Creating a Gene Expression Cassette for the Cyanobacterium Synechococcus sp. PCC 7002

And Using it to Replace the Endogenous Hydrogenase

Christopher Wright Williams

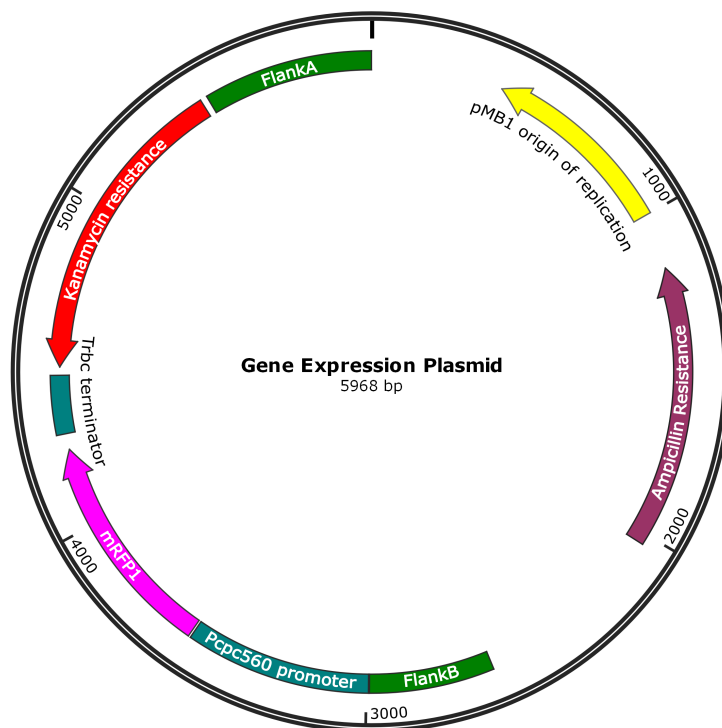# CREATING A GENE EXPRESSION CASSETTE FOR THE CYANOBACTERIUM SYNECHOCOCCUS SP. PCC 7002

CHRISTOPHER WRIGHT WILLIAMS

Gene Expression Plasmid
5968 bp

FlankA
Kanamycin resistance
pMB1 origin of replication
1000
Ampicillin Resistance
2000
Trbc terminator
mRFP1
4000
Pcpc560 promoter
FlankB
3000
5000

And Using it to Replace the Endogenous Hydrogenase

May 2015 – version 1.0

ABSTRACT

Eukaryotic algae and cyanobacteria are a promising source of future biofuels, having many important advantages over 1st and 2nd generation biofuels. *Synechococcus* sp. PCC7002 is the fastest growing oxygen producing photoautotroph known at present, and is capable of surviving a large wide of environments. As such, it is a very interesting organism for the production of biofuels. This is the first work that involves *Synechococcus* sp. PCC7002 at NTNU. In order to develop this organism into a hydrogen-producing organism, the native and inefficient [NiFe]-hydrogenase should be deleted and replaced by an energetically poised efficient [FeFe]-hydrogenase. A codon-optimized gene for expressing a [FeFe]-hydrogenase was designed. A plasmid was designed to allow for easy transformation and expression of exogenous genes, including the codon-optimized [FeFe]-hydrogenase, in *Synechococcus*. It contained a high expression promoter, capable of driving expression of proteins in a related cyanobacteria to 15% of the cell's dry weight. A temporary marker gene, *mRFP1*, was used to verify transformation. However, due to time-constraints, this overexpression system was not evaluated in *Synechococcus.*

In addition, a plasmid was constructed that could delete the native [NiFe]-hydrogenase in *Synechococcus*. Unfortunately, *Synechococcus* could not be succesfully transformed with this deletion plasmid. Further, using the gene expression plasmid, a different hydrogenase could be expressed in this strain.

Modifications to the *Synechococcus* transformation protocol are discussed, to help future researchers at NTNU.

. . .

SAMMENDRAG

Eukaryotiske alger og cyanobakterier er en lovende kilde for fremtidig biobrensel, og har mange viktige fordeler over 1. og 2. generasjons biobrensel. *Synechococcus* sp. PCC7002 er den raskest voksende oksygen-produsrende fotoautotrof kjent i dag, og overlever i mange forskjellige miljøer. Den er derfor en veldig interessant organisme for produksjonen av biobrensel. Dette er første gang *Synechococcus* er blitt brukt på NTNU. For å utvikle denne organismen til å produsere hydrogen, bør den ineffektiv, endogen [NiFe]-hydrogenasen erstattes med en mer energisk fordelaktig og effektiv [FeFe]-hydrogenase. En kodonoptimalisert gen for å uttrykke en [FeFe]-hydrogenase ble designet. Et plasmid ble konstruert for å tillate enkel transformasjon og genuttryk av eksogene gener, inkludert den kodonoptimaliserte [FeFe]-hydrogenasen, i *Synechococcus*. Den inneholdt en sterk promotor, som er i stand til å drive genuttrykk av proteiner i et beslektet cyanobakterie til 15% av cellens tørrvekt. En midlertidig markørgen, *mRFP1*, ble brukt til å verifisere transformasjon. På grunn av tidsbegrensninger ble denne overekspresjons systemet ikke evaluert i *Synechococcus.*

I tillegg, ble et plasmid konstruert som kunne slette den opprinnelige [NiFe]-hydrogenasen i *Synechococcus*. Desverre ble ikke *Synechococcus* transformert med denne plasmiden. Videre, ved hjelp av genekspresjonsplasmidet, skulle et annet hydrogenase ble uttrykket i denne stammen,

Modifikasjoner på *Synechococcus* transformasjonsprotokollen blir diskutert, for å hjelpe fremtidige forskere ved NTNU.

. . .

"Any road followed precisely to its end leads precisely nowhere.

Climb the mountain just a little bit to test that it's a mountain.

From the top of the mountain, you cannot see the mountain."

—Frank Herbert, *Dune*

ACKNOWLEDGMENTS

This master's project was carried out at the Norwegian University of Science and Technology (NTNU), Faculty of Natural Sciences and Technology, in the Department of Biotechnology's PhotoSynLab.

I would firstly like to thank my supervisor, Associate Professor Martin Hohmann-Marriott, for helping me through this. His ideas, support, enthusiasm, and willingness to let me play around with my project, even after my multiple mistakes, helped keep me enjoying it. Giving me the oppurtunity to go to New Zealand for a semester to work on this project was just the icing on the cake.

Other people in Norway were instrumental to the success of this project. Jacob Lamb, a PhD student in Martin's lab, taught me many of the basic molecular biology lab skills I needed. I hope my myriad cloning failures were entertaining. I'd also like to thank Rahmi Lale, a researcher in Martin's lab, for helping me understand some of the finer points of cloning. The other master's students in the lab were great, so thank you for the companionship. And Azra, thanks for the *mRFP1* plasmid, it worked much better then GFP.

In New Zealand, I worked in Professor Julian Eaton-Rye's lab at the University of Otago. I had a wonderful time, and have to thank, first Julian letting me work in his lab, but also Dr. Simon Jackson and Jackie Daniels for all the help with lab work. The social environment in the lab was fantastic, thanks for making me feel welcome! I'd also like to thank lab-mate Julian Taffner for driving me around NZ, a task that was, until only very recently, entirely thankless.

And for keeping me sane outside of school, I'd like to thank my family (especially mom!) and my friends all over the world. You guys rock!

These wonderful people have helped me write through this master's project, but any mistakes left inside are mine.

v

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

PCR Polymerase Chain Reaction

GFP green fluorescent protein

NAD(P)H nicotinamide adenine dinucleotide (phosphate)

mRNA messenger RNA

tRNA transfer RNA

OD optical density

LB lysogeny broth

Part I

INTRODUCTION

SYNECHOCOCCUS AND BIOFUELS

1.1 BIOFUEL PRODUCTION

Renewable and sustainable energy sources are becoming increasingly important in the energy economy of the world, but as of 2013 87% of the world's energy needs are still met by fossil fuels (BP, 2014). In 2012, humans produced 32.7 gigatonnes of carbon dioxide compared to the approximately 12 gigatonnes of carbon dioxide that is removed from the atmosphere each year by natural processes (Bilanovic et al., 2008; EIA, 2015b). In addition to the effects of greenhouse gases on global warming, pollution from fossil fuels has other serious health effects. The World Health Organization estimated in 2012 that 3.7 million deaths worldwide were caused by elevated levels of pollution (WHO, 2014).

Renewable energy sources are the fastest growing sector of energy production, but are mainly focused on producing electricity (BP, 2014). Wind and solar power are the major renewable energy sources today, but biofuels are the main research focus for future renewable energy. Biofuels are derived from biologically produced or converted molecules, mainly from plants. Plant-derived biofuels are carbon neutral, utilizing the sun's light energy for growth and require little additional energy for harvesting, and can be grown by countries all over the globe (Wiebe et al., 2008). The biggest problem with biofuel production has been that they directly compete with food crops, either using food directly or replacing existing agricultural land. First generation biofuels, such as sugarcane and rapeseed oil (also known as canola oil) use the primary crop as the energy source. This has driven up prices for food, causing serious detrimental effects in the third world. Second generation biofuels focus on converting either the whole plant matter, or waste matter left after harvesting food

(Moore, 2008). While some second generation biofuels do not directly compete with food production, the conversion of plant waste into biofuels is not yet economically competitive. Even more importantly, due to the lower energy density in second generation biofuels, transportation costs become a serious drawback, limiting their usefulness (Richard, 2010). The only currently competitive biofuel without subsidies is Brazilian bioethanol production from sugarcane, and even when subsidies are provided, only a few companies in the U.S.A. find it economical (Wiebe et al., 2008). Clearly, biofuels are not currently a realistic alternative to fossil fuels, but with the advent of third generation biofuels, the break-even point is expected in the next decade (Wijffels and Barbosa, 2013).

Ideally, biofuels should be economically competitive with other forms of energy production, have minimal impact on food production, decrease greenhouse gas emissions, and require minimal usage of fresh water (Khosla, 2008). Third generation micro-algal biofuels hope to solve many of these problems more efficiently then previous generations. In the context of biofuel, "algae" refers to prokaryotic cyanobacteria (previously called blue-green algae) as well as microscopic eukaryotic algae, or more generally, to any photosynthetic microbe. Algae are more energy intensive to grow then traditional crops, and require carbon dioxide bubbling, mechanical mixing and higher fertilizer usage. As of 2010, algal biomass is about 500 times more expensive per kilogram of dried biomass compared to palm oil (Wijffels and Barbosa, 2013). Algae do not compete with food crops for land, since they can be grown on non-arable land, like deserts. In addition, algae grow a significantly higher amount of biomass on the same amount of land, and have a higher photosynthetic efficiency (Kebede and Ahlgren, 1996; Zittelli et al., 1996). For production of oils for biofuel, algae can produce more than 50 times more oil than coconut, a high yield crop, on the same area of land. Corn and soybean are more commonly used in the USA for the production of oils for biofuel (Wiebe et al., 2008; EIA, 2015a), and algae can outproduce them with a factor of more than 800 and 300, respectively (Chisti, 2007). Water usage for growing micro-algae per unit of dry weight is comparable to other crops. It is higher than sugarcane (Shih and Gascho, 1980), but much lower then, for example, rapeseed (Fengrui et al., 2000). The biggest advantage in water usage is that micro-algae do not need freshwater, but can grow on seawater or wastewaters from industry, agriculture or sewage (Hu et al., 2008). However, growing on wastewater could provide the required nutrients, reducing the need for fertilizers and clean water (Clarens et al., 2010). Growing on wastewater could see algae serving a dual purpose, bioremediation as well as biofuel production (Oswald, 1963). In fact, growing algae on source separated urine would make them more environmentally friendly then terrestrial crops (Clarens et al., 2010). Additional car-

bon dioxide could be provided to algae from flue gas, such as from coal plants, reducing the carbon impact of other industries while generating energy (Kadam, 2002). And lastly, algal biofuels would lower carbon emissions by replacing fossil fuels. Depending on the end product used, the biofuels could be either carbon neutral, or even carbon sequestering when producing hydrogen.

There are three main types of biofuels that could be created from algae: starch, oils, or hydrogen. Depending on the algae chosen, starch can be produced at up to 50% of dry cell weight (Dismukes et al., 2008), or oils at up to nearly 80% (Chisti, 2007). Different strains have different efficiencies under different conditions, and could be optimized for the end product chosen. There are two main ways to grow algal biomass, either in open ponds or enclosed bioreactors (Singha et al., 2010). Raceway ponds are the cheapest artificial system for growing algae, and consist of an artificial canal that loops on itself, normally with a paddlewheel to help with circulation and mixing. These systems are open to the air, and therefor also to contamination. To combat this, extremophilic algae are used with highly selective conditions, such as high salt content or alkalinity. Bioreactors can be used to lower levels of contaimnation and pollution, as they allow much more control over temperature, carbon dioxide, oxygen and nutrient levels, as well as more choice algal choice due to lower levels of contamination and pollution. Bioreactors also have much higher operating costs (Brennan and Owende, 2010). For starch or oil production, biomass should continuously harvested and dried before further processing. Harvesting and drying are both expensive processes (Bahadar and Khan, 2013). Alternatively, algae can be used to produce hydrogen gas. Hydrogen is a very concentrated fuel source for its weight, and when burned it forms water, emitting no greenhouse gases. As such, it has become a very interesting fuel . Unfortunately, the most common methods for creating hydrogen today involve using fossil fuels to generate the power to split water, nullifying its environmental advantage over other fuels. There is therefore considerable interest in producing hydrogen using biological processes (Show et al., 2012). These would have the added bonus of sequestering carbon from the atmosphere during growth. Using algae to produce hydrogen might be efficient enough to be economical, but once produced there are still many hurdles before hydrogen can be used as general energy carrier (Bossel, 2006).

1.2   PHOTOLYTIC BIOHYDROGEN PRODUCTION

There are three main methods for algal biohydrogen production, photolysis, photofermentation, and dark fermentation,

Using light, algae can either perform photolysis, the splitting of water into hydrogen and oxygen using energy derived from light, or photofermentation (Show et al., 2012). In photofermentation the electron donors can be simple organic acids or hydrogen sulfide instead of water, and in the absence of nitrogen, protons from the proton gradient used by ATP synthetase are the final electron acceptor (Akkermana et al., 2002). Algae can also use be used fermentively, without light, to create hydrogen. This requires energy in the form of carbohydrates or other organic compounds for fermentation, either from other energy crops or from waste. The waste from this process could be used by photofermenters and methanogens to increase the energy yield (Show et al., 2012). This method is closer to commercial realization, but still suffers from low yield and production rates in addition to the fact that it requires energy input in the form of organic matter (Hallenbeck and Ghosh, 2009). Photolysis could offer a better solution, requiring only water and sunlight as substrates, but still has many hurdles to overcome before it becomes commercially feasible. Current methods for producing hydrogen are inefficient, and lose energy in the conversion to hydrogen.

Many algae and bacteria can create hydrogen, either with the help of nitrogenases or hydrogenases. Nitrogenases produce hydrogen as a byproduct of nitrogen fixation, but at a 1000-fold lower production rate then hydrogenase-based systems, and are thus not very interesting for biohydrogen production (Bahadar and Khan, 2013). Hydrogenases catalyze the reaction

$2H^+ + e^- \rightleftharpoons 2H_2$,
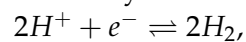
and can be bidirectional or have one direction dominate. The physiological role of the specific hydrogenase determines its directionality. They have diverse functions, and can work as a sensor for H_2 to regulate the activity of other enzymes, generate energy through the one-way "knallgas" reaction, function as apart of nitrogen uptake, or act as an electron dump in anaerobic respiration when other more efficient electron acceptors are not available (Vignais and Billoud, 2007).

There are three major families of hydrogenase enzymes: [NiFe]-, [FeFe]-, and [Fe]-only hydrogenases, named after the metal cofactors in the active site. [Fe]-only hydrogenases are found in methanogenic archaea, and do not produce hydrogen; rather they split hydrogen to form a hydride that is used in the methanogenic reduction of CO_2 (Shima and Thauer, 2006). Additionally, they only seem to be required for growth in nickel limited conditions, when the archaea can

no longer produce [NiFe]-hydrogenases. [NiFe]- and [FeFe]-H_2ases, while phylogenetically unrelated, show significant similarities, such as structure, mechanism, and the usage of the two very unusual ligands carbon monoxide and cyanide at the active site. [NiFe]-hydrogenases are found in all major phyla of bacteria and archaea, while the [FeFe]-hydrogenases are found in some microbial eukaryotes and a few classes of bacteria (Vignais and Billoud, 2007; Meyer, 2007).

[FeFe]-hydrogenases typically mediate the production of hydrogen, and are irreversibly inhibited by the presence of oxygen. It is estimated that [FeFe]-hydrogenases can catalyze the production of hydrogen at rates of 20000 s^{-1} per enzyme (Madden et al., 2012). [FeFe]-hydrogenases show a degree of modularity, with some additional domains in some species that function as electron transfer mediators for various electron donors/acceptors, or binding sites to help localize the protein inside the cell. Chlorophycean algae, such as *Chlamydomonas reinhardtii* (henceforth *C. reinhardtii*), have the simplest [FeFe]-hydrogenases, which contain only the basic H-domain (Mulder et al., 2011).

The H-domain is found in all of the [FeFe]-hydrogenases, and contains the catalytic site, the H-cluster, located in the center of the protein. This cluster contains two subclusters, a [4Fe4S] cubane, and unique, catalytic [2Fe2S] moiety. The two CN^- and two or three CO ligands are found bound to this latter subcluster. The CN^- ligands help increase the alkalinity as well as fine-tune the redox and electronic properties of the site. The CO ligands help stabilize the Fe atoms at low oxidation states. This centrally located H-cluster interfaces with "tunnels" in the protein that allow gas molecules to diffuse into the protein (Winkler et al., 2013).

In contrast, [NiFe]-hydrogenases are commonly bidirectional, generally more active in uptake of hydrogen, and, either O_2-tolerant, or only reversibly inhibited by the presence of oxygen (Ghirardi et al., 2007). The [NiFe]-hydrogenases that are entirely O_2-insensitive are solely uptake enzymes (Schäfer et al., 2013). [NiFe]-hydrogenases are always comprised of many sub-units, with 2 sub-units forming the catalytic heterodimer found in all [NiFe]-hydrogenases. The larger sub-unit contains the [NiFe] active site, and the smaller subunit has three Fe-S clusters which ferry electrons between the active site and other electron donors and acceptors, which are usually flavodoxin or ferredoxin (Gutekunst et al., 2014). Gas diffuses into the enzymatic center via diffusion through protein "tunnels", much as in [FeFe]-hydrogenases. [NiFe]-hydrogenases also use the unique ligands CN^- and CO ligands, bound to the Fe-atom at the active site.

There are 5 families of [NiFe]-hydrogenases. Group 1 contains membrane-bound uptake hydrogenases, which use hydrogen oxidation for respiration, using electron acceptors like nitrate, sulphate, fumarate, and CO_2 or oxygen during aerobic conditions. Group 2 is

split into two subgroups of cytoplasmic hydrogenases; Group 2a includes the cyanobacterial uptake hydrogenases, which are important during nitrogen fixation, and Group 2b contains the H_2-sensor hydrogenases, which up-regulate hydrogenases in certain H_2-oxidizing bacteria in the presence of hydrogen. Group 3 consists of bidirectional cytoplasmic hydrogenases, found in many bacteria and methanogenic Archaea. These bidirectional hydrogenases are theorized to function as electron valves to get rid of excess reducing equivalents. The cyanobacterial hydrogenases are members of this group. Group 4 [NiFe]-hydrogenases are the hydrogen-evolving, energy conserving, membrane-associated hydrogenases, and are found mainly in Archaea. They dispose of the electrons from the fermentation of single carbon compounds, like CO and formate (Vignais and Billoud, 2007). Group 5 hydrogenases are the most recently classified, and are unique in that they oxidize atmospheric levels of hydrogen for energy, using the "knallgas reaction" and are oxygen insensitive (Constant et al., 2011). Species often contain multiple hydrogenases from different families, or even within a family. This has been theorized to grant improved metabolic flexibility, separating hydrogen evolution, consumption, and regulation (Berney et al., 2014).

The photolysis that creates protons used for hydrogen evolution also form oxygen that inhibits the hydrogenase. This inhibition prevents sustained hydrogen production using water as the electron source. There are two popular solutions to this problem, either separating the photolysis from hydrogen production (either spatially or temporally) or finding/bioengineering oxygen tolerant species of hydrogenase. Spatial separation requires the hydrogenase to be in an anaerobic environment, and a system to transport electrons to the enzyme. Some cyanobacteria naturally form heterocysts, anaerobic cells dedicated to nitrogen fixation, that receive organic compounds from neighboring vegetative cells. However, heterocystous systems suffer from low efficiencies, and will probably not be commercially viable (Hallenbeck and Benemann, 2002). A temporal separation has been demonstrated in *C. reinhardtii*, where a photosynthetic, oxygen-evolving growth phase alternates with an anaerobic consumption and hydrogen production stage, also in light. Sulfur deprivation is the stress factor that triggers the change to hydrogen evolution. It does this by reversibly inhibiting photosystem II and oxygen evolution; respiration then depletes the oxygen levels over the course of approximately 1 day, creating an anaerobic environment. In these conditions, the hydrogenase functions to remove the excess electrons (Melis et al., 2000). Using this method in *C. reinhardtii* is currently the best method for producing biohydrogen (Esquível et al., 2011). Another temporal separation has been demonstrated, where photosynthesis creates organic compounds in the light, and organisms ferment these compounds in the dark, creating hydrogen. This has been shown with green algae, and

in a mixture of two different bacteria that split photosynthesis and hydrogen production between them (Miura et al., 1986, 1992).

