role of the S1 protein and its interaction with the ribosome in cyanobacteria.

The structure of the 30S subunit with bound S1 (*E. coli*) was determined in 2001, which identified the S1-30S interface and confirmed that the S1 protein bound the 5'-end of the mRNA bound to the ribosome

(Sengupta, Agrawal, and Frank 2001). Salah et al. (2009) examined the sequence similarities of bacteria. They found that *Synechococcus sp.* (strain not stated) S1 appears to lack domains 1, 4/5, and 6 found in *E. coli* S1, and contains an additional domain not found in *E. coli* S1. It contains only three domains compared with six found in *E. coli* (Salah et al. 2009). In *E. coli*, domain 1 is essential for binding the ribosome. Domains 3, 4, and 5 are all involved in binding RNA. The function of domain 6 is not known, and it is not essential for translation.

S1-mediated translation initiation does not provide a full explanation in species of archaea and bacteria which lack the S1 protein entirely (many gram-positive bacteria), or which lack an S1 protein with activity in translation (Babitzke and O'Connor 2017). In those species with functional S1 protein, what is the mechanism by which it assists translation initiation? Conformational switching in the *Vibrio vulnificus* S1 protein assists with unfolding mRNA in a step-wise manner with domains D3 and D4 operating in tandem (Qureshi et al. 2018). This mechanism contrasts slightly with what has been observed for *E. coli* S1-mediated mRNA unfolding. Qu et al. used optical tweezers to monitor *E. coli* S1 binding mRNA, concluding that individual domains of S1 molecules capture single-stranded mRNA during the brief times secondary structure unfolds, each occupying about 10 nucleotides and preventing it from refolding (Qu et al. 2012).

Romilly et al. conducted an in-depth study to characterize the standby site of the *tisB* mRNA expressed in *E. coli*. Under normal growth, a small RNA (IstRI) hybridizes with the fifteen-nucleotide single-stranded standby site which inhibits 30S ribosome binding and promotes degradation of the mRNA by RNase III. Under stress conditions, IstRI is expressed at a lower level compared with *tisB*, allowing the ribosome to bind the standby site and subsequently initiate translation. The ribosomal S1 protein was essential for 30S subunits to bind the standby site, and S1 proteins were also found (independent of the 30S subunit) to bind the standby site. A stem-loop structure 5' of the standby site also appeared to be essential to 30S binding of the standby site, suggesting to the authors that structures at both ends of the standby

site can stabilize the 30S-mRNA complex. An aspect of the *tisB* standby site which is contrary to the general model of standby sites is its remoteness from the RBS – approximately 100 nucleotides distant (Romilly, Deindl, and Wagner 2019).

70S scanning and leaderless initiation

Yamamoto et al. (2016) demonstrated that 70S scanning can also initiate translation in *E. coli*. Although the 70S ribosome could not bind directly to RBSs in their study, the full 70S ribosome was shown to continue scanning an mRNA after translation termination and re-initiate at secondary open reading frames (Yamamoto et al. 2016). This study also showed that IF3 was bound to the 70S ribosome during initiation of both leaderless mRNA and in re-initiation. This may be the primary mechanism for initiating translation of genes in operons. Further, this lab found that IF3 binding to the 70S ribosome promoted initiation of leaderless transcripts, while IF3 binding of 30S ribosomes prevents initiation of leaderless transcripts. This is used to explain cold shock responses and energy source depletion responses where 70S ribosomes accumulate to a higher concentration, out competing 30S for IF3 binding, and activating translation of leaderless transcripts (Goyal, Belardinelli, and Rodnina 2017).

At least in a few cases, the interactions between the 5'-UTR of the mRNA and ribosome are not used at all because the start codon is coincident with the 5'-end of the mRNA. A few bacterial mRNAs have been identified that have a start codon either coincident with the 5' terminus or within a few nucleotides of it, precluding any SD-antiSD interactions. Only about 40 such mRNA were identified by Moll et al. in 2002. More recently, certain species of bacteria and archaea have been found have many leaderless transcripts making up 26% up to 69% of all transcripts (Brenneis et al. 2007; Shell et al. 2015; Wurtzel et al. 2010).

Fully assembled 70S ribosome including fMet-tRNA binds the start codon. Brock et al. used photo-crosslinking to study initiation of leaderless phage transcripts and the ability of 70S particles to bind the

start codon. They also observed that a synthetic random transcript with an AUG start codon at the 5' end could be bound by the 70S ribosome as well, suggesting that the AUG sequence was sufficient signal to initiate translation of leaderless transcripts (Brock et al. 2008).

Transcription-Translation Coupling

Transcription and translation can become coupled such that a ribosome follows close behind the RNA polymerase, commencing translation before the mRNA is fully transcribed. Interactions that have been observed between the ribosome and RNA polymerase suggest that the ribosome may bind the RBS before secondary structure can form within the mRNA, overcoming a typical barrier to translation initiation. Conn et al. recently reviewed the literature on this topic to construct a model of the mechanisms involved. In their model, transcription pauses 100-150 nucleotides into the transcript, allowing the 30S ribosome to bind. It is hypothesized that the ribosome then assists the RNAP in releasing the pause site and continuing transcription.

In factor-free coupling, the RNAP and 30S ribosomal subunit interact directly, while in factor-mediated coupling, the factor RfaH tethers the two together (Conn et al. 2019). Kohler et al. performed cryo-EM to determine the structures of molecules involved and their interfaces. They described interacting RNAP and ribosomes engaged in coupled transcription-translation as an “expressome” complex (Kohler et al. 2017). The NusG transcription factor is used in what described by Conn et al. as a hybrid between factor-free and factor-mediated coupling mechanisms. NusG forms part of the interface between the ribosome and RNAP in *E. coli* but, unlike RfaH, the NusG arrives to the complex after it leaves the pause site. This interaction between NusG and the RNAP and ribosome is not stable, and the binding of adjacent sequences of mRNA are essential to holding this complex together (Webster et al. 2020). Interestingly, NusG in *S. 6803* has only 41% identity with the NusG in *E. coli*. NusA has also been observed tethering the ribosome to RNAP (C. Wang et al. 2020). In cell measurements have also been performed in

Mycoplasma pneumoniae that show both NusA and NusG are involved in the expressome (O'Reilly et al. 2020)

Polysomes

Multiple ribosomes can bind the standby sites of some mRNA, which can lead to the formation of polysomes which result in high translation rates of those mRNA. Once the first ribosome initiates translation, the next can potentially move to the start codon before mRNA folding can occur. This allows the second, and subsequent, ribosomes to avoid one of the barriers to translation initiation. The ribosomal proteins S1 and S2 were essential for this rapid second initiation (Andreeva, Belardinelli, and Rodnina 2018).

Jacobson and Baldassare estimated that 73% of ribosomes in *E. coli* were within polysomes. However, their definition of polysome consisted simply of any mRNA with more than one ribosome bound. They did suggest a slightly different mechanism for ribosome loading than the standby site model – a second ribosome may bind the unfolded mRNA of the RBS that the first ribosome recently vacated (Jacobson and Baldassare 1976).

Operon Re-initiation

Ma et al. (2002) observed that genes in operons across prokaryotic phyla had a higher likelihood of having SD sequence than other genes (Ma, Campbell, and Karlin 2002).

Burkhardt examined translation efficiency of genome-wide operons in *E. coli*. Of the mRNA signals they looked at, the secondary structure of the entire ORF had the strongest impact on translation efficiency. They found low secondary structure around translation initiation regions, and observed that each ORF had a characteristic level of secondary structure. In downstream ORFs, secondary structure propensity was generally decreased just downstream of the start codon, and some specific mRNA structures were proposed as important signals for translation initiation in those genes (Burkhardt et al. 2017).

Translation initiation in cyanobacteria

Wei and Xia (2019) examined translation initiation elements in *Synechocystis*, concluding that SD sequences were found mainly in highly-expressed genes, and that mRNA that lacked an SD sequence compensated by having reduced secondary structure in the ribosome binding site (Wei and Xia 2019). Wei and Xia also examined the frequency each SD-like sequence was used by *S. 6803*, with GGAG, AGGA, and AGGAG being most common representing around 10% each and many other sequences used. When comparing the protein expression levels of mRNA containing an SD and those that don't, the average expression level of those with an SD was slightly higher, though the distributions significantly overlap. It is not a clear-cut rule that an SD should indicate higher expression. Finally, Wei and Xia calculated the free energy of folding the 40 nucleotides 5' of the start codon and compared the expression levels to that energy for the mRNA that contained an SD sequence and those that did not. They found that sequence that had an SD sequence tended to have stronger folding.

Evidence that things are different in cyanobacteria

It is well known that base pairing between the 3' end of the 16s ribosomal RNA and the Shine-Dalgarno sequence in the 5'-untranslated region of the mRNA is an important step in the process of translation initiation. Perhaps its importance has been overstated. Such interaction may facilitate translation initiation for some genes, in some bacterial species. However, only a fraction of mRNA in a given species actually contains a recognizable Shine-Dalgarno sequence. In addition, prokaryotes that do have a high proportion of SD genes, may also utilize alternative mechanisms for the genes that lack the SD sequence.

Several studies have found reduced prevalence of the SD sequence in the 5'-UTR of cyanobacteria, suggesting a reduced reliance on this mechanism for initiating translation. Ma et al. (2002) examined the translation initiation regions of 30 genomes and found that 57% of *E. coli* genes had a SD sequence,

while in *Synechocystis* only 26% of genes had the SD sequence. They also reported a correlation between the presence of a SD sequence and expression level, although expression levels were predicted based only on codon usage (Ma, Campbell, and Karlin 2002). Starmer et al. (2006) used SD-antiSD hybridization energies to identify SD-initiated genes. In the cyanobacteria *Anabaena* sp. PCC7120 and *Synechocystis* sp. PCC6803, only a small proportion of genes had upstream SD sequences (31.1% and 27.5%, respectively) (Starmer et al. 2006). Cyanobacteria species were among the ~10% of genomes that Hyatt et al. determined must rely on translation mechanisms other than SD-antiSD hybridization because few 5'-UTRs contained the SD motif (Hyatt et al. 2010). This group concluded that "*Synechocystis* PCC6803, like most cyanobacteria, does not seem to use the SD motif at all, and instead favors AT-rich regions upstream of its translation start sites."

The study by Nakagawa et al. (2010), again, showed that cyanobacteria have a much lower proportion of genes with strong SD-antiSD interaction than other phyla. The authors analyzed the proportion of known genes in 277 species of prokaryotes that contain a recognizable Shine-Dalgarno sequence. They observed that three unrelated phyla had much a smaller proportion of mRNA's that contained a Shine-Dalgarno sequence, including Nanoarchaeota, Bacteroidetes, and cyanobacteria. The authors suggested that other mechanisms of translation initiation must be used in these organisms. Two possible mechanisms suggested were initiation mediated by the ribosomal S1 protein and initiation of leaderless transcripts for which the start codon coincides with the transcription start site. Others have observed a distinct lack of Shine-Dalgarno sequences in Bacteroidetes and found evidence that the ribosomal S1 protein is used to identify start codons (Accetto and Avguštin 2011). They also suggested that the SD sequence may not necessarily promote translation initiation in species such as these (Nakagawa et al. 2010).

Scharff et al (2011) examined both the SD-antiSD hybridization energy and the mRNA folding energy of genes in prokaryotes and chloroplasts. They estimated that about 50% of all cyanobacterial genes have

SD sequences. Overall, the genes studied tended to have reduced secondary structure (a peak in free energy of folding) around the start codon, suggesting the accessibility of the start codon may be one mechanism that cells use to select correct start codons. For genes that lack SD sequences, secondary structures were even weaker, and in cyanobacteria this difference in the peak heights was greater than in other prokaryotes. Echoing the “unified view” of Nakamoto, the authors proposed that there may be two unique pathways to translation initiation in prokaryotes: one in which accessibility alone is used to select the start codon, and another in which a combination of SD sequence and accessibility are used together. RBSs with SD sequences are less sensitive to the degree of folding free energy. Free energy of folding was also found to be higher (weaker) around correct start codons than around those internal to the gene (both in and out of frame) (Scharff et al. 2011).

Nakagawa and colleagues followed up their previous genomic analysis of translation initiation in 2017 and examined the nucleotides on both sides of the start codon of genes that lacked SD sequences. They found symmetry in the nucleotide biases which may prevent the formation of secondary structures. In cyanobacteria as in most other prokaryotes, free energy of folding calculations confirmed that secondary structures were weaker in the translation initiation region for genes that lacked SD sequences. This work also confirmed a finding from Starmer et al. (2006), that hybridization between the 16s rRNA and transcripts generally has strong binding right at the start codon. In most species this binding is stronger in non-SD-initiated genes than SD-initiated genes. However, in cyanobacteria, this binding was similarly strong for both types of genes. Strong biases were found in nonSD mRNA in cyanobacteria towards A and U were found in the -10 region (from the start codon A), and a very strong bias towards C/U was also found in the -2 position (So Nakagawa, Niimura, and Gojobori 2017). Instead of C/U, Sazuka and Ohara (1996) found a bias across all genes in *S. 6803* for having two pyrimidines (A/U) just 5' of the start codon, followed by a purine 3' of the start codon (Sazuka and Ohara 1996).

Are there any hints from chloroplasts or cyanophages?

Wei and Xia, 2019 compared translation initiation elements in *Synechocystis* and tobacco chloroplasts. They observed two prevalent species of the 16S rRNA in *S. 6803*: one ending with CACCUCCU-3' and the other, more frequently found, having an additional UU at the 3'-end. Both sequences lacked the AAGGG at the 3'-end that was previously annotated as the end. SD-antiSD base pairing tended to align the 3' end of the 16S rRNA to a nucleotide 16-18 nucleotides 5' of the start codon in *S. 6803*. Chloroplast genes examined used SD sequences less often and when they were used, those genes did not tend to have higher expression as they did in cyanobacteria. Cyanophage genes examined showed a similar pattern of SD sequences and SD-start codon distance as was found in the cyanobacteria (Wei and Xia 2019).

The genes endogenous to chloroplasts mainly consist of those coding for proteins involved in photosynthesis and some involved in protein translation. Homologs to bacterial ribosomal RNAs, tRNAs, initiation factors, elongation factor EF-Tu, and others have been identified in nuclear and chloroplast genomes of plants and algae. Like most bacteria genomes, chloroplast genomes contain both genes with SD sequences and those without. In general, mutations of the SD sequences in chloroplast-translated mRNA only slightly reduced expression levels of proteins (Zerges 2000). Chloroplast mRNA often do have SD sequences, though they have more variable spacing from the start codon than they do in *E. coli* (Sugiura, Hirose, and Sugita 1998). Three elements of the *psbD* gene in the chloroplast of *Chlamydomonas reinhardtii* were identified as being important to translation initiation and the expression level: 1) a stem loop important for mRNA stability, a U-rich region from -20 to -15, and an antiSD -14 to -11 from the start codon (Nickelsen et al. 1999).

Regulation of translation in cyanobacteria

In a couple cases, translation regulatory mechanisms of specific genes are well studied. The translation of photosystem II protein D1 in *S. 6803* is regulated by a complex mechanism involving a pause site

within the mRNA and targeting of the nascent peptide chain to the thylakoid membrane. Ribosomes were found to pause at a specific position in the mRNA under dark conditions. After transition to light conditions, translation resumed and the D1 peptide was targeted to the thylakoid membrane (Tyystjarvi, Herranen, and Aro 2001). In the filamentous cyanobacterium *Fremyella diplosiphon*, the Initiation Factor III encoded by *infC* regulates translation of the *cpeC* operon that encodes photosynthetic light-harvesting proteins, which is part of a chromatic acclimation process. Possibly, this IF3 regulates *cpeC* translation initiation by modulating the degree of transcriptional-translational coupling (Gutu et al. 2013). Riboswitches and riboregulators used in cyanobacterial biotechnology have recently been reviewed (Till et al. 2020).

Importance for metabolic engineering & synthetic biology

Many researchers have been engaged in investigating the use of cyanobacteria as catalysts to convert sunlight and carbon dioxide into fuels and other valuable chemicals. Generally, we have attempted to replicate the successes metabolic engineers have had in controlling heterologous gene expression in *E. coli*. Genetic parts such as promoters and ribosome binding sites have been imported to cyanobacteria, having been first characterized in *E. coli*. Genetic parts native to cyanobacteria have also been developed and characterized following the methods previously used in *E. coli*. Computational design tools such as the RBS Calculator have also been utilized to design genetic parts for use in cyanobacteria. This has been successful in some studies (Markley et al. 2015), and not in others (B. Wang et al. 2018; Sebesta and Peebles 2020). Bacteria may rely more, or less, on each of the many different mechanisms for initiating translation. A more complete understanding of each of those mechanisms is necessary because different mechanisms are all encoded in the same region of mRNA and each change to the sequence invokes trade-offs between the different mechanisms.

Conclusions and Future Directions

Translation initiation is thought to be a rate limiting step in protein synthesis. However, many other processes affect this rate. First and foremost, the rate of transcription, controlled by promoter sequences, as well as the rate of mRNA degradation control the abundance of the mRNA. The translation elongation rate is another variable, controlled by codon usage, and ribosomal pausing within the transcript can reduce the elongation rate. Protein degradation rates also affect the protein concentrations observed in cells. The translation initiation rate and elongation rate determine the occupancy of ribosomes in the mRNA. Regions of mRNA bound by ribosomes (or other proteins) are protected from degradation by RNases such as RNase E. mRNAs with higher ribosome occupancy, therefore, have longer lifetimes (Yarchuk et al. 1992; Deana and Belasco 2005). Widespread regulation of translation and mRNA degradation by anti-sense RNA may also occur in cyanobacteria (Georg et al. 2009). It is not clear at this time how the polyploidy of cyanobacterial chromosomes may impact how cells control translation initiation.

It has been proposed that cyanobacteria use the ribosomal S1 protein to initiate translation using A/U or C/U-rich sequences, but this protein has not been characterized. Detailed study of cyanobacterial S1 protein is needed, similar to what was conducted with *E. coli* S1 protein by Duval et al. (2013). It would be informative to also follow the work by Studer and Joseph and use cyanobacteria ribosomes and mRNA to study binding kinetics and mechanism of initiation. Given the widespread transcription of antisense RNA in *Synechocystis* found by Georg et al. (2009), could translation be widely regulated in cyanobacteria by a mechanisms similar to that of flv4 (Eisenhut et al. 2012) and psbA (Sakurai et al. 2012) in *Synechocystis* sp. PCC6803, and rpsO in *E. coli* (Marzi et al. 2007)? Is such regulation more common in cyanobacteria than it is in *E. coli*?

New studies are needed to determine what controls translation efficiency in cyanobacteria. To benefit biological engineering efforts in cyanobacteria, those studies should identify which genetic elements are necessary for reliably controlling translation efficiency or determine if translation initiation mechanisms can be simplified in order to design reliable parts. Ideally, a large library of random RBS sequences would be characterized and machine learning applied to identify the determinants of translation efficiency in cyanobacteria as has been done by Barendt et al. and Evfratov et al. in *E. coli* (Barendt et al. 2013; Evfratov et al. 2017). Such studies should also take into account the findings of Thiel et al. which suggested that different coding sequences of reporter genes and genes of interest impact the relative “strength” of RBSs (Thiel et al. 2018).

References

- Accetto, Tomaž, and Gorazd Avguštin. 2011. "Inability of *Prevotella bryantii* to Form a Functional Shine-Dalgarno Interaction Reflects Unique Evolution of Ribosome Binding Sites in Bacteroidetes." Edited by Grzegorz Kudla. *PLoS ONE* 6 (8): e22914. <https://doi.org/10.1371/journal.pone.0022914>.
- Andreeva, Irena, Riccardo Belardinelli, and Marina V. Rodnina. 2018. "Translation Initiation in Bacterial Polysomes through Ribosome Loading on a Standby Site on a Highly Translated mRNA." *Proceedings of the National Academy of Sciences* 115 (17): 4411–16. <https://doi.org/10.1073/pnas.1718029115>.
- Babitzke, Paul, and Michael O'Connor. 2017. "Noncanonical Translation Initiation Comes of Age." Edited by Tina M. Henkin. *Journal of Bacteriology* 199 (14): e00295-17. <https://doi.org/10.1128/JB.00295-17>.
- Baers, Laura L., Lisa M. Breckels, Lauren A. Mills, Laurent Gatto, Michael J. Deery, Tim J. Stevens, Christopher J. Howe, Kathryn S. Lilley, and David J. Lea-Smith. 2019. "Proteome Mapping of a Cyanobacterium Reveals Distinct Compartment Organization and Cell-Dispersed Metabolism." *Plant Physiology* 181 (4): 1721–38. <https://doi.org/10.1104/pp.19.00897>.
- Barendt, Pamela A., Najaf A. Shah, Gregory A. Barendt, Parth A. Kothari, and Casim A. Sarkar. 2013. "Evidence for Context-Dependent Complementarity of Non-Shine-Dalgarno Ribosome Binding Sites to *Escherichia coli* rRNA." *ACS Chemical Biology* 8 (5): 958–66. <https://doi.org/10.1021/cb3005726>.
- Barendt, Pamela A., Najaf A. Shah, Gregory A. Barendt, and Casim A. Sarkar. 2012. "Broad-Specificity mRNA–rRNA Complementarity in Efficient Protein Translation." Edited by Diarmaid Hughes. *PLoS Genetics* 8 (3): e1002598. <https://doi.org/10.1371/journal.pgen.1002598>.
- Belinky, Frida, Igor B. Rogozin, and Eugene V. Koonin. 2017. "Selection on Start Codons in Prokaryotes and Potential Compensatory Nucleotide Substitutions." *Scientific Reports* 7 (1). <https://doi.org/10.1038/s41598-017-12619-6>.
- Boni, Irina V., Dilara M. Isaeva, Maxim L. Musychenko, and Nina V. Tzareva. 1991. "Ribosome-Messenger Recognition: mRNA Target Sites for Ribosomal Protein S1." *Nucleic Acids Research* 19 (1): 155–62. <https://doi.org/10.1093/nar/19.1.155>.
- Brenneis, Mariam, Oliver Hering, Christian Lange, and Jörg Soppa. 2007. "Experimental Characterization of Cis-Acting Elements Important for Translation and Transcription in Halophilic Archaea." Edited by William F Burkholder. *PLoS Genetics* 3 (12): e229. <https://doi.org/10.1371/journal.pgen.0030229>.
- Brock, J. E., S. Pourshahian, J. Giliberti, P. A. Limbach, and G. R. Janssen. 2008. "Ribosomes Bind Leaderless mRNA in *Escherichia coli* through Recognition of Their 5'-Terminal AUG." *RNA* 14 (10): 2159–69. <https://doi.org/10.1261/rna.1089208>.
- Burkhardt, David H, Silvi Rouskin, Yan Zhang, Gene-Wei Li, Jonathan S Weissman, and Carol A Gross. 2017. "Operon MRNAs Are Organized into ORF-Centric Structures That Predict Translation Efficiency." *eLife* 6 (January). <https://doi.org/10.7554/eLife.22037>.
- Conn, Adam B., Stephen Diggs, Timothy K. Tam, and Gregor M. Blaha. 2019. "Two Old Dogs, One New Trick: A Review of RNA Polymerase and Ribosome Interactions during Transcription-Translation Coupling." *International Journal of Molecular Sciences* 20 (10): 2595. <https://doi.org/10.3390/ijms20102595>.
- Deana, A., and Joel G. Belasco. 2005. "Lost in Translation: The Influence of Ribosomes on Bacterial mRNA Decay." *Genes & Development* 19 (21): 2526–33. <https://doi.org/10.1101/gad.1348805>.

- Duval, Mélodie, Alexey Korepanov, Olivier Fuchsbauer, Pierre Fechter, Andrea Haller, Attilio Fabbretti, Laurence Choulier, et al. 2013. "Escherichia Coli Ribosomal Protein S1 Unfolds Structured MRNAs Onto the Ribosome for Active Translation Initiation." Edited by John Hershey. *PLoS Biology* 11 (12): e1001731. <https://doi.org/10.1371/journal.pbio.1001731>.
- Eisenhut, Marion, Jens Georg, Stephan Klähn, Isamu Sakurai, Henna Mustila, Pengpeng Zhang, Wolfgang R. Hess, and Eva-Mari Aro. 2012. "The Antisense RNA As1_flv4 in the Cyanobacterium *Synechocystis* Sp. PCC 6803 Prevents Premature Expression of the *Flv4-2* Operon upon Shift in Inorganic Carbon Supply." *Journal of Biological Chemistry* 287 (40): 33153–62. <https://doi.org/10.1074/jbc.M112.391755>.
- Evfratov, Sergey A., Ilya A. Osterman, Ekaterina S. Komarova, Alexandra M. Pogorelskaya, Maria P. Rubtsova, Timofei S. Zatselin, Tatiana A. Semashko, et al. 2017. "Application of Sorting and next Generation Sequencing to Study 5'-UTR Influence on Translation Efficiency in Escherichia Coli." *Nucleic Acids Research* 45 (6): 3487–3502. <https://doi.org/10.1093/nar/gkw1141>.
- Farwell, Mary A., Mark W. Roberts, and Jesse C. Rabinowitz. 1992. "The Effect of Ribosomal Protein S1 from Escherichia Coli and Micrococcus Luteus on Protein Synthesis in Vitro by E. Coli and Bacillus Subtilis." *Molecular Microbiology* 6 (22): 3375–83. <https://doi.org/10.1111/j.1365-2958.1992.tb02205.x>.
- Georg, Jens, Björn Voß, Ingeborg Scholz, Jan Mitschke, Annegret Wilde, and Wolfgang R Hess. 2009. "Evidence for a Major Role of Antisense RNAs in Cyanobacterial Gene Regulation." *Molecular Systems Biology* 5 (September). <https://doi.org/10.1038/msb.2009.63>.
- Gingold, H., and Y. Pilpel. 2011. "Determinants of Translation Efficiency and Accuracy." *Molecular Systems Biology* 7 (1): 481–481. <https://doi.org/10.1038/msb.2011.14>.
- Goyal, Akanksha, Riccardo Belardinelli, and Marina V. Rodnina. 2017. "Non-Canonical Binding Site for Bacterial Initiation Factor 3 on the Large Ribosomal Subunit." *Cell Reports* 20 (13): 3113–22. <https://doi.org/10.1016/j.celrep.2017.09.012>.
- Gualerzi, Claudio O., and Cynthia L. Pon. 2015. "Initiation of mRNA Translation in Bacteria: Structural and Dynamic Aspects." *Cellular and Molecular Life Sciences* 72 (22): 4341–67. <https://doi.org/10.1007/s00018-015-2010-3>.
- Gutu, Andrian, April D. Nesbit, Andrew J. Alverson, Jeffrey D. Palmer, and David M. Kehoe. 2013. "Unique Role for Translation Initiation Factor 3 in the Light Color Regulation of Photosynthetic Gene Expression." *Proceedings of the National Academy of Sciences of the United States of America* 110 (40): 16253–58. <https://doi.org/10.1073/pnas.1306332110>.
- Hockenberry, Adam J, Aaron J Stern, Luís A N Amaral, and Michael C Jewett. 2018. "Diversity of Translation Initiation Mechanisms across Bacterial Species Is Driven by Environmental Conditions and Growth Demands." *Molecular Biology and Evolution* 35 (3): 582–92. <https://doi.org/10.1093/molbev/msx310>.
- Hüttenhofer, A., and H.F. Noller. 1994. "Footprinting mRNA-Ribosome Complexes with Chemical Probes." *The EMBO Journal* 13 (16): 3892–3901. <https://doi.org/10.1002/j.1460-2075.1994.tb06700.x>.
- Hyatt, Doug, Gwo-Liang Chen, Philip F LoCascio, Miriam L Land, Frank W Larimer, and Loren J Hauser. 2010. "Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification." *BMC Bioinformatics* 11 (1). <https://doi.org/10.1186/1471-2105-11-119>.
- Jacobson, L. A., and J. C. Baldassare. 1976. "Association of Messenger Ribonucleic Acid with 70S Monosomes from Down-Shifted Escherichia Coli." *Journal of Bacteriology* 127 (1): 637–43. <https://doi.org/10.1128/JB.127.1.637-643.1976>.
- K.E. van Holde, and W.E. Hill. 1974. "General Physical Properties of Ribosomes." In *Ribosomes*, edited by Masayasu Nomura, Alfred Tissieres, and Peter Lengyel, 1974th ed., 4:53–91. CSH Monographs. Cold Spring Harbor Laboratory.

- Kohler, R., R. A. Mooney, D. J. Mills, R. Landick, and P. Cramer. 2017. "Architecture of a Transcribing-Translating Expressome." *Science* 356 (6334): 194–97. <https://doi.org/10.1126/science.aal3059>.
- Komarova, Anastassia V., Ludmila S. Tchufistova, Elena V. Supina, and Irina V. Boni. 2002. "Protein S1 Counteracts the Inhibitory Effect of the Extended Shine-Dalgarno Sequence on Translation." *RNA (New York, N.Y.)* 8 (9): 1137–47.
- Kozak, Marilyn. 2005. "Regulation of Translation via mRNA Structure in Prokaryotes and Eukaryotes." *Gene* 361 (November): 13–37. <https://doi.org/10.1016/j.gene.2005.06.037>.
- Laursen, Brian Sjøgaard, Hans Peter Sørensen, Kim Kusk Mortensen, and Hans Uffe Sperling-Petersen. 2005. "Initiation of Protein Synthesis in Bacteria." *Microbiology and Molecular Biology Reviews: MMBR* 69 (1): 101–23. <https://doi.org/10.1128/MMBR.69.1.101-123.2005>.
- Ma, Jiong, Allan Campbell, and Samuel Karlin. 2002. "Correlations between Shine-Dalgarno Sequences and Gene Features Such as Predicted Expression Levels and Operon Structures." *Journal of Bacteriology* 184 (20): 5733–45.
- Markley, Andrew L., Matthew B. Begemann, Ryan E. Clarke, Gina C. Gordon, and Brian F. Pflieger. 2015. "Synthetic Biology Toolbox for Controlling Gene Expression in the Cyanobacterium *Synechococcus* Sp. Strain PCC 7002." *ACS Synthetic Biology* 4 (5): 595–603. <https://doi.org/10.1021/sb500260k>.
- Marzi, Stefano, Alexander G. Myasnikov, Alexander Serganov, Chantal Ehresmann, Pascale Romby, Marat Yusupov, and Bruno P. Klaholz. 2007. "Structured MRNAs Regulate Translation Initiation by Binding to the Platform of the Ribosome." *Cell* 130 (6): 1019–31. <https://doi.org/10.1016/j.cell.2007.07.008>.
- Moll, I., M. Huber, S. Grill, P. Sairafi, F. Mueller, R. Brimacombe, P. Londei, and U. Blasi. 2001. "Evidence against an Interaction between the mRNA Downstream Box and 16S rRNA in Translation Initiation." *Journal of Bacteriology* 183 (11): 3499–3505. <https://doi.org/10.1128/JB.183.11.3499-3505.2001>.
- Mutalik, Vivek K, Joao C Guimaraes, Guillaume Cambay, Colin Lam, Marc Juul Christoffersen, Quynh-Anh Mai, Andrew B Tran, et al. 2013. "Precise and Reliable Gene Expression via Standard Transcription and Translation Initiation Elements." *Nature Methods* 10 (4): 354–60. <https://doi.org/10.1038/nmeth.2404>.
- Mutsuda, Michinori, and Masahiro Sugiura. 2006. "Translation Initiation of Cyanobacterial *RbcS* MRNAs Requires the 38-KDa Ribosomal Protein S1 but Not the Shine-Dalgarno Sequence: DEVELOPMENT OF A CYANOBACTERIAL IN VITRO TRANSLATION SYSTEM." *Journal of Biological Chemistry* 281 (50): 38314–21. <https://doi.org/10.1074/jbc.M604647200>.
- Nakagawa, S., Y. Niimura, K.-i. Miura, and T. Gojobori. 2010. "Dynamic Evolution of Translation Initiation Mechanisms in Prokaryotes." *Proceedings of the National Academy of Sciences* 107 (14): 6382–87. <https://doi.org/10.1073/pnas.1002036107>.
- Nakagawa, So, Yoshihito Niimura, and Takashi Gojobori. 2017. "Comparative Genomic Analysis of Translation Initiation Mechanisms for Genes Lacking the Shine-Dalgarno Sequence in Prokaryotes." *Nucleic Acids Research* 45 (7): 3922–31. <https://doi.org/10.1093/nar/gkx124>.
- Nakamoto, Tokumasa. 2006. "A Unified View of the Initiation of Protein Synthesis." *Biochemical and Biophysical Research Communications* 341 (3): 675–78. <https://doi.org/10.1016/j.bbrc.2006.01.019>.
- Nakamoto, Tokumasa, and Brigitta Vogl. 1978. "On the Accessibility and Selection of the Initiator Site of mRNA in Protein Synthesis." *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis* 517 (2): 367–77. [https://doi.org/10.1016/0005-2787\(78\)90203-4](https://doi.org/10.1016/0005-2787(78)90203-4).
- Neidhardt, Frederick C., and Roy Curtiss, eds. 1996. *Escherichia Coli and Salmonella: Cellular and Molecular Biology*. 2nd ed. Washington, D.C: ASM Press.