1.3 *Synechococcus* sp. pcc7002

Synechococcus sp. PCC 7002, previously known as *Agmenellum quadruplicatum* strain PR-6; American Type Culture Collection strain 27167, (henceforth just *Synechococcus)* is considered a prime candidate for biofuel production. It is a cyanobacteria, making it simpler to work with than plants or eukaryotic algae. *Synechococcus* is often used as a model organism for research into photosynthesis, as the photosynthetic apparatus in cyanobacteria is very similar to that in plants. It is thought that the ancestors of today's cyanobacteria were involved in the endosymbiotic that led to photosynthetic plants; and are in fact the precursors to the chloroplast of today's plants (Giovannoni et al., 1988). Cyanobacteria also naturally accumulate organic compounds such as carbohydrates and lipids that could be used for biofuels.

Synechococcus has many advantages over other cyanobacteria and algae: fast accumulation of biomass, tolerance of a large range of temperatures, salt concentrations and light intensities, and its ability to grow photoautotrophically, mixotrophically, or heterotrophically (Sakamoto and Byant, 2002; Baalen et al., 1971; Batterton and Van Baalen, 1971; Ingram et al., 1973). Its doubling time of less then 2.6 hours in optimal conditions is the fastest recorded of any cyanobacteria. *Synechoccocus* is typically found in tidal zones and estuaries, where conditions can vary drastically and quickly, leading to its great adaptability. *Synechococcus* can grow at light intensities up to 5000 μmol photons m^{-2}s^{-1}, or more than twice the light intensity of sunlight. It is saturated at about 250 μmol photons m^{-2}s^{-1}. The organism can grow heterotrophically, and can use glycogen as the sole carbon source. Many cyanobacteria, like *Synechocystis*, must use glucose instead, which is more expensive. *Synechococcus* must be grown in medium containing glycogen for a few days to make a glycogen resistant strain before it can be used. *Synechococcus* is naturally transformable, and alongside its rapid growth rate, this enables researchers to rapidly make new mutant strains, and grow them in sufficient quantities to study (Stevens and Porter, 1980).

Synechococcus has a group 3 [NiFe]-hydrogenase, a bidirectional cytoplasmic hydrogenase. Similar hydrogenases are found in many cyanobacteria. The physiological function of these hydrogenases in cyanobacteria has been uncertain, as knock-out strains grow equally

well as wildtype under most conditions tested, leading to the theory that they might have had importance far earlier in the evolutionary history of cyanobacteria, but have since lost most of their usefulness (Eckert et al., 2012). Hydrogenases are still present in many cyanobacteria, and regularly produced, but only enzymatically active for 60-90s in normal aerobic conditions before becoming inhibited by oxygen. A recent study was the first to show a strong phenotype for a hydrogenase-lacking strain in a cyanobacteria. When growing *Synechocystis* under mixotrophic, nitrate-limiting conditions, the hydrogenase became essential for growth (Gutekunst et al., 2014). Under these conditions, the two main electron sinks are unavailable: the Calvin cycle and nitrate reduction. The presence of glucose downregulates Rubisco, slowing down the Calvin cycle considerably, which accepts electrons from nicotinamide adenine dinucleotide (phosphate) (NAD(P)H). In addition, nitrate can also accept electrons, and limiting nitrate reduction has been shown to increase hydrogen production in *Synechocystis* (Gutthann et al., 2007). It is therefore reasonable to assume that the hydrogenase functions to help rid the cell of excess electrons, especially when other electron sinks are not available. This environment could occur naturally when phytoplankton blooms deplete nitrate and oxygen levels while increasing the concentration of simple sugars (Wetz and Wheeler, 2004; Teeling et al., 2012).

The hydrogenase found in *Synechococcus* is a heteropentamer formed from the subunits HoxEFUYH. HoxH and Y are the large and small subunits of the hydrogenase moiety found in all [NiFe]-hydrogenases. HoxH contains the active site. HoxEFU are the diaphorase part, which, in addition to HoxY, contain the [FeS] clusters that allow electron transport between the active site and flavodoxin and ferredoxin (Eckert et al., 2012; Gutekunst et al., 2014).

In addition to the hydrogenase, there are a number of maturation factors, HypA through F, and HoxW. HypE and HypF catalyze the creation of the cyanide group, which associates with Fe on the HypC-HydD complex. The Fe and its ligands are inserted into HoxH by HypCD. HypA and B deliver nickel into the active site, and HoxW cleaves off the C-terminal extension of HoxH, allowing HoxH and HoxY to collapse around the active site, forming the active hydrogenase.

Synechococcus has already proved itself as a possible biofuel producer, both in production of carbohydrates (Aikawa et al., 2014) and lipids (Ruffing, 2014). Its ability to grow in a wide range of conditions, fast growth rate, and its transformability make it a promising organism for the production of biofuels. Its hydrogen producing capabilities have not been extensively studied, but as the platform is developed, it might even be able to rival the H_2-producing capabilities of *C. reinhardtii.*

1.4 OBJECTIVE 1: DESIGNING AND CREATING A GENE EXPRESSION CASSETTE

To be able to study *Synechococcus* and its use in biohydrogen and other biofuel production, an efficient method for inserting genes into the bacteria was developed in this thesis. The designed gene expression cassette should take advantage of *Synechococcus'* natural transformability, and insert itself into a neutral site in the genome. It should contain an antibiotic marker to allow for selection, and a strong promoter to allow for the constitutive expression of an inserted gene. A gene encoding a fluorescent protein was initially chosen to allow for quick validation of the expression cassette. However, the expression cassette was designed to allow the expression of any protein, but specifically an exogenous hydrogenase.

1.5 OBJECTIVE 2: CREATING THE $\triangle hoxH$ *Synechococcus* STRAIN

A strain without hydrogenase activity was designed, where the native *Synechococcus hoxH* gene was replaced by an antibiotic resistance cassette. The strain without native *Synechococcus* hydrogenase will be used in the future to house the expression cassette for a non-native [FeFe]-synthetic hydrogenase. The hydrogen production of this new strain would be entirely due to the new hydrogenase, and methods to optimize its hydrogen production could be tested.

Part II

HYDROGENASE KNOCK OUT IN
SYNECHOCOCCUS SP. PCC 7002

2

INTRODUCTION

2.1 EXPRESSION SYSTEMS

Expression systems allow for the production of specific proteins. These can be proteins that have been modified, are highly expressed, come from a different organism, or can even be entirely synthetic. The most commonly used expression system is *Escherichia coli* (henceforth: *E. coli).* There are many well-described methods for cloning in *E. coli*, and as a simple bacteria it is often the first choice for many protein expression studies (Sodoyer, 2004).

The choice of expression strain depends on many factors. *E. Coli*, although easily manipulated, is a simple bacteria that has limited capabilities for messenger RNA (mRNA) modification, complex folding or the post-transcriptional modification required for maturation of many eukaryotic proteins (Sodoyer, 2004). Eukaryotes often have mRNA which contains unexpressed sections, called introns, and many proteins require the use of additional maturation factors after translation for correct folding, cleavage, glycolysation, etc. In some cases, proteins and maturation factors can be expressed simultaneously to increase the capabilities of the host cell, it is often easier to choose a different expression system. Eukaryotic proteins often have post-translational modification that requires more complicated machinery than exists in bacteria, and it is common to use eukaryotic cells, like *Saccharomyces cerevisiae*, insect, or mammalian cells to produce these proteins. However, eukaryotic systems increase complexity and cost *(Sodoyer, 2004).* For the production of pharmaceuticals or food components it is often necessary to use a bacteria that is non-pathogenic and does not contain the highly-inflammatory lipopolysaccharide wall that some bacterias, like *E. coli*, have. For example, *Corynebacterium glutamicum* is used for the industrial production of amino acids, and

Figure 1: Planned insert for the gene expression cassette.

has been shown to be industrially useful for the production of human epidermal growth factor (Kimura, 2003; Date et al., 2006).

There are many important factors in choosing which organism to use for protein expression. Ease of transformation, as well as the amount of modification necessary from the original strain, are often the most important factors. However, for some applications, other considerations can become more important: is the product poisonous for this organism? Can it secrete it into the medium, or does it have to be extracted? How efficient is the system?

Cyanobacteria are not commonly used for industrial expression of proteins, but are currently used in the production of some high-value food products, like caretenoids and ω-3 fatty acids (Wijffels and Barbosa, 2013). Cyanobacteria have some advantages over commonly used bacteria, but are very expensive to grow. As photosynthetic organisms, they require little input of chemical energy for growth, and can instead use sunlight as their main energy source. And, as prokaryotes, they can be much simpler to optimize than plants for biomass accumulation. As described earlier, *Synechococcus* is often used for research into biofuel production, where it can be modified to produce high levels of organic molecules, like carbohydrates or lipids (Ruffing, 2014; Aikawa et al., 2014). Using light as their energy source means that, hopefully, humans can take advantage of the organisms to store solar energy, and give us a net increase in energy.

The creation of an effective expression vector for *Synechococcus* could make research using this organism simpler. For the over-expression cassette to be successful, it must transform *Synechococcus*, strongly promote production of a specific protein, and contain a selection marker. Figure 1 shows features of such an over-expression vector. Flanking regions would allow for integration into the genome of *Synechococcus*, and the promoter and terminator would allow for the production of the gene product from the inserted gene. The marker is in the opposite orientation so that it can still be expressed and located downstream of the terminator without the terminator stoppings its transcription.

By taking advantage of neutral sites in the endogenous plasmids found in *Synechococcus*, a powerful method for inserting genes has previously been designed (Xu et al., 2011). The smallest plasmid, pAQ1, contains a neutral site that can be used for insertion of foreign DNA. A linear fragment of DNA must be produced, with the

sequence of interest between two flanking regions. These flanking regions allow for integration into pAQ1, replacing about 1.2 kbp of nonessential sequence, and increase the size of pAQ1 from 4809 bp to 6980 bp. As pAQ1 has the highest copy number of the plasmids naturally found in *Synechococcus*, this means that after transformation, the new strain might contain multiple copies of the inserted DNA, which could also lead to increased expression.

To drive protein production, a promoter and terminator had to be chosen. *E. coli* promoters do not usually result in high levels of mRNA in cyanobacteria, but there are cyanobacterial promoters that work well. In the 560bp immediately upstream of The phycocyanin subunit B, *cpcB*, native to *Synechocystis* sp PCC 6803 (henceforth *Synechocystis)* there are 2 promoters and 14 predicted transcription factor binding sites. This promoter region, named P_{cpc560}, has been shown to be very effective in *Synechocystis* (Zhou et al., 2014). Only the one promoter closest to *cpcB* has been used in *Synechococcus* before, but it is an effective promoter (Xu et al., 2011). Its regulation appears to be modulated by light, where low light increases mRNA production. Nitrogen levels also affects promoter strength; nitrogen starvation stops expression of downstream genes (Gasparich et al., 1987).

For the expression cassette, a well-described terminator was chosen from *Synechocystis.* This terminator is immediately downstream of *rbcL*, which encodes the Rubisco large sub-unit. Named T_{rbc}, it has been successfully used before, in concert with P_{cpc560}, to drive protein production in *Synechocystis.*

The final design can be seen in Figure 2.

2.2 FLUORESCENT PROTEINS

Fluorescent proteins are naturally occurring proteins that are very useful for molecular biologists. Their fluorescence allows them to be easily located and quantified. The green fluorescent protein (GFP) isolated from the jellyfish *Aequorea victoria* was the first commonly used fluorescent protein (Tsien, 1998). Fluorescent proteins are commonly used as expression markers or as a proof-of-concept marker, to easily show if a genetic manipulation has been successful. When placed immediately downstream of a gene, without a stop codon inbetween, this gene and the GFP will produice a single mRNA transcript. After translation, the GFP will be attached to the protein encoded by that gene. This allows the GFP to be used to localize and quantify the concentration of the protein of interest inside the cell. This ability
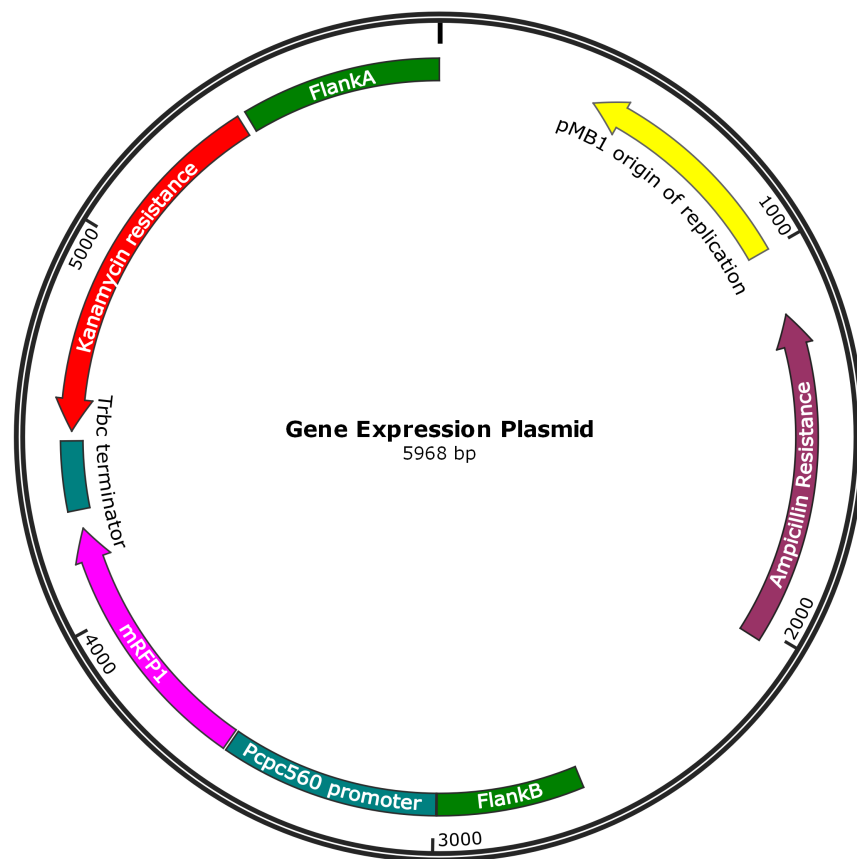
Figure 2: Design for the *Synechococcus* gene expression plasmid.

is dependent on a couple of factors. If the C-terminus is embedded inside the protein, the GFP will not allow proper folding, or if there is post-translational modification of the poly-peptide chain that involves cleaving close to the C-terminus, the GFP will no longer be associated with the protein. GFP can also be used as a "folding reporter" in fusion proteins. The chromophore of GFP is only active if the protein has folded properly, and can be useful for showing if expression system is capable of folding the desired protein correctly (Waldo et al., 1999). The expression of GFP can also be used to quantify various qualities about the expression system, like promotion strength and copies of a gene. GFP expression is proportional to the amount of mRNA transcripts, which is proportional to the number of copies of the GFP gene in the cell (Soboleski et al., 2005).

Other organisms also contain fluorescent proteins, allowing for even more colors. A red fluorescent protein from *Discosoma* sp., named DsRed, has been the modified for use in molecular biology. Originally a tetrameric protein, it was modified to become stable as a monomeric protein. A total of 33 amino acid substitutions were required, both random and directed, to first make the protein form stable monomers, then to recreate the fluorophore, which was lost during this process (Campbell et al., 2002). The resulting monomeric red fluorescent protein, mRFP1, has a couple advantages over GFP. It has a maturation time of about 20 minutes, instead of the 6 hours required by GFP, and does not require the presence of oxygen for its maturation. *mRFP1* has since been the subject of even more mutations, forming several variants with colors from yellow through pink to purple, colors which have not been sucessfully formed with GFP mutants (Shaner et al., 2004).

The fluorescent protein mRFP1 was chosen as a marker for successful transformation. It is a red protein, hopefully allowing it to be seen with the naked eye in both *E. coli* colonies, and in *Synechococcus*.

<div style="text-align: right; font-size: 3em;">3</div>

MATERIALS AND METHODS

3.1 EXPERIMENTAL PROCEDURES

General protocols used throughout this experiment.

Polymerase Chain Reaction (PCR) - For isolating and amplifying DNA fragments, as well as overlap extension PCR. The PCR protocol outlined for Phusion® High-Fidelity DNA Polymerase (New England Biolabs, USA) followed. Reaction mixtures can be seen in Table 1, and thermocycling conditions are shown in Table 2. The thermocycler used was a C1000 Touch® thermal cycler (Bio-Rad Laboratories, USA). For purification of PCR products, a QIAGEN purification kit was used (Qiagen, Germany). Annealing temperatures were estimated using NEB's $T_m$calculator, but were optimized using an annealing temperature gradient were necessary to eliminate non-specific products caused by incorrect primer binding. Primers used can be found in Table 7 in the appendix.

Gel Electrophoresis - used to verify PCR reactions and restriction digests. Results show length of fragments, as compared to a DNA ladder, as well as give a rough estimate of concentration. 5 µL of PCR product was mixed with 5 µL ddH$_2$O and 1 µL 10x loading dye (containing 0.25% bromophenol blue and 0.25% xylene cyanol FF) before loading. GeneRuler 1kb from Thermo Scientific was used as a reference. Samples were loaded into 0.8% (w/v) agarose gels in TAE buffer (a pH 8 tris-acetate buffer containing EDTA), and 100V was applied for 30-60 minutes for separation of DNA fragments based on size. The DNA fragments were visualized on a ChemiDoc XRS+ Imaging System (Bio-Rad, USA).

BioBricks - Originally developed by Tom Knight, BioBricks are short, standardized flanking regions that contain restriction sites. Their use allows for easy restriction cloning and ligating together of BioBrick

Table 1: Reaction mixtures for 100µL PCR reaction.

COMPONENT	VOLUME (µL)	FINAL CONCENTRATION
ddH$_2$O, autoclaved	64	-
5x Phusion HF buffer	20	1x
10µM dNTP	5	0.5 µM
10µM forward primer	5	0.5 µM
10µM reverse primer	5	0.5 µM
template DNA	~1	-
Phusion® DNA Polymerase	1	2 units / 100 µL

(NEB, 2015a)

Table 2: Thermocycling protocol for Phusion HF polymerase.

STEP	CYCLES	TEMPERATURE (^{o}C)	TIME (s)
Initial Denaturation	1	98	30 [1]
Denaturation		98	5-10
Annealing	25-30	45-72	10-30
Extension		72	15 s/kbp [2]
Final Extension	1	72	5 min
Hold	1	4	-

[1]For colony PCR (where the template comes from a bacterial colony) this should be extended to 10 min.
[2]15s per kb is sufficient for plasmid DNA, genomic DNA requires roughly twice that time.
(NEB, 2015a)
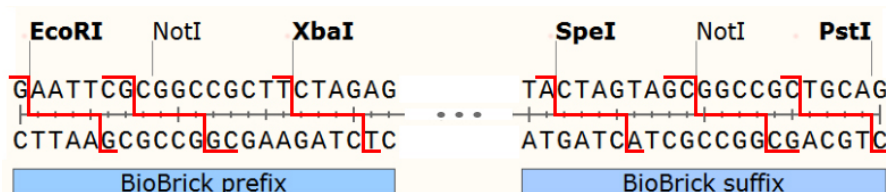
Figure 3: The BioBrick prefix and suffix.

parts. BioBrick cloning creates new pieces that can be further manipulated in the same fashion. Figure 3 shows the BioBrick prefix and suffix (Knight, 2003). In this experiment BioBricks were used on marker genes, like the antibiotic resistance cassettes. They were added to the DNA fragment during PCR; forward primers were designed that started with the BioBrick prefix followed by the part complementary to the gene in question, and likewise for the reverse primers.

Making competent cells - Table 3 shows media and antibiotic concentrations used for growing *E. coli*, while Table 4 shows the reagents used for making competent *E. coli*. Competent *E. coli* cells were made by inoculating 10 mL lysogeny broth (LB) with a single colony of *E. coli* DH5α, incubated overnight at 37°C with agitation. 1 mL of this culture was transferred to 100 mL ψB media at 37°C and grown, with shaking, until the optical density (OD) at 600 nm is between 0.3-0.4. The culture is then incubated on ice for 5 minutes, before being transferred to 2 cold 50 mL falcon tubes and centrifuged at 4000 rpm at 4°C on the Centrifuge 5424 (Eppendorf, Germany). The supernatant was discarded, and the pellet was resuspended in 15 mL cold TfBI before being centrifuged for 10 min at 4°C. Again, the supernatant was discarded, and then the pellet was carefully resuspended in 2 mL chilled TfBII using a 1000 μL pipette. This culture was divided into 200 μL aliquots in sterile microcentrifuge tubes and snap-frozen using either dry ice and ethanol or liquid nitrogen. The competent cells were stored at -80°C.