- Nickelsen, Jörg, Mark Fleischmann, Eric Boudreau, Michele Rahire, and Jean-David Rochaix. 1999. "Identification of *Cis*-Acting RNA Leader Elements Required for Chloroplast *PsbD* Gene Expression in *Chlamydomonas*." *The Plant Cell* 11 (5): 957–70. <https://doi.org/10.1105/tpc.11.5.957>.
- Olsthoorn, R. C., and J. van Duin. 1996. "Random Removal of Inserts from an RNA Genome: Selection against Single-Stranded RNA." *Journal of Virology* 70 (2): 729–36. <https://doi.org/10.1128/JVI.70.2.729-736.1996>.
- Olsthoorn, R.C., N. Lisis, and J. van Duin. 1994. "Leeway and Constraints in the Forced Evolution of a Regulatory RNA Helix." *The EMBO Journal* 13 (11): 2660–68. <https://doi.org/10.1002/j.1460-2075.1994.tb06556.x>.
- Olsthoorn, René C. L., Stephen Zoog, and Jan Duin. 1995. "Coevolution of RNA Helix Stability and Shine-Dalgarno Complementarity in a Translational Start Region." *Molecular Microbiology* 15 (2): 333–39. <https://doi.org/10.1111/j.1365-2958.1995.tb02247.x>.
- O'Reilly, Francis J., Liang Xue, Andrea Graziadei, Ludwig Sinn, Swantje Lenz, Dimitry Tegunov, Cedric Blötz, et al. 2020. "In-Cell Architecture of an Actively Transcribing-Translating Expressome." *Science* 369 (6503): 554–57. <https://doi.org/10.1126/science.abb3758>.
- Qu, Xiaohui, Laura Lancaster, Harry F. Noller, Carlos Bustamante, and Ignacio Tinoco. 2012. "Ribosomal Protein S1 Unwinds Double-Stranded RNA in Multiple Steps." *Proceedings of the National Academy of Sciences of the United States of America* 109 (36): 14458–63. <https://doi.org/10.1073/pnas.1208950109>.
- Qureshi, Nusrat Shahin, Jasleen Kaur Bains, Sridhar Sreeramulu, Harald Schwalbe, and Boris Fürtig. 2018. "Conformational Switch in the Ribosomal Protein S1 Guides Unfolding of Structured RNAs for Translation Initiation." *Nucleic Acids Research*, August. <https://doi.org/10.1093/nar/gky746>.
- Riediger, Matthias, Philipp Spät, Raphael Bilger, Boris Macek, and Wolfgang R. Hess. 2020. "Analysis of a Photosynthetic Cyanobacterium Rich in Internal Membrane Systems via Gradient Profiling by Sequencing (Grad-Seq)." Preprint. *Microbiology*. <https://doi.org/10.1101/2020.07.02.184192>.
- Romilly, Cédric, Sebastian Deindl, and E. Gerhart H. Wagner. 2019. "The Ribosomal Protein S1-Dependent Standby Site in *TisB* mRNA Consists of a Single-Stranded Region and a 5' Structure Element." *Proceedings of the National Academy of Sciences of the United States of America* 116 (32): 15901–6. <https://doi.org/10.1073/pnas.1904309116>.
- Russell, J. B., and G. M. Cook. 1995. "Energetics of Bacterial Growth: Balance of Anabolic and Catabolic Reactions." *Microbiological Reviews* 59 (1): 48–62.
- Saito, Kazuki, Rachel Green, and Allen R Buskirk. 2020. "Translational Initiation in *E. Coli* Occurs at the Correct Sites Genome-Wide in the Absence of mRNA-RRNA Base-Pairing." *eLife* 9 (February). <https://doi.org/10.7554/eLife.55002>.
- Sakurai, I., D. Stazic, M. Eisenhut, E. Vuorio, C. Steglich, W. R. Hess, and E.-M. Aro. 2012. "Positive Regulation of *PsbA* Gene Expression by *Cis*-Encoded Antisense RNAs in *Synechocystis* Sp. PCC 6803." *PLANT PHYSIOLOGY* 160 (2): 1000–1010. <https://doi.org/10.1104/pp.112.202127>.
- Salah, Philippe, Marco Bisaglia, Pascale Aliprandi, Marc Uzan, Christina Sizun, and François Bontems. 2009. "Probing the Relationship between Gram-Negative and Gram-Positive S1 Proteins by Sequence Analysis." *Nucleic Acids Research* 37 (16): 5578–88. <https://doi.org/10.1093/nar/gkp547>.
- Sazuka, T., and O. Ohara. 1996. "Sequence Features Surrounding the Translation Initiation Sites Assigned on the Genome Sequence of *Synechocystis* Sp. Strain PCC6803 by Amino-Terminal Protein Sequencing." *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes* 3 (4): 225–32.

- Scharff, Lars B., Liam Childs, Dirk Walther, and Ralph Bock. 2011. "Local Absence of Secondary Structure Permits Translation of MRNAs That Lack Ribosome-Binding Sites." Edited by Josep Casadesús. *PLoS Genetics* 7 (6): e1002155. <https://doi.org/10.1371/journal.pgen.1002155>.
- Schlax, Paula Jean, and David J. Worhunsy. 2003. "Translational Repression Mechanisms in Prokaryotes: Factors Influencing Translation Repression Efficiency." *Molecular Microbiology* 48 (5): 1157–69. <https://doi.org/10.1046/j.1365-2958.2003.03517.x>.
- Sebesta, Jacob, and Christie AM. Peebles. 2020. "Improving Heterologous Protein Expression in *Synechocystis* Sp. PCC 6803 for Alpha-Bisabolene Production." *Metabolic Engineering Communications* 10 (June): e00117. <https://doi.org/10.1016/j.mec.2019.e00117>.
- Sengupta, J., R. K. Agrawal, and J. Frank. 2001. "Visualization of Protein S1 within the 30S Ribosomal Subunit and Its Interaction with Messenger RNA." *Proceedings of the National Academy of Sciences* 98 (21): 11991–96. <https://doi.org/10.1073/pnas.211266898>.
- Shell, Scarlet S., Jing Wang, Pascal Lapiere, Mushtaq Mir, Michael R. Chase, Margaret M. Pyle, Richa Gawande, et al. 2015. "Leaderless Transcripts and Small Proteins Are Common Features of the Mycobacterial Translational Landscape." Edited by Patrick H. Viollier. *PLOS Genetics* 11 (11): e1005641. <https://doi.org/10.1371/journal.pgen.1005641>.
- Smit, M. H. de, and J. van Duin. 1990. "Secondary Structure of the Ribosome Binding Site Determines Translational Efficiency: A Quantitative Analysis." *Proceedings of the National Academy of Sciences* 87 (19): 7668–72. <https://doi.org/10.1073/pnas.87.19.7668>.
- . 1994. "Translational Initiation on Structured Messengers. Another Role for the Shine-Dalgarno Interaction." *Journal of Molecular Biology* 235 (1): 173–84.
- Smit, Maarten H. de, and Jan van Duin. 2003. "Translational Standby Sites: How Ribosomes May Deal with the Rapid Folding Kinetics of mRNA." *Journal of Molecular Biology* 331 (4): 737–43.
- Sørensen, Michael A, Jens Fricke, and Steen Pedersen. 1998. "Ribosomal Protein S1 Is Required for Translation of Most, If Not All, Natural MRNAs in *Escherichia Coli* in Vivo." *Journal of Molecular Biology* 280 (4): 561–69. <https://doi.org/10.1006/jmbi.1998.1909>.
- Sprengart, Michael L., Hans P. Fatscher, and Eckart Fuchs. 1990. "The Initiation of Translation in *E. Coli* : Apparent Base Pairing between the 16srRNA and Downstream Sequences of the mRNA." *Nucleic Acids Research* 18 (7): 1719–23. <https://doi.org/10.1093/nar/18.7.1719>.
- Starmer, J, A Stomp, M Vouk, and D Bitzer. 2006. "Predicting Shine–Dalgarno Sequence Locations Exposes Genome Annotation Errors." Edited by Vivian Hook. *PLoS Computational Biology* 2 (5): e57. <https://doi.org/10.1371/journal.pcbi.0020057>.
- Studer, Sean M., and Simpson Joseph. 2006. "Unfolding of mRNA Secondary Structure by the Bacterial Translation Initiation Complex." *Molecular Cell* 22 (1): 105–15. <https://doi.org/10.1016/j.molcel.2006.02.014>.
- Sugiura, M., T. Hirose, and M. Sugita. 1998. "Evolution and Mechanism of Translation in Chloroplasts." *Annual Review of Genetics* 32: 437–59. <https://doi.org/10.1146/annurev.genet.32.1.437>.
- Takyar, Seyedtaghi, Robyn P. Hickerson, and Harry F. Noller. 2005. "mRNA Helicase Activity of the Ribosome." *Cell* 120 (1): 49–58. <https://doi.org/10.1016/j.cell.2004.11.042>.
- Thiel, Kati, Edita Mulaku, Hariharan Dandapani, Csaba Nagy, Eva-Mari Aro, and Pauli Kallio. 2018. "Translation Efficiency of Heterologous Proteins Is Significantly Affected by the Genetic Context of RBS Sequences in Engineered Cyanobacterium *Synechocystis* Sp. PCC 6803." *Microbial Cell Factories* 17 (1). <https://doi.org/10.1186/s12934-018-0882-2>.
- Till, Petra, Jörg Toepel, Bruno Bühler, Robert L. Mach, and Astrid R. Mach-Aigner. 2020. "Regulatory Systems for Gene Expression Control in Cyanobacteria." *Applied Microbiology and Biotechnology* 104 (5): 1977–91. <https://doi.org/10.1007/s00253-019-10344-w>.
- Tuller, T., and H. Zur. 2015. "Multiple Roles of the Coding Sequence 5' End in Gene Expression Regulation." *Nucleic Acids Research* 43 (1): 13–28. <https://doi.org/10.1093/nar/gku1313>.

- Tuller, Tamir, Yedael Y. Waldman, Martin Kupiec, and Eytan Ruppin. 2010. "Translation Efficiency Is Determined by Both Codon Bias and Folding Energy." *Proceedings of the National Academy of Sciences* 107 (8): 3645–50. <https://doi.org/10.1073/pnas.0909910107>.
- Tyystjarvi, Taina, Mirkka Herranen, and Eva-Mari Aro. 2001. "Regulation of Translation Elongation in Cyanobacteria: Membrane Targeting of the Ribosome Nascent-Chain Complexes Controls the Synthesis of D1 Protein." *Molecular Microbiology* 40 (2): 476–84. <https://doi.org/10.1046/j.1365-2958.2001.02402.x>.
- Wang, Bo, Carrie Eckert, Pin-Ching Maness, and Jianping Yu. 2018. "A Genetic Toolbox for Modulating the Expression of Heterologous Genes in the Cyanobacterium *Synechocystis* Sp. PCC 6803." *ACS Synthetic Biology* 7 (1): 276–86. <https://doi.org/10.1021/acssynbio.7b00297>.
- Wang, Chengyuan, Vadim Molodtsov, Emre Firlar, Jason T. Kaelber, Gregor Blaha, Min Su, and Richard H. Ebright. 2020. "Structural Basis of Transcription-Translation Coupling." *Science* 369 (6509): 1359–65. <https://doi.org/10.1126/science.abb5317>.
- Webster, Michael William, Maria Takacs, Chengjin Zhu, Vita Vidmar, Ayesha Eduljee, Mo'men Abdelkareem, and Albert Weixlbaumer. 2020. "Structural Basis of Transcription-Translation Coupling and Collision in Bacteria." *Science* 369 (6509): 1355–59. <https://doi.org/10.1126/science.abb5036>.
- Wei, Yulong, and Xuhua Xia. 2019. "Unique Shine–Dalgarno Sequences in Cyanobacteria and Chloroplasts Reveal Evolutionary Differences in Their Translation Initiation." Edited by Ruth Hershberg. *Genome Biology and Evolution* 11 (11): 3194–3206. <https://doi.org/10.1093/gbe/evz227>.
- Wurtzel, O., R. Sapra, F. Chen, Y. Zhu, B. A. Simmons, and R. Sorek. 2010. "A Single-Base Resolution Map of an Archaeal Transcriptome." *Genome Research* 20 (1): 133–41. <https://doi.org/10.1101/gr.100396.109>.
- Yamamoto, Hiroshi, Daniela Wittek, Romi Gupta, Bo Qin, Takuya Ueda, Roland Krause, Kaori Yamamoto, Renate Albrecht, Markus Pech, and Knud H. Nierhaus. 2016. "70S-Scanning Initiation Is a Novel and Frequent Initiation Mode of Ribosomal Translation in Bacteria." *Proceedings of the National Academy of Sciences* 113 (9): E1180–89. <https://doi.org/10.1073/pnas.1524554113>.
- Yarchuk, Oleg, Nathalie Jacques, Jean Guillerez, and Marc Dreyfus. 1992. "Interdependence of Translation, Transcription and mRNA Degradation in the LacZ Gene." *Journal of Molecular Biology* 226 (3): 581–96. [https://doi.org/10.1016/0022-2836\(92\)90617-S](https://doi.org/10.1016/0022-2836(92)90617-S).
- Zerges, W. 2000. "Translation in Chloroplasts." *Biochimie* 82 (6–7): 583–601. [https://doi.org/10.1016/s0300-9084\(00\)00603-9](https://doi.org/10.1016/s0300-9084(00)00603-9).

CHAPTER 4: CONTEXT DEPENDENCE OF RIBOSOME BINDING SITES IN *SYNECHOCYSTIS SP. PCC6803* AND *E. COLI*²

Summary

Adapting gene expression elements to new contexts is an overlooked problem in metabolic engineering. As a starting point, it is assumed that genetic parts such as ribosome binding sites will work similarly regardless of the gene being expressed, and the organism in which the gene is expressed. This assumption has not generally been confirmed, and brute force methods of testing many elements to achieve a specified expression level has been required for success which leads to increased cost in strain development. We tested 13 ribosome binding sites and compared expression of green fluorescent protein (GFP) and yellow fluorescent protein (YFP) in *Escherichia coli* (*E. coli*) and *Synechocystis sp. PCC 6803* (*S. 6803*). Protein leader sequences and bicistronic designs were implemented to address this issue of sequence context. The results confirm previous findings that the bicistronic design can support reliable gene expression in *E. coli*. Interestingly, the bicistronic design was not as reliable in *S. 6803*. The protein leader sequences did not have similar success in either organism. Correlations between the RBS Calculator and the measured fluorescence values were weak in both organisms, and the overall correlation between expression in the two organisms had an R^2 of 0.498 and a Spearman rank

² This chapter has been submitted as:

Sebesta, Jacob, and Christie AM. Peebles. (submitted) "Effect of coding sequences on translation initiation rates of 13 ribosome binding sites in *E. coli* and *Synechocystis sp. PCC6803*."

correlation R value of 0.788. Precise control of gene expression through ribosome binding site sequence variation still generally requires testing multiple sequences in *S. 6803*, though bicistronic designs may be reliably applied to this end in *E. coli*.

Introduction

Motivation

Benefits of reusable/reliable genetic parts including ribosome binding sites

Many metabolic engineering efforts involve the balancing of expression of multiple enzymes in a pathway, requiring genetic elements that can generate expression at a specific level. Ideally, transcription promoter sequences and translation initiation regions or ribosome binding sites (RBSs), that generate known rates of transcription and translation could be applied to this end without significant variation in expression levels in different contexts. At a minimum, they should have the same rank order. However, there are significant differences in observed expression levels depending on both the nucleotide sequences adjacent to these parts and many other differences that arise when using these parts in different organisms. As such, metabolic engineering projects often rely on brute force methods and high-throughput screening to test many combinations of genetic parts simultaneously. Metabolic engineering could benefit from the development of reusable genetic parts that function similarly in different contexts. This reliability is essential for projects that don't use high-throughput methods of cloning and screening. The development of reliable parts could also facilitate projects that do use high-throughput methods in that they might help to focus the search on more interesting problems than simply trying to achieve a certain expression level. In other words, we might be better able to control expression level with reliable parts rather than simply needing to test many sequences for each gene. We focused on improving ribosome binding sites in this work to improve control of translation initiation rates in different contexts. In synthetic biology, the ribosome binding site is a DNA

sequence that encodes part of the 5'-untranslated region of the mRNA which controls the rate that ribosomes initiate translation. This rate is thought to be a rate limiting step in protein production. Translation initiation rates have coding sequence dependence mainly due to secondary structure that may form within the mRNA, spanning the RBS itself and the adjacent sequences. This can inhibit ribosome binding.

Strategies to overcome context dependence of ribosome binding sites

We tested two strategies to overcome these challenges of context dependency in translation initiation. We utilized the 26 plasmids constructed and tested by Thiel et al., (2018) as a starting point. In their tests, GFP and YFP were expressed in S.6803 from replicative plasmids using the Plac promoter, and 13 ribosome binding sites. They found significant differences in the expression profile of the RBSs depending on which coding sequence was used. We repeated their experiment, and constructed additional plasmids designed to reduce the differences in the expression profiles.

First, a leader peptide was inserted at the N-terminus of YFP. The first 24 nucleotides from GFP were inserted, followed by a 'glycine-glycine-serine' linker. This was intended to keep the sequence context of the ribosome binding sites constant between the different proteins. We hypothesized that the expression level of other proteins may be more accurately predicted for an RBS characterized using GFP if the N-terminal leader of GFP was inserted. N-terminal peptide tags are used frequently for purification or detection of proteins and are usually non-disruptive to protein folding.

A second strategy was implemented which does not alter the coding sequence. The bicistronic design developed by Mutalik et al. (2013) was utilized in order to reduce context dependence of RBS function. This design functions by inserting the RBS sequences within a small open reading frame, creating an operon structure with the stop codon of the first ORF partially overlapping with the start codon of the gene of interest. The ribosome has helicase activity during the elongation of the first ORF peptide,

allowing it to unfold secondary structures within the RBS, and re-initiate translation. These designs worked well in *E. coli* in improving the Spearman rank correlation for a set of RBSs when used to express different proteins (generally from ~0.5 to ~0.85). Recently, one bicistronic design (BCD) was successful in translating a terpenoid synthase in *S. 6803* (Dienst et al. 2020). We adapted the 13 RBSs to bicistronic designs by mutating any stop codons in the upstream open reading frame which could allow the ribosome to either disengage the mRNA before reaching the start codon of the gene of interest, or to re-enter the pre-initiation state.

A precisely defined ribosome binding site should have similar function in different contexts. It would be useful to metabolic engineers to be able to reuse well-characterized genetic elements to narrow the immense number of possible sequences that currently must be tested to achieve a defined output. We utilized the correlations statistics (including R^2 and Spearman rank correlation coefficient) of GFP and YFP expression to estimate the coding sequence dependence of expression levels. Utilization of the leader peptide for YFP expression consistently increased the expression of YFP, but did not improve the correlation between GFP and YFP expression in either *E. coli* or *S. 6803*. The correlation between GFP and YFP were much improved by the bicistronic design RBSs in *E. coli*, but not in *S. 6803*.

Methods

Strains and cultivation

E. coli (DH5 α) was used for plasmid construction and was grown at 37 °C in lysogeny broth with shaking at 225 rpm. Twenty-six initial plasmids were generously shared with us by the lab of Dr. Paulo Kallio (Thiel et al. 2018). The basic plasmid map and translation initiation design variants are shown in Figure 4.1. These broad host range replicative plasmids (RSF1010-derived) contained either GFP or YFP (EYFP). Transcription was driven by a lac promoter including a lac operator for inducible expression (plasmids also contained the lacI gene). The 5' UTR for GFP/YFP included an “insulator” sequence (Zelcbuch et al.

2013) and one of 13 different ribosome binding sites. These plasmids were modified by insertion of the cyan fluorescent protein (CFP) gene, mTurquoise2 (Goedhart et al. 2012) using the coding sequence utilized by Gutu et al., 2018 (Gutu, Chang, and O'Shea 2018). Expression of CFP was driven by the lac promoter and the RBS7, one of the 13 applied to expression of GFP and YFP. To prevent transcription readthrough, CFP was flanked by the terminators BBa_B1006 (5' of the lac promoter), and BBa_B0014 (3') from the iGEM Registry of Standard Biological Parts.

Additional ribosome binding sequences, such as the bicistronic designs, were generated *via* restriction enzyme cloning. Overlap extension PCR using New England Biolabs Phusion DNA polymerase was completed to generate bicistronic design DNA fragments which were digested and cloned into the plasmid backbone carrying either GFP or YFP using NdeI and XhoI restriction enzymes. To facilitate cloning, the *mobA* gene in the plasmid backbone was mutated to knock out DNA nicking activity (Taton et al. 2014). This mutation was completed *via* whole-plasmid PCR to generate blunt ended linear DNA which was directly transformed. Correct colonies were screened *via* colony PCR, and then sequenced by Sanger sequencing (GeneWiz). Wild type *S. 6803* was transformed by electroporation (Ludwig et al. 2008) and colonies were confirmed by colony PCR.

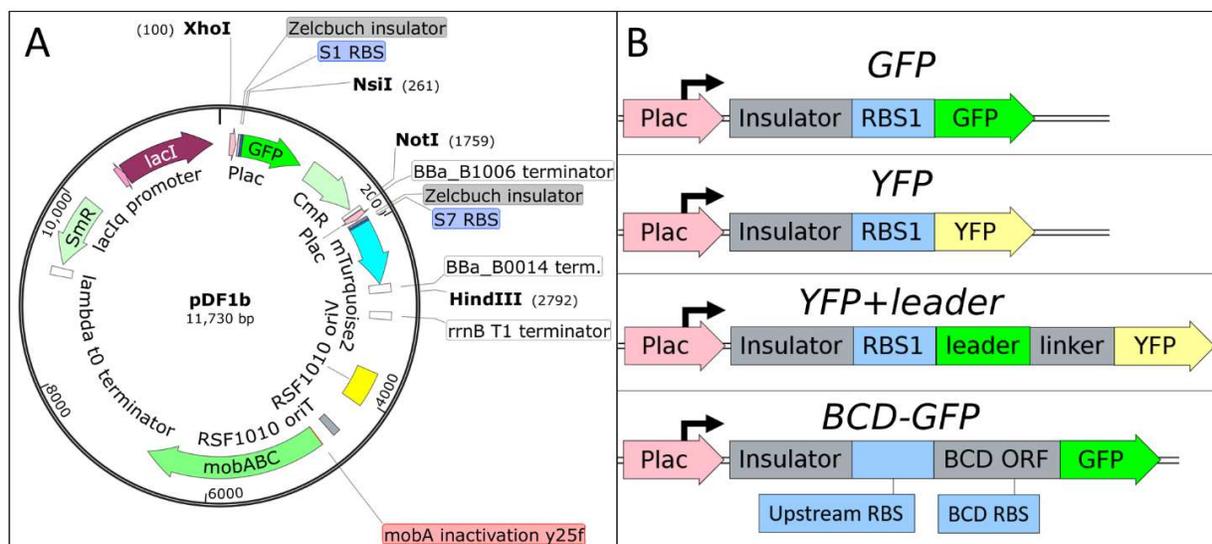


Figure 4.1: Plasmid map (A) for the plasmids used in this study and, (B): depiction of different translation initiation designs used (GFP and YFP, YFP+leader, BCD-GFP, BCD-YFP not shown).

For fluorescence measurements, *E. coli* containing the plasmids of interest were grown in 2 mL of lysogeny broth at 37 °C with shaking at 250 rpm. Two cultures were started from overnight cultures of each strain. After 2 hours of growth, IPTG was added to 1 mM. Optical density and fluorescence were measured at that time, as well as at six hours after the addition of IPTG.

S.6803 strains containing each of the plasmids were grown in 3 mL of BG11 medium at 30 °C with shaking at 225 rpm. Two cultures were started from these cultures of each strain. After 24 hours of growth, IPTG was added to 1 mM, and NaHCO₃ to 20 mM. Optical density and fluorescence were measured at that time, as well as at 48 hours after the addition of IPTG. Fluorescent lighting was supplied continuously with intensity of 195 ± 25 μmol photons/m²/s at the approximate level of the top surface of the cultures.

Absorbance and Fluorescence measurements

At each timepoint, three 100 μL samples of each culture were transferred to black-sided, clear bottom polystyrene 96-well plates (Corning). Fluorescence was measured using a BioTek H1 Synergy plate reader using monochromators to select specific light wavelengths for excitation and emission (Ex./Em.)

for each fluorescent protein: GFP fluorescence was measured using Ex./Em. at 485/525nm, YFP at 505/535nm, and CFP at 425/495 nm. Fluorescence measurements in *S. 6803* were taken using Ex./Em = 490/520 nm for GFP, 505/535 nm for YFP, and 425/477 nm for CFP. Optical density measurements using absorbance at 600 nm for *E. coli* and 730 nm for *S. 6803* in the same BioTek H1 plate reader.

Figures were prepared using pandas (McKinney 2010) and Matplotlib (Hunter 2007), and correlations were calculated using Scipy (SciPy 1.0 Contributors et al. 2020).

Table of sequences

Table 4.1: RBS sequences used here including the ID, the names used by Thiel et al., the original source of the sequence, and the sequence itself. Shine-Dalgarno sequences are shown in red italics and the start codon is shown in bold.

RBS ID	Thiel RBS Name	RBS Source	Sequence	Notes
0			ATG	
1	S1	RBS*	CTAGAGTAGT <i>GGAGG</i> TTACTAG ATG	Originally designed by (Huang and Lindblad 2013)
2	S2	cpcB (7002)	AATATAAGT <i>AGGAG</i> ATAAAAAC ATG	Native to <i>Synechococcus sp.</i> PCC7002
3	S3	cpcB	AGTCAAGT <i>AGGAG</i> ATTAATTCA ATG	Native to <i>S. 6803</i>
4	S4	psbA2	ATACATA <i>AGGA</i> ATTATAACCAA ATG	Native to <i>S. 6803</i>
5	S5	rbcl	TGTTTAT <i>GGAGG</i> ACTGACCTAG ATG	Native to <i>S. 6803</i>
6	S6		TAGCCTAGG <i>AGGAGG</i> AAAAATC ATG	
7	S7	lac	CATTAAAGA <i>GGAG</i> AAAGGTACC ATG	From the <i>E. coli</i> lac gene
8	A	Salis-ZB	AACAAAATG <i>AGGAGG</i> TACTGAG ATG	Originally designed by Salis et al. (2009) and selected for use by (Zelcbuch et al. 2013)
9	B	Salis-ZA	<i>AGGAGG</i> TTTGGA ATG	
10	C	Salis-ZC	AAGTTAA <i>GAGG</i> CAAGA ATG	
11	D	Salis-ZE	TAAGC <i>AGGA</i> CCGGCGGCG ATG	
12	E	Salis-ZD	TTCGCAGGGGGAAG ATG	
13	Z	ZF	CACCATACACTG ATG	"Dead-RBS" from BIOFAB also used by Zelcbuch et al., 2013
Preceding sequence	aattgtgagcggataacaatttcacacagaattcattaaagactagttaatagaaataatTTTTgtttaacttta			Italics indicate the "insulator" used by Zelcbuch et al., 2013

Table 4.2: The bicistronic design sequences used. The constant sequence of the upstream cistron is shown in italics. Mutations to the RBS which were made to avoid stop codons in the upstream cistron are lowercase.

Upstream RBS (constant): <i>GGGCCCAAGTTCACCTTAAAAAGGAGATCAACA</i>	
Name	Upstream open reading frame sequence
BCD0	<i>ATG</i> AAAGCAATTTTCGTACTGAAACATCTTt ATG
BCD1	<i>ATG</i> AAAGCAATTTTCGTACTGAAACATCTTACTAGAGTAGTGGAGGTTACTt aATG
BCD2	<i>ATG</i> AAAGCAATTTTCGTACTGAAACATCTTAAATATAAG aAG GAGATAAAA t a ATG
BCD3	<i>ATG</i> AAAGCAATTTTCGTACTGAAACATCTTAAAGTCAAGTAGGAGATTAATTt aATG
BCD4	<i>ATG</i> AAAGCAATTTTCGTACTGAAACATCTTAAATACATAAGGAATTATAAC ct ATG
BCD5	<i>ATG</i> AAAGCAATTTTCGTACTGAAACATCTTATGTTTATGGAGGAC aGAC CTt aATG
BCD6	<i>ATG</i> AAAGCAATTTTCGTACTGAAACATCTTATAGCC aAG GAGGAGGAAAA t a ATG
BCD7	<i>ATG</i> AAAGCAATTTTCGTACTGAAACATCTTACATTAAGAGGAGAAAGGT a ATG
BCD8	<i>ATG</i> AAAGCAATTTTCGTACTGAAACATCTTAAACAAAATGAGGAGGTACTG t aATG
BCD9	<i>ATG</i> AAAGCAATTTTCGTACTGAAACATCTTATT aAACTT aAAG GAGGTTT Gt ATG
BCD10	<i>ATG</i> AAAGCAATTTTCGTACTGAAACATCTTAACTTTAAAGTTAAGAGGCA at ATG
BCD11	<i>ATG</i> AAAGCAATTTTCGTACTGAAACATCTTATTTATAAGCAGGACCGGCG Gt aATG
BCD12	<i>ATG</i> AAAGCAATTTTCGTACTGAAACATCTTATAACTTTATTCGCAGGGG Gat aATG
BCD13	<i>ATG</i> AAAGCAATTTTCGTACTGAAACATCTTATT aAACTTT ACACCATACACT a ATG

Table 4.3: The five design sets tested in this study.

	'GFP' design set	'YFP' design set	'YFP+leader' design set	'GFP-BCD' design set	'YFP-BCD' design set
Translation initiation	RBS0 – RBS13	RBS0 – RBS13	RBS0 – RBS13	BCD0-BCD13	BCD0-BCD13
Fluorescent protein	GFP	YFP	GFP24+YFP	GFP	YFP

Results

We constructed plasmids for five design sets, each with the lac promoter (Plac) and one of thirteen ribosome binding site sequences driving expression of either GFP or YFP (

Table 4.3). Each plasmid also contained CFP with expression driven by Plac and RBS7. The 'GFP' and 'YFP' design sets simply contain one of the ribosome binding sites from RBS0-RBS13 and GFP or YFP. The 'YFP+leader' set included YFP with the first six nucleotides of the YFP coding sequence replaced with the first 24 nucleotides of the GFP coding sequence plus nine additional nucleotides coding for a serine-glycine-glycine linker peptide. Each of the ribosome binding sites were adapted for use as bicistronic translation initiation sequences by inserting an additional translation initiation sequence and start codon

5' of each RBS and removing any stop codons within the RBS which would interfere with the re-initiation mechanism used by the bicistronic design. These were used to construct the GFP-BCD and YFP-BCD design sets.

We included, for each of these designs, an RBS-0 which simply removed the nucleotides between the insulator and the start codon. Two of the design sets, 'GFP' and 'YFP+leader', also included strains that lacked any 5' UTR, such that the start codon coincided with the transcription start site ('gLL' and 'y+LL'). In *E. coli*, these two had OD-normalized fluorescence divided by CFP fluorescence of 8.24 ± 2.15 and 8.33 ± 3.84 , respectively, which exceeded the expression of the similar RBS0 strains (RBS0-GFP and RBS0-YFP+leader) by a factor of 4.7 and 1.2. In *S. 6803* the leaderless and RBS0 constructs had similar expression levels to one another (1.39 ± 0.06 and 1.60 ± 0.11 for 'gLL' and 'y+LL' versus 1.5 ± 0.09 and 1.55 ± 0.13 for RBS0-GFP and RBS0-YFP+leader). However, the fluorescence of leaderless constructs only represented 2-7% of the highest fluorescence measurements in *E. coli* and 5-20% of those in *S. 6803*.

Comparison to Thiel et al. 2018 data

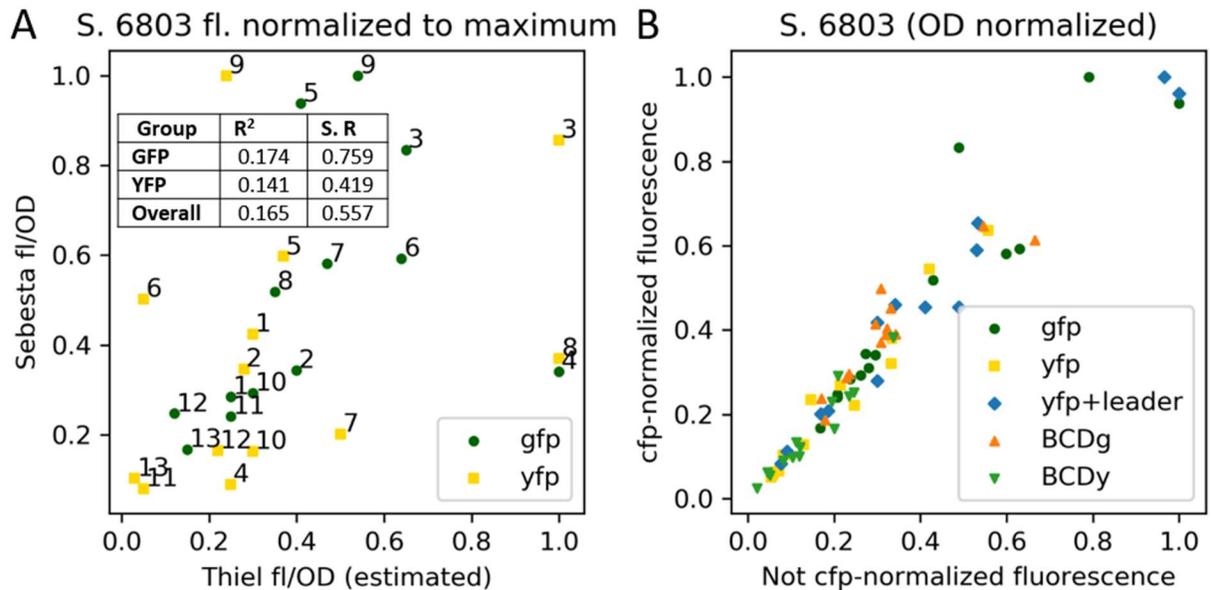


Figure 4.2: A: Our measurements did not correlate well with the GFP and YFP fluorescence measurements reported in Thiel et al. (2018). To compare, the fluorescence measurements were not normalized to CFP fluorescence, but only to OD, following Thiel et al. (error bars not shown). The inset table summarizes the R^2 and Spearman R (S. R) for GFP, YFP, and all strains. B: the CFP-normalized fluorescence (y-axis) for each plasmid tested in this work is compared with the fluorescence without CFP-normalization (x-axis). In this panel, all GFP measurements are normalized to the maximum GFP measurement, and all YFP measurements normalized to the maximum YFP measurement.

We compared our measurements to those made by Thiel et al. (2018), normalized only to optical density and to the maximum GFP or YFP fluorescence (Figure 4.2A). The two datasets were not well correlated with R^2 values of 0.174 for GFP, 0.141 for YFP, and 0.165 overall. Spearman rank correlation coefficients were 0.759 for GFP, 0.419 for YFP, and 0.557 overall. Within the GFP set, RBS4 deviated the most from the apparent correlation, while the YFP set showed several deviations in fluorescence between the two datasets including RBS6, RS8, and RBS9. The deviations in the YFP dataset may be due to a cloning scar in the Thiel sequences that left a second start codon after the NdeI cutsite. These secondary start sites have the potential to have different effects on the translation initiation rates of the ribosome binding sites. Either one, or both, of these start sites may be utilized, likely they have different rates of use. Other differences between our tests should affect all strains similarly. These include the mutation to

mobA, the precise growth conditions, and our fluorescence measurement equipment. These differences should not affect the rank order of the different RBSs because all are impacted similarly.

To gauge the contribution of the CFP normalization to the results, we compared *S.6803* fluorescence with and without CFP normalization (Figure 4.2B). We expect plasmid copy number to be the main driver of variability between the different constructs because the promoter, RBS, and coding sequence for CFP were the same in each one. Fluorescence measurements with and without CFP normalization were well correlated ($R^2 = 0.929$, Spearman $R = 0.982$), with larger deviations found at higher fluorescence values.

A consistent N-terminal leader sequence failed to make expression profiles more similar for GFP and YFP

We hypothesized that inserting the N-terminal coding sequence for GFP at the start codon for eYFP would improve the correlation of YFP fluorescence with GFP fluorescence when using the same 13 ribosome binding sites to control translation initiation. A reliable RBS design would result in expression of different proteins with the same rank order. This would allow the GFP leader sequence to be used in other coding sequence contexts with a predictable trend in protein expression. The YFP-to-CFP fluorescence was normalized to OD and compared to the GFP-to-CFP fluorescence normalized to OD for the same RBS sequences.

In both *E. coli* and *S. 6803*, the YFP fluorescence without the leader correlated better with GFP fluorescence than the YFP constructs which implementation of the leader (

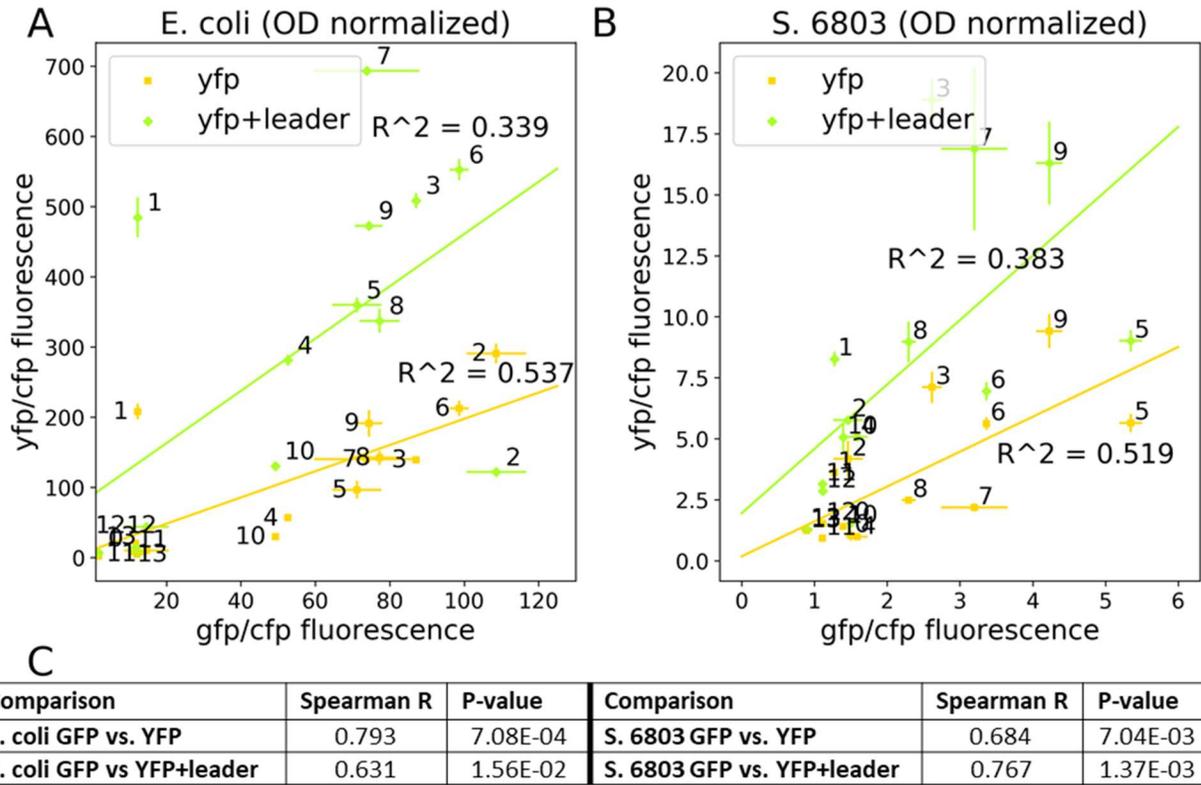
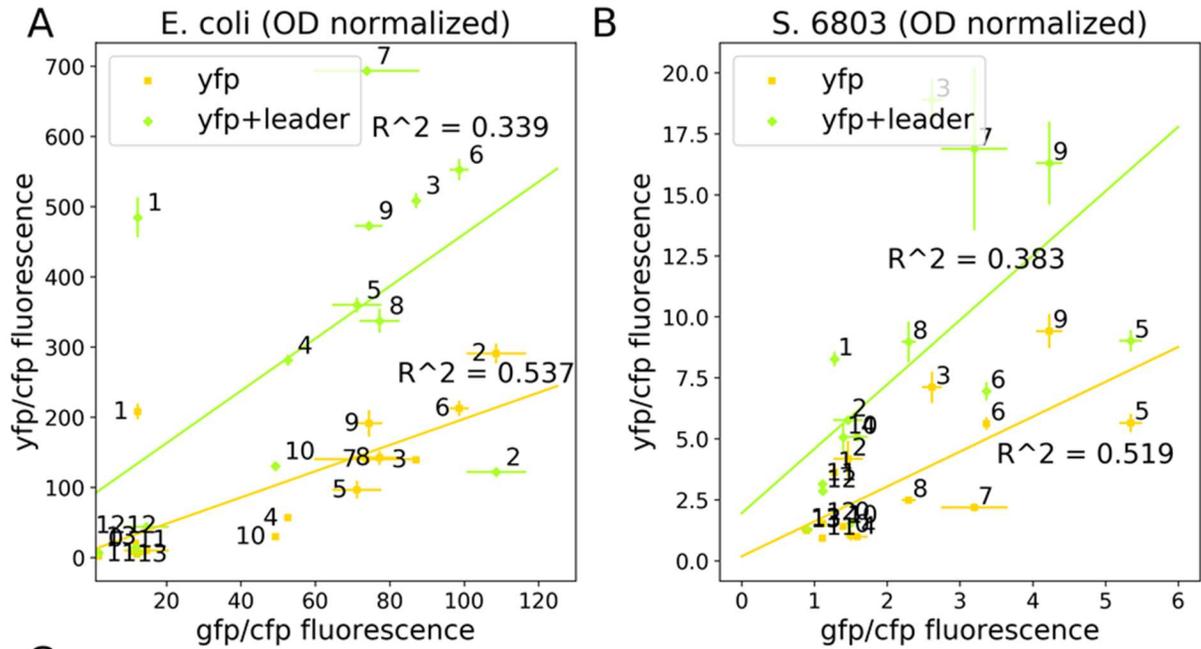


Figure 4.3). In *E. coli*, the Spearman rank order correlation was also decreased by YFP+leader versus GFP (Figure 4.3C), while in *S. 6803* the rank order correlation was slightly improved when the leader was used. In both species, YFP fluorescence was increased by the insertion of the leader peptide for all ribosome binding sites, except for RBS-S2 in *E. coli*. Compared with the measurements in *S. 6803* by Thiel et al. (2018), the Spearman rank correlation coefficient for GFP versus YFP was higher in these results (0.659 versus 0.543).



C

Comparison	Spearman R	P-value	Comparison	Spearman R	P-value
E. coli GFP vs. YFP	0.793	7.08E-04	S. 6803 GFP vs. YFP	0.684	7.04E-03
E. coli GFP vs YFP+leader	0.631	1.56E-02	S. 6803 GFP vs. YFP+leader	0.767	1.37E-03

Figure 4.3: GFP and YFP fluorescence comparisons revealed only weak correlations in both *E. coli* (A) and *S. 6803* (B). The leader peptide generally increased YFP fluorescence for given RBSs, but its insertion did not improve the correlation in either organism. All measurements were normalized to OD and CFP fluorescence. Error bars represent the standard deviation of two biological replicates. Spearman rank correlation coefficients and the corresponding P-values for each design set are shown in the table (C).

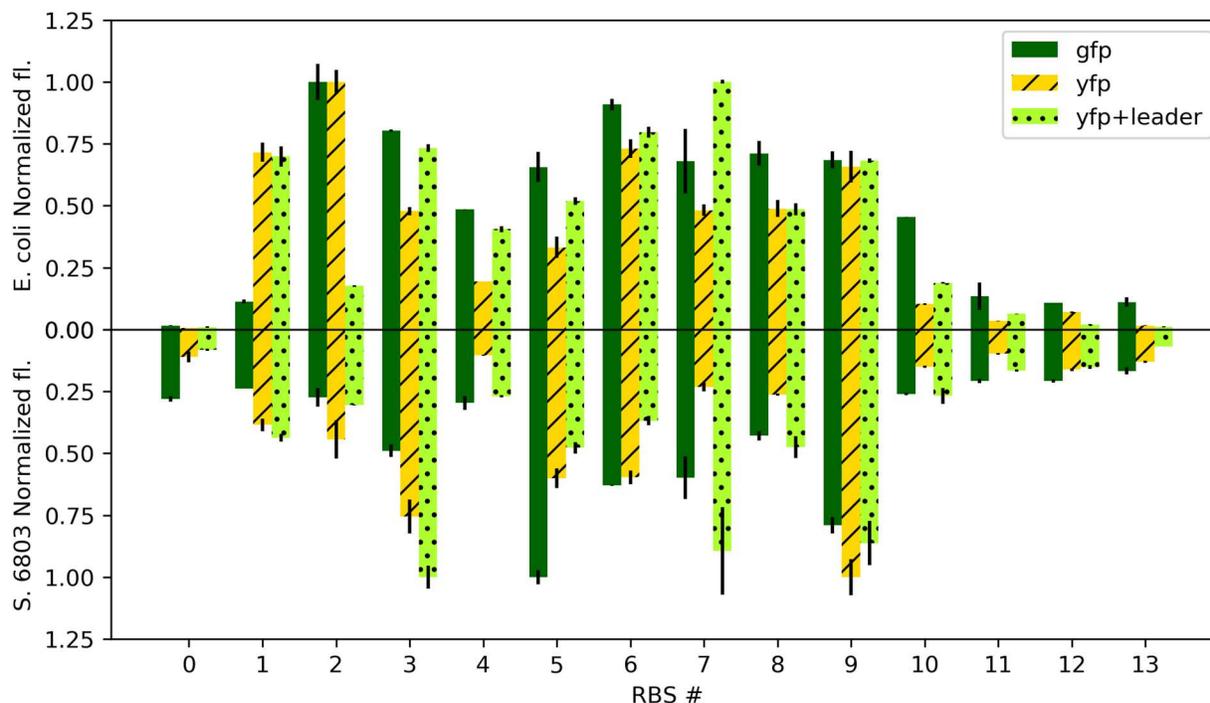


Figure 4.4: The relative expression profiles of *gfp*, *yfp*, and *yfp+leader* design sets differed between the two species: *E. coli* (top) and *S. 6803* (bottom).

We compared the relative expression of three design sets (GFP, YFP, and YFP+leader) normalized to the maximum expression within each design set (Figure 4.4) as measured in *E. coli* (top) and *S. 6803* (bottom). Overall, the expression profiles are similar between the two species (the corresponding Spearman rank correlation coefficient was 0.772). One notable difference was the higher relative fluorescence of the RBS0 in *S. 6803* than in *E. coli*. This is a null RBS in which the insulator sequence leads directly to the start codon. GFP fluorescence differed between the two species by the most for RBS0 and RBS2 (*cpcB* RBS from *S. 7002*). For RBS1, YFP expression with or without the leader was higher in both species than GFP expression. For RBS2, the YFP leader decreased expression in *E. coli*, making the relative expression dissimilar to the GFP relative expression. The leader did not have the same effect in *S. 6803*.

The online tool, RBS Calculator, offers a means to predict translation initiation rates for any nucleotide sequence in prokaryotes. We compared the relative strengths of the ribosome binding sites when used

to express GFP, YFP and YFP+leader to the expression level predicted by the RBS Calculator for *E. coli* and *S. 6803* (Figure 4.5). Based on reports of the RBS Calculator performing poorly for RBSs tested in cyanobacteria (Wang et al. 2018; Sebesta and Peebles 2020), we expected the RBS Calculator to more accurately predict fluorescence in *E. coli* than in *S. 6803*. However, predicted translation initiation rates were not well correlated with the normalized fluorescence measurements in either *E. coli* or *S. 6803*. Within each design set, the correlations are also poor. The overall Spearman R of all designs (excluding bicistronic designs) was 0.520. In *E. coli* the Spearman R was much lower for the ‘YFP+leader’ set than the other two.

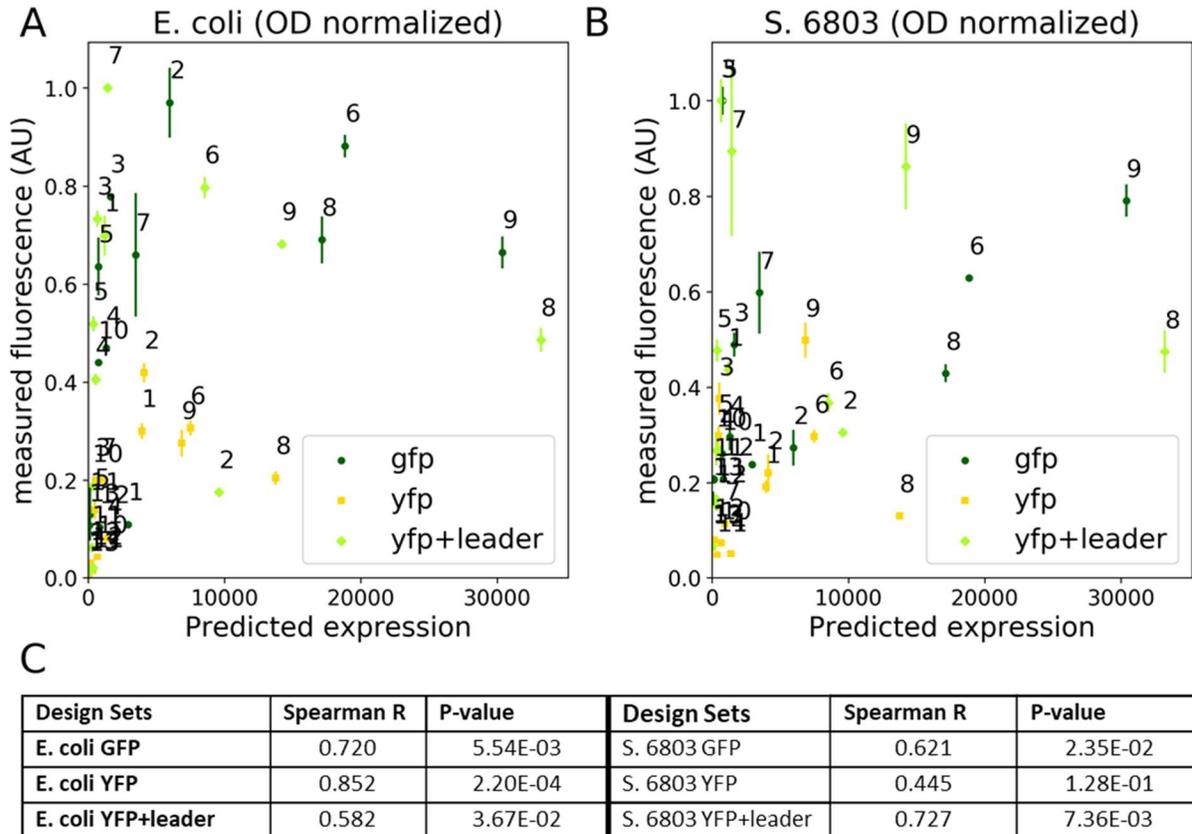


Figure 4.5: The RBS Calculator prediction were not well correlated with fluorescence measurements in *E. coli* (A) or in *S. 6803* (B). Fluorescence measurements are normalized to the maximum measurement for all GFP or YFP measurements in each organism. Y-axis error bars represent the standard deviation of two biological replicates. Spearman rank correlation coefficients and the corresponding P-values for

Bicistronic designs improved the GFP/YFP correlation in *E. coli*, but not in *S. 6803*

Bicistronic designs have previously been demonstrated in *E. coli* to reduce the variation in expression levels when different proteins are expressed using the same BCD sequence when compared with the more commonly used RBSs (Mutalik et al. 2013). We adapted the 13 RBSs used in this study to the bicistronic design by removing any in frame stop codons in the upstream open reading frame and compared GFP to YFP expression in *E. coli* and *S. 6803* (specifically, the TAA codons were mutated to AAA). In contrast to the leader/no-leader constructs, the fluorescent protein expression from bicistronic GFP and YFP constructs matched expression profiles well in *E. coli* (Figure 4.6A); GFP expression was a good predictor of YFP expression. The Spearman rank correlation coefficient of the bicistronic designs was also improved over that of the basic design, to 0.956. GFP and YFP expression were less well correlated in *S. 6803* (Figure 4.6B). The Spearman rank correlation was 0.527 for *S. 6803*, and the R^2 value was just 0.348.

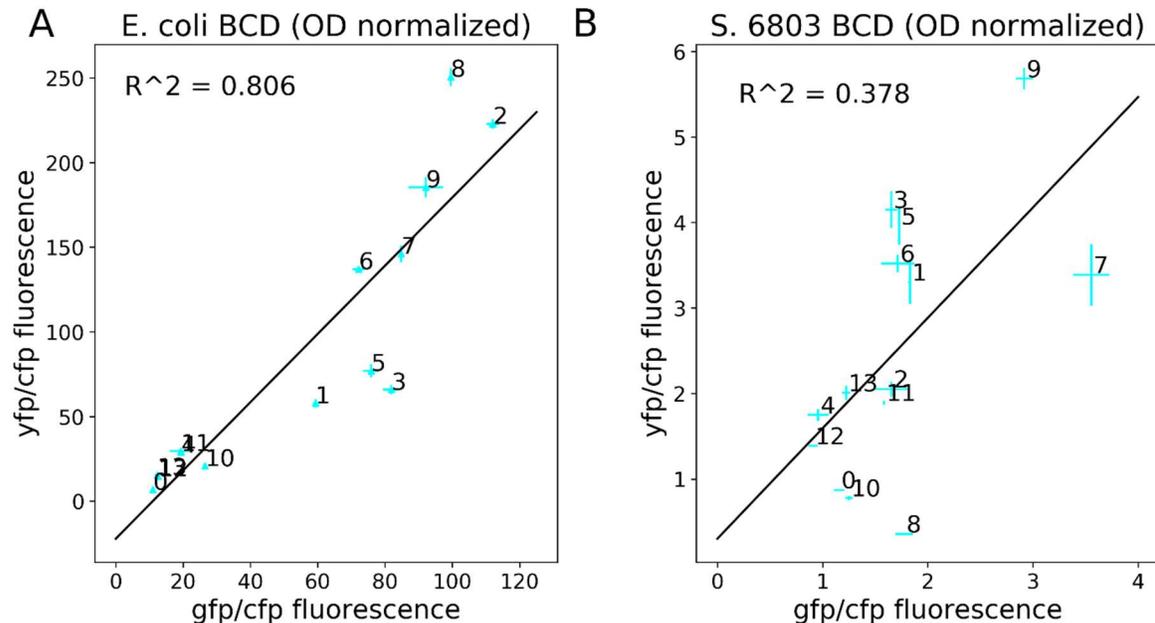


Figure 4.6: Bicistronic designs improved the GFP-YFP correlation in *E. coli* (A), resulting in a Spearman R of 0.956. The bicistronic designs did not improve the correlation in *S. 6803* (B). Measurements are OD and CFP-fluorescence normalized. Error bars represent the standard deviation of two biological replicates.

Comparison between species

Since most genetic parts have been tested in *E. coli*, it is often convenient to utilize those same parts in other species. However, parts such as RBSs, do not always function the same way in different contexts. We compared the measured fluorescence for all plasmids tested in *E. coli* and *S. 6803* (per OD, per CFP, and normalized to the maximum fluorescence for YFP or GFP) (Figure 4.7). We found a correlation with R^2 value of 0.519. The overall Spearman R was 0.812 with values for each design set ranging from 0.415 for BCD_YFP to 0.881 for YFP+leader.

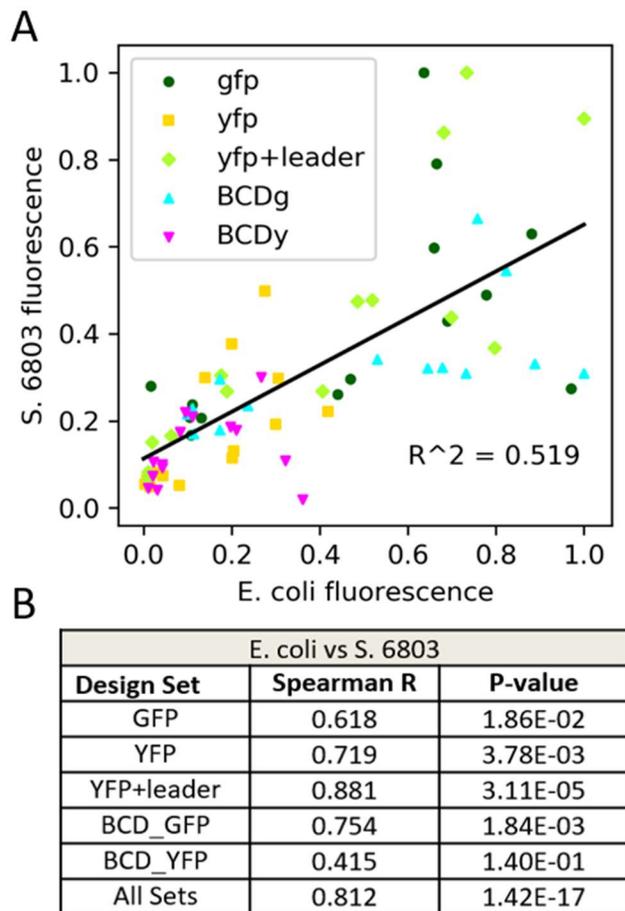


Figure 4.7: Fluorescence measurements were not correlated between *E. coli* and *S. 6803* (A). All GFP-expressing strains were normalized to the maximum GFP measurement, and all YFP-expressing strains were normalized to the maximum YFP measurement in each organism. Spearman rank correlation coefficients and the corresponding P-values between the organisms for each design set are shown in the table (C).

Discussion

Metabolic engineering promises to develop the capability to redesign organisms to modify the profile of metabolites accumulated. A central requirement to achieve this is the ability to precisely and reliably control gene expression. Presently, the design of genetic elements used to control gene expression is not adequate to avoid testing many of these new elements in each new context in which they are used (both at the sequence level, and at the organism level). This work assessed some current methods for predicting gene expression levels across contexts. First, we replicated the work of others who have tested a set of RBS with more than one protein of interest. Second, we tested RBS which were insulated from the gene of interest sequence by including a leader peptide. Third, we adapted the RBS set to the bicistronic design, and tested those with the same to reporter proteins. We compared expression profiles in *E. coli* and *S. 6803* for each design set.

RBSs are not totally modular because their strength depends on the coding sequence

We first replicated the work of Thiel et al. (2018) who compared the expression of GFP and YFP when translation was initiated by the same 13 RBSs. We utilized the same plasmids as a starting point and modified them three ways. We inserted a second Plac promoter, the RBS7, and mTurquoise2 into each of the plasmids as a control for potential variation in the plasmid copy numbers. We also mutated the *mobA* gene to knock out its nicking activity to facilitate cloning, and we removed cloning scars from the plasmids containing YFP which inserted a second potential start codon starting at nucleotide 7 of the coding sequence. This second ATG could code for methionine, but it may also provide an alternative start site for YFP.

The results reported here do not exactly follow the expression profiles previously measured. The GFP measurements are generally well correlated with RBS 4 being an outlier. The YFP measurements are poorly correlated. This may be due to the cloning scar. Overall differences between the data sets could

also be due to different timepoints being used: while we induced at $OD \sim 0.12$ and measured at 48 hrs, while Thiel et al. induced at $OD = 0.5$ and measured fluorescence 6 hr. Six RBSs are the same as those used by Zelcbuch et al. (2013) who also tested expression in *E. coli* using YFP. Our results in *E. coli* and YFP match the rank order found there. Our GFP measurements in *E. coli* differ slightly from that order in that two strongest RBS switched places, and the two weakest RBS switched places.

Like Thiel et al., we did also observe two different expression profiles for GFP and YFP. The linear regression R^2 was 0.537 for *E. coli* and 0.569 for *S. 6803*, while the Spearman R was 0.793 in *E. coli* and 0.684 in *S. 6803*. This is despite the YFP and GFP sequences being quite similar. The full coding sequences are 75% identical (many of these differences are due to codon usage since these two proteins are derived from the same source and share 232 of 241 amino acids). Only nine of the first 35 nucleotides in the two sequences are the different. More sequence diversity at the N-terminus of the coding region could lead to greater difference in expression profiles due to differences in mRNA hybridization between the RBS and coding region and other coding sequences might have worse correlations.

One mechanism that may explain why coding sequences impact relative RBS strength is that secondary structures can span either side of the start codon. Self-hybridization of the mRNA within this region must be broken, or unfolded, to allow the ribosome to properly bind. One mechanism that has been proposed to facilitate initiation in sites with strong secondary structure involves the ribosome binding a nearby, unstructured standby site where it can wait for the unfolding to occur unfolds (de Smit and van Duin 2003; Studer and Joseph 2006). It has been suggested that the SD-antiSD hybridization is another mechanism used to compensate for strong secondary structure (Scharff et al. 2011). All RBSs tested here except RBS0, RBS12, and RBS13 contain SD sequences capable of hybridizing with the 16S rRNA in either *E. coli* or *S. 6803*. If such secondary structures explain differences in expression profiles, our results

suggest that the SD-antiSD hybridization is insufficient for completely overcoming the impacts that those structures have on translation initiation rates.

We expected the insertion of the GFP N-terminus sequence into the 5' end of the YFP coding sequence to make the RBS context more similar and improve the correlation between GFP expression and YFP expression. For most RBSs, the GFP leader peptide increased YFP expression. However, the correlation between GFP and YFP expression was lower in both *E. coli* and *S. 6803*. Alternative designs of the leader peptide may have performed better.