Ligation and Transformation - Competent cells were thawed on ice for 5-60 minutes, and simultaneously the DNA insert was ligated into plasmid backbones. Ligation was done using the Quick Ligation kit, following NEB's protocol but making only 10 μL ligation mixtures (NEB, 2015b). The entire ligation mixture was then added to 100 μL competent cells. An additional positive control, using 1 ng of a standard plasmid, like pUC19, was also transformed simultaneously. The tubes were gently mixed and incubated on ice for 30 minutes. The cells were then transferred to a water bath; for either 2 min at 37°C, or 45 s at 42°C, then transferred back to ice for 3 minutes. To each microcentrifuge tube, 1 mL of LB was added, and the cells were incubated at 37°C, with shaking, for 90 minutes. The cells were then plated onto selective LB plates, using 50 μL of culture. The microcentrifuge tubes were centrifuged for 3 minutes at 12000 x g, most of the

Table 3: Media and antibiotic concentrations for growing *E. coli* and selecting transformants.

	CONCENTRATION	COMPONENT	
LB MEDIA	(g/L)		
	10	bactotryptone	
	5	yeast extract	
	10	NaCl	
	15	agar	(for plates only)
ANTIBIOTICS [1]	(mg/L)		
	50	ampicillin	
	50	kanamycin	
	50	spectinomycin	
	30	chloramphenicol	

[1]Before addition of antibiotics, media was cooled to 55ºC

supernatant was then decanted, and the cell pellet was resuspended in the remaining LB. This was plated onto an additional selective LB plate. This plate contains a higher concentration of colony forming units, in case the first plate has too few.

Plasmid isolation - Overnight cultures of single transformed colonies were prepared, containing approximately 5 mL of LB and the appropriate antibiotic(s). Plasmids were isolated using the Wizard® *Plus* SV Minipreps DNA Purification Kit (Promega, USA) according to protocol.

Table 4: Media and solutions for transforming *E. coli.*

COMPONENT	CONCENTRATION	
ψB MEDIA	(g/L)	
bactotryptone	20	
yeast extract	5	
KCl	0.76	
KOH	until pH 7.6	autoclave, then add:
sterile 1 M MgSO$_4$	17 mL	
TFBI		
potassium acetate	2.94	
MnCl$_2$	9.80	
RbCl$_2$	12.10	
CaCl$_2$	1.48	
glycerol	150 mL	
0.2 M acetic acid	until pH 5.8	
TFBII		
0.1 M MOPS (pH 7)	10 mL	
CaCl$_2$	2.20	
RbCl$_2$	0.24	
glycerol	30 mL	

3.2   CREATING THE GENE EXPRESSION CASSETTE

The expression cassette was created from DNA from a variety of different sources. PCR amplification was used to isolate and amplify these fragments from their respective sources. The primer sequences and source organism or plasmid can be seen in the appendix, Table 7. Primers were designed to overlap with other fragments, so the pieces could be assembled into initially 3 "inserts" using extension PCR. Assembly was done by blunt ligation of the first part into a pUC19 plasmid, then three successive steps of digestion and ligation to insert the next part. An overview of all these restriction and ligation steps can be seen in Figure 7. The figures shown were generated using SnapGene software (GSL Biotech, USA). Each ligation was followed by transformation into *E. coli*, then plasmid isolation and purification.

For overlap extension PCR, first the individual DNA fragments were amplified as described earlier. The flanking regions were amplified with *Synechococcus* as template, and the promoter and terminator with *Synechocystis*. The three antibiotic resistance cassettes containing BioBrick prefix and suffix were amplified from mutant strains of *Synechocystis* carrying the respective resistance genes. Chloramphenical resistance is encoded by *cat*, producing chloramphenicol acetyltransferase. Kanamycin resistance comes from *aph*, for aminoglycoside-3'-O-phosphotransferase, and spectinomycin resistance comes from the aminoglycoside resistance protein, encoded by *aadA1* (Poteete et al., 2006).

To assemble the 2 inserts shown in Figure 4, overlap extension PCR is used. This is a slight modification of the previously described protocol; The outermost primers are used (for insert 1, this would be FlankB_fw and FlankA_rv) and for template ~1μL of each sub-piece is added (for insert 1, this would be flank B, flank A, and the antibiotic resistance cassette.) The cycling protocol is identical to the one shown in Table 2, with elongation times long enough to cover all 3 fragments sequentially. The *mRFP1* gene came as a BioBrick part, named BBa_E1010, found on the plasmid shown in Figure 5.

A slightly modified pUC19 was used as the plasmid backbone. The multiple cloning site was removed to allow for more restriction enzyme choice in the cloning procedure, and two SwaI sites were added. This plasmid backbone was amplified using PCR, creating a linear fragment, and also adding the restrictions sites to each end. The primers used had phosphate groups on their 5'-ends to increase ligation efficiency. The pUC19 plasmid infers resistance to ampicillin through the *bla* gene.

Insert 1 was blunt-ligated into the modified pUC19, then used to transform *E. coli.* This blunt ligation has no directionality, however,

(a) Insert 1 with the kanamycin resistance cassette.

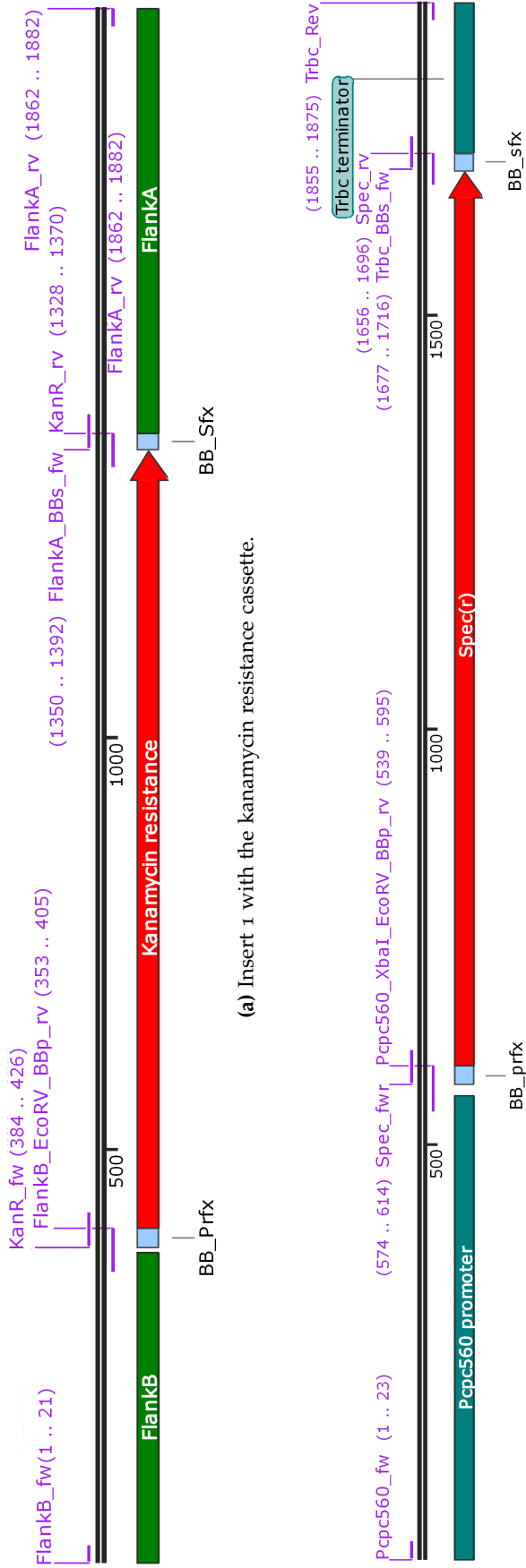**(b)** Insert 2 with the spectinomycin resistance cassette.

Figure 4: Inserts for creation of the *Synechococcus* gene expression plasmid.
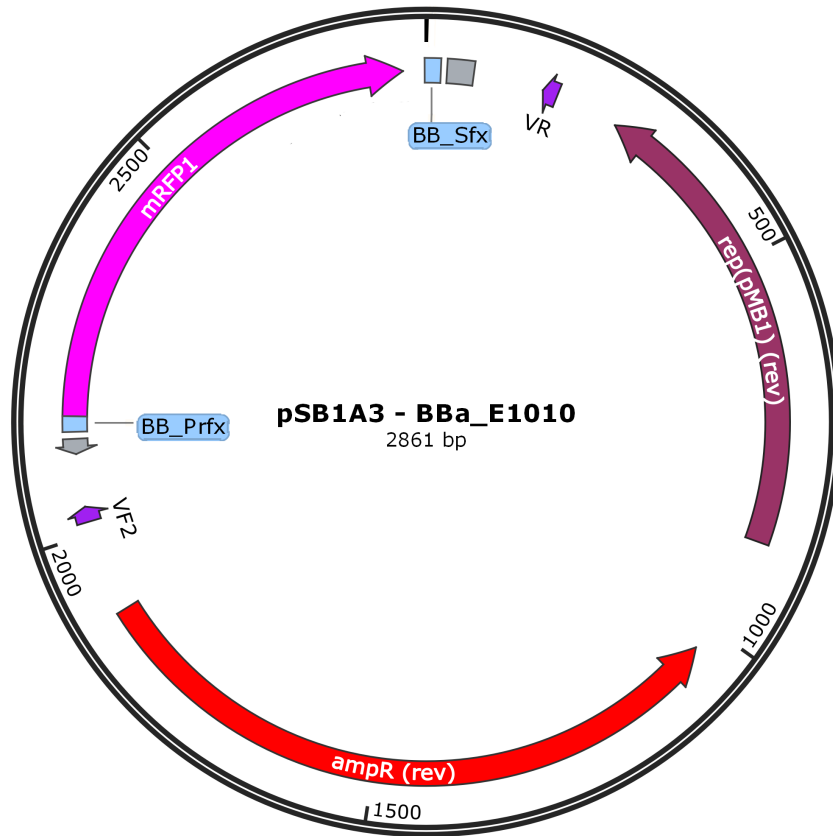
Figure 5: pSB1A3-BBa_E1010, the plasmid containing the *mRFP1* gene as a BioBrick part.
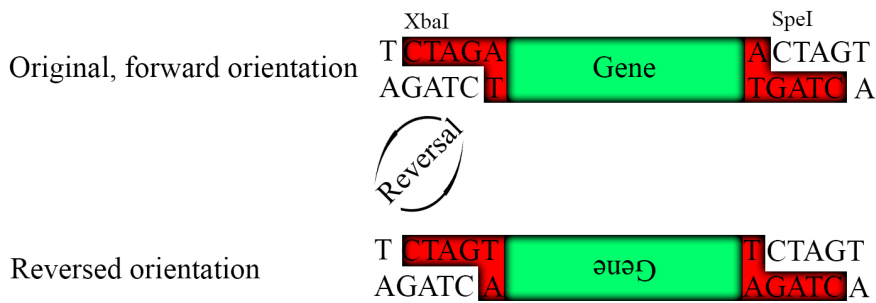
Figure 6: Reversing a gene using BioBricks.

the final DNA fragment used for transformation will not be affected because the pUC19 backbone will not be used to transform *Synechococcus*. From the transformed *E. coli*, single colonies were picked, and cultured overnight. Plasmids were isolated from these cultures, and they were in turn verified by restriction enzyme digestion and gel electrophoresis. The antibiotic cassette then had to be reversed. This both removes two digestion sites in the BioBrick parts, which are needed for later steps, and also ensures that translation of *aph* is not inhibited by the terminator, which would otherwise have been immediately upstream of the antibiotic resistance gene in the final vector and stopped its transcription. The antibiotic resistance chosen here remains throughout the cloning, becoming eventually the only antibiotic resistance marker left for transformation into *Synechococcus*. The reversal is done by digesting the isolated plasmid with XbaI and SpeI, then re-ligating and transforming. After transformation, about 50% of the colonies will have the reversed version, while the other half will still retain the original orientation, as shown in 6. The restriction enzymes will no longer recognize these pieces of DNA, since the innermost base pair will be different. Sequencing was used to verify that only plasmids with the correctly reversed antibiotic cassette were used.

The resulting plasmid was then digested with EcoRV, opening it so that insert 2, containing the promoter, terminator, and an additional resistance cassette, could be inserted. This was a blunt ligation, meaning the insert could be ligated in either orientation. This should have no effect on the final vector's efficacy, and for simplicity's sake all figures show it in the same orientation. This new plasmid contains the "super-promoter," P_{cpc560}, which might be mutated by *E. coli* after transformation. Several transformed colonies were picked, sequenced, and aligned with the original promoter sequence. Only plasmids containing the correct sequence were used, keeping the promoter intact for transformation into *Synechococcus*. This new plasmid was named plasmid 2.

Plasmid 2 was then digested with XbaI and SpeI, the two fragments were separated using gel electrophoresis, and the largest fragment
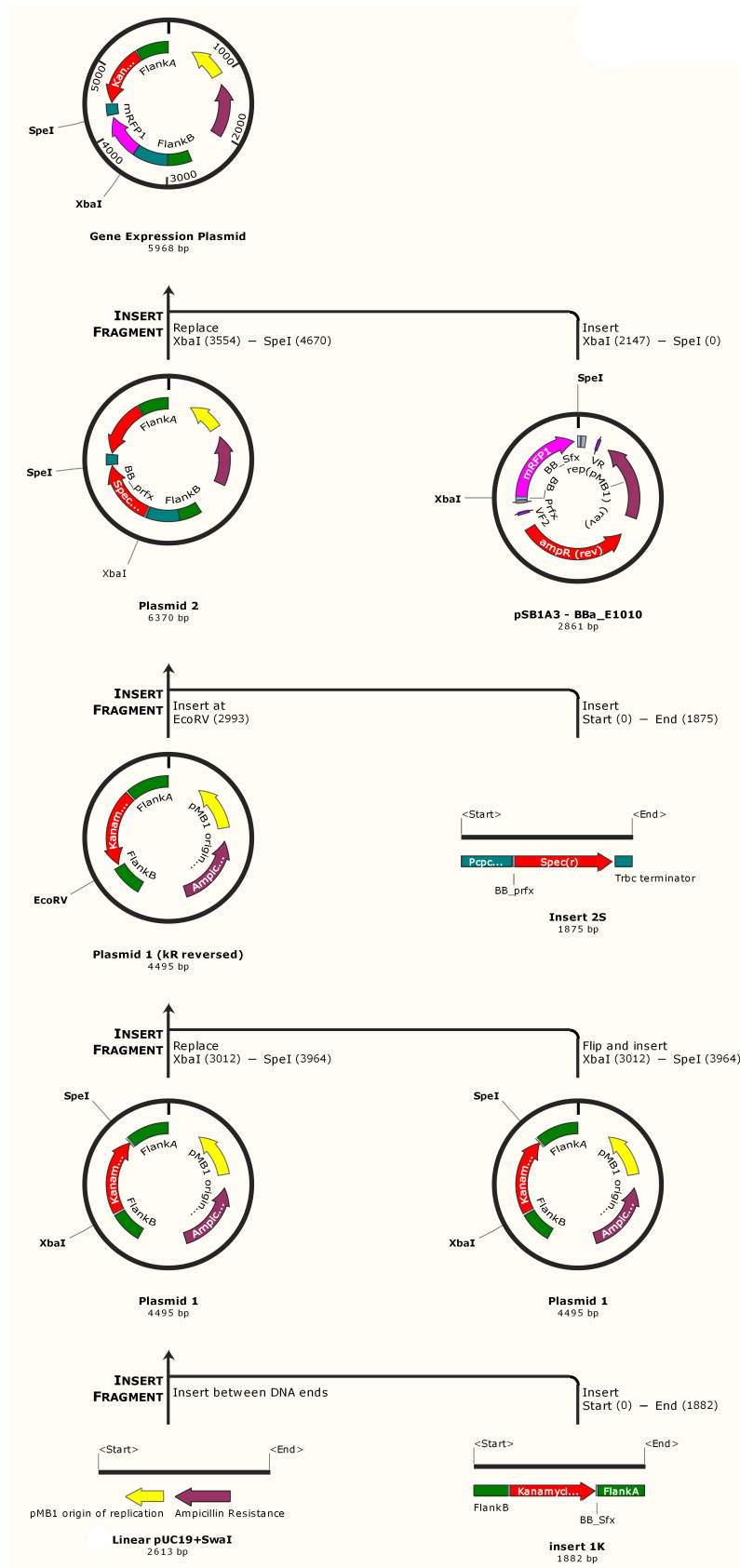
Figure 7: Cloning procedure for creating the *Synechococcus* gene expression cassette.

was isolated and purified. The plasmid containing *mRFP1* was also digested with SpeI and XbaI, and gel electrophoresis was used to isolate it from its plasmid backbone before purification. These two DNA fragments were ligated together, and used to transform *E. coli*. Successful transformants were red, because of the production of mRFP1. Some of these colonies were picked for overnight cultures, to isolate the plasmid.

XbaI and SpeI form compatible sticky ends, so it was necessary to see if *mRFP1* was in the correct orientation. A reversal would block recognition by both XbaI and SpeI, while a ligation in the correct direction would leave those restriction sites. Digesting a sample of either linear or plasmid DNA with XbaI and SpeI and separating the resultant DNA fragments on a gel would show the orientation of *mRFP1*: 2 fragments would be positive, resulting from active restriction sites, while a single fragment would be negative because SpeI and XbaI were unable to recognize the sequence.

To obtain the linear DNA fragment required for transformation into *Synechococcus* there are two strategies. The modified pUC19 contains SwaI sites on either side of the insert, allowing for easy digestion, separation, and purification of the insert. This option leaves four additional bases on either side of the flanking regions. Alternatively, the insert could be PCR amplified from the constructed plasmid, using the FlankB_fw and FlankA_rv primers. Either way, the resulting fragment should look like that seen in Figure 8, and be ready for transforming *Synechococcus*.

3.3 TRANSFORMATION OF *Synechococcus*

Synechococcus is naturally transformable. The linearized insert from the *hoxH* deletion plasmid was used for transformation. The following is a slightly modified version of the protocol outlined by Xu, et al. 2011.

Synechococcus was grown in modified BG-11 media, named mBG-11, the composition of which is shown in Table 5. A culture was bubbled with sterile air, incubated at 30°C with enough light to saturate (>250 μmol photons $s^{-1}m^{-2}$,) and allowed to grow overnight. The culture was concentrated by centrifuging at 4000 x g, and the supernatant was discarded. The pellet was resuspended in fresh media to an OD_{730} of 1-1.5. 2 mL was transferred to a sterile tube, and 5 - 10 μL of purified linearized PCR product was added. The culture
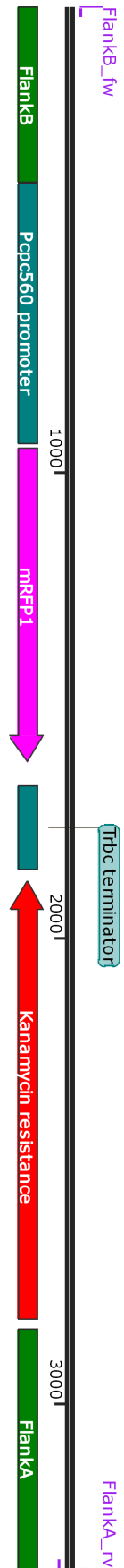
Figure 8: The final gene expression cassette for *Synechococcus*.

was incubated for 5 hours to overnight at the same light and temperature conditions as before. Shaking was employed for mixing instead of bubbling, due the small volumes used and the need to keep the culture sterile.

The culture was then transferred onto mBG-11 plates containing kanamycin at a concentration of 100 µg/mL. These plates were incubated as previously described, until colonies were visible. Restreaking was employed until single colonies were visible.

Colony PCR was used for verification of transformation. The FlankB_fw and FlankA_rv primers should amplify the insert from the pAQ1 plasmid in *Synechococcus*. Naturally this insert should be about 1.2 kbp long, and the transformed version should be 3355 bp. Alternatively, primers from inside the insert could be used, like Pro_fw and Ter_rv.

Table 5: Blue-Green medium and mBG-11 modification.

	COMPONENT	CONCENTRATION
BG-11		(g/L)
	$NaNO_3$	15
	$MgSO_4 \cdot 7H_2O$	0.75
	$CaCl_2 \cdot 2H_2O$	0.36
	Citric acid	0.06
	K_2HPO_4	0.4
	Ferric ammonium citrate	0.06
	Na-EDTA	0.01
	Na_2CO_3	0.2
	Trace metals	1.0 mL/L
MBG-11	(in addition to BG-11 components)	
	NaCl	18
	KCl	0.596
	$MgSO_4 \cdot 7H_2O$	6.162
	Vitamin B12[1]	4 µg/L
SOLID AGAR PLATES	(in addition to liquid media)	
	Na-thiosulphate	3
	bacto-agar	15
TRACE METALS		
	H_3BO_3	2.86
	$MnCl_2 \cdot 4H_2O$	1.81
	$ZnSO_4 \cdot 7H_2O$	0.222
	$Na_2MoO_4 \cdot 2H_2O$	0.39
	$CuSO_4 \cdot 5H_2O$	0.079
	$Co(NO_3)_2 \cdot 6H_2O$	0.0494

[1]Vit. B12 is only added immediately before use.

BG-11 recipe from (Stanier et al., 1971)

mBG-11 modification from Niels-Ulrik Frigaard, University of Copenhagen (unpublished results)

4

RESULTS

4.1 CREATING THE GENE EXPRESSION PLASMID

An expression vector was designed that could be used to promote the expression of mCherry in *Synechococcus.* The final plasmid was designed to be easily modified, replacing *mRFP1* with a gene of the user's choice, such as a gene encoding a codon optimized hydrogenase. To allow for transformation of *Synechococcus*, the expression vector is easily converted to a linear DNA fragment, either by SwaI digestion or PCR amplification.