Secondary structure predictions performed using Nupack (Zadeh et al. 2011) show that the YFP mRNA has a strong hairpin structure with 10 base pairs (8 of 10 G/C) starting 30 nucleotides 3' of the start codon. Our leader design shifted this structure to 57 nucleotides 3' of the start codon without introducing a similar, but slightly weaker structure contained at the same position of the GFP mRNA. Although our leader peptide improved the sequence context similarity between the 'gfp' and 'yfp+leader' designs, this structure context was inadvertently altered. This might have been avoided by using a longer leader sequence that would have included the GFP hairpin structure, or by replacing the N-terminus of YFP with that of GFP rather than inserting the GFP sequence. However, it is not clear what impact this hairpin has on expression of these genes and it may be too far from the RBS to have an impact on translation initiation.

Bicistronic design translation initiation is modular in *E. coli*, but not in *S. 6803*

In BCDs, an upstream RBS initiates translation of an upstream open reading frame (μ ORF). In elongation mode, the ribosome has helicase activity which reduces the impact of mRNA secondary structure on re-initiation. The stop codon of the μ ORF overlaps with the start codon of the gene of interest which is expected to result in re-initiation. In *E. coli*, bicistronic designs provide predictable protein expression across different genes of interest. We found a Spearman R of 0.956 for the BCDs tested here. Mutalik et

al. found a Spearman R of 0.85 between GFP and RFP expressed in *E. coli* for 22 different BCDs. In contrast, when the same constructs were tested in *S. 6803*, the Spearman R was only 0.547. Is it possible that *S. 6803* ribosomes do not reinitiate in precisely the same way that they do in *E. coli* at coupled stop-start codons (with sequence TAA_{tg})? Our BCD measurements were not generally higher than the MCD measurements in *E. coli*, in contrast to the findings of Nieuwkoop et al. (2019) (Nieuwkoop, Claassens, and van der Oost 2019).

Ribosome binding sites are not modular across species

In this work, the GFP- and YFP-to-CFP ratios were generally much higher in *E. coli* than they were in *S. 6803*. This is partly due to much higher CFP fluorescence measurements in *S. 6803*, approximately 15-fold higher. Although the mechanisms and components involved in translation initiation are thought to be well conserved across prokaryotic species, differences may have been overlooked. For all sequences tested, the correlation between expression in *E. coli* and *S. 6803* had an R^2 of just 0.519. Most detailed studies of prokaryotic translation initiation have been carried out in *E. coli*. The results have been extrapolated to other prokaryotes *via* bioinformatics studies which examine genomes assuming that things are similar in other species.

In the cyanobacteria, *Synechocystis sp. PCC6803* (*S. 6803*), only about 26% of genes have a SD sequence compared with 57% in *E. coli* (Ma, Campbell, and Karlin 2002). Analysis of the non-SD translation initiation regions in cyanobacteria found bias towards A and U around -10 from the start codon as well as a strong bias for A and U at the -2 position. In gamma-proteobacteria, such genes displayed a bias towards A-rich sequences -2 to -15, while cyanobacteria have a bias against G in this region (Nakagawa, Niimura, and Gojobori 2017). Recently, ribosome profiling indicated the presence of 70S ribosomes 5' of the SD region, providing more evidence that cyanobacterial ribosome assembly may occur in an A/U-rich region -15 to -45 rather than directly at the SD (Karlsen et al. 2018).

In *E. coli*, S1 promotes mRNA unfolding by binding transiently formed single stranded regions (Qu et al. 2012). S1 functions independent of GTP or ATP, in contrast to helicases, and is therefore characterized as an RNA melting protein. Two smaller proteins in *S. 6803*, homologs A and B, have 49% and 43% sequence identity with *E. coli* S1. Mutsuda and Sugiura showed that these were essential to the translation of *rbcS* in *Synechococcus sp.* PPC6301 and 7942 (Mutsuda and Sugiura 2006), but no further studies of these proteins have been reported. Takyar et al. (2005) demonstrated the helicase activity of ribosomes in elongation mode and attributed it to the ribosomal proteins S3 and S4 (Takyar, Hickerson, and Noller 2005). *S. 6803* *rps3* and *rps4* have 53% and 42% identity to *rpsC* and *rpsD* in *E. coli* and no studies on their activity have been reported. S1 ribosomal proteins, which bind A/U-rich RNA, are required for initiation of *rbcS* mRNA in *Synechococcus sp.* PCC 6301 and 7942 (when SD sequence mutated) (Mutsuda and Sugiura 2006). There are two analogs to the *E. coli* S1 protein which are 31 and 38 kDa compared with 61 kDa for the *E. coli* S1. *S. 6803* also has two sequence homologs of similar size to those in *S. 6301*, *rbcS1A* and *rbcS1B* which have not been characterized at this time.

Conclusions

Ideally, metabolic engineers would have a design tool that can provide an RBS sequence with a given expression level based on upstream and downstream sequence and the organism used. For reliable translation initiation in *E. coli*, bicistronic designed sequences are supported as the best option at this time. Our results don't support a similar recommendation for the use of BCD in *S. 6803*. While they were functional in *S. 6803* translation initiation, the expression levels were not similarly ranked for YFP and GFP.

In both *E. coli* and *S. 6803*, inserting an N-terminal GFP sequence at the beginning of the YFP sequence (YFP+leader sets) increased the measured YFP fluorescence for most RBSs. However, inserting the leader did not improve the correlation in expression between GFP and YFP. Bicistronic designs worked well in *E.*

coli in that they resulted in similarly ordered strengths for GFP and YFP. The rank orders of the same BCDs were not very similar in *S. 6803*.

The RBS Calculator has been widely used to predict translation initiation rates of RBS sequences and to design RBS sequences and libraries. It has been updated and improved over the years since 2009 when it was introduced. Here we found it did not accurately predict relative expression levels in either *E. coli* or *S. 6803*. Since it is not designed to predict translation initiation by the re-initiation mechanism used by the BCD, we only used the RBS Calculator to predict the monocistronic design translation initiation rates. The Spearman rank correlation coefficients were 0.720 and 0.852 for the GFP and YFP design, but only 0.582 for the YFP+leader set when tested in *E. coli*. In *S. 6803* the rank correlations were lower for GFP and YFP, at 0.621 and 0.445, but higher for YFP+leader at 0.648.

The overall correlation between *E. coli* and *S. 6803* for all tested plasmids had an R^2 of 0.519 with a Spearman rank correlation coefficient of 0.812. This suggests that RBSs that have been characterized in *E. coli* can usually be used in *S. 6803* with a similar relative translation initiation rate, but not always precisely the same expression level. Different cultivation temperatures of the two organisms could contribute to the difference in expression profiles. Different expression profiles for RBSs have been shown for *E. coli* grown at different temperature and in different medias (Vimberg et al. 2007). Design of modular RBSs faces a difficult challenge from such differences which have an unknown basis.

Synthetic design of ribosome binding sites for use in cyanobacteria should take into account the sequence biases found by Nakagawa et al., (2010) which identified a strong nucleotide bias in cyanobacteria in the -2 position for C and for C or U in the -1 position. Of the RBSs tested here, RBS7 has CC in those positions, while RBS3 has CA and RBS11 has CG. Vimberg et al. also showed that an A/U rich region between 8 and 27 nucleotides 5' of the SD sequence could enhance translation initiation rates (Vimberg et al. 2007). This suggests that this region should also be considered part of the RBS. Eight to

27 nucleotides 5' of the SD sequence generally corresponds to the insulator sequence used in this work (for RBS that had a SD sequence), which is 89% A/U.

Future work

Further research is needed to reduce the context dependence of ribosome binding sites in both sequence context and organism context. Meeting this objective would benefit metabolic engineering because such context dependence makes it difficult to predict how an RBS will perform. Our results suggest that bicistronic designs work well in *E. coli*. In *S. 6803*, however, precise control of translation initiation is elusive, and it remains essential to test several sequences to ensure the desired expression level in *S. 6803*. Improved understanding of translation initiation mechanisms and determinants of initiation rates in cyanobacteria is needed. Future improvements in RBS design might include utilization of RiboJ, which is a self-cleaving ribozyme which has been used to insulate ribosome binding sites from the sequence further 5' by cutting it off following transcription (Lou et al. 2012). Longer leader sequences should also be tested. Kudla et al. (2009), for example, used 28 codons designed to have low secondary structure as a leader which dramatically reduced variability in expression of different codon optimizations of GFP (Kudla et al., 2009). Further efforts to improve control of protein expression in could follow the work of Bernstein et al. (2007), which demonstrated that modification of the native ribosomal proteins could improve heterologous protein expression (Bernstein et al. 2007).

References

- Bernstein, Jeffrey R., Thomas Bulter, Claire R. Shen, and James C. Liao. 2007. "Directed Evolution of Ribosomal Protein S1 for Enhanced Translational Efficiency of High GC *Rhodospseudomonas Palustris* DNA in *Escherichia Coli*." *Journal of Biological Chemistry* 282 (26): 18929–36. <https://doi.org/10.1074/jbc.M701395200>.
- Dienst, Dennis, Julian Wichmann, Oliver Mantovani, João S. Rodrigues, and Pia Lindberg. 2020. "High Density Cultivation for Efficient Sesquiterpenoid Biosynthesis in *Synechocystis* Sp. PCC 6803." *Scientific Reports* 10 (1). <https://doi.org/10.1038/s41598-020-62681-w>.
- Goedhart, Joachim, David von Stetten, Marjolaine Noirclerc-Savoye, Mickaël Lelimosin, Linda Joosen, Mark A. Hink, Laura van Weeren, Theodorus W.J. Gadella, and Antoine Royant. 2012. "Structure-Guided Evolution of Cyan Fluorescent Proteins towards a Quantum Yield of 93%." *Nature Communications* 3 (1). <https://doi.org/10.1038/ncomms1738>.
- Gutu, Andrian, Frederick Chang, and Erin K. O'Shea. 2018. "Dynamical Localization of a Thylakoid Membrane Binding Protein Is Required for Acquisition of Photosynthetic Competency: Vipp1 Patterning Enables Photosynthetic Metabolism." *Molecular Microbiology* 108 (1): 16–31. <https://doi.org/10.1111/mmi.13912>.
- Huang, Hsin-Ho, and Peter Lindblad. 2013. "Wide-Dynamic-Range Promoters Engineered for Cyanobacteria." *Journal of Biological Engineering* 7 (1): 10. <https://doi.org/10.1186/1754-1611-7-10>.
- Hunter, John D. 2007. "Matplotlib: A 2D Graphics Environment." *Computing in Science & Engineering* 9 (3): 90–95. <https://doi.org/10.1109/MCSE.2007.55>.
- Karlsen, Jan, Johannes Asplund-Samuelsson, Quentin Thomas, Michael Jahn, and Elton P. Hudson. 2018. "Ribosome Profiling of *Synechocystis* Reveals Altered Ribosome Allocation at Carbon Starvation." Edited by David F. Savage. *MSystems* 3 (5). <https://doi.org/10.1128/mSystems.00126-18>.
- Kudla, G., A. W. Murray, D. Tollervey, and J. B. Plotkin. 2009. "Coding-Sequence Determinants of Gene Expression in *Escherichia Coli*." *Science* 324 (5924): 255–58. <https://doi.org/10.1126/science.1170160>.
- Lou, Chunbo, Brynne Stanton, Ying-Ja Chen, Brian Munsky, and Christopher A Voigt. 2012. "Ribozyme-Based Insulator Parts Buffer Synthetic Circuits from Genetic Context." *Nature Biotechnology* 30 (11): 1137–42. <https://doi.org/10.1038/nbt.2401>.
- Ludwig, Alfred, Thomas Heimbucher, Wolfgang Gregor, Thomas Czerny, and Georg Schmetterer. 2008. "Transformation and Gene Replacement in the Facultatively Chemoheterotrophic, Unicellular Cyanobacterium *Synechocystis* Sp. PCC6714 by Electroporation." *Applied Microbiology and Biotechnology* 78 (4): 729–35. <https://doi.org/10.1007/s00253-008-1356-y>.
- Ma, Jiong, Allan Campbell, and Samuel Karlin. 2002. "Correlations between Shine-Dalgarno Sequences and Gene Features Such as Predicted Expression Levels and Operon Structures." *Journal of Bacteriology* 184 (20): 5733–45.
- McKinney, Wes. 2010. "Data Structures for Statistical Computing in Python." In , 56–61. Austin, Texas. <https://doi.org/10.25080/Majora-92bf1922-00a>.
- Mutalik, Vivek K, Joao C Guimaraes, Guillaume Cambay, Colin Lam, Marc Juul Christoffersen, Quynh-Anh Mai, Andrew B Tran, et al. 2013. "Precise and Reliable Gene Expression via Standard Transcription and Translation Initiation Elements." *Nature Methods* 10 (4): 354–60. <https://doi.org/10.1038/nmeth.2404>.
- Mutsuda, Michinori, and Masahiro Sugiura. 2006. "Translation Initiation of Cyanobacterial *RbcS* MRNAs Requires the 38-KDa Ribosomal Protein S1 but Not the Shine-Dalgarno Sequence:

- DEVELOPMENT OF A CYANOBACTERIAL IN VITRO TRANSLATION SYSTEM.*” *Journal of Biological Chemistry* 281 (50): 38314–21. <https://doi.org/10.1074/jbc.M604647200>.
- Nakagawa, So, Yoshihito Niimura, and Takashi Gojobori. 2017. “Comparative Genomic Analysis of Translation Initiation Mechanisms for Genes Lacking the Shine–Dalgarno Sequence in Prokaryotes.” *Nucleic Acids Research* 45 (7): 3922–31. <https://doi.org/10.1093/nar/gkx124>.
- Nieuwkoop, Thijs, Nico J. Claassens, and John van der Oost. 2019. “Improved Protein Production and Codon Optimization Analyses in *Escherichia Coli* by Bicistronic Design.” *Microbial Biotechnology* 12 (1): 173–79. <https://doi.org/10.1111/1751-7915.13332>.
- Qu, X., L. Lancaster, H. F. Noller, C. Bustamante, and I. Tinoco. 2012. “Ribosomal Protein S1 Unwinds Double-Stranded RNA in Multiple Steps.” *Proceedings of the National Academy of Sciences* 109 (36): 14458–63. <https://doi.org/10.1073/pnas.1208950109>.
- Scharff, Lars B., Liam Childs, Dirk Walther, and Ralph Bock. 2011. “Local Absence of Secondary Structure Permits Translation of MRNAs That Lack Ribosome-Binding Sites.” Edited by Josep Casadesús. *PLoS Genetics* 7 (6): e1002155. <https://doi.org/10.1371/journal.pgen.1002155>.
- SciPy 1.0 Contributors, Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, et al. 2020. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python.” *Nature Methods* 17 (3): 261–72. <https://doi.org/10.1038/s41592-019-0686-2>.
- Sebesta, Jacob, and Christie AM. Peebles. 2020. “Improving Heterologous Protein Expression in *Synechocystis* Sp. PCC 6803 for Alpha-Bisabolene Production.” *Metabolic Engineering Communications* 10 (June): e00117. <https://doi.org/10.1016/j.mec.2019.e00117>.
- Smit, Maarten H. de, and Jan van Duin. 2003. “Translational Standby Sites: How Ribosomes May Deal with the Rapid Folding Kinetics of mRNA.” *Journal of Molecular Biology* 331 (4): 737–43.
- Studer, Sean M., and Simpson Joseph. 2006. “Unfolding of mRNA Secondary Structure by the Bacterial Translation Initiation Complex.” *Molecular Cell* 22 (1): 105–15. <https://doi.org/10.1016/j.molcel.2006.02.014>.
- Takyar, Seyedtaghi, Robyn P. Hickerson, and Harry F. Noller. 2005. “MRNA Helicase Activity of the Ribosome.” *Cell* 120 (1): 49–58. <https://doi.org/10.1016/j.cell.2004.11.042>.
- Taton, Arnaud, Federico Unglaub, Nicole E. Wright, Wei Yue Zeng, Javier Paz-Yepes, Bianca Brahamsha, Brian Palenik, et al. 2014. “Broad-Host-Range Vector System for Synthetic Biology and Biotechnology in Cyanobacteria.” *Nucleic Acids Research* 42 (17): e136. <https://doi.org/10.1093/nar/gku673>.
- Thiel, Kati, Edita Mulaku, Hariharan Dandapani, Csaba Nagy, Eva-Mari Aro, and Pauli Kallio. 2018. “Translation Efficiency of Heterologous Proteins Is Significantly Affected by the Genetic Context of RBS Sequences in Engineered Cyanobacterium *Synechocystis* Sp. PCC 6803.” *Microbial Cell Factories* 17 (1). <https://doi.org/10.1186/s12934-018-0882-2>.
- Vimberg, Vladimir, Age Tats, Maido Remm, and Tanel Tenson. 2007. “Translation Initiation Region Sequence Preferences in *Escherichia Coli*.” *BMC Molecular Biology* 8 (1): 100. <https://doi.org/10.1186/1471-2199-8-100>.
- Wang, Bo, Carrie Eckert, Pin-Ching Maness, and Jianping Yu. 2018. “A Genetic Toolbox for Modulating the Expression of Heterologous Genes in the Cyanobacterium *Synechocystis* Sp. PCC 6803.” *ACS Synthetic Biology* 7 (1): 276–86. <https://doi.org/10.1021/acssynbio.7b00297>.
- Zadeh, Joseph N., Conrad D. Steenberg, Justin S. Bois, Brian R. Wolfe, Marshall B. Pierce, Asif R. Khan, Robert M. Dirks, and Niles A. Pierce. 2011. “NUPACK: Analysis and Design of Nucleic Acid Systems.” *Journal of Computational Chemistry* 32 (1): 170–73. <https://doi.org/10.1002/jcc.21596>.
- Zelcbuch, Lior, Niv Antonovsky, Arren Bar-Even, Ayelet Levin-Karp, Uri Barenholz, Michal Dayagi, Wolfram Liebermeister, et al. 2013. “Spanning High-Dimensional Expression Space Using

Ribosome-Binding Site Combinatorics." *Nucleic Acids Research* 41 (9): e98–e98.
<https://doi.org/10.1093/nar/gkt151>.

CHAPTER 5: DISCUSSION

Conclusions

We set out first to increase expression of bisabolene synthase in *S. 6803* by varying the coding sequence and the sequences controlling translation initiation. The designed RBS sequences significantly impacted the expression of bisabolene synthase and, therefore, the measured titers of bisabolene. Those impacts were unpredictable. The RBS Calculator was used to design the sequences to have 10-fold higher expression than an initial design. However, only one design of ten designed RBSs resulted in nearly a 10-fold higher expression, while many designs reduced expression.

We hypothesized that the challenge in predicting expression levels was partly due to secondary structures that may form in the translation initiation region and span the 5'-UTR and the coding sequence. Using 13 RBS sequences and two reporter genes with different coding sequences, we showed that the two reporter genes had different expression profiles. This was the case in both *E. coli* and *S. 6803*. We tested two alternative designs for RBS sequences and adapted the 13 RBSs to these designs. Including a leader peptide sequence as part of the RBS sequence was expected to reduce the variability in the expression profiles by maintaining the mRNA secondary structure formation propensity around the start codon for different genes. In both *E. coli* and *S. 6803* this leader peptide failed to improve the correlation between expression of the two reporters. The second design, known as bicistronic design, locates the RBS within a short upstream open reading frame which allows the ribosome to re-initiate from elongation mode, and reduces the impact of mRNA folding on initiation of the gene of interest. This design was successful in *E. coli*, resulting in a strong correlation between the expression of the two reporters. It was not successful in *S. 6803*.

In contrast to previous findings in the Peebles Lab (Cheah et al. 2015), bisabolene specific productivity wasn't reduced by light-dark cycles. Further research to determine why FFA accumulation was decreased in light:dark cycles in the Δ as mutant that over-accumulated FFA in continuous light compared to wildtype. Could in situ extraction of FFA with a dodecane layer improve FFA accumulation in light:dark cycles?

Precise control of translation initiation rates hasn't yet been achieved in cyanobacteria. In Chapter 3, one out of ten ribosome binding sites designed by the RBS Calculator to increase expression 10-fold above the base case actually achieved that in *S. 6803*. We interpreted this as consequence of the RBS Calculator utilizing statistical fitting to measurements in *E. coli*. However, the predictions produced by the Calculator for the RBSs tested in Chapter 5 were not accurate in either species. Instead, the bicistronic designs (BCD) present a more reliable tool for predicting expression in *E. coli*. The set of BCD sequences used here resulted in similar expression profiles for two different reporter genes, and the results from one reporter may be used to predict expression of another protein. This was a significant improvement over the typical RBS designs as well as over the predictions made by the RBS calculators. However, the BCD did not improve the correlation between the two reporters when implemented in *S. 6803*. Many of the promoters that Dr. Allison Werner tested to identify those that oscillate with light:dark cycles failed to produce luxAB luminescence significantly different than the wild type *S. 6803*. Potentially, translation initiation of the luxAB protein was inhibited by secondary structures that formed that do not form when the native genes are expressed from those promoters. Improved understanding of translation initiation in cyanobacteria could potentially improve tool development for biotechnology in these organisms.

Future directions

The work presented in this dissertation focused on improving heterologous protein expression in *S. 6803*. This is only one strategy used to maximize production in metabolic engineering. We considered many other approaches that I did not pursue, but could be implemented in combination to reach higher titers. In light of the results presented here, it is presently essential to test multiple variants of ribosome binding sites and verify predicted protein expression levels. High throughput screening methods may allow us to test thousands of constructs simultaneously. Improved understanding of translation initiation would reduce the sequence space dedicated to this one variable. Further metabolic engineering methods could be applied toward increasing the accumulation of bisabolene in *S. 6803* cultures.

Directed evolution and high-throughput screening

Directed evolution could be potentially improve expression of bisabolene synthase in *S. 6803*. Error-prone PCR could be used to generate sequence diversity in the RBS region. A good starting point for generating mutants would be either of the two strong promoters demonstrated in *S. 6803*: Pcp560 (Zhou et al. 2014) or PpsbA2* (Wang et al. 2018). A C-terminal fusion of GFP to bisabolene synthase could provide a quicker way to estimate relative expression than measuring bisabolene concentrations. Selection of the some of the strains with the highest fluorescence could then be subjected to another round of mutagenesis.

The substrate for bisabolene synthase is farnesyl pyrophosphate, a 15-carbon terpene backbone, which is also a precursor for geranylgeranyl pyrophosphate (GGPP). GGPP is an important precursor for forming photosynthetic pigments including both the carotenoids and chlorophyll. In the strains we tested, no growth defect was observed. Two possible explanations for this are 1) the cells compensated for the redirection of FPP to bisabolene by increasing the total metabolic flux through the MEP pathway,

to GGPP, or 2) the low expression levels of bisabolene synthase were not capable of redirecting enough FPP to produce a growth defect that could be reliably measured. It would be interesting to investigate the effect of the bisabolene synthase on the gene expression of the genes within the MEP pathway and continuing into GGPP synthase. Does this whole path get upregulated when taxed by a new metabolic sink?

Metabolic valve

Since geranylgeranyl pyrophosphate (GGPP) is a precursor for photosynthetic pigments it is presumably essential to cell growth, and it may not be possible to knockout the gene responsible for its production. Instead, attenuation of this gene could be pursued to reduce the competition for FPP. A metabolic valve could be designed such that GGPP synthase expression is reduced to a minimal after a certain cell density is reached. Finding the optimal timing and reduced level of expression would be an interesting experiment. Several possibilities exist for mechanisms that could be used to affect the expression reduction, including CRISPR interference, interference by antisense RNA, or the use of degradation tags.

In vitro part characterization

To accelerate the characterization of genetic elements in bacterial species other than *E. coli*, one important tool that may be used is cell-free transcription-translation systems. Such systems require upfront development, but would facilitate testing by removing the slow process of transforming cells (at least 2 weeks), and growing them to test them. At this time, no cell-free systems for cyanobacteria have been reported. Most papers published utilize *E. coli* extracts, though a few such systems have been developed for other organisms. We worked to adapt those systems to *S. 6803* on a shoestring budget and were not able to demonstrate expression of reporter genes. Initially, this was attempted without using any additives such as ATP, amino acids, and components needed for energy regeneration. While troubleshooting, more components that are typically used in these systems (which are also expensive)

were gradually added, but still without success. The conclusion of Abraham Martinez, the REU (Research Experience for Undergraduates) program student who worked on this for a summer, was that it was essential to use a French press to lyse the cells in order to generate membrane vesicles which contain some membrane-bound proteins needed for the energy regeneration system. Development of a cell-free system would accelerate progress in prototyping genetic parts and circuits in *S. 6803*.

High-density cultivation and product extraction

The cultivation conditions for bisabolene production can be modified to increase the titer. A recent paper demonstrated high density cultivation of *S. 6803* engineered to produce bisabolene could dramatically increase the titer (Dienst et al. 2020). High density was enabled by growing the cyanobacteria in cultivator vessels (HDC 6.10 starter kit from CellDEG) where the cell cultures are separated by a hydrophobic membrane from a carbonate buffer to continuously provide high CO₂ concentrations. This setup may have more efficient capture of CO₂ than bubbling CO₂-enriched air through the culture and doesn't require the energy input of compressing the gas. Scale-up of this type of system, however, may present other challenges. The optical density reached 42.6 after 192 hours with a bisabolene titer of 179 mg/L. This was achieved using the P_{petE} promoter and a bicistronic design RBS from Mutalik et al. (Mutalik et al. 2013). In this report bisabolene was recovered from a dodecane overlay of the cultures. It is unclear how the bisabolene is transported from inside the cell to the dodecane. Several publications have explored more active extraction of microalgae cultures.

A series of papers studying continuous extraction of carotenoids from the microalgae *Dunaliella salina* (summarized in (Kleinegris, Janssen, et al. 2011b)) found that contact with dodecane ruptured cells, allowing extraction of the hydrophobic molecules and causing cell death. An overlay like what we used to recover bisabolene did not measurably impact growth rates, but sparging the dodecane through the culture did reduce the growth rate (Kleinegris, van Es, et al. 2011). After long periods of exposure to

dodecane, cell membrane components can accumulate and act as surfactants that result in the formation of an emulsion which can complicate recovery of products (Kleinegris, Janssen, et al. 2011a). Can cyanobacteria be engineered to excrete hydrophobic products and reduce the need for cell lysis for extraction?

Pathway engineering

As originally proposed in the grant application that funded the bisabolene project, the metabolic pathway from central metabolism to isopentenyl pyrophosphate could be modified to increase metabolic flux to farnesyl pyrophosphate. This would include just increasing expression of the FPP synthase. FPP is expected to be toxic as it is in *E. coli* (Martin et al. 2003), so the RBS of FPPS in operon to increase expression until growth defect found. While prokaryotes generally use the MEP pathway for production of DMAPP and IPP precursors for terpenoids, eukaryotes and archaea use the MVA pathway. Researchers have introduced the entire MVA pathway to *E. coli* to increase production of terpenoids (Martin et al. 2003). This has been done in cyanobacteria as well, and even though the pathway expression was not optimized to maximize flux through it (or balanced as in *E. coli* as the previously mentioned paper did), the titer was increased by 2.5-fold (Bentley, Zurbriggen, and Melis 2014). To save money on codon optimized gene synthesis, plant mRNA could be used as in Englund et al. (2015). Further, it would be interesting to test some 5'-UTRs from plants to see if they function for translation initiation in cyanobacteria. For genes that lack Shine-Dalgarno sequences (which is most genes in *S. 6803*), Nakagawa et al. found a strong nucleotide bias in *S. 6803* genes to have 'ACU' or 'ACC' before the start codon similar to part of the Kozak sequence ((Nakagawa, Niimura, and Gojobori 2017). Bisabolol would be less hydrophobic than bisabolene which may facilitate higher productivity. However, Dienst et al. (2020) recently compared similar constructs expressing either bisabolene or bisabolol synthase, and they both achieved similar titers. I constructed plasmids to express bisabolol synthase in

S. 6803, but did not have time to test them. Those plasmids would allow comparison between a codon optimized gene and the sequence cloned from *Maricaria chamomillia* mRNA.

Translation initiation

Further work is needed to address the context dependence of translation initiation rates of ribosome binding sites. Our study showed that neither the leader peptide nor the bicistronic designs reduced the context dependence. It is unclear why the bicistronic designs might not work in S. 6803. Longer leader peptide sequences could reduce more of the potential secondary structures that may prevent ribosome binding. In addition, the leader peptides could be designed to have weaker secondary structure by making them A/T rich. Nakagawa et al. found that S. 6803 genes that lacked 5'-UTR Shine-Dalgarno sequences were A-rich and G-depleted (Nakagawa, Niimura, and Gojobori 2017). Since nearly all of the ribosome binding sites characterized in S. 6803 contain an SD sequence, researchers should examine some of the native RBSs that lack the SD sequence as well as based on the consensus sequence identified by Nakagawa et al. (2017).

The free energy of folding mRNA is used by the RBS Calculator. However, the accuracy of that free energy calculation is questionable. First, those calculations are completed assuming 30 °C, and the change in free energy of folding may be different at the optimal growth temperature of S. 6803, 37 °C. Second, the free energy of folding calculation likely assumes a salt concentration, an RNA concentration, and concentrations of other nucleotides which may not be consistent with cyanobacteria. Third, the folding kinetics also impact translation initiation rates. All these may need to be addressed to improve predictions of translation initiation rates from mRNA sequences in cyanobacteria.

References

- Bentley, Fiona K., Andreas Zurbriggen, and Anastasios Melis. 2014. "Heterologous Expression of the Mevalonic Acid Pathway in Cyanobacteria Enhances Endogenous Carbon Partitioning to Isoprene." *Molecular Plant* 7 (1): 71–86. <https://doi.org/10.1093/mp/sst134>.
- Cheah, Yi Ern, Allison J. Zimont, Sunny K. Lunka, Stevan C. Albers, Sei Jin Park, Kenneth F. Reardon, and Christie A.M. Peebles. 2015. "Diel Light:Dark Cycles Significantly Reduce FFA Accumulation in FFA Producing Mutants of *Synechocystis* Sp. PCC 6803 Compared to Continuous Light." *Algal Research* 12 (November): 487–96. <https://doi.org/10.1016/j.algal.2015.10.014>.
- Dienst, Dennis, Julian Wichmann, Oliver Mantovani, João S. Rodrigues, and Pia Lindberg. 2020. "High Density Cultivation for Efficient Sesquiterpenoid Biosynthesis in *Synechocystis* Sp. PCC 6803." *Scientific Reports* 10 (1). <https://doi.org/10.1038/s41598-020-62681-w>.
- Kleinegris, Dorinde M. M., Marjon A. van Es, Marcel Janssen, Willem A. Brandenburg, and René H. Wijffels. 2011. "Phase Toxicity of Dodecane on the Microalga *Dunaliella Salina*." *Journal of Applied Phycology* 23 (6): 949–58. <https://doi.org/10.1007/s10811-010-9615-6>.
- Kleinegris, Dorinde M.M., Marcel Janssen, Willem A. Brandenburg, and René H. Wijffels. 2011a. "Continuous Production of Carotenoids from *Dunaliella Salina*." *Enzyme and Microbial Technology* 48 (3): 253–59. <https://doi.org/10.1016/j.enzmictec.2010.11.005>.
- . 2011b. "Two-Phase Systems: Potential for in Situ Extraction of Microalgal Products." *Biotechnology Advances* 29 (5): 502–7. <https://doi.org/10.1016/j.biotechadv.2011.05.018>.
- Martin, Vincent J J, Douglas J Pitera, Sydnor T Withers, Jack D Newman, and Jay D Keasling. 2003. "Engineering a Mevalonate Pathway in *Escherichia Coli* for Production of Terpenoids." *Nature Biotechnology* 21 (7): 796–802. <https://doi.org/10.1038/nbt833>.
- Mutalik, Vivek K, Joao C Guimaraes, Guillaume Cambray, Colin Lam, Marc Juul Christoffersen, Quynh-Anh Mai, Andrew B Tran, et al. 2013. "Precise and Reliable Gene Expression via Standard Transcription and Translation Initiation Elements." *Nature Methods* 10 (4): 354–60. <https://doi.org/10.1038/nmeth.2404>.
- Nakagawa, So, Yoshihito Niimura, and Takashi Gojobori. 2017. "Comparative Genomic Analysis of Translation Initiation Mechanisms for Genes Lacking the Shine–Dalgarno Sequence in Prokaryotes." *Nucleic Acids Research* 45 (7): 3922–31. <https://doi.org/10.1093/nar/gkx124>.
- Wang, Bo, Carrie Eckert, Pin-Ching Maness, and Jianping Yu. 2018. "A Genetic Toolbox for Modulating the Expression of Heterologous Genes in the Cyanobacterium *Synechocystis* Sp. PCC 6803." *ACS Synthetic Biology* 7 (1): 276–86. <https://doi.org/10.1021/acssynbio.7b00297>.
- Zhou, Jie, Haifeng Zhang, Hengkai Meng, Yan Zhu, Guanhui Bao, Yanping Zhang, Yin Li, and Yanhe Ma. 2014. "Discovery of a Super-Strong Promoter Enables Efficient Production of Heterologous Proteins in Cyanobacteria." *Scientific Reports* 4 (March). <https://doi.org/10.1038/srep04500>.

Appendix 1: METHODS FOR GENETICALLY MODIFYING *SYNECHOCYSTIS*

SP. PCC 6803 AND CHARACTERIZING ENGINEERED STRAINS³

Summary

Synechocystis sp. PCC 6803 is a model cyanobacterium which has been investigated to produce a variety of fuels and chemicals. Genetic mutations are of interest for studying photosynthesis and engineering chemical production. Methods for culturing, preserving, and genetically transforming *Synechocystis* sp. PCC 6803 are detailed here, complete with methods to test promoter strength using the green fluorescent protein reporter. Furthermore, a method for markerless transformation of chromosomal DNA is presented. Sufficient detail is described to enable application by the novice. This chapter updates the publication listed above, adding methods for transformation by electroporation, Western blotting, and affinity tag protein purification.

Introduction

Cyanobacteria, formerly called blue-green algae, perform oxygenic photosynthesis using solar light energy and atmospheric carbon dioxide. *Synechocystis* sp. PCC 6803 (hereafter *Synechocystis* 6803) is a

³ Portions of this chapter were co-written with Allison Werner and have been published as:

Sebesta, Jacob, Allison Werner, and Christie Ann Marie Peebles. 2019. "Genetic Engineering of Cyanobacteria: Design, Implementation, and Characterization of Recombinant *Synechocystis* Sp. PCC 6803." In *Microbial Metabolic Engineering*, edited by Christine Nicole S. Santos and Parayil Kumaran Ajikumar, 1927:139–54. Methods in Molecular Biology. New York, NY: Springer New York.

https://doi.org/10.1007/978-1-4939-9142-6_10.

model cyanobacterium by virtue of a fully-sequenced, publicly-available genome (Kaneko et al. 1996; Nakao et al. 2010). *Synechocystis* 6803 performs natural DNA uptake and homologous recombination, making the process of modifying chromosomal DNA relatively easy. Compared to common heterotrophic production chassis such as yeast and *E. coli*, cyanobacteria have a major advantage of not requiring expensive sugar feedstocks (which results in competition with food supply systems) while maintaining high growth rates than other photosynthetic organisms. Engineered cyanobacteria have great potential to be more efficient than plant feedstocks which require multiple conversion steps and subsequent fermentation. Many researchers are working to extend metabolic engineering principles developed in *E. coli* and yeast to cyanobacteria. Taken together, these advantages have resulted in *Synechocystis* 6803 gaining wide attention as a potential chassis for the production of a variety of chemicals (Angermayr, Gorchs Rovira, and Hellingwerf 2015). The objective of this chapter is to describe the protocols necessary for obtaining, culturing, preserving, and genetically modifying *Synechocystis* 6803 with sufficient detail so that new investigators can apply them successfully. An example is provided: constructing and transforming a promoter:GFP cassette into the *Synechocystis* 6803 *slr0168* neutral site using markerless transformation and testing promoter strength by measuring fluorescence.

Materials

Cell Culture

1. Glass Erlenmeyer flasks, 250 mL (see **Note 1**)
2. Foam Stoppers
3. Glass storage bottles, 1 L
4. BG-11 Stock Solutions (see **Table A1.1**), sterile
5. 3M Hydrochloric acid, sterile

6. 3M Sodium hydroxide, sterile
7. Antibiotic solutions (see **Table A1.2**)
8. Culture plates
9. Agar
10. Sodium thiosulfate, sterile
11. *N*-Tris(hydroxymethyl)methyl-2-aminoethanesulfonic Acid, sterile
12. Glass beads or cell spreaders, sterile
13. Surgical tape
14. 50% Dimethyl sulfoxide, sterile
15. Centrifuge tubes, 15-50 mL
16. Microcentrifuge tubes, 1.5 mL
17. Gloves
18. 75% ethanol in spray bottle
19. Laminar flow hood or biosafety cabinet
20. Temperature-controlled growth chamber with lights and platform shaker
21. -80°C freezer

DNA Design for Homologous Recombination

1. Software for visualizing DNA sequences (see **Note 2**)

Plasmid Construction

1. Cloning vectors (see *Methods §3.2.3.*)
2. Chemically-competent *E. coli* strain

3. RNA polymerase
4. Nuclease-free water
5. KpnI restriction enzyme (Thermofisher #ER0521)
6. NdeI restriction enzyme (Thermofisher #ER0581)
7. T4 DNA ligase

Natural DNA Uptake and Homologous Recombination

1. Falcon tubes, 14 mL
2. Cell spreaders, sterilized
3. Inoculation loops or toothpicks, sterilized
4. Petri dishes
5. Glass Erlenmeyer flasks, 250mL

Cell Culture

All handling of cyanobacteria and media should be performed using sterile technique in a laminar flow hood. Wild-type *Synechocystis* 6803 [N-1] can be obtained from the American Type Culture Collection or the Pasteur Culture Collection (*see Note 3*).

1. Prepare BG-11 components as specified in **Table A1.1**. Autoclave components A, B, and C on liquid cycle (121°C, 15 PSIG) for >30 minutes. Filter sterilize the Trace Metals component. Store at 4°C.
2. To prepare BG-11 liquid media, autoclave 1 L deionized water in a 2 L glass bottle. Using sterile technique in a laminar flow hood, add 10 mL 100X A, 10 mL 600X B (*see Note 4*), 10 mL 100X C,

and 1 mL 1000X Trace Metals. Adjust the pH to 8.0 using HCl and NaOH as needed. Add 10 mL 1M TES-NaOH, pH 8.0. Store at 4°C or room temperature.

3. To prepare BG-11 agar plates (60 mm diameter petri dishes), autoclave 700 mL deionized water with 7 g agar and a stir bar. Allow to cool, while gently stirring. When the media temperature reaches ~55°C, add 7 mL 100X A, 7 mL 600X B, 7 mL 100X C, and 0.7 mL 1000X Trace Metals. Add 7 mL 1M TES-NaOH, pH 8.0. In addition, add 14.6mL 1M sodium thiosulfate (to 3.3 g/L) and antibiotics as needed (see **Table A1.2** and **Note 5**). Pour into plates and let sit until solidified. Agar plates can be stored at 4°C; however, plates less than 1 week old allow faster cyanobacteria growth.
4. To inoculate and maintain a liquid cyanobacterial culture, add liquid BG-11 and appropriate antibiotics (see **Table A1.2**) to an Erlenmeyer flask. Inoculate liquid culture by pipetting in cells from another culture (*e.g.* from freezer stock) or scraping solid culture (*e.g.* individual colony, agar stab) to 20-30 mL of liquid BG-11. Incubate at 26-30°C, 225 rpm, 50-200 $\mu\text{mol photons m}^{-2} \text{ s}^{-2}$ for 2-5 days (see **Note 6**).
5. Monitor cell growth by performing cell counts or optical density measurements at 730 nm. Routine culture maintenance should remain below $\text{OD}_{730} \sim 2.0$. Passage number (the number of times a colony is re-streaked or transferred to a new flask) should be kept as low as possible to minimize genetic drift of the strain.
6. To transfer a liquid cyanobacterial culture to solid media, pipette 10-100 μL liquid culture onto a BG-11 agar plate and spread using autoclaved glass beads or cell spreaders (see **Note 7**). Seal plate with surgical tape and incubate at 26-30°C, ~40% humidity, 50-200 $\mu\text{mol photons m}^{-2} \text{ s}^{-2}$ for 5-15 days (see **Note 8**).
7. To freezer stock a cyanobacterial culture, grow culture in BG-11 supplemented with appropriate antibiotics to $\text{OD}_{730} 0.6-0.9$. Centrifuge culture at 2,760g for 5 minutes. Discard supernatant and

re-suspend cell pellet in 1/10th volume BG-11 liquid media (no antibiotic). Add DMSO (see **Table A1.2**) to final concentration of 5%. Aliquot into 1.5 mL microcentrifuge tubes and store at -80°C (Section 3.1.4, above, describes how to revive a culture from a DMSO stock).

Counterselection for Markerless Transformation

1. Same materials as *Materials §2.4. plus* nickel sulfate hexahydrate

Methods

This section describes methods for:

1. Inoculating, maintaining, and preserving *Synechocystis* 6803 cultures;
2. Designing and constructing plasmids for transformation of *Synechocystis* 6803 chromosomal DNA;
3. Transforming the chromosomal DNA of *Synechocystis* 6803 using an antibiotic selection marker; and
4. Transforming the chromosomal DNA of *Synechocystis* 6803 using a counterselection marker.

Plasmid Design and Construction

1. Retrieve the DNA sequence of interest. Native *Synechocystis* 6803 genes can be accessed from Cyanobase (<http://genome.microbedb.jp/cyanobase/>), and non-native genes can be found on NCBI GenBank or an organism-specific database (see **Note 9**). Save the sequence as a new file in your choice of DNA editing software. If a genetic knock-out is the end goal, the gene of interest should be an antibiotic resistance cassette or other selectable marker.
2. Determine which expression control elements will be applied. Promoters and ribosome bindings sites (RBSs) are important control elements to consider for strain design. Several native

Synechocystis 6803 promoters, as well as modified *E. coli* promoters that perform in *Synechocystis* 6803, have been previously characterized (Albers, Gallegos, and Peebles 2015a; H. H. Huang et al. 2010a; Camsund and Lindblad 2014; Albers and Peebles 2016). The RBS Calculator is a valuable tool for assessing native RBSs and designing sequences for strong translation (Salis 2011). Codon optimization is frequently employed to optimize heterologous gene expression, especially with genes derived from eukaryotes. Commercial gene synthesis companies typically offer codon optimization free of charge when synthetic genes are ordered. However, codon optimization is not fully understood, and the wide variety of mechanisms by which codon usage impacts gene expression introduces a degree of uncertainty to optimization. A good review of codon usage in this context has been provided by Quax and colleagues (Quax et al. 2015).

3. Selection of a cloning vector to use for assembling the DNA construct in *E. coli* should be informed by the vector copy number, selection marker, multiple cloning site (MCS), and presence of homologous regions to *Synechocystis* 6803 neutral sites (see **Note 10**). For example, pIGA4 is a plasmid used in our lab to test promoter strengths. It uses the pUC origin of replication to achieve a high copy number, contains the ampicillin antibiotic resistance marker, and homologous regions (HRs) for the transformation into the *slr0168* neutral site (Albers, Gallegos, and Peebles 2015a). Between the homologous regions, the plasmid contains the gene sequence for a green fluorescent protein variant mutated to degrade more quickly than the wild-type, facilitating measurement of rapid changes in expression (Albers, Gallegos, and Peebles 2015a). The plasmid backbone has been mutated to remove an NdeI cut site from the backbone so that KpnI and NdeI cut sites can be used to clone in and test different promoters. If a vector is used that does not have HRs, they need to be introduced flanking the gene of interest (or MCS) and selection marker.

4. Select a cloning method to insert the gene of interest into the cloning vector. Many options exist, including “cut-and-paste” using restriction enzymes and ligase, GoldenGate (Engler and Marillonnet 2014), ligase cycling (Kok et al. 2014), Gibson assembly (Gibson et al. 2009) and others. In the simplest cases, a gene can be inserted into the MCS of a vector using a cut-and-paste approach, explained here.
5. Design PCR primers to copy the gene or DNA region of interest (see **Note 2** for DNA editing software options). In general, primers should be 15-30 bases long, have a melting temperature (T_m) between 50-60°C, finish with a G or C at the 3’ end, and have minimal predicted primer dimers or secondary structure. Additionally, the forward and reverse primers melting temperatures should be within 5°C of each other. If using cut-and-paste into the MCS, add restriction endonuclease recognition sequences to 5’ end of both primers. In this case, four bases should be added on the far 5’ end to facilitate enzyme binding.
6. Follow standard cloning practices for the method of choice to assemble the plasmid.

Natural DNA Uptake and Homologous Recombination

6803 is naturally competent (*i.e.* naturally uptakes DNA) due to the presence of a type IV-like pilus structure (Yoshihara et al. 2001). Outlined here is a protocol for transforming *Synechocystis* 6803. This protocol takes into consideration some of the findings of Zang et al. (2007), which previously examined the effects of variables such as homologous region length and the time of incubation with plasmid DNA on transformation efficiency (Zang et al. 2007). Though it is not included in this protocol, Wang et al. (2015) showed that pre-methylation of the plasmid DNA used to transform by methyltransferase C from *Synechocystis Sp.* PCC6803 improved transformation efficiency (Wang et al. 2015a).

1. To concentrate *Synechocystis Sp.* PCC6803 cells for natural DNA uptake, grow cells to mid-log phase. For each transformation, aliquot 10 mL of cell culture into a 15 mL conical centrifuge tube on ice. Centrifuge each tube at 4000g for 6 minutes at 4°C. In addition to the tubes

destined for transformation, prepare a no-DNA negative control. Discard the supernatant, re-suspend the cell pellet in 100 μL of BG-11, and transfer suspension to a 14 mL round-bottom falcon tube.

2. Add 1 μg of the target plasmid to the suspension and incubate cells for 5 hours under low light (20-100 $\mu\text{mol photons m}^{-2}\text{s}^{-1}$) at 30°C. Tubes should be gently shaken by hand after \sim 2.5 hours to until the cells are re-suspended.
3. Pipette 100 μL of the transformation mixture onto an BG-11 agar plate without antibiotics and spread using a cell spreader. Seal the plate with surgical tape and incubate under low light at 30°C for 48 hours for recovery. The plate should be covered with a dense green film of cells at this point.
4. After 48 hours, select for transformants (colonies that have the desired mutation). Use a bacterial loop spreader to perform a triple streak procedure from the recovery plate to a BG-11 plate supplemented with antibiotics (see **Table A1.2**). To isolate individual colonies, perform a triple streak by touching the loop spreader to the plate and tracing a long, serpentine path through one third of a fresh plate containing antibiotics. Flame the spreader to sterilize. Once the spreader is cool, the loop is streaked through part of the first serpentine path to gather some cells but less than was initially gathered from the plate lacking antibiotics. These cells are then spread through a second third of the plate in a second serpentine path. The loop is flamed and cooled again, then dragged through the second serpentine path to collect cells and streaked in the last third of the plate in a third serpentine path. Seal the plate with surgical tape and incubate under low light at 30°C. Single colonies should be observed in 5-14 days.
5. To verify that your gene of interest has been inserted by homologous recombination, perform a colony PCR (cPCR) with primers that bind-to and amplify the region of modification. Briefly, design primers which bind to the homologous region outside of the DNA inserted into the

cloning vector (see *Methods 3.2.5* for primer design). As a negative control, perform a cPCR on wild-type *Synechocystis Sp. PCC6803* (see **Note 11**).

6. To achieve full segregation, re-streak single colonies on a fresh BG-11 plate with antibiotics and place in 30°C lighted incubator. Repeat this process once a week for 3-4 weeks or until full segregation is achieved (see **Note 12**). Sequence confirmation of segregated mutants is highly recommended.

Counterselection for Markerless Recombination

To remove selection markers, counterselection methods have been developed for transformation of *Synechocystis Sp. PCC6803*. This allows the accumulation of multiple mutations without the potential burden of expressing multiple antibiotic resistance genes. Described here is a counterselection method that utilizes nickel-inducible expression of *mazF*, a toxic endoribonuclease, and *aphII*, kanamycin resistance, into the locus of interest (Cheah, Albers, and Peebles 2013). Nickel induction of the *mazF* gene is due to the use of the native, nickel-responsive *nrsB* promoter.

To perform markerless recombination, two transformations are necessary: (1) insert the *mazF/kanR* cassette into the HR of interest, and (2) replace *mazF/kanR* with the modification/gene of interest. The following protocol describes the steps necessary to select against cells retaining the *mazF/aphII* cassette and for the cells that obtained the modification of interest (see **Note 13**). Here we assume that the starting strain already contains the *mazF/KanR* cassette inserted into the locus of interest (see *Methods 3.3.* for transformation technique) and that a new plasmid has been constructed to replace the cassette using the same homologous regions (see *Methods 3.2.* for plasmid design and construction).

1. To prepare *Synechocystis Sp. PCC6803* for natural DNA uptake, follow steps 1-2 in *Section 3.3*, above.

2. Following the 5 hour incubation with the plasmid DNA, centrifuge the cells at 4000g for 6 minutes. Discard the supernatant and resuspend the cell pellet in a small amount of fresh BG-11 (100-500 μ L).
3. Add the resuspended cells to 10 mL of fresh BG-11 in a shake flask and incubate for an additional 24 hours.
4. With a cell spreader, spread 100 μ L of each culture on BG-11 agar plates supplemented with 20 μ M nickel sulfate (see **Table A1.2**). Seal the sides of the plates with surgical tape. Incubate the plates under low light at 30°C, 40-50% humidity, for 5-7 days.
5. Once colonies form, streak single colonies on both BG-11 plus nickel plates and BG11 plus kanamycin plates. The colonies with the expected insert should survive on BG-11 plus nickel but not on BG-11 plus kanamycin plates. Colonies may need to be re-streaked two or more times on nickel plates before the chromosomal copies are completely segregated (see **Note 14**).
6. Colony PCR (cPCR) may be used to confirm the presence of the mutations as well as the absence of the wild-type fragment. Generally, primers that anneal to the homologous region and amplify the region where the insertion is expected. This allows the detection of the wild-type fragment and/or the mutation. If the cPCR results in multiple bands, both versions of the chromosome are present. However, if the insert and the fragment it replaces are not significantly different in length, two reactions for each colony may be necessary to determine the presence/absence of the two versions. In such a case, one reaction would check for the presence of the correct mutation by including one primer that binds within the homologous region and the second binding with the gene being inserted, while the second would confirm the absence of the mazF/kanR cassette by including one primer that binds within the homologous region and the second binding within the mazF/kanR cassette. For such dual reaction tests, full segregation is indicated when only the first of these reactions produces a band.

7. To start a liquid culture, take a swab of cells from the BG-11 plus nickel plate and place it in 1 mL of BG-11 liquid media supplemented with 20 μ M nickel sulfate in a 14 mL round-bottom Falcon tube. Incubate at 30°C with shaking. Once cells reach a sufficient cell density (about 2 days), inoculate 50 mL of BG-11+20 μ M nickel with the cell suspension. The culture should be transferred to fresh media every 5-7 days.

Synechocystis sp. PCC6803 Electroporation Protocol (version 2)

Based on the following publication:

Ludwig et al. "Transformation and gene replacement in the facultatively chemoheterotrophic, unicellular cyanobacterium *Synechocystis* sp. PCC6714 by electroporation." DOI: 10.1007/s00253-008-1356-y <https://www.ncbi.nlm.nih.gov/pubmed/18286280>

1. Grow *Synechocystis* Sp. PCC6803 (S. 6803) in 50 mL BG11 to an OD_{730nm} of approximately 0.7
2. Centrifuge at 6000 x g for 5 minutes¹
3. Repeat 3x:
 - a. Resuspend in 10 mL 1 mM HEPES buffer, pH 7.5 (approx.)
 - b. Centrifuge at 6000 x g for 5 minutes
4. Resuspend in 300 μ L 1mM HEPES² and aliquot 60 μ L into 5 PCR tubes
5. Add 4 μ L (1-8 μ g) plasmid DNA to 60 μ L of washed cells for each electroporation³
6. Set electroporator to following settings:
 - Field strength: 2400 V (2 mm gap electroporation cuvettes)⁴
 - Resistance: 400 Ohm
 - Capacitor: 25 μ F
 - Time constant (target): \sim 9 ms⁵

7. Pipette the cells plus DNA into the electroporation cuvette
8. Have BG11 ready to go
9. Electroporate and quickly add 1 mL of BG11
10. Incubate in light with shaking at 30 °C for 48 hours
11. Streak cells onto BG11 plus antibiotics corresponding to the marker in the plasmid transformed
12. Colonies should appear in 1-2 weeks

Notes:

- 1: These steps may be completed at room temperature
- 2: Paper resuspended in 120 μ L and completed one transformation per 50 mL culture.
- 3: Paper found optimal DNA amount as 1-10 μ g DNA in 60 μ L of cells
- 4: I used 2 mm cuvettes because that's what I had. Cuvettes with a 1 mm gap are usually used for bacteria. Adjust voltage for different cuvette gaps to meet 1200 V/mm. The resistance may also need to be adjusted so that the time constant is around 9 ms.
- 5: The time constant is somewhat variable. Don't worry about it being exactly 9 ms every time

Western blot protocol for quantifying proteins in *Synechocystis sp.* PCC6803

1. Thaw cell pellets on ice (from 30ml of culture)
2. Resuspend with 500 μ L ice cold PBS
3. Centrifuge at 4°C at 6000g for 5 minutes
4. Discard supernatant
5. For total soluble protein, resuspend in 450 μ L ice cold lysis buffer

- a. Lysis buffer: 1x PBS, pH 8.0, 1% TritonX100, 1mM DTT, 1x HALT protease inhibitor
 - b. If insoluble, protein aggregates desired, 8M urea can be used as part of the lysis buffer
6. Sonicate (on ice) with microtip probe
 - a. Amplitude = 25, 3s on 10s off for 2min processing time
7. Centrifuge at 4°C 10,000g for 30 minutes
8. Supernatant should be emerald green
9. If it is not, and a dark green pellet remains, resuspend and sonicate, centrifuge again
10. Measure protein concentration with BCA
 - a. 5 µL sample + 250 µL Bradford reagent (at room temp)
 - b. Make a new standard curve if using different lysis buffer. Otherwise, use this standard curve equation: $\text{Conc [ng/}\mu\text{L]} = 3,154.9 \cdot \text{Abs} - 116.1$
 - c. It's a good idea to make new standard curves occasionally, or at least verify the accuracy of this one using bovine serum albumin standard solutions.
11. Mix samples with water, loading dye, DTT according to "gels.xls" (use consistent total volume)
12. Boil samples at 95°C for 5 minutes, cool on ice
13. Centrifuge briefly
14. Load gel (mini-protean TGX stainfree, 4-15% 10x50 µL wells) and run 150 V for 1hr15min
 - a. Use standard Tris-glycine-SDS 1x buffer
15. Prepare Towbin buffer and chill in fridge while gel runs
 - a. 1x Tris/glycine/SDS, pH 8.3, 20% methanol
16. Break open gel cassette and image on gel-imager
17. Place gel on imager
18. Set excitation to 302 nm and filter to SYBR green
19. Develop gel by exposing to 302 nm light for 1min

20. Image gel (usually capture 3s, 6s, and 12s exposures)
21. Place gel in chilled Towbin buffer to equilibrate while setting up western
22. Cut out PVDF to gel size (nick one corner to correspond to the large-MW end of lane 1 of the gel)
23. Soak PVDF in 100% methanol for 15-60 s, then place in Towbin buffer
24. Put some ice in a smaller autoclave tray, then place the transfer bath in with the lid on. Fill the autoclave tray around the sides of the transfer bath with ice
25. Pour in the chilled Towbin buffer
26. Wet two transfer sandwich sponges and two filter papers in the Towbin buffer
27. Assemble sandwich in the following order
 - a. Black plate of cassette – sponge – filter paper – gel – PVDF – filter paper – sponge – red plate of cassette
28. Place cassette in transfer bath with red plate on the side with the red terminal
29. Place autoclave tray with the loaded transfer bath in the walk-in fridge
30. Connect to an electrophoresis control box (with the correct wiring!) and transfer overnight, 16 hrs. at 30 V
31. Image gel and PVDF with the gel imager (302nm and SYBR green filter) to ensure full transfer of protein to the PVDF
32. Block membrane for 1hr at room temperature (RT) with gentle shaking
33. 40 ml PBS + 0.05% Tween20 + 5 mM EDTA + 5% (w/v) dehydrated skim milk
34. Pour out blocking buffer and add primary antibody solution
35. 15 mL PBS + 0.05% Tween20 + 5 mM EDTA + 0.05% (w/v) dehydrated skim milk + 1:2500 mouse anti-histag antibody
36. Incubate at RT with gentle shaking for 2hr

37. Recover primary antibody solution for future use
38. Wash membrane with PBS + 0.05% Tween20 1x 15 min., 2x 5 min.
39. Remove wash buffer and add secondary antibody
40. 40 mL PBS + 0.05% Tween20 + 0.05% (w/v) dehydrated skim milk + 1:25000 goat anti-mouse
+HRP antibody
41. Incubate at RT with gentle shaking for 2 hr
42. Recover secondary antibody solution for future use
43. Wash membrane with PBS + 0.05% Tween20 1x 15 min., 2x 5 min.
44. Mix 0.5 mL of each substrate solution (Thermo WestFemto super sensitive substrate)
45. Place membrane on plastic film and add substrate, fold film over so that it fully contacts
membrane and incubate at RT for 5 min.
46. Remove membrane from film and place in gel box
47. Set filter to clear, and lights to off
48. Capture 1 min., 3 min., and 6 min. exposures
49. Turn white light on and capture one image to facilitate alignment
50. Quantify band intensity using ImageJ

Histag purification of protein from *E. coli* using spin columns

Using BIO-RAD Profinity IMAC resin

Capacity: up to 15mg/mL resin

Compatibility: < 5mM DTT, <50mM TRIS, <5% Triton/Tween, <1% SDS, <8M urea

Cell growth stage

1. Grow overnight culture of *E. coli* expression strain (E.g., BL21 or Rosetta (DE3), lemo) in LB or 2xyt (EMBL says grow at 30°C)
2. Inoculate 1:100 into large culture with antibiotics
3. EMBL: "For good aeration, don't use more medium than **20%** of the total flask volume."
4. During mid-exponential phase (OD600 between 0.6 and 1.0), chill on ice, and move growth to cooler temperature (20°C used for bisabolene synthase) and add inducer (0.05 to 2 mM IPTG, 1mM is typical)
5. Harvest cells after 12 hours (centrifuge 10 min. at 6000g at 4°C)
6. Discard supernatant and flash freeze in liquid nitrogen or continue with lysis

Lysis stage

Lysis buffer: 50mM sodium phosphate pH 8.0, 0.3M NaCl, 5mM imidazole, 1x HALT protease inhibitor

1. Weigh pellet and resuspend in 1:10 ratio (w/v) lysis buffer + protease inhibitor
2. Vortex to resuspend
3. If not sonicating, DNase treatment may be necessary to reduce viscosity
4. Sonicate: 3s on/3s off for 2 min. 50% amplitude
5. Centrifuge at 15000g for 20 minutes at 4°C
6. For inclusion bodies: resuspend in 1:10 ratio (w/v) in PBS/8M urea, pH 7.5
7. Sonicate if needed to redissolve
8. Measure concentration (Bradford or BCA assay) and run gel to determine, roughly, the proportion of target protein present

Purification stage

This section closely adapted from BioRad Profinity IMAC Resins Instruction Manual.

Resin

Spin column

Binding Buffer: 50mM sodium phosphate, 0.3M NaCl, pH 8.0

Washing buffer: 50mM sodium phosphate, 0.3M NaCl, pH 8.0

Elution Buffer: 50mM sodium phosphate, 0.3M NaCl, 0.5M imidazole, pH 8.0

Charge Column

Using Biorad profinity resin with 15 mg histagged protein/ml resin capacity

1. Equilibrate the column with 5 column volumes of 50 mM sodium acetate, 0.3 M NaCl, pH 4.0.
After slurry packing is complete (see Sections 4 and 5), the column is ready for application of metal ions.
2. Make a 0.1–0.3 M solution of the metal ion of choice. For best results, the pH of the solution should be <7 (neutral to weakly acidic).
3. Apply 3–5 column volumes of the metal ion solution.
4. Wash with 5 column volumes of 50 mM sodium acetate, 0.3 M NaCl, pH 4.0. Remove excess ions by washing.
5. Wash with 10 column volumes of deionized water.
6. Equilibrate with at least 5 column volumes of starting buffer, for example, 50 mM sodium phosphate, 0.3 M NaCl, pH 7–8.

Part 1: Binding of Sample

1. Start with a prepacked spin column, charged with the metal ion of choice.
2. Place prepacked spin column in an appropriate spin collection tube.
3. Pre-equilibrate the spin column with 5 column volumes of binding buffer.
4. Add an appropriate amount of lysate (≤ 0.5 ml) to the micro spin column.

5. Mix by pipetting up and down 5 times.
6. Incubate for up to 5 min in micro spin column.
7. Centrifuge at 1,000 x g for 1 min.

Part 2: Washing the Resin

1. Insert micro spin into new collection vessel.
2. Wash the resin with at least 5 column volumes of binding buffer containing imidazole.
3. Pipet up and down at least 5 times.
4. Centrifuge at 1,000 x g for 1 min.
5. Remove remaining unbound proteins by centrifuging. The wash step can be repeated if necessary.

Part 3: Eluting the His-Tagged Protein

1. Insert micro spin column into new, clean collection vessel.
2. Elute bound proteins with 5 column volumes of elution buffer.
3. Pipet up and down at least 5 times and incubate for up to 5 min.
4. Analyze fractions from above steps by A280, SDS-PAGE, ELISA, etc.
5. Storage of protein at -20°C is okay in the short term. For longer term storage, buffer exchange may be necessary to remove imidazole and add glycerol, and storage at -80°C is preferred.

Example: Testing promoter strength in *Synechocystis* 6803

In this example, step-by-step instructions will be provided for creating a strain of *Synechocystis* 6803 which contains a promoter driving green fluorescent protein (GFP) within the *slr0168* neutral site *via* markerless homologous recombination. A flow-diagram of these methods is provided (*see* **Figure A1.1**).

Obtain parent plasmids

Two plasmid backbones are needed for this protocol: pMK0168 and pIGA4 (Addgene). pMK0168 contains a nickel-inducible *mazF* and kanamycin resistance marker targeted to the *slr0168* neutral site (see **Figure A1.1A**). pIGA4 contains the GFP marker and flanking *slr0168* homologous regions; the promoter of interest will be inserted into the MCS immediately upstream of GFP (see **Figure A1.1A**).