Sequencing was used to verify that the plasmids shown in Figure 7 were successfully created. The plasmids were sequenced using primers that started in the pUC19 backbone, the SEQ primers seen in Table 7. The SEQ_fw and SEQ_rv primers bind outside the insert on either side. Figure 9a shows the sequencing results for the plasmid containing kanamycin resistance. The arrows show the sequencing results in each direction, where gaps and mismatches are seen as holes in the arrow. After about 1,000 bp the sequencing information becomes unreliable, and in the figure the alignment seems to diverge from the expected sequence after about that distance. Figure 12 shows this in more detail, but for a different plasmid. The gray bars show the quality, or reliability, of the individual base assignments. We can see that the ligation of plasmid 1, and the subsequent reversal of its antibiotic cassette, was successful for the kanamycin resistant variant. The reversals of the spectinomycin and chloramphenicol resistance cassettes in their respective plasmids were unsuccessful. Figure 9b shows the spectinomycin resistant variant still has the *aadA1* gene in the original orientation. Notice that in the reversed version, the XbaI and SpeI sites are no longer present, while they are

still there in Figure 9b. Only the kanamycin resistance variant was used to further develop the over-expression plasmid.

During the next ligation, the promoter, terminator, and an additional antibiotic resistance cassette were inserted at the EcoRV site of plasmid 1. This was a blunt ligation, so that the insert could be oriented in either direction. Four colonies were picked for plasmid isolation and sequencing using the same technique as earlier. The aligned sequence for the plasmid isolated from colony 2 can be seen in Figure 10a. Note that in this ligation the insert was in the opposite orientation from the one shown in Figure 7, but as stated earlier, this should have no effect on the final expression plasmid's performance. Colonies 3 & 4 had inserts in the original direction, as seen in Figure 10b. Colony 1 was not used, as the sequencing data from one of the primers chosen was of low quality, as can be seen in comparison to the sequencing for colony 3, shown in Figure 11. The plasmid isolated from colony 3 was used for the next step. After truncation of low quality base calls, approximately the first 50 bp and everything after 1000 bp, it had nearly 100% identity to the planned sequence. There were two inserted bases, seen in Figure 10b as a small gap above the promoter, but the changes at this position in the promoter were deemed insignificant for its performance. The possible base insertions are at about 900 bp into the sequencing results and therefore of reduced reliability, but also follows CCGGGGG, which might cause problems with sequencing. Furthermore, the sequencing results from the opposite direction have 100% identity with the desired sequence after truncation of low quality base assignments. More detailed sequence alignment for this plasmid can be seen in the appendix, section A.5.

The plasmid pSBA3-BBa_E1010, containing *mRFP1*, and plasmid 2 were digested with XbaI and SpeI, then ligated together, creating the final gene expression plasmid. Figures 13a & 13b shows *E. coli* colonies after transformation with the constructed plasmid growing on an LB plate. Because of the compatible sticky ends of the XbaI and SpeI digestion, there are three likely plasmids created: *mRFP1* can be ligated in the forward (1) or reverse (2) direction, or the backbone can self-ligate, with no insert (3). Plasmids were isolated from colonies that were red, as production of mRFP1 is a positive marker for transformation, removing any option (3) plasmids from being used further.

As the expected ligation products could have contained the insert in either direction, ten colonies were picked and further characterized. The resultant plasmids were digested with XbaI and SpeI, as shown in Figure 14. The mRFP1 gene is 711 bp long, and the remaining fragment should be about 5.3 kbp. All 10 plasmids seem to have the correct restriction sites. It is worth noting that the DNA ladder does not seem to agree with these results, but this ladder has been consistently giving incorrect sizes with the dye system and could be used only for rough size estimation.
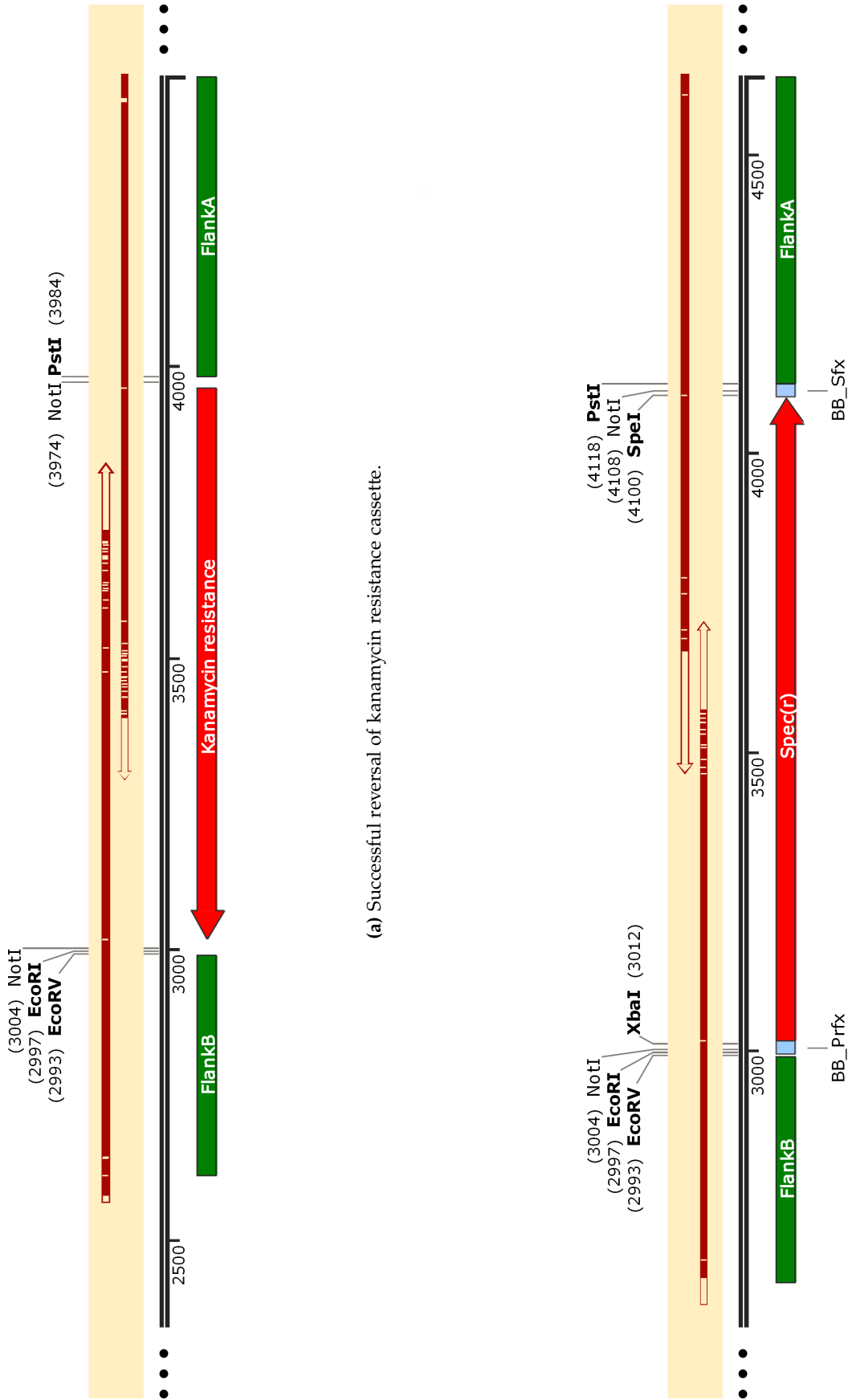
(a) Successful reversal of kanamycin resistance cassette.

(b) Unsuccessful reversal of spectinomycin resistance cassette.

Figure 9: Comparison of planned and actual sequence for two variants plasmid 1 after attempted reversal of the antibiotic resistance cassette.
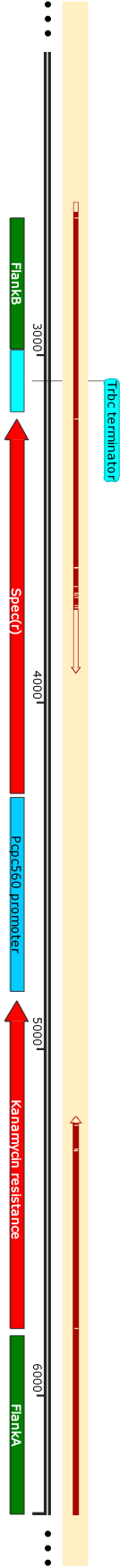
(a) Plasmid 2 isolated from colony 2.
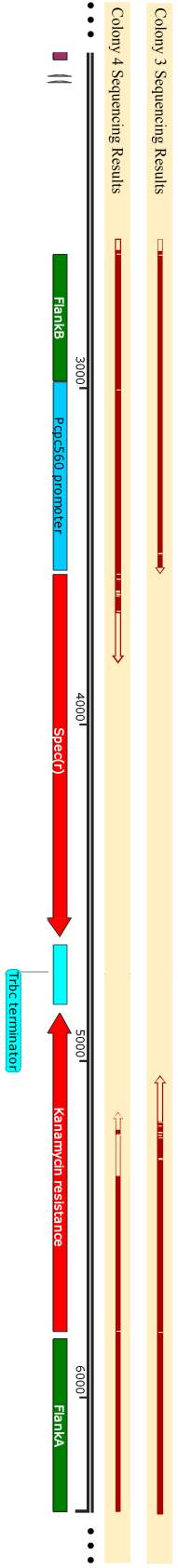
(b) Plasmid 2 isolated from colonies 3 and 4.

Figure 10: Sequencing results for plasmid 2 isolated from 3 different colonies, aligned against the planned insert.
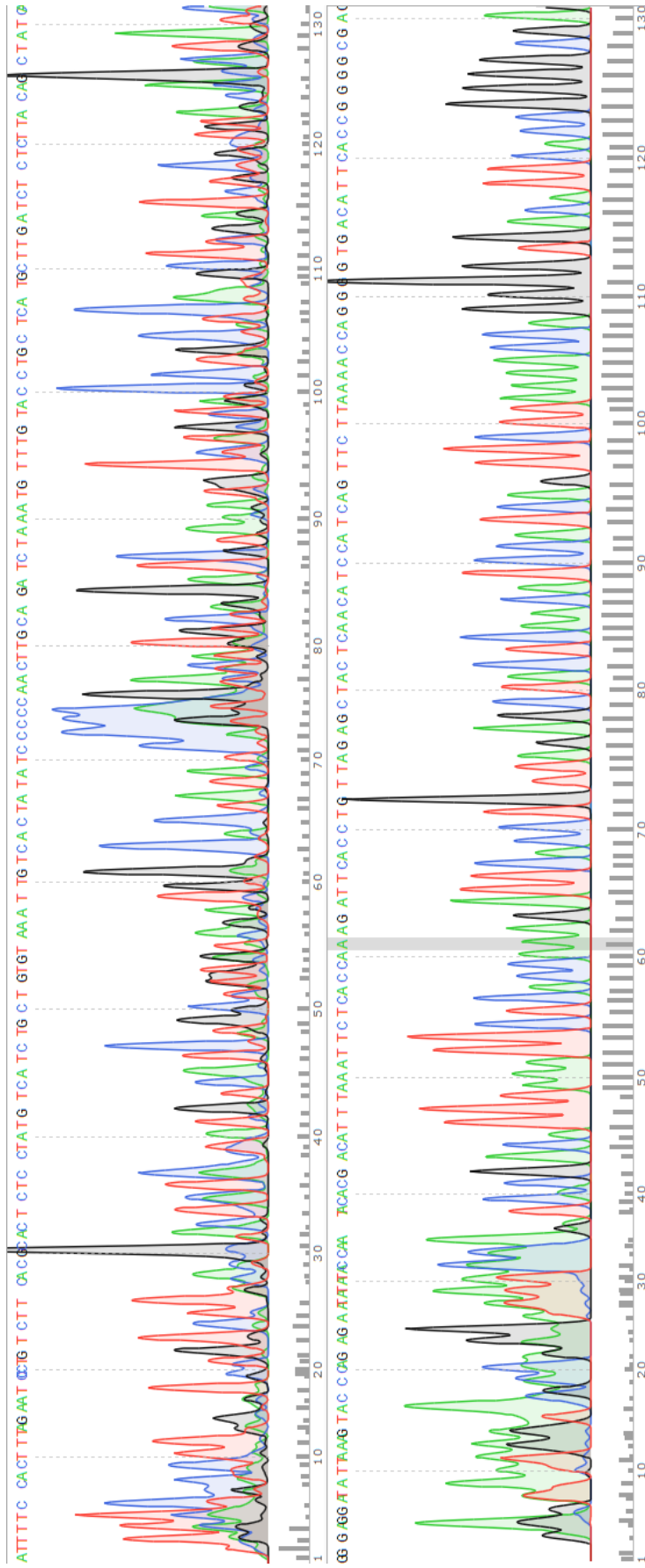
Figure 11: Comparison between the quality of sequencing results for the first 100 bp of plasmid 2 isolated from colony 1 (top) and colony 3 (bottom).
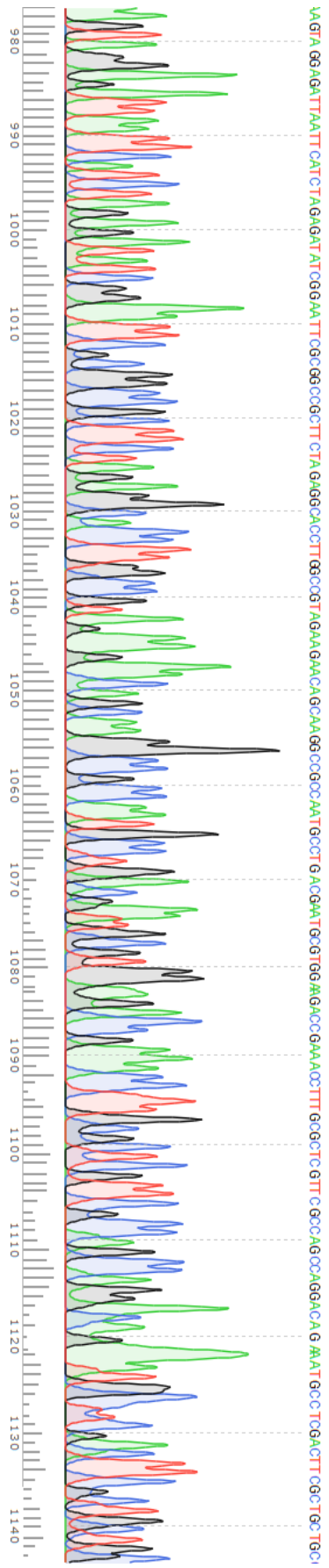
Figure 12: Last reliable base assignments from the forward sequencing results for plasmid 2 isolated from colony 3.

(a) Low concentration
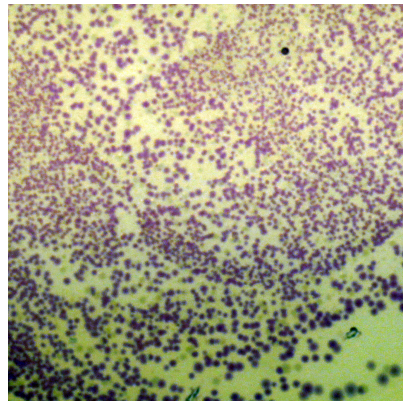
(b) Magnified and false color rendering of image shown in 13a
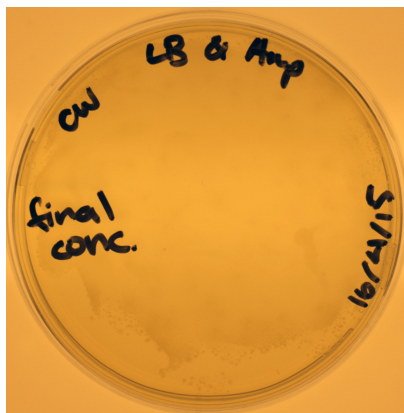
(c) High concentration
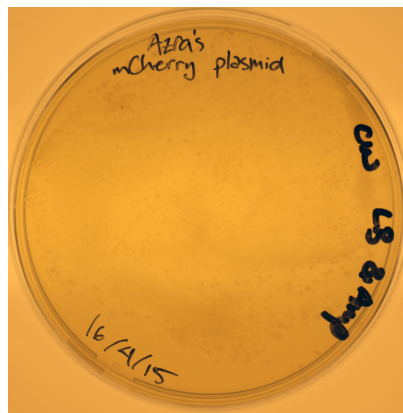
(d) pSB1A2 - BBa_E1010

Figure 13: *E. coli* colonies transformed by (a) and (b) the *Synechococcus* gene expression plasmid plated at low and high concentration; (d) pSB1A2 containing *mRFP1*. (b) shows a false color rendering of (a), where colonies producing mRFP1 are significantly darker then those that do not produce it.
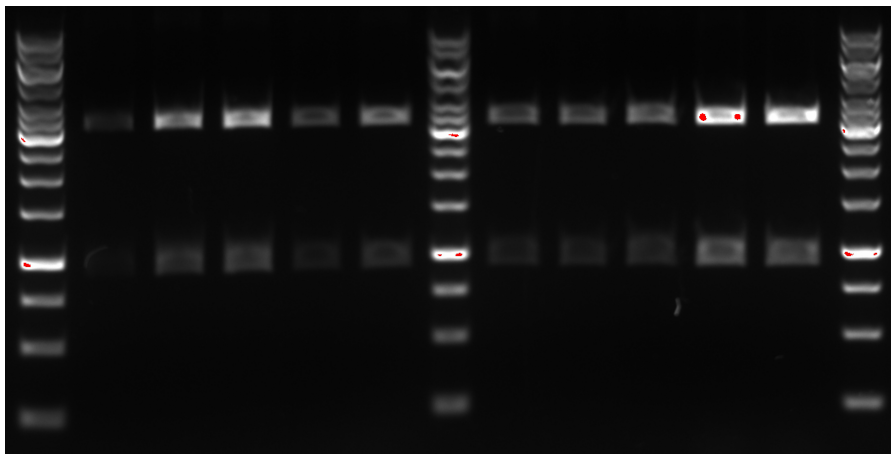
Figure 14: SpeI and XbaI double digest of the gene expression plasmid.
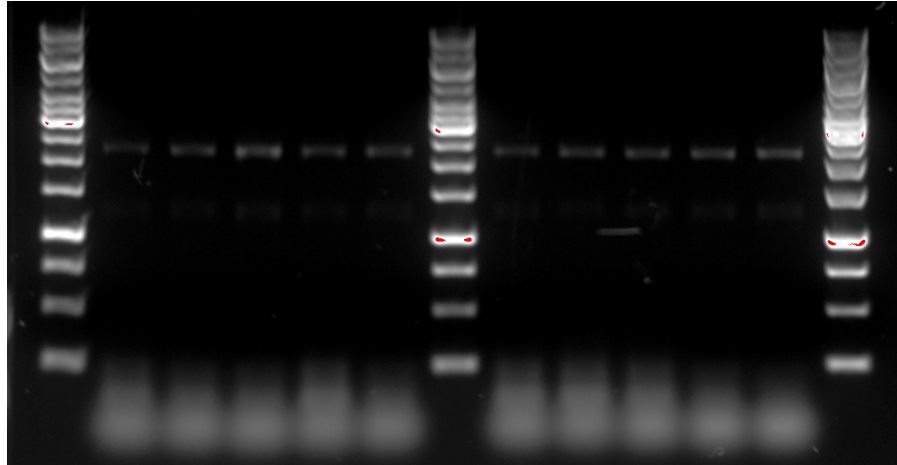
Figure 15: Gene expression cassette amplification and gel separation.

The insert was then amplified from the plasmid containing *mRFP1* in the correct orientation using PCR, with FlankB_fw and FlankA_rv as primers, creating the gene expression cassette. Gel electrophoresis was used to verify the amplification, seen in Figure 15.

For additional verification that the final plasmid was correct, the entire insert (flank B, promoter, *mRFP1*, terminator, and flank A) was sequenced. Given the large size of the insert, about 3.4 kbp, multiple primers were used to allow for overlapping coverage: FlankB_fw, P_{cpc560}_fw, T_{rbc}_rv, and FlankA_rv. The plasmid that presumably carries the entire gene expression cassette was used for transformation before the plasmid was verified by sequencing. The sequencing results were inconclusive, of very low quality, and could not be aligned with the expected plasmid.

4.2 *Synechococcus* TRANSFORMATION AND *m*RFP1 PRODUCTION

Successfully transformed *Synechococcus* should contain *aph*, providing kanamycin resistance, and produce mRFP1, hopefully giving the algae a reddish color. *Synechococcus* was transformed using the gene expression cassette, and transferred to selective plates. The transformants were given a week to grow, however no colonies were seen, of any color, signifying a lack of both *aph* and *mRFP1*. The transformation was therefore not considered successful.

<div style="text-align: right; font-size: 3em; color: gray;">5</div>

DISCUSSION

5.1 GENE EXPRESSION CASSETTE IN *Synechococcus*

Powerful promoters in *E. coli* do not work well in cyanobacteria, creating very little protein (Zhou et al., 2014). In *Synechocystis*, the promoter P_{cpc560} has previously been shown to produce protein at up to 15% of cell's dry weight. Sequencing of the plasmid after each transformation showed that *E. coli* was not likely to mutate P_{cpc560}, as was initially feared. The promoter also works in *E. coli*, as shown in Figure 13a, where the transformed *E. coli* is producing mRFP1 that is visible to the eye. In Figure 13d, *mRFP1* is not downstream of a promoter, and does not lead to red colonies. This also explains why all 10 red colonies chosen for restriction digest had *mRFP1* in the correct orientation, shown in Figure 14 by the retained XbaI and SpeI sites. There does, however, seems to be a concentration-dependent effect on the production of mRFP1. Colonies in Figure 13c were plated at a much higher bacterial concentration, and are able to grow on selective media, but do not produce mRFP1 at an easily discernible level. As mentioned earlier, P_{cpc560}'s promotion strength is regulated by light intensity and completely stopped during nitrogen starvation. The colonies were grown in the dark, so maybe nitrogen starvation, light intensity, or some other factor, like interplay between *E. coli* regulators and the promoter, stopped expression of *mRFP1*.