Amplify promoter of interest

Design primers to amplify the promoter of interest. The forward primer should be designed to contain a KpnI recognition site (GGTACC) and the reverse primer should be designed to contain a NdeI recognition site (CATATG). Perform PCR on the promoter of interest and purify the amplified promoter fragment. The resulting fragment will contain KpnI and NdeI restriction enzyme cut sites (see **Figure A1.1B**). For example, to clone the promoter from principal RNA Polymerase Sigma factor *PsigA*, use forward primer (5'-TATAAGGTACCCCGTCATCGATTC-3') and reverse primer (5'-TATAACCATATGGCCGTTTCCTCGTTAAC - 3') where the underlined bases denote restriction enzyme recognition sites.

Construct promoter:GFP plasmid

Digest both pIGA4 and the promoter fragment with KpnI and NdeI at 37°C for 20 minutes to 1 hour followed by heat inactivation at 80°C for 20 minutes. Dephosphorylate the sticky ends of pIGA4 using CIP; incubate at 37°C for 20 minutes to 1 hour. Purify all products. Ligate the digested promoter into pIGA4 using T4 DNA Ligase overnight at 16°C. Transform the ligation product into chemically competent DH5α *E. coli* via heat shock. Plate transformed *E. coli* onto LB agar plates supplemented with Ampicillin and incubate at 37°C for ~16 hours. Perform colony PCR on colonies and extract the plasmid from correct colonies. Confirm the plasmid is constructed correctly with sequencing (see **Figure A1.1C**).

Transform *Synechocystis* 6803 with counterselection plasmid pMK0168

Follow the steps described in *Methods 3.3* to insert *mazF/kanR* cassette into the *slr0168* neutral site (see **Figure A1.1D**). Perform colony PCR (cPCR) of *Synechocystis* colonies to confirm proper transformation using the primers listed in **Table A1.3**. The same protocols used to complete cPCR of *E. coli* may be used for *Synechocystis*. **Figure A1.1F** shows the binding sites of the primers used for cPCR. If *mazF/kanR* is properly inserted, cPCR with primers #3 and #4 will produce a 4.4 kb band. If proper insertion was not achieved, Primers #3 and #4 should produce a 0.6 kb band. Incomplete segregation would be indicated if both a 4.4 kb band and a 0.6 kb band are observed.

Synechocystis 6803 transformation with promoter:GFP plasmid

Follow the steps described in *Methods 3.4* to replace the *mazF/kanR* cassette with the gene of interest in the *slr0168* neutral site (see **Figure A1.1E**). Transformants should grow on BG-11+Ni agar plates but not on BG-11+Kan agar plates. Ensuring complete segregation of the transformed cyanobacteria by completing cPCR with Primers #3 and #4 listed in **Table A1.3**. If the GFP expression cassette is properly inserted, cPCR with primers #3 and #4 will produce a 2.6 kb band. If the *mazF/kanR* cassette remains in the *slr0168* site, primers #3 and #4 will produce a 4.4kb band and primers. (see **Note 15**). Sequence verify either the PCR fragment produced from primers #3 and #4, or the genomic DNA, using primer #3.

Test promoter by measuring fluorescence

Cultivate cells using the steps described in *Methods 3.1*. Inoculate wild-type and transformed cultures to $OD_{730} \sim 0.1$ from seed cultures at $OD_{730} \sim 0.6-1.0$ in triplicate. To measure fluorescence, transfer 300 μL of culture to a black 96-well flat-bottom plate. Using a plate reader with fluorescence capabilities, measure fluorescence with excitation wavelength 485 nm and emission wavelength of 525 nm. Measure optical density and fluorescence every 6-12 hours for as many days as desired. Relative fluorescence will provide a measure of promoter strength and activity over the test period.

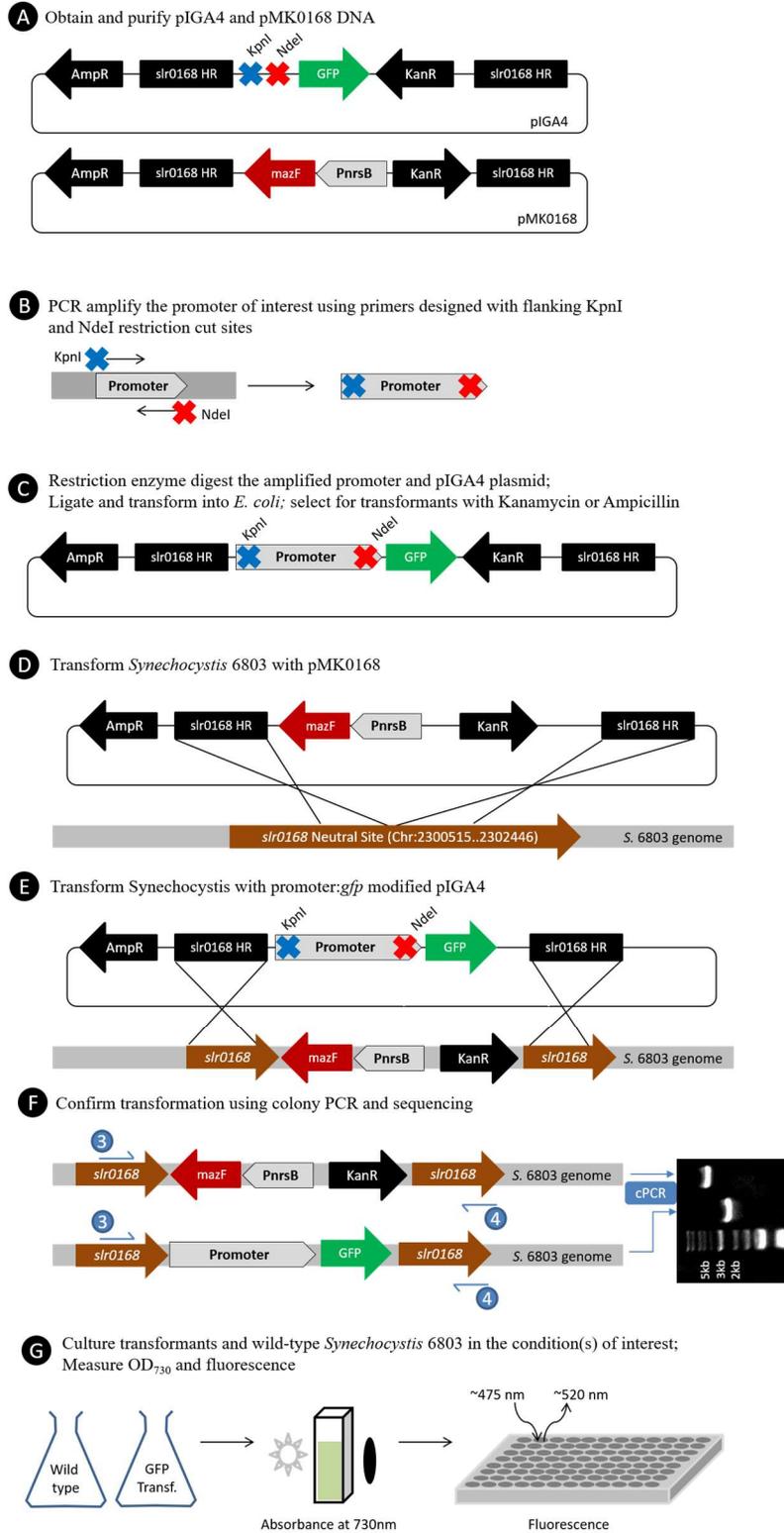


Figure A1.1. Design, implementation, and characterization of a *Synechocystis* 6803 transformants with a promoter of interest driving GFP production.

Table A1.1. BG-11 Media Component Recipes

Part Name	Chemical Formula and Name	Amount in 200 mL
100X Comp. A	NaNO ₃ , Sodium nitrate	30.00 g
	MgSO ₄ -7H ₂ O, Magnesium sulfate heptahydrate	1.498 g
	CaCl ₂ -2H ₂ O, Calcium chloride dehydrate	0.720 g
	C ₆ H ₈ O ₇ , Citric acid	0.120 g
	NH ₃ FeCitrate, Ammonium ferric citrate	0.120 g
	EDTA IDRANAL	0.020 g
600X Comp. B	K ₂ HPO ₄ -3H ₂ O, Potassium phosphate trihydrate	4.800 g
100X Comp. C	Na ₂ CO ₃ , Sodium carbonate	0.400 g
1000X Trace Metals	H ₃ BO ₃ , Boric acid	0.572 g
	MnCl ₂ -4H ₂ O, Manganese II chloride tetrahydrate	0.362 g
	ZnSO ₄ -7H ₂ O, Zinc sulfate heptahydrate	0.044 g
	Na ₂ MoO ₄ -2H ₂ O, Sodium molybdate dehydrate	0.078 g
	CuSO ₄ -5H ₂ O, Copper II sulfate pentahydrate	0.016 g
	Co(NO ₃) ₂ -6H ₂ O, Cobalt II nitrate hexahydrate	0.010 g

--	--	--	--

Table A1.2: BG-11 Media Additions

Chemical Name	Final Concentration	Stock Concentration	Notes
Kanamycin	50 µg/mL	50 mg/mL	Prepare 1000x stocks by dissolving 500 mg kanamycin salt in 10 mL deionized water. Filter sterilize using a syringe and 0.2µm filter into autoclaved 1.5 mL microcentrifuge tubes. Store in 1 mL aliquots at -20°C.
Spectinomycin	50 µg/mL	50 mg/mL	Prepare 1000x stocks by dissolving 500 mg spectinomycin salt in 10 mL deionized water. Filter sterilize using a syringe and 0.2 µm filter into autoclaved 1.5 mL microcentrifuge tubes. Store in 1 mL aliquots at -20°C.
Sodium thiosulfate	3.3 g/L	1 M	Prepare 300 mL 1 M stock by dissolving 47.43 g sodium thiosulfate in 200mL using a magnetic stir bar on a stir plate. Once dissolved, adjust volume to 300 mL and filter sterilize into a sterile bottle.
TES ^a	10 mM	1 M	Dissolve 68.78 g TES in 200 mL deionized water. Adjust the pH to 8.0 using NaOH. Adjust volume to 300 mL. Filter sterilize using a syringe and 0.2 µm filter and store at 4°C protected from light.
DMSO ^b	5%	50%	Filter sterilize a 50% DMSO in water solution and store at room temperature.
Nickel sulfate	20 µM	20 mM	Prepare 1000x stock of 20 mM nickel sulfate by dissolving 0.526 g of nickel sulfate hexahydrate (MW=262.848) in 100 mL deionized water. Filter sterilize into an autoclaved glass bottle. Store at room temperature.

^a N-Tris(hydroxymethyl)methyl-2-aminoethanesulfonic acid

^b Dimethyl sulfoxide

Table A1.3: Primers

Primer #	Primer name	5'→3' sequence
1	PsigA forward	5'-TATAAGGTACCCCGTCATCGATTC-3'
2	PsigA reverse	5'-TATAACATATGGCCGTTTTCTCGTTAAC -3'
3	Slr0168 up forward	5'- TCCAGGCCACATTGTTG -3'
4	Slr0168 dn reverse	5'- AGTGACCTATTCAATCAGGAAGG -3'
5	MK reverse	5'- CAGAAAGACTTAAAATTATTGCCG -3'
6	GFP reverse	5'- TTATTATTTGTATAGTTCATCCATGCC -3'

Notes:

1. We recommend an acid wash of flasks after culturing *Synechocystis* 6803.
2. APlasmidEditor (Ape, <http://biologylabs.utah.edu/jorgensen/wayned/ape/>) is a free tool for visualizing DNA sequences. SnapGene (<http://www.snapgene.com/>) can produce better plasmid images and may be more intuitive to use, but requires a license to use. IDT Oligo Analyzer (<https://www.idtdna.com/calc/analyzer>) is a tool and service for analyzing and ordering primers.
3. Several strains of *Synechocystis* 6803 are available for purchase and have different characteristics, including the ability to grow with light-activated heterotrophic growth. The strain of cyanobacteria referred to here (ATCC® 27184™) can be purchased here: <https://www.atcc.org/Products/All/27184.aspx>. *Synechocystis* 6803 is also available through the Pasteur Culture Collection at https://brclims.pasteur.fr/crbip_catalogue/faces/recherche_catalogue.xhtml.
4. Since BG-11 was previously found to be limiting in phosphate [17], Component B can be prepared as a 600x solution, but still added to the BG-11 media preparation at a 100x dilution.

5. Although chloramphenicol is reported as an appropriate antibiotic for selection in *Synechocystis* 6803 and other cyanobacteria, in our hands, selection is poor. We recommend Kanamycin or Spectinomycin if given a choice of antibiotic resistance.
6. Commercially available fluorescent lights are commonly used. LED lights may also be used. “Grow lights” targeting specific wavelengths should not be used as these trigger the reorganization of pigment-protein complexes in *Synechocystis* 6803 which may alter growth.
7. If the culture is sufficiently dense, adding small volumes (e.g. 10 μ L) will be difficult to spread. Add liquid BG-11 up to 100 μ L (e.g. 90 μ L) on plate before spreading.
8. Using surgical tape, as opposed to Parafilm[®], is advantageous for incubation of BG-11 plates because it allows for more efficient gas transfer into the plate and thus faster cell growth.
9. The reader is referred to the Eaton-Rye methods paper for details on how to locate and download sequences from Cyanobase [18]. *Synechocystis* 6803 sequences can be accessed at Cyanobase: <http://genome.microbedb.jp/cyanobase/Synechocystis>. Gene sequences from many organisms can be obtained at NCBI GenBank: <https://www.ncbi.nlm.nih.gov/genbank/>.
10. Several neutral sites have been identified in *Synechocystis* 6803 [19].
11. If expected PCR product is similar to the negative control, then a second primer specific to the DNA insert can be used.
12. *Synechocystis* 6803 often maintains multiple copies of its chromosome. Full segregation means that the genetic modification of interest was achieved on each copy of the chromosome. Interestingly, the first round of transformation rarely results in full segregation, but it can be achieved by repeatedly re-streaking the same colony onto selective media.
13. To perform markerless recombination as these protocols are written, the investigator should apply the protocols in *Section 3.3* using nickel-inducible *mazF*/*KanR* cassette as the gene interest. To select for the successful transformants, BG-11 agar plates supplemented with only

kanamycin should be used. Then, this base strain should be applied for further modification using the protocols in *Section 3.4* using the gene of interest.

14. In our laboratory, we have observed the formation of both small and large colonies. To increase the probability of observing segregated transformants, screen a mixture of both large and small colonies.
15. Colonies may be screened for the desired mutation as soon as they appear. The presence of multiple copies of the chromosome in *Synechocystis* 6803 complicates screening for transformants as some copies may carry the mutation while others retain the wild-type sequence. Fully segregated transformants are those that carry the mutation and have completely lost the wild-type sequence (or the *mazF/kanR* cassette in the case of a markerless selection transformation) between the homologous regions used. Screening colonies using colony PCR should be conducted with primers that bind within the homologous regions in order to capture both the mutated and wild-type sequences (if both are present). Alternatively, two PCR reactions may be completed for each colony, one with primers that amplify only the mutated sequence, and a second with primers that amplify only the wild-type sequence.

References

- Abe, Koichi, Kotone Miyake, Mayumi Nakamura, Katsuhiko Kojima, Stefano Ferri, Kazunori Ikebukuro, and Koji Sode. 2014. "Engineering of a Green-Light Inducible Gene Expression System in *Synechocystis* Sp. PCC6803: Engineered Green Light Gene Expression System." *Microbial Biotechnology* 7 (2): 177–83. <https://doi.org/10.1111/1751-7915.12098>.
- Albers, Stevan C, Victor A Gallegos, and Christie A M Peebles. 2015a. "Engineering of Genetic Control Tools in *Synechocystis* Sp. PCC 6803 Using Rational Design Techniques." *Journal of Biotechnology* 216 (December): 36–46. <http://dx.doi.org/10.1016/j.jbiotec.2015.09.042>.
- Albers, Stevan C., Victor A. Gallegos, and Christie A.M. Peebles. 2015b. "Engineering of Genetic Control Tools in *Synechocystis* Sp. PCC 6803 Using Rational Design Techniques." *Journal of Biotechnology* 216 (December): 36–46. <https://doi.org/10.1016/j.jbiotec.2015.09.042>.
- Albers, Stevan C, and Christie A M Peebles. 2016. "Evaluating Light-Induced Promoters for the Control of Heterologous Gene Expression in *Synechocystis* Sp. PCC 6803." *Biotechnology Progress*, November, n/a-n/a. <https://doi.org/10.1002/btpr.2396>.
- Angermayr, S Andreas, Alex Gorchs Rovira, and Klaas J Hellingwerf. 2015. "Metabolic Engineering of Cyanobacteria for the Synthesis of Commodity Products." *Trends in Biotechnology* 33 (6): 352–61. <http://dx.doi.org/10.1016/j.tibtech.2015.03.009>.
- Balibar, Carl J., Xiaoyu Shen, and Jianshi Tao. 2009. "The Mevalonate Pathway of *Staphylococcus Aureus*." *Journal of Bacteriology* 191 (3): 851–61. <https://doi.org/10.1128/JB.01357-08>.
- Beck, Christian, Stefanie Hertel, Anne Rediger, Robert Lehmann, Anika Wiegard, Adrian Kölsch, Beate Heilmann, Jens Georg, Wolfgang R. Hess, and Ilka M. Axmann. 2014. "Daily Expression Pattern of Protein-Encoding Genes and Small Noncoding RNAs in *Synechocystis* Sp. Strain PCC 6803." Edited by H. L. Drake. *Applied and Environmental Microbiology* 80 (17): 5195–5206. <https://doi.org/10.1128/AEM.01086-14>.
- Benedetti, Manuel, Valeria Vecchi, Simone Barera, and Luca Dall'Osto. 2018. "Biomass from Microalgae: The Potential of Domestication towards Sustainable Biofactories." *Microbial Cell Factories* 17 (1). <https://doi.org/10.1186/s12934-018-1019-3>.
- Bentley, Fiona K., Andreas Zurbruggen, and Anastasios Melis. 2014. "Heterologous Expression of the Mevalonic Acid Pathway in Cyanobacteria Enhances Endogenous Carbon Partitioning to Isoprene." *Molecular Plant* 7 (1): 71–86. <https://doi.org/10.1093/mp/sst134>.
- Briggs, Linda M., Vincent L. Pecoraro, and Lee McIntosh. 1990. "Copper-Induced Expression, Cloning, and Regulatory Studies of the Plastocyanin Gene from the Cyanobacterium *Synechocystis* Sp. PCC 6803." *Plant Molecular Biology* 15 (4): 633–42. <https://doi.org/10.1007/BF00017837>.
- Burgard, Anthony P., Priti Pharkya, and Costas D. Maranas. 2003. "Optknock: A Bilevel Programming Framework for Identifying Gene Knockout Strategies for Microbial Strain Optimization." *Biotechnology and Bioengineering* 84 (6): 647–57. <https://doi.org/10.1002/bit.10803>.
- Callahan, S. M., and W. J. Buikema. 2001. "The Role of HetN in Maintenance of the Heterocyst Pattern in *Anabaena* Sp. PCC 7120." *Molecular Microbiology* 40 (4): 941–50. <https://doi.org/10.1046/j.1365-2958.2001.02437.x>.
- Camsund, Daniel, Thorsten Heidorn, and Peter Lindblad. 2014. "Design and Analysis of LacI-Repressed Promoters and DNA-Looping in a Cyanobacterium." *Journal of Biological Engineering* 8 (1): 4. <https://doi.org/10.1186/1754-1611-8-4>.
- Camsund, Daniel, and Peter Lindblad. 2014. "Engineered Transcriptional Systems for Cyanobacterial Biotechnology." *Frontiers in Bioengineering and Biotechnology* .

- Chae, Tong Un, So Young Choi, Je Woong Kim, Yoo-Sung Ko, and Sang Yup Lee. 2017. "Recent Advances in Systems Metabolic Engineering Tools and Strategies." *Current Opinion in Biotechnology* 47 (October): 67–82. <https://doi.org/10.1016/j.copbio.2017.06.007>.
- Chaves, Julie E., Paloma Rueda Romero, Henning Kirst, and Anastasios Melis. 2016. "Role of Isopentenyl-Diphosphate Isomerase in Heterologous Cyanobacterial (*Synechocystis*) Isoprene Production." *Photosynthesis Research*, July. <https://doi.org/10.1007/s11120-016-0293-3>.
- Cheah, Yi Ern, Stevan C Albers, and Christie A M Peebles. 2013. "A Novel Counter-Selection Method for Markerless Genetic Modification in *Synechocystis* Sp. PCC 6803." *Biotechnology Progress* 29 (1): 23–30. <https://doi.org/10.1002/btpr.1661>.
- Cui, Jinyu, Tao Sun, Lei Chen, and Weiwen Zhang. 2020. "Engineering Salt Tolerance of Photosynthetic Cyanobacteria for Seawater Utilization." *Biotechnology Advances* 43 (November): 107578. <https://doi.org/10.1016/j.biotechadv.2020.107578>.
- Davies, Fiona K., Victoria H. Work, Alexander S. Beliaev, and Matthew C. Posewitz. 2014. "Engineering Limonene and Bisabolene Production in Wild Type and a Glycogen-Deficient Mutant of *Synechococcus* Sp. PCC 7002." *Frontiers in Bioengineering and Biotechnology* 2 (June). <https://doi.org/10.3389/fbioe.2014.00021>.
- Dueber, John E, Gabriel C Wu, G Reza Malmirchegini, Tae Seok Moon, Christopher J Petzold, Adeeti V Ullal, Kristala L J Prather, and Jay D Keasling. 2009. "Synthetic Protein Scaffolds Provide Modular Control over Metabolic Flux." *Nature Biotechnology* 27 (8): 753–59. <https://doi.org/10.1038/nbt.1557>.
- Elhai, Jeff, and C.Peter Wolk. 1988. "[83] Conjugal Transfer of DNA to Cyanobacteria." In *Methods in Enzymology*, 167:747–54. Elsevier. [https://doi.org/10.1016/0076-6879\(88\)67086-8](https://doi.org/10.1016/0076-6879(88)67086-8).
- Engler, Carola, and Sylvestre Marillonnet. 2014. "Golden Gate Cloning." In *DNA Cloning and Assembly Methods*, edited by Svein Valla and Rahmi Lale, 119–31. Totowa, NJ: Humana Press. https://doi.org/10.1007/978-1-62703-764-8_9.
- Englund, Elias, Johan Andersen-Ranberg, Rui Miao, Björn Hamberger, and Pia Lindberg. 2015. "Metabolic Engineering of *Synechocystis* Sp. PCC 6803 for Production of the Plant Diterpenoid Manoyl Oxide." *ACS Synthetic Biology* 4 (12): 1270–78. <https://doi.org/10.1021/acssynbio.5b00070>.
- Englund, Elias, Feiyan Liang, and Pia Lindberg. 2016. "Evaluation of Promoters and Ribosome Binding Sites for Biotechnological Applications in the Unicellular Cyanobacterium *Synechocystis* Sp. PCC 6803." *Scientific Reports* 6 (November): 36640. <https://doi.org/10.1038/srep36640>.
- Eriksson, Jan, Gaza F. Salih, Haile Ghebramedhin, and Christer Jansson. 2000. "Deletion Mutagenesis of the 5' *PsbA2* Region in *Synechocystis* 6803: Identification of a Putative Cis Element Involved in Photoregulation." *Molecular Cell Biology Research Communications* 3 (5): 292–98. <https://doi.org/10.1006/mcbr.2000.0227>.
- Ferreira, Eunice A, Catarina C Pacheco, Filipe Pinto, José Pereira, Pedro Lamosa, Paulo Oliveira, Boris Kirov, Alfonso Jaramillo, and Paula Tamagnini. 2018. "Expanding the Toolbox for *Synechocystis* Sp. PCC 6803: Validation of Replicative Vectors and Characterization of a Novel Set of Promoters." *Synthetic Biology* 3 (1). <https://doi.org/10.1093/synbio/ysy014>.
- Gale, Grant A. R., Alejandra A. Schiavon Osorio, Lauren A. Mills, Baojun Wang, David J. Lea-Smith, and Alistair J. McCormick. 2019. "Emerging Species and Genome Editing Tools: Future Prospects in Cyanobacterial Synthetic Biology." *Microorganisms* 7 (10): 409. <https://doi.org/10.3390/microorganisms7100409>.
- Gao, Xiang, Fang Gao, Deng Liu, Hao Zhang, Xiaoqun Nie, and Chen Yang. 2016. "Engineering the Methylerythritol Phosphate Pathway in Cyanobacteria for Photosynthetic Isoprene Production from CO₂." *Energy & Environmental Science* 9 (4): 1400–1411. <https://doi.org/10.1039/C5EE03102H>.

- Gibson, Daniel G, Lei Young, Ray-Yuan Chuang, J Craig Venter, Clyde A Hutchison, and Hamilton O Smith. 2009. "Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases." *Nat Meth* 6 (5): 343–45.
- Gordon, Gina C., Travis C. Korosh, Jeffrey C. Cameron, Andrew L. Markley, Matthew B. Begemann, and Brian F. Pfeleger. 2016. "CRISPR Interference as a Titratable, Trans -Acting Regulatory Tool for Metabolic Engineering in the Cyanobacterium *Synechococcus* Sp . Strain PCC 7002." *Metabolic Engineering* 38 (November): 170–79. <https://doi.org/10.1016/j.ymben.2016.07.007>.
- Griese, Marco, Christian Lange, and Jörg Soppa. 2011. "Ploidy in Cyanobacteria." *FEMS Microbiology Letters* 323 (2): 124–31. <https://doi.org/10.1111/j.1574-6968.2011.02368.x>.
- Guerrero, Fernando, Verónica Carbonell, Matteo Cossu, Danilo Correddu, and Patrik R. Jones. 2012. "Ethylene Synthesis and Regulated Expression of Recombinant Protein in *Synechocystis* Sp. PCC 6803." Edited by Brett Neilan. *PLoS ONE* 7 (11): e50470. <https://doi.org/10.1371/journal.pone.0050470>.
- Gupta, Apoorv, Irene M Brockman Reizman, Christopher R Reisch, and Kristala L J Prather. 2017. "Dynamic Regulation of Metabolic Flux in Engineered Bacteria Using a Pathway-Independent Quorum-Sensing Circuit." *Nature Biotechnology* 35 (3): 273–79. <https://doi.org/10.1038/nbt.3796>.
- Heidorn, Thorsten, Daniel Camsund, Hsin-Ho Huang, Pia Lindberg, Paulo Oliveira, Karin Stensjö, and Peter Lindblad. 2011. "Synthetic Biology in Cyanobacteria." In *Methods in Enzymology*, 497:539–79. Elsevier. <http://linkinghub.elsevier.com/retrieve/pii/B978012385075100024X>.
- Huang, Chun-Hung, Claire R. Shen, Hung Li, Li-Yu Sung, Meng-Ying Wu, and Yu-Chen Hu. 2016. "CRISPR Interference (CRISPRi) for Gene Regulation and Succinate Production in *Cyanobacterium* S. *Elongatus* PCC 7942." *Microbial Cell Factories* 15 (1). <https://doi.org/10.1186/s12934-016-0595-3>.
- Huang, H H, D Camsund, P Lindblad, and T Heidorn. 2010a. "Design and Characterization of Molecular Tools for a Synthetic Biology Approach towards Developing Cyanobacterial Biotechnology." *Nucleic Acids Res* 38. <https://doi.org/10.1093/nar/gkq164>.
- Huang, H. H., D. Camsund, P. Lindblad, and T. Heidorn. 2010b. "Design and Characterization of Molecular Tools for a Synthetic Biology Approach towards Developing Cyanobacterial Biotechnology." *Nucleic Acids Research* 38 (8): 2577–93. <https://doi.org/10.1093/nar/gkq164>.
- Huang, Hsin-Ho, and Peter Lindblad. 2013. "Wide-Dynamic-Range Promoters Engineered for *Cyanobacteria*." *Journal of Biological Engineering* 7 (1): 10. <https://doi.org/10.1186/1754-1611-7-10>.
- Imamura, Sousuke, Munehiko Asayama, and Makoto Shirai. 2004. "In Vitro Transcription Analysis by Reconstituted Cyanobacterial RNA Polymerase: Roles of Group 1 and 2 Sigma Factors and a Core Subunit, RpoC2: Roles of Cyanobacterial Sigma Factors & RpoC2." *Genes to Cells* 9 (12): 1175–87. <https://doi.org/10.1111/j.1365-2443.2004.00808.x>.
- Ivanikova, Natalia V., R. Michael L. McKay, and George S. Bullerjahn. 2005. "Construction and Characterization of a Cyanobacterial Bioreporter Capable of Assessing Nitrate Assimilatory Capacity in Freshwaters: Cyanobacterial Nitrate Bioreporter." *Limnology and Oceanography: Methods* 3 (2): 86–93. <https://doi.org/10.4319/lom.2005.3.86>.
- Jin, Haojie, Peter Lindblad, and Devaki Bhaya. 2019. "Building an Inducible T7 RNA Polymerase/T7 Promoter Circuit in *Synechocystis* Sp. PCC6803." *ACS Synthetic Biology* 8 (4): 655–60. <https://doi.org/10.1021/acssynbio.8b00515>.
- Kaczmarzyk, Danuta, Ivana Cengic, Lun Yao, and Elton P. Hudson. 2018. "Diversion of the Long-Chain Acyl-ACP Pool in *Synechocystis* to Fatty Alcohols through CRISPRi Repression of the Essential Phosphate Acyltransferase PlsX." *Metabolic Engineering* 45 (January): 59–66. <https://doi.org/10.1016/j.ymben.2017.11.014>.