Sequencing results were inconclusive. The plasmid containing all the fragments except *mRFP1* was successfully created, as shown in Figure 10, but the final gene expression plasmid was not successfully sequenced. After the final ligation, *E. coli* colonies were red, meaning that *mRFP1* was ligated in, and in the correct orientation downstream of the promoter. This is a good reason to assume that the last step was successful. However, *Synechococcus* was not transformed by the

gene expression cassette generated from this plasmid. This might be due to some problems with the sequence of the DNA that were not previously seen, caused during the digestion, ligation, or *E. coli* transformation. The spectinomycin resistance cassette had previously been placed downstream of the promoter, and this had successfully transformed *E. coli* without getting mutated, so the promoter was probably not causing *E. coli* to mutate the plasmid. The flanking regions chosen for incorporation into *Synechococcus* have been used before with success, and were probably still intact.

If the gene expression plasmid was not correct, the final plasmid could be recreated and tested again. The method for producing this plasmid has already been through a number of modifications to streamline and effectivize its production and usefulness, but since it still hasn't viably produced the final plasmid, it might still need some changes. The final plasmid could be made using an alternative method. This involves amplifying the promoter and terminator from the previously made "Insert 2" (or, of course, from *Synechocystis*, if one so desires). *mRFP1* is amplified from its plasmid using normal BioBrick insert primers. Because the BioBrick sites are contained on all 3 pieces, overlap extension PCR can be used to create 1 contiguous DNA fragment, a new Insert 2-*mRFP1*, as shown in Figure 16. This could then be ligated into plasmid 1 (with *aph* reversed), as shown in Figure 17. This step is similar to a previously successful ligation and transformation, but removes the spectinomycin resistance cassette from the method, using instead *mRFP1* immediately. This new cloning procedure has one less ligation and transformation, but would still allow for the selection of successful transformants, using *mRFP1* instead of the spectinomycin resistance cassette as the selection marker; *mRFP1* downstream of the promoter makes visibly red colonies.

If plasmid itself was correct, then the transformation of *Synechococcus* was unsuccessful because of a problem with one of the modifications made to the protocol. Assuming this was the case, the use of mBG-11 instead of A$^+$medium might have reduced the transformation efficiency of the bacteria. The two media are very similar, as can be seen in Table 8 in the appendix. The trace metal concentrations are the same, even if the anions for the salts used vary between the two media. The biggest difference is in the increased concentration of NaNO$_3$ in mBG-11. Nitrate starvation makes the cyanobacteria enter the stationary phase, lowering transformation efficiency by 90%, but the effect of increasing nitrate has not been studied (Essich et al., 1990). *Synechococcus* was very capable of growing in mBG-11, but maybe the different concentration of salts changed the cell's permeability to DNA and resulting transformation efficiency. The amount of DNA used for the transformation might have been incorrect. The protocol stated that 1-5 µg DNA or 5-10 µL PCR product should be
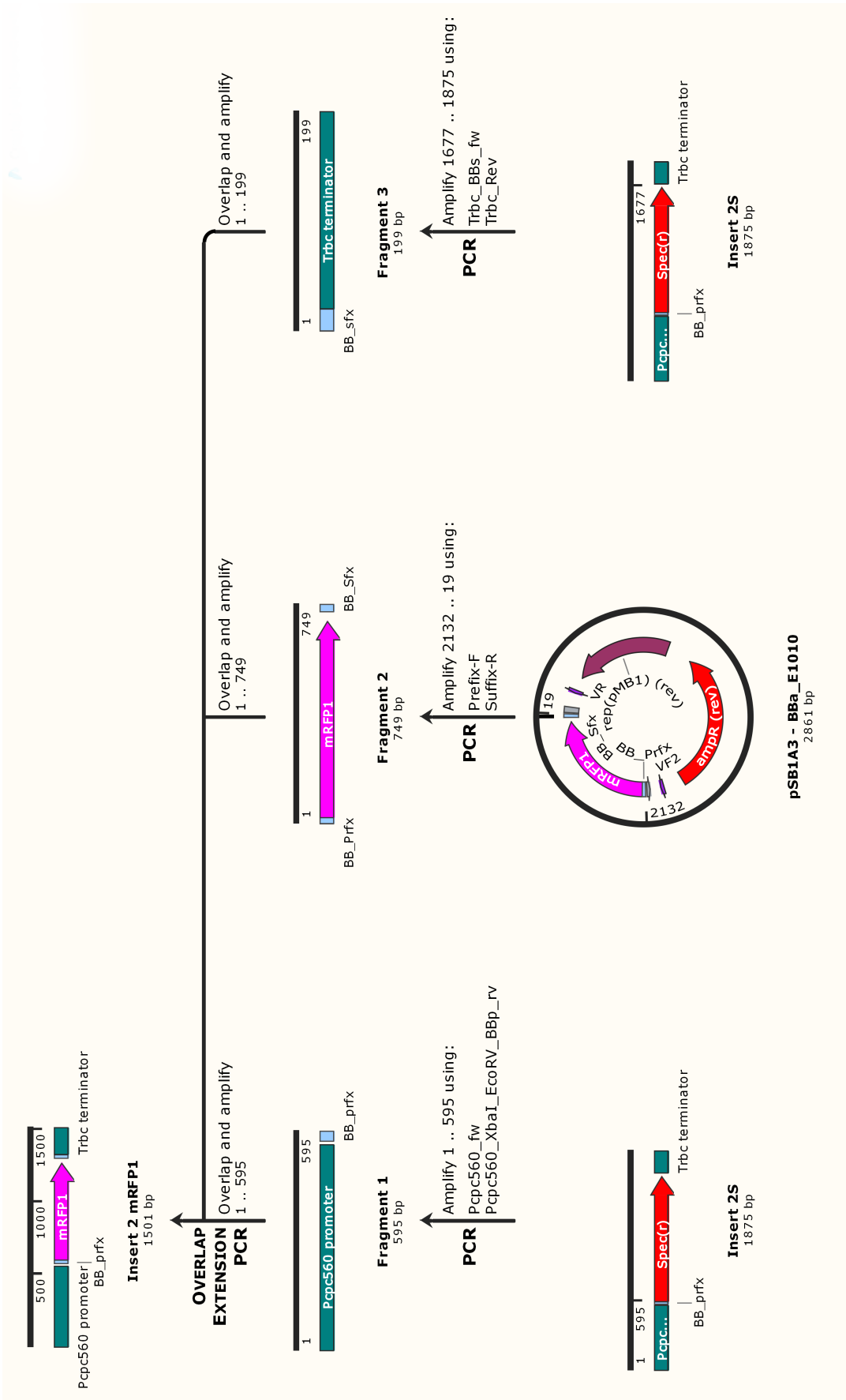
Figure 16: Creation of the new Insert 2 - *mRFP1*.
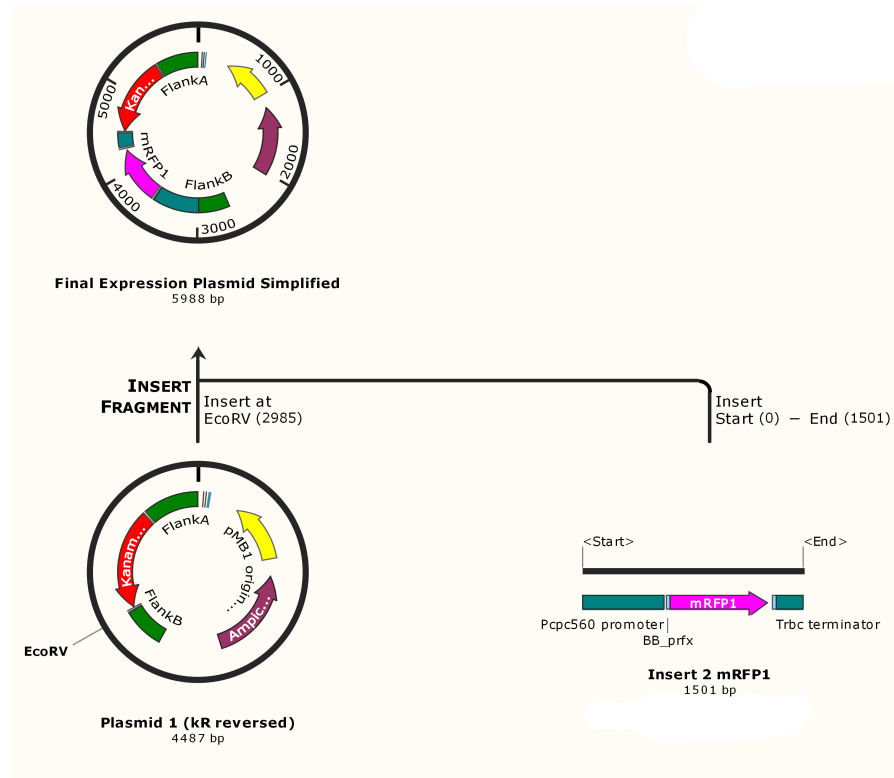
Figure 17: Simplified method for creating the final expression plasmid using Insert 2-*mRFP1*.

used (Xu et al., 2011). The cleaned PCR product used for transformation contained DNA at a concentration of 40 ng/µL, and this was not sufficient to obtain 1 µg DNA with 10 µL of the PCR product. Other studies have successfully used 0.5 µg of transforming DNA, but have other modifications to the protocol as well (Ruffing et al., 2015). Another modification employed in this thesis involved the use of shaking *Synechococcus* during transformation, instead of bubbling. There was no satisfactory method of keeping a 2 mL culture sterile and bubbling, so shaking was chosen. Shaking might have been either too vigorous or too relaxed to enable transformation. Shaking has been successfully used for transformation in other studies, at 150 rpm, and this rpm could be used for future transformations (Ruffing et al., 2015).

Concentration of *Synechococcus* cultures used for transformation vary between publications, but recently, an optimum was found between 0.7 and 1.0 at OD_{730} (Ruffing et al., 2015). This is lower than recommended in previously published protocols, for example the OD_{730} of 2-3 (Frigaard et al., 2004), or OD_{730} of 1-1.5 (Xu et al., 2011). It is also very important that the cells are in the exponential growth phase, as the transformation efficiency drops in the stationary phase. In this experiment, the *Synechococcus* culture was allowed to grow for 3 days at 1500 µmol photons $m^{-2}s^{-1}$. This is not longer then used in other

protocols, but is at a much higher light intensity, and so the culture might have grown quicker, and reached the maximum sustainable population and entered stationary phase quicker.

The light intensity used in the transformation was also much higher then reported light intensities from other transformations. Due to other experiments being run in the incubator at the same time, the transformations were done at approximately 1500 µmol photons m^{-2}s^{-1}. This is well within the range that *Synechococcus* can tolerate and grow under, but it might affect transformation by stressing the cell or affecting its DNA uptake mechanisms. There is a link between phototaxis and transformability in cyanobacteria, and motility and transformation in most gram-negative bacteria. In *Synechocystis*, many genes necessary for motility are also required for transformation. The uptake machinery seems to be inherently linked to the pili and motility (Yoshihara et al., 2001). However, *Synechococcus* and *Synechocystis* take up DNA differently; *Synechococcus* takes up double-strand DNA, while *Synechocystis* digests one strand during uptake, leaving single stranded DNA in the cytosol for transformation (Essich et al., 1990). It is highly probable that the uptake machinery is very similar otherwise. Growing *Synechococcus* at very high light intensities might therefore be problematic for transformation, as phototaxis would be down-regulated, possibly stopping DNA from binding to the cell and being transported across the cell membrane and into the cytoplasm by the pili and associated proteins. The high light intensity might cause the cell to up-regulate DNA repair systems, and these might also inhibit transformation. The author could find no research attempting to optimize the light intensity used during transformation of any cyanobacteria. Other studies report light intensities from 60 to 250 µmol photons m^{-2}s^{-1} (Ruffing et al., 2015; Xu et al., 2011), and this might be a more reasonable intensity to use for future transformations. Lower light intensities might increase transformation efficiency by upregulating phototaxis and its associated transformation machinery, but could decrease growth rates. Therefore, low light would be most necessary during the actual transformation stage, right after DNA is mixed with the *Synechococcus* culture. Before and after, the light intensity could be at, or even above, 250 µmol photons m^{-2}s^{-1}, although care should be taken that light intensities are not too high prior to transformation as there might not be enough time to upregulate phototaxis enough to take up enough DNA. Some light does seem to be important for successful transformation, *Synechococcus* incubated in the dark before transformation has decreased transformation efficiency, related to the length of dark incubation (Essich et al., 1990).

To increase the chance of future transformations in *Synechococcus* working, a few modifications to the protocol can be implemented. The concentration of exponentially growing cells should be between

0.7 and 1.0 at OD_{730}, an increased amount of linear DNA should be used, and *Synechococcus* should be transformed at a reduced light intensity, between 60 and 250 μmol photons $m^{-2}s^{-1}$.

If transformation into *Synechococcus* is successful at a later date using the designed expression cassette, it can be modified to promote the production of other proteins. There are three ways to change out *mRFP1* for another gene: firstly, the BioBrick restriction sites still exist flanking *mRFP1*, and can be used to excise it and create sticky ends for a new ligation. Unfortunately, NotI and PstI sites are also present in the plasmid at other locations, so SpeI and XbaI must be used. This re-creates a problem seen earlier: the insert will be compatible in either direction, so after ligation and transformation, only about 50% of the colonies will have the plasmid with the correct orientation. This requires another round of digestion and gel electrophoresis to check for conserved restriction sites. In the second method, a new Insert 2 could be made with the new gene and ligated into plasmid 1, as shown in Figures 16 and 17. The final method involves using overlap extension PCR using primers that overlap with the expression cassette and the new gene; forward primers would have the last ~20 bases of the promoter and the first ~20 of the new gene, while the reverse primers would be on the opposite strand and overlap with the end of the gene, and the start of the terminator. Using this method, new restriction sites could be introduced, replacing the BioBrick prefix and suffix. This woukd allow for easier restriction cloning later, instead of the reversion problems seen with XbaI and SpeI and the BioBrick parts.

Interesting genes to insert into *Synechococcus* include those related to biofuel production. As stated earlier, *Synechococcus'* fast growth and adaptability make it a good choice for biofuel production, and increasing its capabilities could finally make biofuel production a viable alternative to fossil fuel based energy production.

Part III

OVER-EXPRESSION IN *SYNECHOCOCCUS*

6

INTRODUCTION

6.1 SYNTHETIC HYDROGENASE

Chlamydomonas reinhardtii is currently viewed as one of the most efficient eukaryotic hydrogen producers. *C. reinhardtii* is a green algae, a single-celled eukaryote. It can produce hydrogen in a short burst from water, or continuously via photofermentation, using a [FeFe]-H$_2$ase, called Hyd1, as described earlier (Esquível et al., 2011). An engineered strain of *Synechococcus* that can produce Hyd1 could consolidate the advantages of both algae into one organism. The hydrogen production efficiency of *C. reinhardtii* coupled with the fast growth rates of *Synechococcus* could lead to an extremely flexible platform for biohydrogen production. However, cloning *hyd1* directly into *Synechococcus* might not lead to the formation of functional protein, and so some additional genetic engineering is required.

C. reinhardtii is a eukaryotic algae, while *Synechococcus* is a prokaryotic cyanobacteria. The transcription/translation machinery in eukaryotes is much more complex then in prokaryotes. For example, in humans, only 1.5% of the genome codes for proteins while most of the rest is involved in regulation. In contrast, 88% of the *E. coli* genome encodes proteins (Blattner et al., 1997). Bacteria generally can not recognize and splice eukaryotic introns. Eukaryotic mRNA can also contain untranslated regions at the start and end of the transcript, or be modified by other RNA editing (Mignone et al., 2002). While mRNA modifications can also play an important role in bacteria and archaea, it is less extensive than in eukaryotes (Dalgaard et al., 1995). There is also post-translational machinery that can be important for maturation of eukaryotic proteins that are not present in prokaryotes (Sodoyer, 2004).

Hydrogenases are complicated enzymes, and require many maturation factors. The hydrogenases in *C. reinhardtii* and *Synechococcus* come from different families, Hyd1 is an [FeFe]-H_2ase and Hox is a [NiFe]-H_2ase. They have different maturation factors required for their production. *E. coli* has previously been modified to produce recombinant [FeFe]-H_2ases, but even though *E. coli* has the maturation factors for its own [NiFe]-H_2ase, the exogenous enzyme was not functional. The production of mature hydrogenases in *E. coli* required the simultaneous expression of three maturation factors, *hydE*, *hydF*, and *hydG* (King et al., 2006). This might not be necessary for maturation in *Synechococcus.* Studies have shown the ability of two cyanobacteria, *Synechocystis* sp. PCC6803 and *Synechococcus elongatus* sp. PCC7942, to produce mature [FeFe]-H_2ase without the addition or expression of exogenous maturation factors (Asadaa et al., 2000; Bertoa et al., 2011).

Another hurdle in the production of exogenous protein is codon bias. There are 64 codons that code for 20 amino acids, leading to many synonymous codons. Which of the synonymous codons is used more frequently by an organism is a reflection of the GC content of its genome (Knight et al., 2001). The total GC content in the genome can vary from 72% to 20% between organisms (Bentley et al., 2002; Gardner et al., 2002). Organisms can show extreme preference for a codon that might be largely omitted in another organism. This preference for some codons is more pronounced in highly expressed proteins (Gustafsson et al., 2004). This codon bias is mirrored in the transfer RNA (tRNA) concentrations within the cell, rare codons match tRNA molecules that are only found at low concentrations. Trying to produce exogenous proteins that use a rare codon in a transfected organism can lead to a variety of problems, from misfolding to incorrect peptide sequences to halted translation (Kane, 1995).

Synechococcus has a genome wide GC content of 49.2%, while *C. reinhardtii* is at 64%. This is a good indication that the two have different codon usage. A comparison of the most common codons used for each amino acid in both organisms can be seen in Table 10 in the appendix. Half of the amino acids have a different most common codon, not including amino acids encoded by a single codon. *Synechococcus* would probably not be able to produce a protein from *C. reinhardtii* with the original codons. To solve this problem, there are two solutions used today: the rare tRNA molecules can be introduced into an organism, usually by genetically engineering the bacteria to produce them at higher concentrations; or a modified version of the gene can be designed, reflecting the codon use of the host cell (Sodoyer, 2004).

Creating a synthetic gene would sidestep problems such as the presence of introns and differences in codon usage. Back-translating the amino acid sequence into genomic DNA would the most efficient way of doing this. The designed gene would contain no introns that

Synechococcus can not properly recognize and excise, and the gene can be designed to mimic the codon usage seen in *Synechococcus*. The synthetic *hyd1* could then be placed inside the previously described gene expression plasmid (Chapters 3 & 4) to allow it to be expressed at high levels.

7

MATERIALS AND METHODS

7.1 CREATING THE *hoxH* KNOCK-OUT PLASMID

The *hoxH* gene encodes the large subunit of the [NiFe]-hydrogenase, which contains the active site. Knocking out this gene inhibits hydrogen production in *Synechococcus* (Eckert et al., 2012), and it was therefore chosen as the target in order to generate a background strain for the expression of a non-native [FeFe]-hydrogenase.

The *hoxH* gene is 1425 nucleotides long, and is found on the main chromosome of the *Synechococcus* genome. The 240 bp to either side of this gene was used to create flanking regions for creating the knock out strain, as shown in Figure 18, and in more detail in the appendix, Figure 25. The knock-out plasmid was ordered from GeneArt (Life Technologies, Germany) containing the two flanking regions with a NotI site separating them. NotI is a restriction enzyme with an 8bp recognition sequence.

Kanamycin resistance was chosen as the antiobiotic resistance marker, encoded by *aph*. The *aph* gene was modified during PCR amplification to contain BioBrick sites on both sides of the cassette. The plasmid was digested with NotI, and a NotI digested kanamycin resistance cassette was ligated into it, as shown in Figure 19. This ligation was then used to transform supercompetent *E. coli* DH5-α, and transformants were subsequently grown on media containing kanamycin.

The plasmid was isolated from succesfully transformed *E. coli* colonies. The insert, containing both flanking regions and antibiotic resistance cassette, was PCR amplified from this isolated plasmid. The primers used can be seen in the appendix, Table 9. This extended the flanking regions, and created a linear DNA fragment. This PCR product was then used for transforming *Synechococcus.*
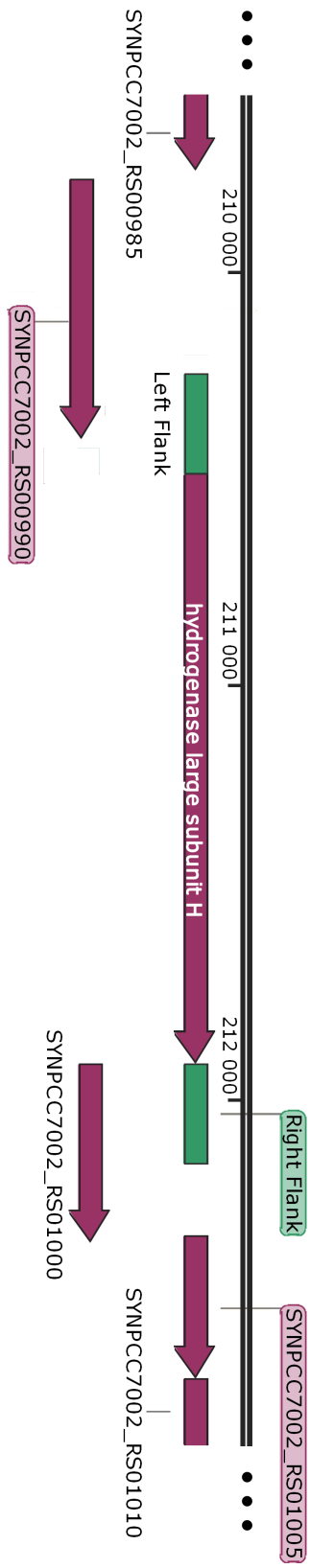
Figure 18: The *hoxH* gene and planned flanking regions for the *Synechococcus* △*hoxH* strain (Nakao et al, 2010).
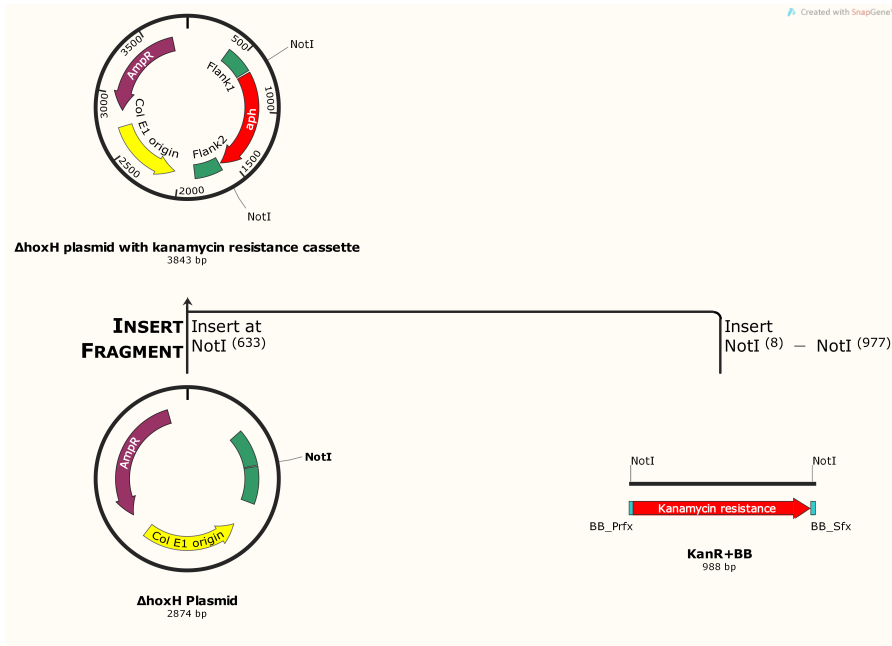
Figure 19: Cloning procedure for making the *hoxH* deletion plasmid.

7.2 TRANSFORMATION OF *Synechococcus*

Synechococcus was transformed following the protocol outlined in section 3.3. Only succesful transformants should contain the *aph* gene, and be resistant to kanamycin.

7.3 FINDING THE $\triangle hoxH$ MUTANT PHENOTYPE

Previous studies have only recently found a phenotype for $\triangle hoxH$ strains in cyanobacteria. In *Synechocystis* sp PCC 6803, the first bacteria shown to have a conclusive $\triangle hoxH$ phenotype, the hydrogenase functions as an electron acceptor. It is crucial for growth during nitrate limiting, mixotrophic conditions (Gutekunst et al., 2014). To mimic these conditions, the $\triangle hoxH$ strain should be grown in a further modified version of mBG-11. This medium should contain no $NaNO_3$, and instead have 10 mM arginine as the sole nitrate source. Mixotrophic conditions are achieved by the addition of 10 mM glucose to *Synechocystis*, which could replaced with glycerol in glycerol-resistant strains of *Synechococcus*.

```
  1 MSALVLKPCA AVSIRGSSCR ARQVAPRAPL AASTVRVALA TLEAPARRLG NVACAAAAPA
 61 AEAPLSHVQQ ALAELAKPKD DPTRKHVCVQ VAPAVRVAIA ETLGLAPGAT TPKQLAEGLR
121 RLGFDEVFDT LFGADLTIME EGSELLHRLT EHLEAHPHSD EPLPMFTSCC PGWIAMLEKS
181 YPDLIPYVSS CKSPQMMLAA MVKSYLAEKK GIAPKDMVMV SIMPCTRKQS EADRDWFCVD
241 ADPTLRQLDH VITTVELGNI FKERGINLAE LPEGEWDNPM GVGSGAGVLF GTTGGVMEAA
301 LRTAYELFTG TPLPRLSLSE VRGMDGIKET NITMVPAPGS KFEELLKHRA AARAEAAAHG
361 TPGPLAWDGG AGFTSEDGRG GITLRVAVAN GLGNAKKLIT KMQAGEAKYD FVEIMACPAG
421 CVGGGGQPRS TDKAITQKRQ AALYNLDEKS TLRRSHENPS IRELYDTYLG EPLGHKAHEL
481 LHTHYVAGGV EEKDEKK
```

Figure 20: Amino acid sequence for Hyd1, the [FeFe]-hydrogenase in *C. rein-hardtii* (Merchant et al., 2007).

7.4 DESIGNING THE SYNTHETIC HYDROGENASE GENE

To make *Synechococcus* a more efficient biohydrogen producer, a hydrogenase gene was designed. The [FeFe]-hydrogenase from *C. reinhardtii* was used as the template for the synthetic hydrogenase. An [FeFe]-hydrogenase was chosen because this group of enzymes has a much higher turnover rate for hydrogen, and *C. reinhardtii* was chosen because it has already proved itself to be an efficient biohydrogen producer. *Synechococcus* might not be able to produce mature hyd1 from the original DNA sequence. The amino acid sequence for this hydrogenase can be seen in Figure 20. To optimize the production of the non-native [FeFe]-hydrogenase from *C. reinhardtii* in *Synechococcus,* a synthetic gene was designed. This was done using codon usage tables, choosing the most common codon used in *Synechococcus* for each amino acid in the hydrogenase. The codon usage table for *Synechococcus* can be seen in the appendix, Table 11. The codon optimized version of *hyd1* can be seen in Figure 21.

The synthetic *hyd1* gene was designed so it could be inserted into the gene expression plasmid created earlier, replacing *mRFP1*. However, because of time constraints, this synthetic hydrogenase gene was not placed into the expression vector.

```
   1 ATG TCC GCC TTC GTG TTC AAA CCC TGT GCC GCC GTG TCC ATT CGC GGC TCC TCC TGT CGC
  61 GCC CGC CAA GTG GCC CCC CGC GCC CCC TTC GCC GCC TCC ACC GTG CGC GTG GCC TTC GCC
 121 ACC TTC GAA GCC CCC GCC CGC CGC TTC GGC AAT GTG GCC TGT GCC GCC GCC GCC CCC GCC
 181 GCC GAA GCC CCC TTC TCC CAT GTG CAA CAA GCC TTC GCC GAA TTC GCC AAA CCC AAA GAT
 241 GAT CCC ACC CGC AAA CAT GTG TGT GTG CAA GTG GCC CCC GCC GTG CGC GTG GCC ATT GCC
 301 GAA ACC TTC GGC TTC GCC CCC GGC GCC ACC ACC CCC AAA CAA TTC GCC GAA GGC TTC CGC
 361 CGC TTC GGC TTT GAT GAA GTG TTT GAT ACC TTC TTT GGC GCC GAT TTC ACC ATT ATG GAA
 421 GAA GGC TCC GAA TTC TTC CAT CGC TTC ACC GAA CAT TTC GAA GCC CAT CCC CAT TCC GAT
 481 GAA CCC TTC CCC ATG TTT ACC TCC TGT TGT CCC GGC TGG ATT GCC ATG TTC GAA AAA TCC
 541 TAT CCC GAT TTC ATT CCC TAT GTG TCC TCC TGT AAA TCC CCC CAA ATG ATG TTC GCC GCC
 601 ATG GTG AAA TCC TAT TTC GCC GAA AAA AAA GGC ATT GCC CCC AAA GAT ATG GTG ATG GTG
 661 TCC ATT ATG CCC TGT ACC CGC AAA CAA TCC GAA GCC GAT CGC GAT TGG TTT TGT GTG GAT
 721 GCC GAT CCC ACC TTC CGC CAA TTC GAT CAT GTG ATT ACC ACC GTG GAA TTC GGC AAT ATT
 781 TTT AAA GAA CGC GGC ATT AAT TTC GCC GAA TTC CCC GAA GGC GAA TGG GAT AAT CCC ATG
 841 GGC GTG GGC TCC GGC GCC GGC GTG TTC TTT GGC ACC ACC GGC GGC GTG ATG GAA GCC GCC
 901 TTC CGC ACC GCC TAT GAA TTC TTT ACC GGC ACC CCC TTC CCC CGC TTC TCC TTC TCC GAA
 961 GTG CGC GGC ATG GAT GGC ATT AAA GAA ACC AAT ATT ACC ATG GTG CCC GCC CCC GGC TCC
1021 AAA TTT GAA GAA TTC TTC AAA CAT CGC GCC GCC GCC CGC GCC GAA GCC GCC GCC CAT GGC
1081 ACC CCC GGC CCC TTC GCC TGG GAT GGC GGC GCC GGC TTT ACC TCC GAA GAT GGC CGC GGC
1141 GGC ATT ACC TTC CGC GTG GCC GTG GCC AAT GGC TTC GGC AAT GCC AAA AAA TTC ATT ACC
1201 AAA ATG CAA GCC GGC GAA GCC AAA TAT GAT TTT GTG GAA ATT ATG GCC TGT CCC GCC GGC
1261 TGT GTG GGC GGC GGC GGC CAA CCC CGC TCC ACC GAT AAA GCC ATT ACC CAA AAA CGC CAA
1321 GCC GCC TTC TAT AAT TTC GAT GAA AAA TCC ACC TTC CGC CGC TCC CAT GAA AAT CCC TCC
1381 ATT CGC GAA TTC TAT GAT ACC TAT TTC GGC GAA CCC TTC GGC CAT AAA GCC CAT GAA TTC
1441 TTC CAT ACC CAT TAT GTG GCC GGC GGC GTG GAA GAA AAA GAT GAA AAA AAA
```

Figure 21: Synthetic, codon-optimized sequence for *Hyd1*, for expression in *Synechococcus*.

8

RESULTS

8.1 BIOINFORMATICS

The amino acid sequences of several [FeFe]- and [NiFe]-H_2ases were aligned by the MUSCLE algorithm (Edgar, 2004), using UGENE software (Okonechnikov et al., 2012).

The catalytic site of both the [NiFe]- and [FeFe]-hydrogenases are formed by two metal atoms, which are coordinated to the protein by four cysteine residues. The positions of these cysteines is conserved in each family of hydrogenase, but is different between the two families. Figure 22 shows an alignment of 10 [NiFe]-hydrogenases across all 5 groups of [NiFe]-H_2ases. Residues that are conserved across at least 75% of the species analysed are highlighted, where darker means more conserved. Two conserved motifs contain the four cysteines involved in coordinating the [NiFe] core. The first of the two conserved motifs is $RxCGxCx_3H$, and is usually found close to the N-terminal end of the protein. For 9 of the proteins shown, the first arginine is at position 53 or 54. HycE from *E. coli* contains an additional 177 amino acids towards the N-terminus that seem to be involved in linking the hydrogenase to NAD(P)H oxidation. The second conserved motif is $DPCx_2Cx_2H$, and contains the second half of the [NiFe] binding site, and is located close to the C-terminal end of the protein.

The diiron cluster of [FeFe]-H_2ases is also coordinated by 4 cysteins. The H-cluster cysteines are spread across 3 motifs, visible in Figure 23. The three motifs are L1: TSCCPxW, where the first cysteine residue is involved in catalytic site coordination, L2: $MPCx_2Kx_2E$, and L3: $ExMACx_2GCx_2G$, which contains the final two cysteine residues. [FeFe]-H_2ases tend to be monomeric and are all bidirectional. They
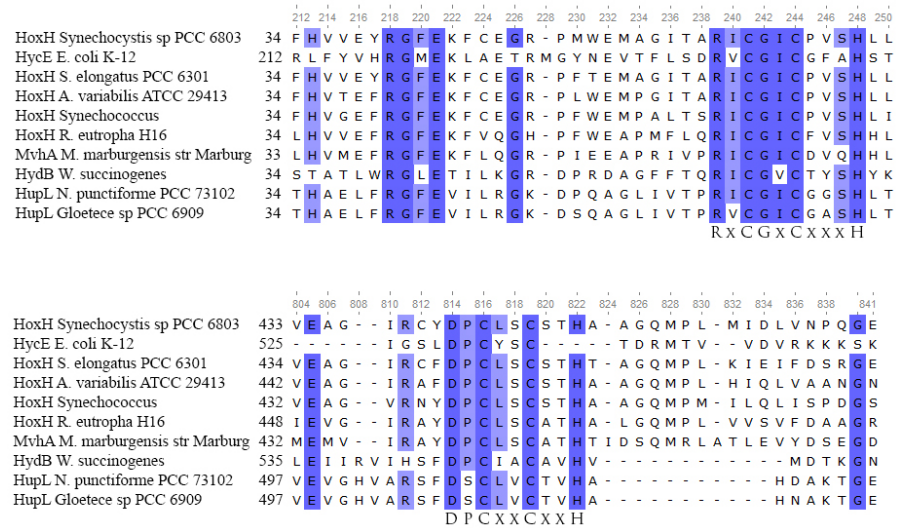
Figure 22: Alignment of [NiFe]-hydrogenases, showing the Ni-binding motif at the active site.
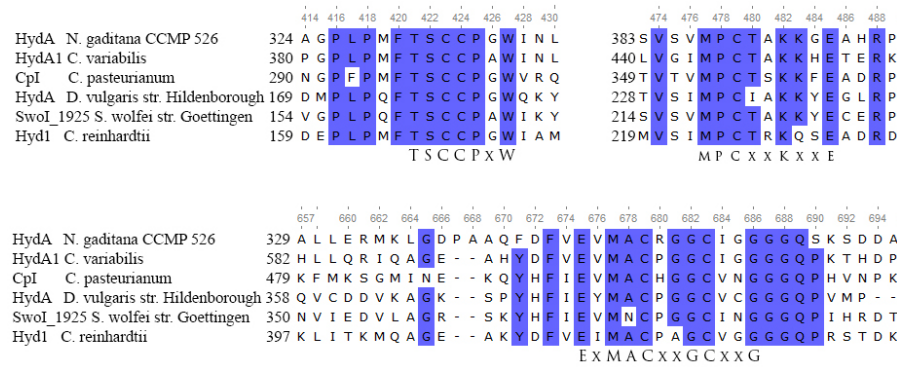
Figure 23: Alignment of [FeFe]-hydrogenases, showing the H-cluster motifs L1, L2, and L3.

have higher sequence similarity then [NiFe]-H$_2$ases, and are less functionally diverse.

8.2 EXPERIMENTAL

The *hoxH* subunit, which contains the active site, was the target for creating a knock-out strain of *Synechococcus* for the native hydrogenase.

To generate the deletion strain, the *hoxH* sequence was replaced with an antibiotic resistance cassette, functioning as a selection marker.
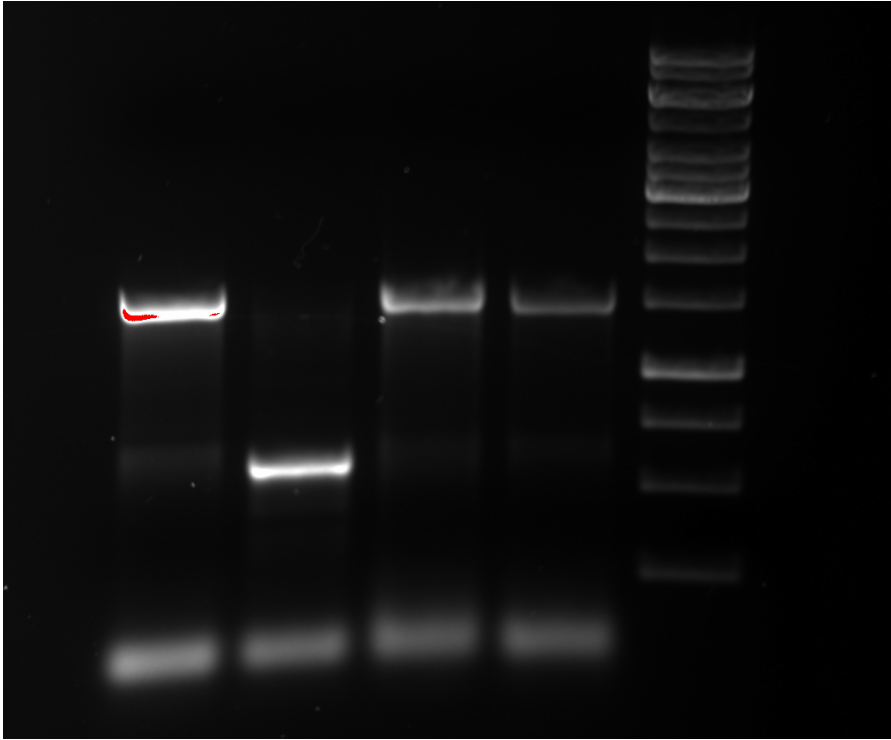
Figure 24: Gel separation of four PCR amplifications of the *hoxH* deletion plasmid insert.

The *hoxH* deletion plasmid was sequenced after inserted of the kanamycin resistance cassette, but the results were of low quality and could not be aligned against the planned plasmid. However, the primers used did enable the amplification of the approximately 1.6 kbp insert, as seen in three out of four lanes in Figure 24. Transformation of *Synechococcus* with the linearized DNA fragment was attempted.

Only successfully transformed algae should be able to grow on the selective media. *Synechococcus* was allowed to grow for 1 week on mBG-11 media containing kanamycin, however no colonies were seen. The transformation was therefore considered unsuccessful.

9

DISCUSSION

9.1 $\Delta hoxH$ *Synechococcus*

Transformation was unsuccessful, and no $\Delta hoxH$-strain of *Synechococcus* was created. For transformation into the chromosome of *Synechococcus,* it should be noted that the flanking regions chosen were shorter then ideal. Previous literature has recommended at least 500 bp flanking regions on either side of the gene (Xu et al., 2011). The synthetic knock-out plasmid that was ordered contained flanking regions only 250 bp in length. These flanking regions were extended using extra-long primers, making the final length of the flanking regions just under 300 bp each. Successful transformations have been accomplished with flanking regions of 250 bp, but longer homologous regions is proportional to increased transformation efficiency (Ruffing et al., 2015). Transformation most probably failed due to the problems with the *Synechococcus* transformation protocol mentioned earlier.

Theoretically, the synthetic hydrogenase could be inserted into a $\Delta hoxH$-strain of *Synechococcus* using a functional copy of the gene expression plasmid designed in this thesis. The *hoxH* knock-out plasmid contained a resistance cassette for kanamycin as its marker, and so did the gene expression plasmid. Remaking the *hoxH* plasmid with a different marker would allow for selection of a double mutant that only contains the new, synthetic [FeFe]-hydrogenase instead of the native [NiFe]-hydrogenase.

The creation of a highly hydrogen-producing organism depends on many parameters, including hydrogenase expression and oxygen tolerance. While there is evidence that cyanobacteria can produce enzymatically active [FeFe]-H$_2$ases despite not having the associated maturation proteins, this would still not create a strain capable of sus-

tained biohydrogen production. The increased growth rate of *Synechococcus* might make it a more viable alternative for biohydrogen production then *C. reinhardtii* (currently the most efficient hydrogen producer,) and as such the experiment might be worth doing. This *hyd1-Synechococcus* might be able to produce H_2 using the same two step process that works in *C. reinhardtii*, alternating between photosynthesis and biomass accumulation, and anaerobic consumption and H_2-production.

One possible advantage of expressing an aerotolerant [NiFe]-H_2ase in cyanobacteria instead of an [FeFe]-H_2ase is creating a one step system for producing hydrogen. Cyanobacteria seem to constantly express their hydrogenases, in aerobic and anaerobic conditions, but the presence of oxygen immediately stops their activity (Eckert et al., 2012). There does not seem to be any other regulation on the activity of the hydrogenases. If hydrogenases could be produced that were active in the presence of oxygen, a one step process for biohydrogen production might be developed, allowing some of the electrons produced during photosynthesis to be immediately shuttled to hydrogen production instead of the Calvin cycle or nitrate fixation, the other two important electron sinks (Gutekunst et al., 2014). In spite of their slower turnover rate than [FeFe]-H_2ases, the increased O_2-tolerance of [NiFe]-H_2ases might make them more potent for bio-hydrogen production.