- Kaneko, Takakazu, Shusei Sato, Hirokazu Kotani, Ayako Tanaka, Erika Asamizu, Yasukazu Nakamura, Nobuyuki Miyajima, et al. 1996. "Sequence Analysis of the Genome of the Unicellular Cyanobacterium *Synechocystis* Sp. Strain PCC6803. II. Sequence Determination of the Entire Genome and Assignment of Potential Protein-Coding Regions." *DNA Research* 3 (3): 109–36. <https://doi.org/10.1093/dnares/3.3.109>.
- Kim, Wook Jin, Sun-Mi Lee, Youngsoon Um, Sang Jun Sim, and Han Min Woo. 2017. "Development of SyneBrick Vectors As a Synthetic Biology Platform for Gene Expression in *Synechococcus Elongatus* PCC 7942." *Frontiers in Plant Science* 8: 293. <https://doi.org/10.3389/fpls.2017.00293>.
- Kleinegris, Dorinde M. M., Marjon A. van Es, Marcel Janssen, Willem A. Brandenburg, and René H. Wijffels. 2011. "Phase Toxicity of Dodecane on the Microalga *Dunaliella Salina*." *Journal of Applied Phycology* 23 (6): 949–58. <https://doi.org/10.1007/s10811-010-9615-6>.
- Kleinegris, Dorinde M.M., Marcel Janssen, Willem A. Brandenburg, and René H. Wijffels. 2011. "Two-Phase Systems: Potential for in Situ Extraction of Microalgal Products." *Biotechnology Advances* 29 (5): 502–7. <https://doi.org/10.1016/j.biotechadv.2011.05.018>.
- Kok, Stefan de, Leslie H Stanton, Todd Slaby, Maxime Durot, Victor F Holmes, Kedar G Patel, Darren Platt, et al. 2014. "Rapid and Reliable DNA Assembly via Ligase Cycling Reaction." *ACS Synthetic Biology* 3 (2): 97–106. <https://doi.org/10.1021/sb4001992>.
- Liu, Deng, Virginia M. Johnson, and Himadri B. Pakrasi. 2020. "A Reversibly Induced CRISPRi System Targeting Photosystem II in the Cyanobacterium *Synechocystis* Sp. PCC 6803." *ACS Synthetic Biology* 9 (6): 1441–49. <https://doi.org/10.1021/acssynbio.0c00106>.
- Liu, Deng, and Himadri B. Pakrasi. 2018. "Exploring Native Genetic Elements as Plug-in Tools for Synthetic Biology in the Cyanobacterium *Synechocystis* Sp. PCC 6803." *Microbial Cell Factories* 17 (1). <https://doi.org/10.1186/s12934-018-0897-8>.
- Markley, Andrew L., Matthew B. Begemann, Ryan E. Clarke, Gina C. Gordon, and Brian F. Pfleger. 2015. "Synthetic Biology Toolbox for Controlling Gene Expression in the Cyanobacterium *Synechococcus* Sp. Strain PCC 7002." *ACS Synthetic Biology* 4 (5): 595–603. <https://doi.org/10.1021/sb500260k>.
- Martin, Vincent J J, Douglas J Pitera, Sydnor T Withers, Jack D Newman, and Jay D Keasling. 2003. "Engineering a Mevalonate Pathway in *Escherichia Coli* for Production of Terpenoids." *Nature Biotechnology* 21 (7): 796–802. <https://doi.org/10.1038/nbt833>.
- Máté, Z., L. Sass, M. Szekeres, I. Vass, and F. Nagy. 1998. "UV-B-Induced Differential Transcription of *PsbA* Genes Encoding the D1 Protein of Photosystem II in the Cyanobacterium *Synechocystis* 6803." *The Journal of Biological Chemistry* 273 (28): 17439–44. <https://doi.org/10.1074/jbc.273.28.17439>.
- Mazouni, K., S. Bulteau, C. Cassier-Chauvat, and F. Chauvat. 1998. "Promoter Element Spacing Controls Basal Expression and Light Inducibility of the Cyanobacterial *SecA* Gene." *Molecular Microbiology* 30 (5): 1113–22. <https://doi.org/10.1046/j.1365-2958.1998.01145.x>.
- Muramatsu, M., and Y. Hihara. 2007. "Coordinated High-Light Response of Genes Encoding Subunits of Photosystem I Is Achieved by AT-Rich Upstream Sequences in the Cyanobacterium *Synechocystis* Sp. Strain PCC 6803." *Journal of Bacteriology* 189 (7): 2750–58. <https://doi.org/10.1128/JB.01903-06>.
- Nakao, Mitsuteru, Shinobu Okamoto, Mitsuyo Kohara, Tsunakazu Fujishiro, Takatomo Fujisawa, Shusei Sato, Satoshi Tabata, Takakazu Kaneko, and Yasukazu Nakamura. 2010. "CyanoBase: The Cyanobacteria Genome Database Update 2010." *Nucleic Acids Research* 38 (Database issue): D379–81. <https://doi.org/10.1093/nar/gkp915>.
- Ng, Andrew H., Bertram M. Berla, and Himadri B. Pakrasi. 2015. "Fine-Tuning of Photoautotrophic Protein Production by Combining Promoters and Neutral Sites in the Cyanobacterium

- Synechocystis Sp. Strain PCC 6803." Edited by K. E. Wommack. *Applied and Environmental Microbiology* 81 (19): 6857–63. <https://doi.org/10.1128/AEM.01349-15>.
- Nielsen, Jens, and Jay D. Keasling. 2016. "Engineering Cellular Metabolism." *Cell* 164 (6): 1185–97. <https://doi.org/10.1016/j.cell.2016.02.004>.
- Pade, Nadin, Sabrina Erdmann, Heike Enke, Frederik Dethloff, Ulf Dühning, Jens Georg, Juliane Wambutt, et al. 2016. "Insights into Isoprene Production Using the Cyanobacterium *Synechocystis* Sp. PCC 6803." *Biotechnology for Biofuels* 9 (1). <https://doi.org/10.1186/s13068-016-0503-4>.
- Peca, Loredana, Péter B. Kás, Zoltan Máté, Andrea Farsang, and Imre Vass. 2008. "Construction of Bioluminescent Cyanobacterial Reporter Strains for Detection of Nickel, Cobalt and Zinc." *FEMS Microbiology Letters* 289 (2): 258–64. <https://doi.org/10.1111/j.1574-6968.2008.01393.x>.
- Peralta-Yahya, Pamela P., Mario Ouellet, Rossana Chan, Aindrila Mukhopadhyay, Jay D. Keasling, and Taek Soon Lee. 2011. "Identification and Microbial Production of a Terpene-Based Advanced Biofuel." *Nature Communications* 2 (September): 483. <https://doi.org/10.1038/ncomms1494>.
- Pitera, Douglas J., Chris J. Paddon, Jack D. Newman, and Jay D. Keasling. 2007. "Balancing a Heterologous Mevalonate Pathway for Improved Isoprenoid Production in *Escherichia Coli*." *Metabolic Engineering* 9 (2): 193–207. <https://doi.org/10.1016/j.ymben.2006.11.002>.
- Plegaria, Jefferson S, and Cheryl A Kerfeld. 2018. "Engineering Nanoreactors Using Bacterial Microcompartment Architectures." *Current Opinion in Biotechnology* 51 (June): 1–7. <https://doi.org/10.1016/j.copbio.2017.09.005>.
- Quax, Tessa E.F., Nico J. Claassens, Dieter Söll, and John van der Oost. 2015. "Codon Bias as a Means to Fine-Tune Gene Expression." *Molecular Cell* 59 (2): 149–61. <https://doi.org/10.1016/j.molcel.2015.05.035>.
- Reed, R. H., and W. D. P. Stewart. 1985. "Osmotic Adjustment and Organic Solute Accumulation in Unicellular Cyanobacteria from Freshwater and Marine Habitats." *Marine Biology* 88 (1): 1–9. <https://doi.org/10.1007/BF00393037>.
- Salem, K., and L. G. van Waasbergen. 2004. "Light Control of HliA Transcription and Transcript Stability in the Cyanobacterium *Synechococcus Elongatus* Strain PCC 7942." *Journal of Bacteriology* 186 (6): 1729–36. <https://doi.org/10.1128/JB.186.6.1729-1736.2004>.
- Salis, Howard M. 2011. "Chapter Two - The Ribosome Binding Site Calculator." In *Synthetic Biology, Part B Computer Aided Design and DNA Assembly*, edited by Christopher Voigt B T - Methods in Enzymology, Volume 498:19–42. Academic Press. <http://dx.doi.org/10.1016/B978-0-12-385120-8.00002-4>.
- Salis, Howard M, Ethan A Mirsky, and Christopher A Voigt. 2009. "Automated Design of Synthetic Ribosome Binding Sites to Control Protein Expression." *Nature Biotechnology* 27 (10): 946–50. <https://doi.org/10.1038/nbt.1568>.
- Satoh, Soichirou, Masahiko Ikeuchi, Mamoru Mimuro, and Ayumi Tanaka. 2001. "Chlorophyll b Expressed in Cyanobacteria Functions as a Light-Harvesting Antenna in Photosystem I through Flexibility of the Proteins." *Journal of Biological Chemistry* 276 (6): 4293–97. <https://doi.org/10.1074/jbc.M008238200>.
- Sengupta, Annesha, Avinash Vellore Sunder, Sujata V. Sohoni, and Pramod P. Wangikar. 2019. "Fine-Tuning Native Promoters of *Synechococcus Elongatus* PCC 7942 To Develop a Synthetic Toolbox for Heterologous Protein Expression." *ACS Synthetic Biology* 8 (5): 1219–23. <https://doi.org/10.1021/acssynbio.9b00066>.
- Sergeyenko, T. V., and D. A. Los. 2000. "Identification of Secreted Proteins of the Cyanobacterium *Synechocystis* Sp. Strain PCC 6803." *FEMS Microbiology Letters* 193 (2): 213–16. <https://doi.org/10.1111/j.1574-6968.2000.tb09426.x>.

- Stal, Lucas J. 2015. "Nitrogen Fixation in Cyanobacteria." In *ELS*, edited by John Wiley & Sons Ltd, 1–9. Chichester, UK: John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470015902.a0021159.pub2>.
- Thiel, Kati, Edita Mulaku, Hariharan Dandapani, Csaba Nagy, Eva-Mari Aro, and Pauli Kallio. 2018. "Translation Efficiency of Heterologous Proteins Is Significantly Affected by the Genetic Context of RBS Sequences in Engineered Cyanobacterium *Synechocystis* Sp. PCC 6803." *Microbial Cell Factories* 17 (1). <https://doi.org/10.1186/s12934-018-0882-2>.
- Thykaer, Jette, and Jens Nielsen. 2003. "Metabolic Engineering of Beta-Lactam Production." *Metabolic Engineering* 5 (1): 56–69. [https://doi.org/10.1016/s1096-7176\(03\)00003-x](https://doi.org/10.1016/s1096-7176(03)00003-x).
- Ungerer, Justin, and Himadri B. Pakrasi. 2016. "Cpf1 Is A Versatile Tool for CRISPR Genome Editing Across Diverse Species of Cyanobacteria." *Scientific Reports* 6 (1). <https://doi.org/10.1038/srep39681>.
- Vasudevan, Ravendran, Grant A. R. Gale, Alejandra A. Schiavon, Anton Puzorjov, John Malin, Michael D. Gillespie, Konstantinos Vavitsas, et al. 2019. "CyanoGate: A Modular Cloning Suite for Engineering Cyanobacteria Based on the Plant MoClo Syntax." *Plant Physiology* 180 (1): 39–55. <https://doi.org/10.1104/pp.18.01401>.
- Vermaas, Josh V., Gayle J. Bentley, Gregg T. Beckham, and Michael F. Crowley. 2018. "Membrane Permeability of Terpenoids Explored with Molecular Simulation." *The Journal of Physical Chemistry B* 122 (45): 10349–61. <https://doi.org/10.1021/acs.jpcc.8b08688>.
- Wang, Bo, Carrie Eckert, Pin-Ching Maness, and Jianping Yu. 2018. "A Genetic Toolbox for Modulating the Expression of Heterologous Genes in the Cyanobacterium *Synechocystis* Sp. PCC 6803." *ACS Synthetic Biology* 7 (1): 276–86. <https://doi.org/10.1021/acssynbio.7b00297>.
- Wang, Bo, Jianping Yu, Weiwen Zhang, and Deirdre R. Meldrum. 2015a. "Premethylation of Foreign DNA Improves Integrative Transformation Efficiency in *Synechocystis* Sp. Strain PCC 6803." *Applied and Environmental Microbiology*, October, AEM.02575-15. <https://doi.org/10.1128/AEM.02575-15>.
- . 2015b. "Premethylation of Foreign DNA Improves Integrative Transformation Efficiency in *Synechocystis* Sp. Strain PCC 6803." Edited by M. J. Pettinari. *Applied and Environmental Microbiology* 81 (24): 8500–8506. <https://doi.org/10.1128/AEM.02575-15>.
- Wendt, Kristen E., Justin Ungerer, Ryan E. Cobb, Huimin Zhao, and Himadri B. Pakrasi. 2016. "CRISPR/Cas9 Mediated Targeted Mutagenesis of the Fast Growing Cyanobacterium *Synechococcus* *Elongatus* UTEX 2973." *Microbial Cell Factories* 15 (1). <https://doi.org/10.1186/s12934-016-0514-7>.
- Werner, Allison, Katelyn Oliver, Alexander Dylan Miller, Jacob Sebesta, and Christie A.M. Peebles. 2018. "Discovery and Characterization of *Synechocystis* Sp. PCC 6803 Light-Entrained Promoters in Diurnal Light:Dark Cycles." *Algal Research* 30 (March): 121–27. <https://doi.org/10.1016/j.algal.2017.12.012>.
- Williams, John G.K. 1988. "[85] Construction of Specific Mutations in Photosystem II Photosynthetic Reaction Center by Genetic Engineering Methods in *Synechocystis* 6803." In *Methods in Enzymology*, 167:766–78. Elsevier. <http://linkinghub.elsevier.com/retrieve/pii/0076687988670881>.
- Yao, Lun, Ivana Cengic, Josefine Anfelt, and Elton P. Hudson. 2016. "Multiple Gene Repression in Cyanobacteria Using CRISPRi." *ACS Synthetic Biology* 5 (3): 207–12. <https://doi.org/10.1021/acssynbio.5b00264>.
- Yoshihara, Shizue, XiaoXing Geng, Shinobu Okamoto, Kei Yura, Takashi Murata, Mitiko Go, Masayuki Ohmori, and Masahiko Ikeuchi. 2001. "Mutational Analysis of Genes Involved in Pilus Structure, Motility and Transformation Competency in the Unicellular Motile Cyanobacterium *Synechocystis* Sp. PCC6803." *Plant and Cell Physiology* 42 (1): 63–73. <https://doi.org/10.1093/pcp/pce007>.

- Zang, Xiaonan, Bin Liu, Shunmei Liu, K. K. I. U. Arunakumara, and Xuecheng Zhang. 2007. "Optimum Conditions for Transformation of *Synechocystis* Sp. PCC 6803." *Journal of Microbiology (Seoul, Korea)* 45 (3): 241–45.
- Zess, Erin K., Matthew B. Begemann, and Brian F. Pfleger. 2016. "Construction of New Synthetic Biology Tools for the Control of Gene Expression in the Cyanobacterium *Synechococcus* Sp. Strain PCC 7002: Construction of New Synthetic Biology Tools." *Biotechnology and Bioengineering* 113 (2): 424–32. <https://doi.org/10.1002/bit.25713>.
- Zhou, Jie, Haifeng Zhang, Hengkai Meng, Yan Zhu, Guanhui Bao, Yanping Zhang, Yin Li, and Yanhe Ma. 2014. "Discovery of a Super-Strong Promoter Enables Efficient Production of Heterologous Proteins in Cyanobacteria." *Scientific Reports* 4 (March). <https://doi.org/10.1038/srep04500>.

Appendix 2: SUPPLEMENTARY INFORMATION FOR CHAPTER 2

Table A2.S1: RBS Calculator predicted translation initiation rate for each construct using versions 1.1 and 2.1 of the Calculator

	v1.1	v2.1
GS1	113	125.8
GS10a	2705	161.4
GS10b	2526	406.4
GS100	20607	3512.2
2.0-1	94.5	129.2
2.0-10a	2337	661.8
2.0-10b	2466	3681.1
IDT1	44	32.9
IDT10a	2580	411.6
IDT10b	2600	168.8
EuH1	194	134.5
EuH10a	2600	2401.1
EuH10b	2363	1032.3
HCR1	259	229.9
HCR10a	2155	371.6
HCR10b	2720	772.6

Figure A2.S1: Growth of wild-type and engineered strains in shake flasks under continuous light. No growth defect due to genetic modifications was found.

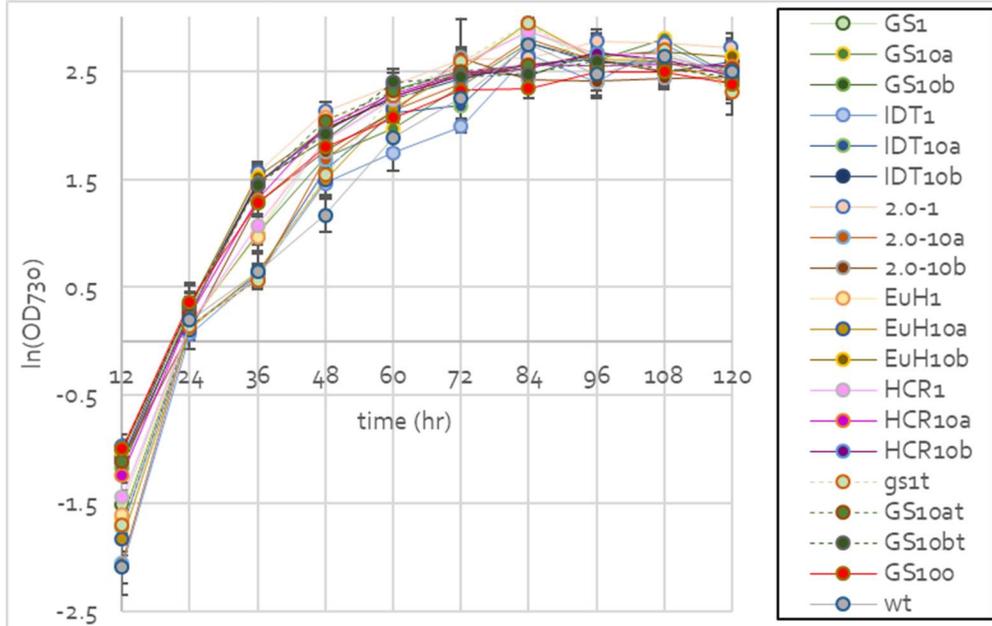


Table A2.S2. Gene sequences used in this study. C-terminal histag and linker sequences are highlighted.

Gene name abbreviation	Gene name	Sequence
GS	Bisabolene synthase GenScript codon optimization	ATGGCTGGAGTGTCTGCCGTGAGCAAAGTGTCTTCTCTGGTGTGTGATCT GTCTAGCACCTCTGGGCTGATTCGCCGGACCGCCAACCCCATCCCAATG TGTGGGGCTATGATTTGGTTCACCTCTTAAAAAGTCCCTACATTGACTCCA GTTATCGCGAACGTGCCGAAGTGTTAGTTAGTGAAATTAAGCCATGTTA AACCCCGCTATTACCGGCGATGGTGAATCCATGATTACTCCAGTGCCTA TGACACTGCTTGGGTGGCTCGCGTCCCGCTATTGATGGTAGCGCTCGTC CCCAATTTCCCAGACCGTGGATTGGATTCTGAAAAATCAATTGAAAGAC GGCAGCTGGGGTATTCAGTCTCATTCTGTTATCCGACCGCCTGTTGGCC ACCTTATCCTGTGTGCTGGTTTTACTGAAATGGAATGTGGGGGATTTGCA AGTTGAACAGGGAATTGAATTTATTAATCTAAGTACTGAACTGGTGAAG ATGAAACCGATCAAGACTCCCTGGTTACTGACTTTGAAATATTTTTCCA GTTTGTACGGGAAGCCAATCCTTGCCTTAGGCCTGCCCTATGATTTG CCCTACATTCATTTGTTGCAAACCAAACGTCAGGAACGGCTGGCTAAATT GAGTCGGGAAGAAATTTACGCCGTGCCAGTCCCTTACTGTATAGCTTGG AAGGGATTCAAGACATTGTGGAATGGGAACGCATTATGGAAGTTCAATC CCAGGATGGCTCCTTTTTGTCTCCCCGCCAGCACCGCTTGTGTGTTTAT GCACACTGGCGATGCCAAATGCTTAGAATTTTTGAATAGTGTATGATTA AATTTGGTAACTTTGTGCCCTGTTTGTACCCCGTTGATTTGTTAGAACGCC TGTTGATTGTGGACAATATTGTCGCTTGGGTATTTACCGTCATTTTGAAA AAGAAATTAAGAAGCCTTAGATTATGTGTACCGGCACTGGAATGAACG CGGGATTGGATGGGGCCGTTGAACCCATTGCCGATTTAGAAACCACT

		<p>GCTCTGGGGTTTCGCTTACTGCGTTTACATCGGTATAATGTGTCCCCCGCT ATTTTTGATAATTTTAAAGACGCCAACGGCAAATTTATTTGCAGTACCGGT CAATTTAACAAAGATGTGGCCTCTATGTTGAACTTATACCGTGCTTCCCAG CTGGCTTTTCCCGGCGAAAACATTTTGGATGAAGCCAAAAGTTTTGCTAC CAAATATTTGCGTGAAGCCTTGGAAAAAGCGAAAACCTCCAGTGCCTGG AATAACAAACAAAACCTGTCTCAGGAAATTAATACGCCCTGAAAACCG CTGGCATGCTTCTGTGCCCGTGTGAAGCCAAACGGTATTGTCAAGTGT ACCGCCCCGATTATGCCCGTATTGCTAAATGCGTGTACAAATTGCCCTAT GTTAACAAACGAAAAATTTCTGGAATTGGGTAAATTGGATTTAACATTATT CAATCCATTACCAGGAAGAAATGAAAAACGTGACCTCTGGTTTCGTGA TTCCGGCTTACCCTGTTTACTTTTGCCCGGAACGGCCCTTGGAAATTTA TTTTCTGGTGCCGCGGCACCTATGAACCCCAATATGCTAAATGTCGGT TTTTGTTTACCAAAGTGGCCTGCTTGCAAACCTGTTTGGATGACATGTATG ATACCTACGGCACTTTAGACGAATTGAAATTGTTTACCGAAGCCGTGCGG CGCTGGGATTTATCCTTTACTGAAAATCTGCCGACTACATGAAATTGTGT TACCAAATTTATTACGATATTGTGCATGAAGTTGCCTGGGAAGCTGAAAA AGAACAGGGTTCGTGAATTGGTGTCTTTTTCCGGAAAGGGTGGGAAGAT TATTTGTTAGGATACTACGAAGAAGCCGAATGGTTGGCCGCTGAATACGT GCCCACCTTAGATGAATACATTA AAAACGGGACTACTAGTATTGGACAAC GGATTCTGTTGTTAAGCGGCGTTCTGATTATGGACGGTCAACTGTTGTCC CAGGAAGCCTTAGAAAAAGTGGATTATCCCGGCCGGCGGGTTTTGACCG AATTAATAGCCTGATTTCTCGGTTGGCCGATGACACAAAACCTTACAAA GCCGAAAAAGCTCGCGGAGAATTAGCCAGCTCTATTGAATGTTATATGAA AGACCATCCGAATGCACCGAAGAAGAAGCCTTGATCACATTTATAGTA TTCTGGAACCCGCCGTGAAAGAAGTACCCGTGAATTTTTGAAACCCGAT GACGTTCCCTTTCCTGTAAGAAAATGTTGTTTGAAGAAACCCGCGTGAC TATGGTTATTTTTAAAGATGGTGACGGGTTTGGAGTTAGCAAATTGGAAG TGAAAGACCATATTAAGAATGCTTGATTGAACCCCTTACCCTTG ggcggtagtcatcaccatcatcaccatTAA</p>
2.0	Bisabolene synthase DNA2.0 codon optimization	<p>ATGgCCGGTGTGAGCGCAGTGAGTAAAGTGAGCAGTCTAGTATGTGATT TGAGTAGTACGTGGGGTTGATTCGGCGCACCGCCAACCCCATCCCAAT GTTTGGGGGTATGATTTGGTCCATTCTCTCAAGTCTCCCTACATTGACAGT AGTTATCGGGAACGGGCCGAAGTACTTGTATCTGAAATCAAAGCCATGTT AAATCCCGCGATACCGGTGATGGCGAATCTATGATTACGCCATCGGCTT ATGATACCGCCTGGGTAGCGCGGGTGCCTGCGATCGATGGCAGCGCCCCG TCCACAGTTCCCCCAAACCGTGGATTGGATTTTAAAAATCAACTGAAAG ATGGATCGTGGGGCATCCAATCGCACTTTCTGCTCTCTGACCGCTTGCTC GCGACCCTGTCTGTGTCTCTGCTCCTGAAATGGAATGTAGGGGATCT CCAAGTAGAACAAGGGATTGAGTTCATTAAGTCCAATCTCGAACTGGTGA AAGATGAAACTGATCAAGATTCTCTCGTTACCGACTTTGAAATCATTTTTC CCAGTCTTTACCGGAGGCACAGTCCTTGGGTTGGGACTGCCCTATGAC TTGCCTTACATCCATCTCCTGCAGACCAAACGCCAAGAACGGCTCGCGAA GCTCTCCCGCGAAGAGATTTATGCCGTGCCAAGTCTCTGCTGTATTCCCT CGAAGGTATTCAAGATATCGTTGAGTGGGAGCGGATCATGGAAGTGCAA AGTCAGGATGGTCTTTTCTGTCCAGTCCGGCTAGCACCGCGTGTGTTTTC ATGCACACTGGTGTGCGAAGTGCCTCGAATTTCTCAATAGCGTTATGAT CAAATTTGGAATTTTCGTTCCCTGCTTATACCCCGTTGATCTTCTGGAGCG GTTATTGATTGTTGATAACATTGTCCGCTGGGGATTTATCGCCATTTTGA</p>

		<p>GAAAGAGATTAAGAAGCCCTCGATTATGTCTACCGGCATTGGAACGAA CGGGGCATTGGTTGGGGTCGTCTGAACCCGATCGCGGATCTCGAAACAA CCGCATTAGGCTTCCGCCTTCTCCGCTTACATCGGTATAATGTGAGCCCAG CGATTTTTGACAACCTTTAAAGATGCTAACGGCAAATTCATTTGTAGCACC GGCCAGTTCAATAAAGACGTTGCCTCGATGCTCAACCTGTACCGTGCTAG TCAGTTGGCATTTCGGGGCGAAAACATTTTGGACGAAGCAAATCTTTTG CTACGAAATATCTCCGGGAAGCCCTTAAAAGAGTGAGACAAGTTCCGGC ATGGAATAACAAGCAGAATCTGTCCAGGAAATCAAATACGCACTTAAA ACGAGCTGGCACGCGAGCGTACCCCGTGTGAAGCCAAACGGTATTGCC AGGTGTACCGTCCCGATTATGCTCGTATCGCCAAGTGTGTGTACAAACTC CCTTACGTCAATAACGAAAAGTTCCTGGAGCTGGGGAAACTCGACTTTAA CATTATTCAATCCATTCACCAAGAAGAAATGAAGAATGTCACCTCTTGTT TCGCGACTCTGGCCTGCCCTGTTTACGTTTGC GCGTGAACGCCCTTAGA GTTTTACTTTCTGGTGGCAGCCGGAACGTACGAGCCCCAATATGCAAAGT GTCGGTTTTGTTTACTAAAGTGGCCTGCCTGCAGACGGTCTTGGATGAT ATGTATGACTTATGGCACATTGGACGAAGTGAAGCTGTTACCCGAAGC GGTGCCTCGCTGGGATCTGAGCTTACCGAAAACTTACCAGATTACATGA AACTCTGTTATCAAATCTATTACGATATTGTCCATGAAGTGGCGTGGGAA GCCGAGAAAGAGCAAGGTGCGGAATTAAGTGCCTTTTTCCGCAAAGGCT GGGAAGATTACCTCTTAGGCTACTACGAAGAAGCGGAATGGCTCGCTGC GGAATATGTGCCACCCTGGATGAGTATATCAAGAATGGTATCACCAGTA TTGGTCAACGGATCTTGTGTTGAGCGGTGTGCTGATTATGGATGGACAA CTTCTCTCGCAGGAAGCCTTGAAAAGGTTGACTATCCCGGACGCCGTGT TCTCACCGAAGTGAATAGTCTGATTTCTCGCTCGCCGATGATACAAAAAC GTATAAGGCCGAAAAGCTCGCGGGGAATTGGCATCCTCTATCGAATGC TATATGAAAGATCATCCCGAATGCACCGAAGAAGAGGCCTTAGACCACA TTACAGCATTCTGGAACCGGCGGTGAAAGAATTGACACGTGAATTTCTT AAACCTGATGACGTGCCATTTGCCTGTAAGAAAATGTTATTTGAAGAAAC CCGCGTACTATGGTGATCTTCAAAGATGGGGATGGGTTTGGCGTCAGC AAATTAGAGGTGAAGGACCACATCAAAGAGTGTCTCATCGAACCGTTGC CGTTAgcggttagtcatcaccatcatcaccatTAA</p>
IDT	Bisabolene synthase IDT codon optimization	<p>ATGGCTGGAGTCTCCGCGGTGAGTAAAGTTTCTCCCTAGTTTGGCAGCTT ATCTAGTACGTCTGGTTTAAATTCGGCGCACCCGCAATCCGCATCCAAACG TCTGGGGGTATGATCTCGTACATAGCCTCAAATCCCCATACATTGATAGC AGCTATCGAGAACGGGCGGAAGTACTCGTGTCCGAGATTAAGCCATGC TAAATCCCGCCATTACCGGGGATGGAGAAAGCATGATTACCCGAGTGC TTACGACACTGCATGGGTGGCACGCGTACCCGCAATCGACGGCTCCGCA CGCCCCAATTTCTCAGACTGTGCACTGGATCTTGAAAAATCAATTA GACGGTAGTTGGGGCATCCAATCCCCTTTTTGTTAAGTGATCGCTTATT GGCTACATTATCCTGTGTACTGGTACTATTGAAATGGAACGTAGGGGACT TACAGGTGGAACAAGGGATCGAATTTATTAATCAACTTAGAATTGGTG AAAGACGAAACCGACCAAGATAGTTTAGTGACCGACTTTGAAATCATTTT TCCCTCCTTGCTGCGGGAAGCCCAATCTCTACGCCTAGGCTTGCCTTACGA CTTACCCTATATCCATTTACTGCAGACGAAACGCCAGGAACGCCTGGCCA AGCTGAGCCGGGAAGAAATCTACGCAGTTCCCTCTCCCTTACTATACTCC CTGGAAGGCATCCAGGATATCGTGGAATGGGAACGCATTATGGAAGTGC AATCTCAGGACGGCTCTTTTTAAGTAGTCCCGCCTCTACGGCCTGCGTTT TCATGCATACCGGTGATGCCAAATGTCTAGAATTTCTCAATAGCGTGATG</p>

		<p>ATCAAATTTGGCAATTTTGTGCCCTGCTTATACCCTGTCGATTTGCTAGAA CGGTTATTAATTGTGGATAACATTGTTCCGGCTGGGAATCTATCGCCATTTT GAAAAGGAAATTAAGAAGCTCTCGATTATGTGTATCGGCATTGGAATG AACGGGGCATTGGATGGGGGCGACTGAATCCTATTGCAGACTTGGAGAC CACTGCCTTGGGCTTTTCGTCTACTGCGGTTACATCGGTATAATGTAAGTCC TGCCATCTTTGACAACTTCAAAGACGCCAATGGCAAATTTATTTGCAGCAC CGGCCAATCAATAAAGATGTCGCTTCCATGTTAACTTATACCCTGCTTC CCAACCTCGCGTTTCCAGGCGAGAATATTCTAGACGAAGCCAAAAGTTTTG CGACCAAATACCTGCGCGAAGCCCTAGAAAAGAGTGAACCAGTAGTGC GTGGAACAACAAGCAGAACCTGTCCCAGGAAATCAAATATGCGCTAAAA ACGAGCTGGCACGCTCCGTTCCCCGGTTGAAGCTAACGCTACTGCCA AGTGTATCGGCCAGACTATGCGCGGATCGCTAAGTGTGTGTACAACTCC CCTACGTGAATAACGAGAAATTCTTGAATTGGGGAAGTTAGATTTTAAC ATTATCCAGAGTATTCATCAGGAGGAGATGAAAAATGTTACCAGTTGGTT TCGGGATAGTGGTTTGCCTTATTTACGTTTGCTCGAGAACGACCCTGG AGTTTTATTTCTCGTAGCCGCGGGGACCTACGAACCGCAGTATGCCAAA TGCCGCTTTTTATTACCAAAGTGGCCTGTCTCAAACCGTATTGGATGAT ATGTATGATACTTATGGGACCTTGGATGAACTCAAATTATTTACCGAAGC TGTACGGCGTTGGGATTTAAGCTTTACTGAAAACCTACCCGATTATATGA AGCTGTGCTACCAGATCTACTATGACATTGTTTCATGAAGTGGCCTGGGAA GCTGAGAAAGAACAAGGCCGTGAACTCGTGTCTTTTTTTTCGGAAGGGCT GGGAGGACTATCTATTAGGCTATTATGAAGAAGCAGAATGGTTGGCCGC GGAGTACGTGCCTACACTCGACGAATATATTA AAAACGGTATTACTTCCA TTGGGCAACGCATTTTGTCTCTATCCGGAGTGTTAATCATGGATGGCCAG TTGTTGAGTCAGGAAGCACTGAAAAAGTGGACTATCCCGGCCGACGAG TGCTGACCGAGTTGAACTCCCTAATCTCCCGCCTCGCCGATGATACAAAA ACTTACAAGGCGGAAAAAGCTCGTGGAGAGTTGGCCTCTTCCATTGAGT GTTATATGAAGGATCATCTGAGTGTACAGAAGAAGAAGCTTTGGACCA TATTTACTCCATTTTAGAACCGGCAGTTAAGAATTA ACTCGGGAATTTTT GAAGCCTGACGATGTACCGTTTGCCTGTAAAAAGATGCTGTTTCGAAGAA ACACGCGTAACCATGGTCATTTTTAAGGATGGTGATGGCTTCGGTGTTAG CAAACCTAGAGGTTAAAGATCATATTAAGGAATGTTTAATTGAACCACTGC CTTTGggcgtagtcatcacatcatcacatTAG</p>
EuH	Bisabolene synthase Eugene - Harmonizatio n codon optimization	<p>ATGGCCGGAGTGAGTGCCGTCAGCAAGGTGAGCTCTTTGGTGTGCGATT TGTCATCTACGTCTGGATTGATCCGTCGGACC GCGAATCCCCATCCTAAT GTCTGGGGGTATGATTTAGTTCATAGTTTTAAAAGCCCCTATATCGATAG TTCTTACCGGGAAAGAGCAGAGGTCTTAGTGTCTGAGATCAAAGCAATG TTAAATCCCGCCATCACTGGCGATGGCGAAAGCATGATCACCCCACTGTC CTATGACACTGCCTGGGTGCGACGTGTTCTGCGATCGATGGAAGTGCCA GACCACAATTTCTCAAACCTGTGGACTGGATCTTGAAAACCACTCAAAA GATGGGAGCTGGGGCATCCAGAGCCACTTTCTGCTGAGCGACCGTTTATT AGCGACCTTAAGTTGCGTGTTAGTTCTTAAAATGGAACGTGGGTGATC TGCAAGTCGAGCAGGGCATCGAATTCATCAAGTCTAATCTGGAACCTGTG AAGGATGAAACGGATCAAGATTCCTTTGGTCACTGACTTTGAGATCATCTT TCCCAGTCTGCTCCGGGAAGCCCAAAGTCTGAGACTCGGCTTACCTTACG ACCTGCCCTATATCCATCTGTTGCAGACCAAAAAGACAGGAACGGCTCGCC AAATTAAGCCGTGAGGAAATCTATGCAGTGCCATCACCCCTGTTGTATAG TCTCGAGGGCATCCAAGATATCGTGAATGGGAACGTATCATGGAAGTG</p>