Class 5 [NiFe]-H_2ases are aerotolerant, and some are completely O_2-insensitive. The class 5 [NiFe]-hydrogenases are solely involved in H_2 uptake, but they are closely related to the H_2-producing hydrogenases in cyanobacteria. Recent studies have shown two properties of [NiFe]-hydrogenases that are important for their aerotolerance. The gas diffusion tunnels of some aerotolerant hydrogenases have different amino acids at the entrance, replacing valine and leucine with isoleucine and phenylalanine. These much larger amino acids seem to function to reduce the size of the entrance, allowing H_2 to pass but blocking O_2 diffusion. For example , the amino acids at positions 62 and 110 of the hydrogenase large subunit, HoxC, of the *R. eutropha* H_2-sensing [NiFe]-H_2ase have been shown to be crucial for its oxygen tolerance (Buhrke et al., 2005). Some O_2-tolerant [NiFe]-H_2ases have a different method of insensitization, containing a unique [4Fe-3S] subcluster close to the active site, like the membrane-bound "knallgas" hydrogenase also found in *R. eutropha*. This cluster replaces the [4Fe-4S] cubane closest to the active site, and has two roles; Like the more common cubane, it functions to ferry electrons in and out of the active site during normal hydrogen oxidation, but it also has the ability to reduce oxygen, forming water. It does this before oxygen can inhibit the nearby catalytic center. Additional differences in the hydrogenase structure allow the newly formed water to be transported out of the active site in a directionally controlled manner (Fritsch et al., 2011).

For bio-engineering an efficient hydrogen-producing H_2ase, restricting the entrances to the diffusion tunnels to exclude oxygen from interacting with the active site might be the easiest method to implement. Comparing the 3d models of a cyanobacterial HoxH and *R. eutropha*'s H_2- sensing H_2ase HoxC subunit, the amino acids that infer oxygen tolerance at the entrance to the gas diffusion channels in *R. eutropha's* H_2ase could be mapped to those in the cyanobacterial ones. A modified HoxH could be designed to replace the one native to the cyanobacteria. The opposite of this experiment has already been done; modifying *R. eutropha*'s HoxC to become oxygen sensitive (Buhrke et al., 2005), and a more oxygen tolerant H_2-evolving [NiFe]-hydrogenase has been expressed in *E. coli* (Huang et al., 2015). Restricting oxygen from entering the active site seems simpler then modifying the cubane sub-cluster. Enzymes with the modified cubane also have gated water-transport channels, meaning even more mutations would be necessary compared to the native cyanobacterial hydrogenase, increasing the complexity of the experiment. In addition, the reduction of oxygen to water requires energy, reducing the efficiency of such a system.

Part IV

CONCLUSION

10

CONCLUSION

A plasmid was designed and created to allow for the expression of genes in *Synechococcus*. This plasmid successfully promoted the expression of a marker gene, *mRFP1*, in *E. coli*, but transformation into *Synechococcus* was unsuccessful. Additionally, a hydrogenase deficient strain of *Synechococcus* was designed, and the deletion plasmid created. Unfortunately, using this plasmid to transform *Synechococcus* was also unsuccessful. The [FeFe]-hydrogenase gene, *hyd1*, from *C. reinhardtii*, the most efficient biohydrogen producer, was chosen for to replace the endogenous [NiFe]-hydrogenase in *Synechococcus*. A synthetic, codon-optimized version of *hyd1* was designed, but no further progress was made.

10.1 OBJECTIVE 1: DESIGNING AND CREATING A GENE EXPRESSION CASSETTE

The gene expression cassette was able to successfully transform *E. coli*, forming red colonies. Sequencing of this plasmid was unsuccessful, however, the red colonies of the *E. coli* transformants and their resistance to kanamycin indicate that the gene expression plasmid might have been correct. The transformation of the expression cassette into *Synechococcus* was unsuccessful, nor were any other *Synechococcus* transformations.

A number of differences between the protocol described earlier and those used in the literature were noted. The very high light intensities used might have had a number of deleterious effects on transformation, including making the culture enter stationary phase early and

decreasing the activity of the DNA uptake machinery. This might have been the most important difference from the literature, and a new protocol should take that into account. Transformation efficiency as a factor of light intensity has never been measured in *Synechococcus*, or other related cyanobacteria. Optimization of light intensity during DNA uptake might be an interesting future experiment, although successful transformations have been reported at light intensities from 60 - 250 μmol photons m^{-2}s^{-1}.

10.2   OBJECTIVE 2: CREATING THE $\triangle hoxH$ *Synechococcus* STRAIN

A plasmid containing the flanking regions around *hoxH* was ordered, and *aph* was placed inside. *E. coli* was successfully transformed with this new plasmid. However, the linearized DNA fragment was not able to transform *Synechococcus*.

To allow for the selection of a double mutant, the knock-out plasmid should be re-ligated with a different marker, like *aadA1*, before transforming *Synechococcus*.

The gene for an [FeFe]-hydrogenase, *hyd1* from *C. reinhardtii*, was modified for expression in *Synechococcus*. Sequence analysis was used to identify features of the [FeFe]-hydrogenase such as the active site and introns. The amino acid sequence for Hyd1 was reverse-translated into nucleotides using the most codons for *Synechococcus*. The designed gene is both codon optimized, and lacks eukaryotic introns. These modifications should hopefully allow for efficient expression in *Synechococcus*.

10.3   FUTURE WORK

Getting *Synechococcus* to transform was the biggest problem encountered, and methods for optimizing the transformation protocol were discussed. The gene expression cassette was finalized, and is ready to be amplified and used to transform *Synechococcus*. The knock-out plasmid should be re-ligated with a different antibiotic cassette, then used to create the $\triangle hoxh$ strain with the new transformation protocol. A new synthetic hydrogenase with a much higher turn-over

rate was designed to replace the endogenous [NiFe]-hydrogenase, but other possible synthetic hydrogenases were discussed. Modifying a [NiFe]-hydrogenase for aerotolerance looks like a promising method to increase biohydrogen production in *Synechococcus*.

The biggest challenge in biohydrogen production is due to the oxygen intolerance of the H_2-evolving H_2ases. Either hydrogen has to be produced during anaerobic conditions, or the hydrogenase has to be engineered to tolerate oxygen.

The current most efficient method for producing hydrogen in a photosynthetic microbe is the alternating, two-step process in *C. reinhardtii*. This process separates photosynthetic oxygen evolution and accumulation of biomass from anaerobic hydrogen production and consumption of photosynthetic products (Esquível et al., 2011). This method might be effective in *Synechococcus*. Combining the hydrogenase native to *C. reinhardtii* and *Synechococcus'* much higher accumulation of biomass might lead to an even more efficient method for producing biohydrogen. The shift from aerobic to anoxic conditions in *C. reinhardtii* takes 24 hours, but this might be much quicker in *Synechococcus* due to its quicker growth rate. The two step-process is necessary in *C. reinhardtii* because of the oxygen intolerance of the [FeFe]-hydrogenase, and a similar problem would manifest itself with cyanobacterial [NiFe]-hydrogenases.

Engineering a hydrogen-evolving, oxygen-tolerant strain of hydrogenases might allow for a simpler one-step process, where some of the electrons formed during photosynthesis can be immediately used to produce hydrogen. [FeFe]-H_2ases are all energetically poised for hydrogen production, but are irreversibly inhibited by the presence of oxygen. The [NiFe]-H_2ases are a more diverse group, being able to oxidize or reduce $H_2$with varying degrees of aerotolerance. The cyanobacterial H_2ases are H_2-evolving, but sensitive to the presence of oxygen. There are several oxygen-tolerant [NiFe]-H_2ases, but these all oxidize H_2. A bioengineered [NiFe]-H_2ase might be able to combine these properties into one oxygen tolerant, H_2-evolving H_2ase.

HoxC from *R. eutropha* is oxygen-tolerant, but H_2-oxidizing. The aerotolerance of this enzyme seems to come from the presence of the large amino acids isoleucine and phenylalanine at the entrance to the gas diffusion channels (Buhrke et al., 2005). Studies have shown that modifying the amino acids at the entrances to these diffusion channels affects the oxygen tolerance, larger amino acids decrease the probability that O_2 can diffuse into the center and reach with the catalytic site, inhibiting the enzyme. This has been used to increased the oxygen tolerance of a hydrogen-evolving [NiFe]-hydrogenase before (Huang et al., 2015). A similar approach could be used to increase the oxygen tolerance of the native [NiFe]-hydrogenase of *Synechococcus*.

Once an oxygen-tolerant species of H_2-producing [NiFe]-H_2ase has been developed, then methods for increasing its electron flux could

be designed. In cyanobacteria, the H_2ases appear to function as a method for disposing of electrons generated from photosynthesis when other, more efficient methods are unavailable. It is active only when the Calvin cycle and nitrogen fixation are blocked, liked during algal bloom. An ideal H_2-producing cyanobacteria would have an initial growth phase, before permanently switching to a H_2-producing stationary phase. During the production phase, the majority of electrons should be shuttled to the hydrogenase to maximize efficiency. In wild type cyanobacteria, the presence of oxygen effectively downregulates the hydrogenase, but how a mutant with an aerotolerant H_2ase would handle competing electron valves is unknown. Conditions could be optimized to down-regulate nitrogen fixation and the Calvin cycle, potentially by the addition of glucose or glycogen and nitrate-starvation, but care should be taken to allow the cyanobacteria to stay in the stationary phase.

Although algal biofuels are not yet currently competitive with fossil fuels, researchers expect them to become so in the next decade or so (Wijffels and Barbosa, 2013). *Synechococcus* in particular looks like a very promising source for producing biofuels, but there are many other interesting algae and cyanobacteria. As these platforms get developed further, expect these advances to be reflected in biohydrogen production. Biohydrogen and its production still has many disadvantages and challenges to overcome, but the advantages of a sustainable, carbon-sequestering biofuel that doesn't compete with food or water, and is possible to grow world-wide, is an extremely exciting prospect that deserves to be researched.

Part V

APPENDIX

A

APPENDIX

A.1 SUPPLEMENTARY INFORMATION FOR CREATING THE GENE EXPRESSION CASSETTE

Including strains used throughout the study.

Table 6: Strains and Plasmids used in this study.

BACTERIAL STRAIN OR PLASMIDS	PURPOSE
pUC19	expression vector backbone
pSBA3 - BBa_E1010[1]	cloning *mRFP1*
Synechocystis mutants	cloning antibiotic resistance cassettes
E. Coli DH5-α	plasmid transformation
Synechocystis sp. PCC6803[2]	cloning P_{cpc560} and T_{rbc}
Synechococcus sp. PCC 7002[3]	transformation

[1] From the iGEM Registry of Standard Biological Parts
[2] Donated by Julian Eaton-Rye, Department of Biochemistry, University of Otago
[3] Donated by Niels-Ulrik Frigaard, Photosynthetic Microbes Laboratory, University of Copenhagen

Table 7: Primers for synthesizing the *Synechococcus* over-expression vector.

PRIMER	SEQUENCE (5'-3')
M13_SwaI_Long_fw[1]	**ATTTAAAT**GTCGTGACTGGGAAAACC
M13_SwaI_Long_rv[1]	**ATTTAAAT**CATGGTCATAGCTGTTTCC
ChloR_fw	GAATTCGGCGCGCCGCTTCTAGAGATGAGACGTTGATCGGCACG
ChloR_rv	CTGCAGCGGCCGCCGCTAGTAATTCAGGCGTAGCACCAGGC
KanR_fw	GAATTCGGCGCGCCGCTTCTAGAGAGCACGTTGTGTCTCAAAATCTC
KanR_rv	CTGCAGCGGCCGCCGCTACTAGTATACAACCAATTAACCAATTCTG
SpecR_fw	GAATTCGGCGCGCCGCTTCTAGAGAAAACGGATGAAAGGCACGAA
SpecR_rv	CTGCAGCGGCCGCCGCTACTAGTATTATTTGCCGACTACCTTGG
FlankB_fw	CTCTCACCAAAGATTCACCTG
FlankB_EcoRV_BBp_rv	CTCTAGAAGCGGGCGCCGCGAATTC**GATATC**GCCTCCTGAATAAAATCTATTTATTTATAC
FlankA_BBs_fw	ATACTAGTAGCGGCCGCGCTGCAGCTTTCTTCTTATGCACAGATGG
FlankA_rv	GGGGTTTCTCGTGTTTAGGC
P_cpc560_fw	ACCTGTAGAGAAGAGTCCCTGAA
P_cpc560_XbaI_EcoRV_BBp_rv	CTCTAGAAGCGGGCGCCGCGAATTCC**GATATC TCTAGA**TGAATTAATCTCCTACTTGACT
T_rbc_BBs_fw	ACTAGTAGCGGCCGCTGCAGACCGGTGTTTGGATTGTCGG
T_rbc_rv	GCTGTCGAAGTTGAACATCAG
SEQ_fw	GCTGGGCGAAAGGGGGATGTGCTGC
SEQ_rv	CGTATGTTGTGTGGAATTGTGAGC

[1]Phosphorylated on the 5'-end to increase ligation efficiency
Introduced restriction sites in **bold**, BioBrick sequences in *italics*.
Biobrick sequences from T Knight 2003
Antibiotic primer sequences modified from Poteete *et al.* 2006
Synechococcus flanking region primers modified from Xu *et al.* 2011
P_cpc560 and T_rbc primers modified from Zhou *et al.* 2013
Primers were ordered from either Integrated DNA technologies (IDT, Singapore) or Sigma Life Sciences (Sigma Aldrich, Germany)

Table 8: A^+medium for growing *Synechococcus* and differences from the mBG-11 medium used.

COMPONENT	CONCENTRATION		(g/L)
	A^+	mBG-11	Deviance from A^+
NaCl	18	18	-
KCl	0.6	0.596	-
$NaNO_3$	1.0	15	+ 14
$MgSO_4 \cdot 7H_2O$	5.0	7.92	+ 2.92
KH_2PO_2	0.5	0.4	- 0.1
$CaCl_2 \cdot 2H_2O$	0.36	0.36	-
NaEDTA	0.03	0.01	- 0.02
vit B_{12}	4 μg/L	4 μg/L	-
Trace metals	1.0 mL/L[1]	1.0 mL/L[2]	-
Ferric ammonium sulphate	-	0.06	+ 0.06
$FeCl_3 \cdot 6H_2O$	3.89 mg/L	-	- 3.89 mg/L
Tris	1g	-	- 1

[1]Similar to the trace metals required for mBG-11

[2]Trace metal mixture shown in Table 5

A.3   SUPPLEMENTARY INFORMATION FOR CREATING THE $\triangle hoxH$ *Synechococcus* STRAIN

```
GAATTCGCAATCAAAAGGGAGATGCCCAAACCATTTTAGAACGAGGCTATAAAGAACTCACCGAAG
AACACCGATTACCCCAACAAATTACCGGAGGAATTTTGCCGCCTTTATTACCCAGAGTTTTGCCCA
TCCATGAAGTCGTTGATATCGATTTATTTCTCCCCGGTTGCCCCCCCGATGCCGACCGAATTAAAG
CGGCGATCGCCCCCTTATTAGAAGGAAAATTGCCGGAAATGGAAGGACGAGAAATGATCAAATTCG
GCTAGATTTTATGTCCCAAACATTTGCCATGCTGACTGCCGCCGATATCATGAATCCCAATGTGGT
GACGATCAAAGGTTTAGCGACCATTGCCAGTGCCACCCAATGTATGCGGGTGAATAAAACCCGTGT
ATTAATAGTTGATCGTCGCCATGTTCACGATGCCTACGGCATTTTGACCGCCACCGATATCGTCAG
TAAAGTGATCGCCTATGGTCGTGATCCAAGGGCAATACGGGTCTATGAAATTATGACCAAACCTTG
TATTTTCGTGAGTCCTGATCTGGCGGTTGAATATGTTGCCCGCCTATTTAGTCAATGGAATTTGCA
CAGTGCCCCCGTCATGACTGACAAACTTTTAGGGATAATTACCGTTGAAGATTTAATTAGCAAAAG
TGATTTTTTAGAACGTCCCAAAGAACTATTATTTGCCGCAGAAATGCAAGCCGCAATTCAAAAGAC
AAAATTAATCTGTCAAGAAAAAGGACATGATTCGAGTGATTGTATTCAAGCATGGGCAATGGTGGA
AGAACTACAAGCTAAAACCGCCTACCAACAATCTACAAAAGTAGATAAAACCGCCCTTGAAGAATA
TTTAGAAAAAAATCCAGAAGCCATCGATCATTTAATGGTTGATAATTGGTGTAGTGGTTAACAGTC
ATAATTCGTTAAAATAAAAACTATTTTAATCAACACATTTTTAGCGTTCGTTTGAAAAATCATCTA
GGAGTAAAAAATGTCTAAAACAATCATTATTGATCCCATGACTCGCATCGAAGGACATGCCAAAAT
TTCCATTTATCTAGATGATCAAGGAGAAGTGCATGAAGCACGTTTTCATGTCGGGGAATTTCGGGG
GTTTGAGAAATTTTGTGAAGGTCGTCCTTTCTGGGAAATGCCTGCCTTAACGTCCCGAATCTGTGG
TATTTGTCCCGTCAGTCATCTAATTGCATCCTCAAAAACAGGCGATCAATTATTAGCTGTAAAGAT
TCCTGTTGGGGCAGGAAAATTACGGCGAATGATTAATCTTGCCCAGATCACTCAATCCCATGCCCT
CAGTTTTTTCCACCTTAGTAGTCCTGATTTAATCTTTGGTTGGGACAGTGATCCGAAGACGAGAAA
TGTATTGGTTTAATTGCCGCTGATCCTGATTTAGCACGGGGTGGAATTCGTCTGAGAAAATTTGG
ACAGGACATTATTAAAATCCTTGGGAGCCAAAAAGTTCACCCCGCTTGGTCAATCCCCGGTGGCGT
GCGATCGCCGTTAACAAAGAAGGACAAACTTATATTAAAGAGGGTTTGCCGGAAGCAAAAAATAC
CGTTAAAAATGCCCTGAGTTTATTTAAAAAAATTCTTGACTCCCATCAGGAAGAAGTGACAGTTTT
CGGTAATTTTCCGAGTCTTTATCTGGGTTTAATAGGCGAGTCCGGCGCATGGGAACATTATGGCGG
CATTTTGCGGATGATCGATAGCCACGGCAATATTGTGGGCGATCGCCTTGATCCGAACAATTACCA
TGAATTTCTGGGGGAAAAATTACAGGAAGATTCCTACTTAAAATCCCCTTACTATAAGCTTTTTGG
GGAGACAGGGGTATATCGCGTCGGGCCTTTAGCACGGTTAAATATTTGCGAACATTTTGGCACCGA
AGCCGCCGATCAAGAACTCATTGAATATCGGCAACGACACGGCAGAATTGTCCAAGCTTCCTTTGT
TTATCACCACGCGAGACTCATTGAAATATTAGGTTCCCTCGAAAGAATTGAGCGGATGATCGATGA
TCCAGATCTTTTTTCTAACCGTTTGCAAGCAGAGGCAGGGGTTAATCAAACCGAAGCCGTTGGCGT
GAGTGAGGCACCACGAGGAACCCTTTTTCATCATTATCAAGTGGATGAAATGGACTGCTCAAAAA
GATAAATTTAGTGATTGCGACAGGACAAAATAACTTTGCTATTAACCGCACCGTGACCCAAATTGC
CAAGCATTATATTCATGGTGAAACTGTTGCAGAAGGCATTTTAAATCGAGTGGAGGCTGGGGTGAG
AAATTATGATCCTTGTTTAAGTTGTTCGACCCATGCCGCCGGACAAATGCCGATGATTTTACAATT
AATTTCCCCTGATGGTTCCATTGTCAAAGAATTCGTCGTGATAGTTAATGAAACAAACTTTAATG
ATTGGTTACGGCAATACATTACGTTCTGATGATGGGGCAGGACAAAAGGTTGCGGAAGCTTTTTTT
GATCAGGAAAATATAACGGCGATCGCTACCCATCAACTCACCCCTGAATTAGTAGAAGATTTGGTA
CAAGTGGAGCAGGTTTACTTTATCGATGCGGCTCCCATTGAGACTGTAACGATTAAACCTATTCAA
CGGAGAGATAATGAACACAATTTTGGACATTTTATTGACCCCAAAAGCCTTCTGAATTTAGCTCAA
GAAATTTATCATTATGCGCCCGATGCTTACTTGGTTTTAATCCCTGCTCAAGATTTTAAACTGGGA
GAAAACTATTCTGAAATAACCCAAAAAGCAATCGAAACAGCGATTCATCTTTTGCAAGAACGACTC
ACACCATGCATGAAGTAGCCATTATGACAGAAACCGTGGCGATCGCCAATGCAGCAGCAGAGCGAC
AAAATGCCACAAAAATTGTAGGTTTAACCATGCGAATTGGTGCAATCAGTGGTGTCGTTCCTGAAG
CCCTGAGTTTTGCCTTTGAAGCGGTGGCAGGCGGAACCCTTGCCGAACAGGCTCAACTAATCATTG
AAACGGTTCCTGTGACCTGTTATTGTGCCCAATGCGATCGCCCTTTACACCCCCTGATTTATTTT
ATGAATGTCCGTTATGTAGTAGTTTAAGTCAACATATTTTAAGTGGTAAAGAAGTCGAATTAAAAT
CCTTAGAGGTGATTTAAATTATGTGTGGAAATTGCCGGTTGTAACGCTGTTGAAAAACCTGTGGAAA
TTCATACCCACGCCCATGATCATTCCCACGGGCATCATGACCATCATCACCCCCATGACCATGAAC
ATCACCATCACCATAATCATGATGGGAACCATAGCCATGCCATTGATATTAATCAATCTATTTTTG
CAAAAAATGATCGTCTAGCTGAGCGTAATCGCGGCTATTTTTTGGCGAAAGATTTATTT
```

Figure 25: Section of *Synechococcus'* genome immediately surrounding *hoxH*, bases ["209491"]-["212915"], containing 1000bp to either side of *hoxH.* Flanking regions used are shown in green, *hoxH* is shown in magenta, and primers are shown underlined. (Nakao et al., 2010)

Table 9: Primers for amplifying the △*hoxH* *Synechococcus* transformation vector.