		<p>CAATCTCAGGATGGTAGTTTCCTCTCTAGCCCCGCCAGTACCGCGTGTGT GTTTCATGCACACTGGCGACGCAAAATGTTTAGAATTCTGAACTCTGTTAT GATCAAGTTTGGCAATTTTGTGCCTTGTCTGTATCCCGTTGATCTGCTGGA AAGACTGTTGATCGTCGATAATATCGTCAGATTAGGCATCTATCGGCACT TTGAAAAGGAAATCAAGGAAGCCTTAGATTATGTGTACCGTCATTGGAAC GAACGGGGCATCGGTTGGGGACGGCTTAATCCTATCGCCGATTTAGAGA CGACCGCCTTGGGCTTTCGTTTGTAAAGACTGCATCGTTACAATGTCAGTC CCGCGATCTTGGACAACCTCAAAGATGCGAATGGTAAATTCATCTGTTCA ACGGGGCAATTCACAAAGATGTCGCCTCTATGCTGAATTTATATCGGGC CAGCCAGCTCGCCTTTCCTGGCGAAAACATCTTAGATGAAGCCAAATCTT TCGCCACCAAATATTTGCGGGAAGCCTTAGAGAAATCTGAGACCAGCTCT GCCTGGAACAACAAACAAAACCTGTCTCAAGAGATCAAATACGCACTGA AGACCAGTTGGCATGCGTCTGTGCCACGGTTGAAGCCAAGCGGTACTG CCAAGTTTATAGACCCGATTATGCCAGAATCGCCAAATGTGTGTACAAGC TTCCTTACGTTAACAATGAAAAGTTTCTCGAGCTGGGCAAACCTCGATTTCA ACATCATCCAGAGCATCCACCAAGAAGAAATGAAGAATGTGACGTCTTG GTTTCGGGATTCAGGTTTGGCCCTTTTACGTTTCGCCAGAGAGCGTCCACT GGAATTCTACTTCTCGTCGCAGCAGGTACGTATGAACCTCAGTATGCGA AATGTCGTTTCTCTTTACTAAAGTTGCCTGTTTGCAGACCGTGTGGACG ATATGTATGACACCTATGGCACGCTTGAATTGAAGCTTTTACCAGAG GCCGTTTCGGCGGTGGGACCTCAGCTTACTGAAAACCTACCCGACTATAT GAACTTTGCTACCAAATCTATTATGACATCGTGCAGAGGTTGCCTGGG AGGCCGAGAAGGAACAGGGTCGTGAATTGGTCTTTTTTCCGGAAGGG CTGGGAGGATTATTTACTGGGGTATTATGAAGAAGCCGAATGGCTCGCC GCCGAGTATGTTCCACGTTGGACGAGTACATCAAGAATGGCATCACTAG TATCGGACAACGTATCTTACTGTTGTCTGGCGTTTTGATCATGGATGGTCA ACTCTTATACAAGAGGCCCTCGAGAAAGTCGATTATCCCGGCCGGCGTG TGCTCACTGAGCTGAATTCTCTCATCAGCAGACTGGCAGATGACACAAAG ACTTATAAAGCCGAGAAGGCCCGTGGCGAATTGGCAAGCTCTATCGAAT GCTACATGAAAGACCATCCCGAATGCACTGAGGAAGAGGCCCTCGATCA CATCTATTCTATCCTGGAGCCAGCAGTTAAGGAAGTACTCGGGAGTTTC TGAAGCCTGACGACGTCCCCTTCGCGTGAAGAAGATGTTATTCGAGGA GACTCGGGTTACAATGGTTATCTTCAAGGATGGCGATGGCTTCGGGGTG AGCAAACCTCGAAGTCAAAGATCATATCAAAGAGTGCCTCATGAACCACT GCCCTGggcgtagtcatcaccatcatcaccatTAG</p>
EuHCR	Bisabolene synthase Eugene – Harmonization/Codon Context/Remove Repeats codon optimization	<p>ATGGCGGGGGTATCGGCGTTTCCAAGGTTTCTCTTTGGTTTGTGACCT ATCATCAACTTCTGGTTTAAATCCGCCGCACTGCTAATCCTCATCCTAATGT CTGGGGCTATGACCTAGTCCATCCCTCAAATCCCTTACATCGACAGTTC TTACCGGGAAAGGGCAGAGGTAAGTGTGCGGAGATAAAAAGTAAATGCTC AATCCGGCGATCACCGGGGATGGGGAATCCATGATCACTCCTTCCGCTA TGACACCGCTTGGGTAGCAAGGGTACCGGCGATCGATGGCAGTGCCAGA CCTCAATTTCCCAAACGGTGGACTGGATATTA AAAAACCAGCTAAAAGA TGGCAGTTGGGGCATCCAGAGCCATTTTCTGCTGTGCGACAGATTACTGG CAACTCTGAGTTGTGTGCTGGTTCTACTCAAATGGAATGTGGGGGATTTA CAGGTAGAGCAGGGCATAGAGTTCATCAAGTCTAACCTGGAAGTACTCA AGGATGAAACGGATCAAGATTCTTTGGTCACCGACTTTGAGATAATCTTT CCTTCCCTACTGCGGGAAGCCCAATCCCTCAGGCTTGGTTTACCTTACGAC CTACCTTACATCCATCTGCTCCAGACTAAAAGACAGGAAAGATTAGCTAA</p>

		<p>ACTTCCAGGGAGGAAATATATGCAGTGCCGTCGCCATTGCTCTACAGCC TGGAGGGCATCCAGGATATAGTGAATGGGAAAGAATAATGGAAGTGC AAAGCCAGGATGGTTCTTCTTCTTCCCCTGCTTCCACTGCTTGCCTTTT CATGCACACCGGCGATGCAAATGCCTGGAATTCCTCAACTCAGTAATGA TCAAGTTTGGCAATTTTGTCCCTTGCCTTTATCCGGTGGATTTACTGGAAA GATTATTAATAGTAGATAATATCGTCAGACTGGGCATCTATCGCCATTTG AAAAGGAAATCAAGGAAGCCCTTGATTATGTTTACCGCCATTGGAACGA ACGGGGCATCGGTTGGGGAAGACTTAATCCGATCGCCGATTTGGAGACA ACAGCATTGGGTTTTCGTTTACTGAGACTCCACCGTTACAACGTCAGCCC AGCGATCTTTGACAACTTCAAAGATGCTAATGGTAAATTCATCTGCTCAAC GGGACAATTCAACAAAGATGTAGCTTCTATGCTCAATCTTTATCGGGCCA GCCAGCTAGCTTTTCTGGGGAAAACATCCTCGATGAAGCAAAATCTTTC GCCACTAAATATCTGCGGGAAGCCCTGGAGAAAATCAGAGACTTCTTCTGC CTGGAACAACAAGCAAAACCTCAGCCAGGAGATAAAATATGCTCTGAAG ACTTCTGGCATGCTTCAGTGCCAAGGGTTGAAGCCAAGCGTACTGCCA AGTTTATCGACCCGATTATGCCAGAATCGCCAAATGCGTTTATAAGCTTCC TTACGTTAACAATGAAAAGTTTCTGGAGTTGGGAAAATTAGATTTCAACA TCATCCAGTCAATCCACCAGGAAGAAATGAAGAATGTCACTTCTTGGTTC CGGGATTCTGGTTTGCCCTTTTCACTTTCCGCAGGGAGCGTCCTCTGGA ATTCTATTTCTGGTAGCGGCGGGAAGTTATGAACCACAGTATGCTAAAT GCCGTTTTCTTTTACCAAAGTTGCTTGCCTGCAAACGGTGTCTGCAGGATA TGTATGACACCTATGGCACACTAGATGAATTAAGCTTTTCACTGAAGCA GTACGCCGTTGGGACGTATCTTTCACGGAAAATTTACCGGACTATATGAA ACTTTGTTACCAAATCTATTACGACATCGTCCACGAGGTAGCTTGGGAGG CGGAGAAGGAACAGGGAAGGGAATTGGTCAGCTTTTCCGTAAGGGTT GGGAGGATTATTTACTGGGTTACTACGAAGAAGCAGAATGGTTAGCGGC GGAGTATGTCCCTTCTTGGACGAGTACATCAAGAATGGCATAACTTCCA TCGGTCAAAGAATATTACTGCTGTCCGGGGTTTTAATCATGGATGGGCAA CTACTCAGCCAAGAGGCCCTGGAGAAAAGTAGATTATCCAGGGCGGCGAG TGCTCACGGAGTTAAACAGCCTAATCAGCCGACTGGCAGATGACACAAA GACCTATAAAGCCGAGAAGGCTAGGGGAGAATTAGCATCATCAATAGAA TGTTACATGAAAGACCATCCGGAATGCACAGAGGAAGAGGCCCTAGACC ACATCTATTCAATACTGGAGCCGGCGGTTAAGGAATTAAGTGGGAGTTT CTAAGCCAGACGACGTACCTTTCGCCTGCAAGAAGATGCTTTTCGAGGA GACGGGGGTAAGTATGGTTATCTTCAAGGATGGCGATGGTTTCGGAGTT TCTAAACTGGAAGTAAAAGACCATATCAAGGAGTGCCTAATCGAACCGTT ACCCCTTggcggtagtcacatcacatcacatTAA</p>
FPPS	Farnesyl pyrophosphate synthase GenScript codon optimization	<p>ATGGATTTTCCCAACAACCTGGAAGCCTGCGTTAAACAAGCCAACCAAGC CCTGTCCCCTTTATTGCCCCCTGCCCTTTCAGAACACTCCCGTGGTTGA AACTATGCAATATGGGGCCTTGTAGGCGGTAACGCTTACGTCCCTTTC TGGTGTACGCCACCGCCACATGTTTGGTGTAGCACCAATACTTTAGAT GCTCCCCTGCTGCCGTGGAATGTATTCATGCCTATTCTCTGATTCACGAT GACTTGCCCGCTATGGATGACGATGACTTGGCGCGGGTTACCCACCTG CCACGTGAAATTTGGCGAAGCTAACGCCATTTTAGCTGGTGATGCCTTGC AACTTTAGCTTTTAGCATTTTGTCTGATGCCGACATGCCGAAGTGTCCG ATCGGGACCGCATTTCCATGATTAGTGAATTAGCTTCCGCCAGTGGCATT GCCGGCATGTGTGGGGGACAAGCCTTGGATTTGGATGCCGAAGGCAAA CATGTGCCCTGGATGCCTTGAACGTATTCATCGGCACAAAACCGGCGC</p>

		CTTGATTGCGCTGCCGTTGTTTTAGGTGCTCTGTCCGCCGGGGATAAAG GACGTCGGGCCTTGCCCGTGTTAGATAAATACGCTGAAAGTATTGGGCT GGCCTTCAAGTTCAGGATGACATTCTGGATGTGGTTGGAGACACCGCTA CTTTGGGGAAACGTCAAGGAGCCGATCAACAGTTGGGGAAAAGCACCTA CCCCGCCCTGTTGGGATTAGAACAGGCTCGGAAAAAAGCCCGCGATCTG ATTGATGACGCTCGCCAATCCCTGAAACAGTTGGCCGAACAATCTCTGGA TACTAGTGCTCTGGAAGCTCTGGCCGATTACATTATTCAGCGGAACAAAT AA
--	--	---

Table A2.S3: Primers used in this study

Sequencing primers		
Primer #	Primer name	Sequence
14	Ag1 down check v2	CACATTTATAGTATTCTGGAACCC
17	0168A down check new	CTACCCACAGAATTTGAACG
18	Ag1 int check1	CAGTCTCATTTTCTGTTATCCG
19	Ag1 int check rev	AAACCAGCACACAGGATAAGG
20	Ag1 int check2	CGCTATTTTTGATAATTTTAAAGACG
21	Ag1 int check3	TGAATTGGTGCCTTTTTCC
22	Ag1 down check	AGAAAATGTTGTTTGAAGAAACC
24	lacY down check	AGTTTGCTAATTTCTTTACTTCG
25	lacI down check	ATGCAAATGCTGAATGAGG
26	0168B down check	TATTGCGTAATCCCTTCAG
27	ispA int check	GCCCGTGTAGATAAATACG
28	Ag1 2.0 int check1	TCTCCCGCGAAGAGATTTATGC
29	Ag1 2.0 int check2	TGTACCGTCCCGATTATGCTC
30	Ag1 2.0 int check3	CGATGATACAAAAACGTATAAGGC
57	Ag1 IDT intchk1	CATGGGTGGCAGCGTACC
58	Ag1 IDT intchk2	ATTGGATGGGGGCGACTGAATCC
59	Ag1 IDT intchk3	GGCGTTGGGATTTAAGCTTTACTG
60	Ag1 EuH intchk1	CCTGGGTCGCACGTGTTT
61	Ag1 EuH intchk2	CGGTTGGGGACGGCTTAATC
62	Ag1 EuH intchk3	CGGTGGGACCTCAGCTTTACT
63	Ag1 HCR intchk1	AGCAAGGGTACCGGCGATC

64	Ag1 HCR intchk2	GTTGGGGAAGACTTAATCCGATCG
65	Ag1 HCR intchk3	CGTTGGGACGTATCTTTCACGG
131	0168up check	aactccaagcggaagatattacgg
	check BS ispA	gcaccgaagaagaagccttgg
	chck HR mazF	ggagaccaagcccaatttcg
	chck ispA lacY	gttgtttgtcgggtaacgctctc
	chck lacI HR	ccacttccacataggagactttgg
	new check 0168A down	GCCAATTTGTTTGGCG
	3' back-ispA	cctttcagcaaaaaaccctcaagtctagacacctgcaccagac
	chck lacY lacI	ctggctgtttctgcttcttaagc
	3' AgBIS-ispA	gttgttggggaaaaatccatattggtttcctcgttattacaaggta
	5' AgBIS-Back	gtctgtgcaggtgtctagacttgaggggtttttgctgaagg
	5' back-AgBIS	gcctttctcgtttataggtaacagttcttagatttctagag
	3' ispA-Back	ctctagaatcgtaagaactgttacctataaacgcagaaaggc
	mazF5'	cagaatagaagtgagttagtaac
	mazF 3'	tctagagaagaagggtgtgctga
	V2 mazChkDw	gtcaaaatcaaccacaaatcagatcg
	V3-kanChkUp	gcaatgtaacatcagagattttgagacac
	V4-KanChkDw	caccttcttcacgaggcagac
Cloning primers		
Primer #	Primer name	Sequence
1	For Ag1 GS	agttactctagaAACCCGGGGTACCAATTGTG
2	Rev Ag1 GS	TGGTTTCACTCTAGAGTGGTTTTT
3	Ag1 2.0 5'	ATGGCCGGTGTGAGCGCAG
4	Ag1 2.0 3'	TTATAACGGCAACGGTTCGATGAGACAC
5	Ptic2op3'	ATGATTCTGTTTCCTGTGTGAAATTG
6	Ag1 term 5'	CCAGGCATCAAATAAACGAAAGG
7	Ptic2op+Ag1 2.0	CGGATAACAATTTACACAGGAAACAGAATCATatggccggtgtgagcgag
8	Ag1 2.0+term	GTGTCTCATCGAACCGTTGCCGTTATAAccaggcatcaataaacgaaaggctcag
9	Ag1 2.0+ispA	GTGTCTCATCGAACCGTTGCCGTTATAAaacgaggaaaaccatattgattttcccaac
23	lacY 5' (actually lacY5')	TCTAGAATGTACTATTTAAAAAACAC
33	Ag1 IDT-CO Gibson I +	TTCACACAGGAAACAGAATCATATGGCTGGAGTCTCCGC
34	Ag1 IDT-CO Gibson I -	GCGGAGACTCCAGCCATATGATTCTGTTTCCTGTGTGAA
35	Ag1 IDT-CO Gibson II +	CCAACCTCGCGTTTCCAGGCGAGAATATTCTAGACGAAGCCA

36	Ag1 IDT-CO Gibson II -	TGGCTTCGTCTAGAATATTCTCGCCTGGAAACGCGAGTTGG
37	Ag1 IDT-CO Gibson III +	CCACTGCCTTTGTAGTAACGAGGAAAACCATATGGATTTTCCCAA
38	Ag1 IDT-CO Gibson III -	TTGGGGAAAATCCATATGGTTTTCTCGTTACTACAAAGGCAGTGG
39	Ag1 COharmon Gibson I +	TTCACACAGGAAACAGAATCATATGGCCGGAGTGAGTG
40	Ag1 COharmon Gibson I -	CACTCACTCCGGCCATATGATTCTGTTTCCTGTGTGAA
41	Ag1 COharmon Gibson II +	CCTTTCCTGGCGAAAACATCTTAGATGAAGCCAAATCTTTCGC
42	Ag1 COharmon Gibson II -	GCGAAAGATTTGGCTTCATCTAAGATGTTTTCGCCAGGAAAGG
43	Ag1 COharmon Gibson III +	GAACCACTGCCCTGTAGTAACGAGGAAAACCATATGGATT
44	Ag1 COharmon Gibson III -	AATCCATATGGTTTTCTCGTTACTACAGGGGCAGTGGTTC
45	Ag1 HarmonConte xt Gibson I +	TTCACACAGGAAACAGAATCATATGGCGGGGGTATCGG
46	Ag1 HarmonConte xt Gibson I -	CCGATACCCCCGCCATATGATTCTGTTTCCTGTGTGAA
47	Ag1 HarmonConte xt Gibson II +	CTGGGGAAAACATCCTCGATGAAGCAAATCTTTCGCCACTA
48	Ag1 HarmonConte xt Gibson II -	TAGTGGCGAAAGATTTTGCTTCATCGAGGATGTTTTCCCCAG
49	Ag1 HarmonConte xt Gibson III +	ATCGAACCGTTACCCCTTAATAACGAGGAAAACCATATGGATT
50	Ag1 HarmonConte xt Gibson III -	AATCCATATGGTTTTCTCGTTATTAAGGGGTAACGGTTCGAT
66	Plac del 5' b	CTATGACCATGATTACGCCAAGC
67	ispA Histag add 3'	TTAgtgatgatggtgatggtgggagccaccTTTGTTCCGCTGAATAATGTAATCG
68	ispA Histag+term	GCTCCCACCATCACCATCATCACTAA CCAGGCATCAAATAAAACGAAAGGCTCAG

69	Ag1 Histag add 3'	TTAatggtgatgatggtgatgactaccgccCAAGGGTAAGGGTTCAATCAAGCA
70	Ag1 Histag+ispA	gCGGTAGTCATCACCATCATCACCATTAAtAACGAGGAAAACCATATGGATTTTC CCCA
77	Ag1 HCR Histag add 3'	TTAatggtgatgatggtgatgactaccgccAAGGGTAACGGTTCGATTAGG
78	Ag1 HCR Histag+ispA	GCGGTAGTCATCACCATCATCACCATTAAtAACGAGGAAAACCATATGGATTTTC CCCA
79	Plac del bridge b	ACTCATTAGGCACCCCAGGCCTATGACCATGATTACGCCAAGCTTG
80	Ag1 GS RBS 5'	TATCGTACGTTTCAAAAACCTTATTAAGGAAAGAATGGCTGGAGTGTCTGCCG
81	Ag1 2.0 RBS 5'	GAGGAGACGGACCCTTTCCAAGACGTTTAGGTAAGATGGCCGGTGTGAGCGC
82	Ag1 IDT RBS 5'	CAATAGCATCTATATAAAACATATCGGTAAAAATGGCTGGAGTCTCCGCG
83	Ag1 EuH RBS 5'	AACAGGAATATACTATTTAGAGGTACGGTAAACATATGGCCGGAGTGAGTGCC
84	Ag1 HCR RBS 5'	GCGCAGCACATCGCAACAATAAAAGGGCTATATGGCGGGGGTATCGGC
85	lacO 3'	AATTGTTATCCGCTCACAATTCCACAC
86	Ag1 GS RBS bridge	TGTGGAATTGTGAGCGGATAACAATTTATCGTACGTTTCAAAAACCTTATTAAGG AAAGA
87	Ag1 2.0 RBS bridge	TGTGGAATTGTGAGCGGATAACAATTGAGGAGACGGACCCTTTCCAA
88	Ag1 IDT RBS bridge	TGGAATTGTGAGCGGATAACAATTCAATAGCATCTATATAAAACATATCGGTAA AAAT
89	Ag1 EuH RBS bridge	TGTGGAATTGTGAGCGGATAACAATTAACAGGAATATACTATTTAGAGGTACGG TAAACA
90	Ag1 HCR RBS bridge	TGTGGAATTGTGAGCGGATAACAATTGCGCAGCACATCGCAACA
91	Ag1 GS RBS2 5'	ataacatagggcaggagtgtccgagCAAAAGTGTCTTCTCTG
92	Ag1 2.0 RBS2 5'	tattataggaagagattATGGCCGGTGTGAGCGCA
93	Ag1 IDT RBS2 5'	aacgaggcaagaatggcaGGAGTCTCCGCGGTGAGT
94	Ag1 EuH RBS2 5'	aagtcagaaaacaaATGGCCGGAGTGAGTGCC
95	Ag1 HCR RBS2 5'	agtttagcgtcaacATGGCGGGGGTATCGGCG
96	Ag1 GS RBS2 3'	ttccgtaagtttttagtattgagaAATTGTTATCCGCTCACAATTC
97	Ag1 2.0 RBS2 3'	gtagatttagaataaAATTGTTATCCGCTCACAATTCCACACATTATAC
98	Ag1 IDT RBS2 3'	cttgtgctaggtaggcctAATTGTTATCCGCTCACAATTCCAC
99	Ag1 EuH RBS2 3'	ctccttctgtgtgAATTGTTATCCGCTCACAATTCCAC

100	Ag1 HCR RBS2 3'	aaaagaattttgtgaaAATTGTTATCCGCTCACAATTCCACACATTATAC
101	Ag1 GS RBS2 bridge	AATTTCTCAATAACTAAAACTTACGGGAAATAACATATGGCAGGAGTGTCCG
102	Ag1 2.0 RBS2 bridge	GAGCGGATAACAATTTTATTCTAAAATCTAACTATTATAGGAAGAGATTATGGCC GGTG
103	Ag1 IDT RBS2 bridge	GGATAACAATTAGGCCTACCTAGCACAGAACGAGGCAAGAATGGCAGGA
104	Ag1 EuH RBS2 bridge	CGGATAACAATTCACACAGAAAGGAGAAGTCAGAAAACAATGGCCGGAG
105	Ag1 HCR RBS2 bridge	GAGCGGATAACAATTTTCACAAAATTCTTTTAGTTTAGGCGTCAACATGGCG
106	ispA(sigA RBS) 5' v2	taacgaggaaaacccatggttttc
107	Ag1 GS RBS2 5' v2	ataacatggtgaggagtgccggtGAGCAAAGTGTCTTCTC
179	Ag1-GS RBS20k	ATCCCCCAAACCAAAGGGAGGTTTAAGAATGGCTGGAGTGTCTGCCG
180	tic2op+GS20k bridge	TGTGGAATTGTGAGCGGATAACAATTATCCCCCAAACCAAAGGGAGG
130	0168 B 3'	ctaagtcagcgtaaatctgacaatgatg
	pUC5'	gcatgcggtgtaacatg
	pUC3'	gagctcttacccaacttaacg
	0168-5'	atgactattcaatacacccttagc
	0168-3'	atgactattcaatacacccttagc
	BS5'	cggatccgtgggatcc
	BS3'	ttacaaggtaagggtcaatcaagc
	ispA5'	taacgaggaaaacccatggt
	ispA3'	tctagaaatcgtaagaactgttacc
	lac3'	tactgcccgtttccag
	lac5'	gaattcctcgacctgcagg
	0168 B5'	ttggggctggcggattgg
	0168 B3'	ctaagtcagcgtaaatctgacaatgatgctg
	lacY 3' v2	atcactccgttatgatatgttggtc
	lacY 3' v3	atcactccgttatgatatgttggtcggataagg
	lacY 5' Ncut	gattacacatggcatggatgaac
	lacY 5' cut v2	agttaggtcaatcatcatcggaataatttaagtctttctg
	lacY 5' cut	agttaggtcaatcatcggaataatttaag
	Ag1 2.0 5'	ATGGCCGGTGTGAGCGCAG
	Ag1 2.0 3'	TTATAACGGCAACGGTTCGATGAGACAC
	Ptic2op3'	ATGATTCTGTTTCTGTGTGAAATTG
	Ag1 term 5'	CCAGGCATCAAATAAACGAAAGG
	Ptic2op+Ag1 2.0	CGGATAACAATTTTCACACAGAAACAGAATCATatggccggtgtgagcgcag
	Ag1 2.0+term	GTGTCTCATCGAACCGTTGCCGTTATAAccaggtcaataaaaacgaaaggctcag

pUC+slr0168	ctgcaaggcgattaagttgggtaGAGTCatgactattcaatacacccccctagccg
slr0168+BS	gctttggtctggtgcaggtgtctagaCGGATCCGTGGGGATCCGGCTGC
BS+ispA	atgcttgattgaacccttaccctaccctttaaTAACGAGGAAAACCATAT
ispA+lacY	gcgtttataggaacagttcttacgattTCTAGAGaattcctcgaCCTGCAGGA
lacY+pUC19	CCTTATCCGACCAACATATCATAACGGAGTGATgcatgcgcgtaatcat

Table A2.S4: Bisabolene Synthase Percent Identity Matrix (<https://www.ebi.ac.uk>) (Li et al., 2015)

```
#
#
# Percent Identity Matrix - created by Clustal2.1
#
#
1: EuH      100.00  82.80  75.55  76.77  76.20
2: HCR      82.80  100.00  74.86  73.43  74.45
3: IDT      75.55  74.86  100.00  78.73  77.26
4: GS       76.77  73.43  78.73  100.00  80.15
5: 20       76.20  74.45  77.26  80.15  100.00
```

Figure A2.S2: Peptides from bisabolene synthase and farnesyl pyrophosphate synthase that were identified by LC-MS/MS from a sample of *Synechocystis* sp. PCC 6803, strain 2.0-10B are highlighted below.

>Bisabolene synthase amino acid sequence

MAGVSAVSKVSSLVCDLSSTSGLIRR TANPHPNVWGYDLVHSLKSPYIDSSYRE **RAEVLVSEIKA** MLNPAITGDGESMIT
 PSAYDTAWVARVPAIDGSARPFQPTVDWILKNQLKDGSWGIQSHFLSDRLLATLSCVLVLLKWNVGDQLQVEQIEFI
 KSNLELVKDETDQDSLVTDFEIIFPSLLREAQSL **RLGLPYDLPYIHLLQTKR** QERLAKLSREEIYAVPSPLLYSLEGIQDIVWE
 RIMEVQSQDGSFLSSPASTACVFMHTGDAKCLEFLNSVMIKFGNFVPCLYPVDLLER **LLIVDNIVRL** GIYRHFEKEIKEALD
 YVYRHWNERGIGWG **RLNPIADLETTALGFRLRLHRYNVSPAIFDNFKDANG** **KFICSTGQFNKD** VASMLNLYRASQLAF
 PGENILDEAKSFAT **KYLREALEKSETSSAWNNKQ** NLSQEIKYALKTSWHASVPRVEAKRYCQVYRPDYARIAKCVYKLPY
 VNNEKFLELGKLDFNIIQSIHQEEM **KNVTSWFRD** **SGLPLFTFARE** RPLFEYFLVAAGTYEPQYAKCRFLFTKVAACLQTVLD
 DMYDTYGTLDELKLFTEAVRRWDLSTENLPDYMKLCYQIYYDIVHEVAWEAEKEQG **RELVSFFRK** GWEDYLLGYEEA
 EWLAAEYVPTLDEYIKNGITSIGQRILLSGVLIMDGQLLSQEALEKVDYPGR **RVLTELSLISRL** ADDTKTYKAEKARGEL
 ASSIECYMKDHPECTEEALDHIYSILEPAVKELT **REFLKPDDVPFACKK** MLFEETRVTMVIFKDGDFGVSKLEVKDHIKE
 CLIEPLPLGGSHHHHHH

>Farnesyl pyrophosphate synthase amino acid sequence

MDFPQQLEACVKQANQALSRIAPLPFQNTPVVETMQYGALLGGKRLRPFLVYATGHMFGVSTNTLDAPAAAVECIH
 AYSLIHDDL PAMDDDDLRRLPTCHVKFGEANILAGDALQTLAFSILSDADMPEVSDRDRISMISELASASGIAGMCG

GQALDLDAEGKHVPLDALERIHRHKTGALIRAAVRLGALSAGDKGRRALPVLDKYAESIGLAFQVQDDILDVVGDTATL
GKRQGADQQLGKSTYPALLGLEQARKKARDLIDDARQSLKQLAEQSLDTSALEALADYIIQRNK

List of Abbreviations

aSD: The anti-Shine-Dalgarno sequence of the 3'-end of 16S rRNA

BCD: bicistronic design

BG11: Blue-green algae media 11

GFP: Green fluorescent protein

RBS: Ribosome binding site

S. 6803: *Synechocystis Sp.* PCC6803

S. 7002: *Synechococcus Sp.* PCC7002

S. 7942: *Synechococcus Sp.* PCC7942

SD: Shine-Dalgarno sequence

YFP: Yellow fluorescent protein