PRIMER	SEQUENCE (5′-3′)
KO_fw	GCAAGCCGCAATTCAAAAGACAAAATTAATCTGTCAAGA AAAAGGACATGATTCGAGTGATTGTATTCAAGCATGGGC
KO_rv	GCATAATGATAAATTTCTTGAGCTAAATTCAGAAGGCTTT TGGGGTCAATAAAATGTCCAAAATTGTGTTCATTATCTCT CCG

Table 10: A comparison of the most common codon used for each amino acid in *C. reinhardtii* and *Synechococcus*.

AMINO ACID	*C. reinhardtii*		*Synechococcus*		DIFFERENT
	Codon	Fraction	Codon	Fraction	
A	GCC	0.43	GCC	0.42	
R	CGC	0.62	CGC	0.35	
N	AAC	0.91	AAU	0.61	x
D	GAC	0.86	GAU	0.69	x
C	UGC	0.9	UGU	0.60	x
Q	CAG	0.9	CAA	0.59	x
E	GAG	0.95	GAA	0.76	x
G	GGC	0.72	GGC	0.34	
H	CAC	0.89	CAC	0.53	
I	AUC	0.75	AUU	0.42	x
L	CUG	0.73	CUC	0.27	
K	AAG	0.95	AAA	0.66	x
M	AUG	1	AUG	1	
F	UUC	0.84	UUC	0.67	
P	CCC	0.47	CCC	0.47	
S	AGC	0.35	AGU	0.21	x
T	ACC	0.52	ACC	0.48	
W	UGG	1	UGG	1	
Y	UAC	0.9	UAU	0.44	x
V	GUG	0.67	GUG	0.37	

Data tabulated from (Nakamura et al., 2000)

Table 11: Codon usage in *Synechococcus*.

CODON	AMINO ACID	FRACTION	CODON	AMINO ACID	FRACTION	CODON	AMINO ACID	FRACTION	CODON	AMINO ACID	FRACTION
UUU	F	0.67	UCU	S	0.18	UAU	Y	0.56	UGU	C	0.60
UUC	F	0.33	UCC	S	0.20	UAC	Y	0.44	UGC	C	0.40
UUA	L	0.16	UCA	S	0.07	UAA	*	0.56	UGA	*	0.15
UUG	L	0.19	UCG	S	0.13	UAG	*	0.28	UGG	W	1.00
CUU	L	0.10	CCU	P	0.16	CAU	H	0.47	CGU	R	0.21
CUC	L	0.27	CCC	P	0.47	CAC	H	0.53	CGC	R	0.35
CUA	L	0.09	CCA	P	0.15	CAA	Q	0.59	CGA	R	0.10
CUG	L	0.19	CCG	P	0.22	CAG	Q	0.41	CGG	R	0.28
AUU	I	0.56	ACU	T	0.15	AAU	N	0.61	AGU	S	0.21
AUC	I	0.42	ACC	T	0.48	AAC	N	0.39	AGC	S	0.20
AUA	I	0.02	ACA	T	0.15	AAA	K	0.66	AGA	R	0.05
AUG	M	1.00	ACG	T	0.22	AAG	K	0.34	AGG	R	0.03
GUU	V	0.24	GCU	A	0.18	GAU	D	0.69	GGU	G	0.30
GUC	V	0.27	GCC	A	0.42	GAC	D	0.31	GGC	G	0.34
GUA	V	0.11	GCA	A	0.14	GAA	E	0.76	GGA	G	0.11
GUG	V	0.37	GCG	A	0.26	GAG	E	0.24	GGG	G	0.25

Data tabulated from (Nakamura et al., 2000).

A.5 SEQUENCING RESULTS

Sequencing results of plasmid 2, isolated from colony 3, using primer FlankB_fw (called S for subject) against the planned plasmid sequence (Q). Notice the inserted bases at positions 889 and 891 of S (Zhang et al., 2000).

>Plasmid 2, colony 3 FW [50-1030]
Score = 1788 bits (968), Expect = 0.0
Identities = 973/975 (99%), Gaps = 2/975 (0%)

```
Q   2607   TCTCACCAAAGATTCACCTGTTAGAGCTACTCAACATCCATCAGTTCTTAAAACCAGGGG

           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

S   1      TCTCACCAAAGATTCACCTGTTAGAGCTACTCAACATCCATCAGTTCTTAAAACCAGGGG


Q   2667   TGACATTCACCGGGGCGAGCCTTGAAGGGTTCAAGGAAAATTGTTTGCGGTATGCCAAGC

           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

S   61     TGACATTCACCGGGGCGAGCCTTGAAGGGTTCAAGGAAAATTGTTTGCGGTATGCCAAGC


Q   2727   CGATCAAGTGGATTCTTGGCAGAACGATCACCGACAAAATGAGCCCGCTCGAAATTGCTC

           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

S   121    CGATCAAGTGGATTCTTGGCAGAACGATCACCGACAAAATGAGCCCGCTCGAAATTGCTC


Q   2787   AGGCGCTCCTAGGCAAGCTTGACCGGAAATTGGAATACAAGGGGCGCTTTGGATCGCGGG

           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

S   181    AGGCGCTCCTAGGCAAGCTTGACCGGAAATTGGAATACAAGGGGCGCTTTGGATCGCGGG


Q   2847   ATAACCGTCAGCGGGTCTATGAGGCGATCGCCCCTAACGATCAGCGCGAAAAGGTCTTTG

           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

S   241    ATAACCGTCAGCGGGTCTATGAGGCGATCGCCCCTAACGATCAGCGCGAAAAGGTCTTTG


Q   2907   CTCATTGGTTACAGCGTGACCAAGCAAAATTAGGGGCCGTGTCCAACCCCTGTATAAATA

           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

S   301    CTCATTGGTTACAGCGTGACCAAGCAAAATTAGGGGCCGTGTCCAACCCCTGTATAAATA


Q   2967   GATTTATTCAGGAGGCGATACCTGTAGAGAAGAGTCCCTGAATATCAAAATGGTGGGATA

           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

S   361    GATTTATTCAGGAGGCGATACCTGTAGAGAAGAGTCCCTGAATATCAAAATGGTGGGATA


Q   3027   AAAAGCTCAAAAAGGAAAGTAGGCTGTGGTTCCCTAGGCAACAGTCTTCCCTACCCCACT

           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

S   421    AAAAGCTCAAAAAGGAAAGTAGGCTGTGGTTCCCTAGGCAACAGTCTTCCCTACCCCACT


Q   3087   GGAAACTaaaaaaaaCGAGAAAAGTTCGCACCGAACATCAATTGCATAATTTTAGCCCTAA
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
S    481    GGAAACTAAAAAAACGAGAAAAGTTCGCACCGAACATCAATTGCATAATTTTAGCCCTAA

Q   3147    AACATAAGCTGAACGAAACTGGTTGTCTTCCCTTCCCAATCCAGGACAATCTGAGAATCC
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
S    541    AACATAAGCTGAACGAAACTGGTTGTCTTCCCTTCCCAATCCAGGACAATCTGAGAATCC

Q   3207    CCTGCAACATTACTTAACAAAAAAGCAGGAATAAAATTAACAAGATGTAACAGACATAAG
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
S    601    CCTGCAACATTACTTAACAAAAAAGCAGGAATAAAATTAACAAGATGTAACAGACATAAG

Q   3267    TCCCATCACCGTTGTATAAAGTTAACTGTGGGATTGCAAAAGCATTCAAGCCTAGGCGCT
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
S    661    TCCCATCACCGTTGTATAAAGTTAACTGTGGGATTGCAAAAGCATTCAAGCCTAGGCGCT

Q   3327    GAGCTGTTTGAGCATCCCGGTGGCCCTTGTCGCTGCCTCCGTGTTTCTCCCTGGATTTAT
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
S    721    GAGCTGTTTGAGCATCCCGGTGGCCCTTGTCGCTGCCTCCGTGTTTCTCCCTGGATTTAT

Q   3387    TTAGGTAATATCTCTCATAAATCCCCGGGTAGTTAACGAAAGTTAATGGAGATCAGTAAC
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
S    781    TTAGGTAATATCTCTCATAAATCCCCGGGTAGTTAACGAAAGTTAATGGAGATCAGTAAC

Q   3447    AATAACTCTAGGGTCATTACTTTGGACTCCCTCAGTTTATCCGGGGG-AA-TTGTGTTTA
            ||||||||||||||||||||||||||||||||||||||||||||||| || ||||||||||
S    841    AATAACTCTAGGGTCATTACTTTGGACTCCCTCAGTTTATCCGGGGGGAATTGTGTTTA

Q   3505    AGAAAATCCCAACTCATAAAGTCAAGTAGGAGATTAATTCATCTAGAGATATCGGAATTC
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
S    901    AGAAAATCCCAACTCATAAAGTCAAGTAGGAGATTAATTCATCTAGAGATATCGGAATTC

Q   3565    GCGGCCGCTTCTAGA
            |||||||||||||||
S    961    GCGGCCGCTTCTAGA
```

Alignment for sequencing from reverse side of insert, using FlankB_rv, is not shown. However, the score is shown below, and it has 100% identity to the expected sequence (Zhang et al., 2000).

>Plasmid 2, colony 3 RV [40-1000]
Score = 1772 bits (959), Expect = 0.0
Identities = 959/959 (100%), Gaps = 0/959 (0%)

BIBLIOGRAPHY

S. Aikawa, A. Nishida, S. H. Ho, J. S. Chang, T. Hasunuma, and A. Kondo. Glycogen production for biofuels by the euryhaline cyanobacteria synechococcus sp. strain pcc 7002 from an oceanic environment. *Biotechnology for Biofuels*, 7, 2014. (Cited on pages 10 and 16.)

I. Akkermana, M. Janssenb, J. Rochac, and R. H. Wijffelsd. Photobiological hydrogen production photochemical efficiency and bioreactor design. *International Journal of Hydrogen Energy*, 27, 2002. (Cited on page 6.)

Y. Asadaa, Y. Koikeb, J. Schnackenbergb, M. Miyakee, I. Uemurac, and J. Miyaked. Heterologus expression of clostridial hydrogenase in the cyanobacterium synechococcus pcc7942. *Biochimica et Biophysica Acta*, 1490, 2000. (Cited on page 52.)

C. V. Baalen, D. S. Hoare, and E. Brandt. Heterotrophic growth of blue-green algae in dim light. *Journal Of Bacteriology*, 105, 1971. (Cited on page 9.)

A. Bahadar and M. B. Khan. Progress in energy from microalgae: A review. *Renewable and Sustainable Energy Reviews*, 27, 2013. (Cited on pages 5 and 6.)

J. C. Batterton and C. Van Baalen. Growth responses of blue-green algae to sodium chloride concentration. *Archives of Microbiology*, 76, 1971. (Cited on page 9.)

S. D. Bentley, K. F. Chater, C.-T. A. M., G. L. Challis, N. R. Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C. H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, O. S, E. Rabbinowitsch, M. A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B. G. Barrell, J. Parkhill, and D. A. Hopwood. Complete genome sequence of the model actinomycete streptomyces coelicolor a3(2). *Nature*, 417, 2002. (Cited on page 52.)

M. Berney, C. Greening, K. Hards, D. Collins, and G. M. Cook. Three different [nife] hydrogenases confer metabolic flexibility in the obligate aerobe mycobacterium smegmatis. *Environmental Microbiology*, 16, 2014. (Cited on page 8.)

P. Bertoa, S. D'Adamoa, E. Bergantinoa, F. Vallesea, G. M. Giacomet-
tia, and P. Costantini. The cyanobacterium synechocystis sp. pcc
6803 is able to express an active [fefe]-hydrogenase without ad-
ditional maturation proteins. *Biochemical and Biophysical Research
Communications*, 405, 2011. (Cited on page 52.)

D. Bilanovic, A. Andargatchew, T. Kroeger, and G. Shelef. Freshwa-
ter and marine microalgae sequestering of co2 at different c and
n concentrations - response surface methodology analysis. *Energy
Conversion and Management*, 50, 2008. (Cited on page 3.)

F. R. Blattner, G. P. III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley,
J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor,
N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau,
and Y. Shao. The complete genome sequence of escherichia coli
k-12. *Science*, 5, 1997. (Cited on page 51.)

U. Bossel. Does a hydrogen economy make sense? *Proceedings of the
IEEE*, 94, 2006. (Cited on page 5.)

BP. British petroleum statistical review of world energy 2014, 2014.
(Cited on page 3.)

L. Brennan and P. Owende. Biofuels from microalgae - a review of
technologies for production, processing, and extractions of biofuels
and co-products. *Renewable and Sustainable Energy Reviews*, 14, 2010.
(Cited on page 5.)

T. Buhrke, O. Lenz, N. Krauss, and B. Friedrich. Oxygen tolerance
of the h2-sensing [nife] hydrogenase from ralstonia eutropha h16 is
based on limited access of oxygen to the active site. *The Journal of
Biological Chemistry*, 280, 2005. (Cited on pages 66, 67, and 73.)

R. E. Campbell, O. Tourand, A. E. Palmer, P. A. Steinbach, G. S. Baird,
D. A. Zacharias, and R. Y. Tsien. A monomeric red fluorescent
protein. *Proceedings of the National Academy of Sciences of the United
States of America*, 99, 2002. (Cited on page 19.)

Y. Chisti. Biodiesel from microalgae. *Biotechnology Advances*, 2007.
(Cited on pages 4 and 5.)

A. F. Clarens, E. P. Resurreccion, M. A. White, and L. M. Colosi. En-
vironmental life cycle comparison of algae to other bioenergy feed-
stock. *Environmental Science and Technology*, 44, 2010. (Cited on
page 4.)

P. Constant, S. P. Chowdhury, L. Hesse, J. Pratscher, and R. Conrad.
Genome data mining and soil survey for the novel group 5 [nife]-
hydrogenase to explore the diversity and ecological importance of
presumptive high-affinity h2-oxidizing bacteria. *Applied and Envi-
ronmental Microbiology*, 77, 2011. (Cited on page 8.)

M. B. Dalgaard, M. E. Reaban, T. Coetzee, and J. Z. Prokaryotic introns and inteins: a panoply of form and function. *Journal of Bacteriology*, 177, 1995. (Cited on page 51.)

M. Date, M. Itaya, H. Matsui, and Y. Kikuchi. Secretion of human epidermal growth factor by corynebacterium glutamicum. *Letters in Applied Microbiology*, 42, 2006. (Cited on page 16.)

G. C. Dismukes, D. Carrieri, N. Bennette, G. M. Ananyev, and M. C. Posewitz. Aquatic phototrophs: efficient alternatives to land-based crops for biofuels. *Current Opinion in Biotechnology*, 19, 2008. (Cited on page 5.)

C. Eckert, M. Boehm, D. Carrieri, J. Yu, A. Dubini, P. J. Nixon, and P.-C. Maness. Genetic analysis of the hox hydrogenase in the cyanobacterium synechocystis sp. pcc 6803 reveals subunit roles in association, assembly, maturation, and function. *Journal of Biological Chemistry*, 287, 2012. (Cited on pages 10, 55, and 66.)

R. C. Edgar. Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32, 2004. (Cited on page 61.)

EIA. Monthly biodiesel production report, 2015a. (Cited on page 4.)

EIA. International energy statistics, 2015b. (Cited on page 3.)

M. G. Esquível, H. M. Amaro, T. S. Pinto, P. S. Fevereiro, and F. X. Malcata. Efficient h2 production via chlamydomonas reinhardtii. *Trends in Biotechnology*, 29, 2011. (Cited on pages 8, 51, and 73.)

E. Essich, E. S. Stevens, and R. D. Porter. Chromosomal transformation in the cyanobacterium agmenellum quadruplicatum. *Journal of Bacteriology*, 172, 1990. (Cited on pages 44 and 47.)

L. Fengrui, Z. Songling, and G. T. Geballe. Water use patterns and agronomic performance for some cropping systems with and without fallow crops in a semi-arid environment of northwest china. *Agriculture, Ecosystems & Environment*, 79, 2000. (Cited on page 4.)

N. U. Frigaard, Y. Sakuragi, and D. A. Bryant. Gene inactivation in the cyanobacterium synechococcus sp. pcc7002 and the green sulfur bacterium chlorobium tepidum using in vitro-made dna constructs and natural transformation. In R. Carpentier, editor, *Methods in Molecular Biology*, volume 274: Photosynthesis Research Protocols. Humana Press, 2004. (Cited on page 46.)

J. Fritsch, P. Scheerer, S. Frielingsdorf, S. Kroschinsky, B. Friedrich, O. Lenz, and C. M. Spahn. The crystal structure of an oxygen-tolerant hydrogenase uncovers a novel iron-sulphur centre. *Nature*, 479, 2011. (Cited on page 66.)

M. J. Gardner, N. Hall, E. Fung, O. White, M. Berriman, R. W. Hyman, J. M. Carlton, A. Pain, K. E. Nelson, S. Bowman, I. T. Paulsen, K. James, J. A. Eisen, K. Rutherford, S. L. Salzberg, A. Craig, S. Kyes, M. S. Chan, V. Nene, S. J. Shallom, B. Suh, J. Peterson, S. Angiuoli, M. Pertea, J. Allen, J. Selengut, D. Haft, M. W. Mather, A. B. Vaidya, D. M. Martin, A. H. Fairlamb, M. J. Fraunholz, D. S. Roos, S. A. Ralph, M. GI, L. M. Cummings, G. M. Subramanian, C. Mungall, J. C. Venter, D. J. Carucci, S. L. Hoffman, C. Newbold, R. W. Davis, C. M. Fraser, and B. Barrell. Genome sequence of the human malaria parasite plasmodium falciparum. *Nature*, 419, 2002. (Cited on page 52.)

G. E. Gasparich, J. Buzby, D. A. Bryant, R. D. Porter, and J. SE Stevens. The effects of light intensity and nitrogen starvation on the phycocyanin promoter in the cyanobacterium synechococcus pcc 7002. In J. Biggins, editor, *Progress in Photosynthesis Research*, pages 761–764. Springer Netherlands, 1987. ISBN 978-94-017-0521-9. (Cited on page 17.)

M. L. Ghirardi, M. C. Posewitz, P. C. Maness, A. Dubini, J. Yu, and M. Seibert. Hydrogenases and hydrogen photoproduction in oxygenic photosynthetic organisms. *Annual Review of Plant Biology*, 58, 2007. (Cited on page 7.)

S. J. Giovannoni, S. Turner, G. J. Olsen, S. Barns, D. J. Lane, and N. R. Pace. Evolutionary relationships among cyanobacteria and green chloroplasts. *Journal Of Bacteriology*, 170(8), 1988. (Cited on page 9.)

C. Gustafsson, S. Govindarajan, and J. Minshull. Codon bias and heterologous protein expression. *TRENDS in Biotechnology*, 22, 2004. (Cited on page 52.)

K. Gutekunst, X. Chen, K. Schreiber, U. Kaspar, S. Makam, and J. Appel. The bidrectional nife-hydrogenase in synechocystis sp. pcc 6803 is reduced by flavodoxin and ferredoxin and is essential under mixotrophic, nitrate-limiting conditions. *The Journal of Biological Chemistry*, 289, 2014. (Cited on pages 7, 10, 57, and 66.)

F. Gutthann, M. Egert, A. Marques, and J. Appel. Inhibition of respiration and nitrate assimilation enhances photohydrogen evolution under low oxygen concentrations in synechocystis sp. pcc 6803. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1767(2):161 − 169, 2007. ISSN 0005-2728. (Cited on page 10.)

P. C. Hallenbeck and J. R. Benemann. Biological hydrogen production; fundamentals and limiting processes. *International Journal of Hydrogen Energy*, 27, 2002. (Cited on page 8.)

P. C. Hallenbeck and D. Ghosh. Advances in fermentative biohydrogen production: the way forward? *Trends in Biotechnology*, 27, 2009. (Cited on page 6.)

Q. Hu, M. Sommerfeld, E. Jarvis, M. Ghirardi, M. Posewitz, M. Seibert, and A. Darzins. Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. *The Plant Journal*, 54, 2008. (Cited on page 4.)

G. F. Huang, X. B. Wu, L. P. Bai, K. Liu, L. J. Jiang, M. N. Long, and Q.-X. Chen. Improved o-2-tolerance in variants of a h-2-evolving [nife]-hydrogenase from klebsiella oxytoca hp1. *FEBS Letters*, 589, 2015. (Cited on pages 67 and 73.)

L. O. Ingram, C. Van Baalen, and J. A. Calder. Role of reduced exogenous organic compounds in the physiology of the blue-green bacteria (algae): photoheterotrophic growth of an "autotrophic" blue-green bacterium. *J Bacteriol*, 114, 1973. (Cited on page 9.)

K. L. Kadam. Environmental implications of power generation via coal-microalgae cofiring. *Energy*, 27, 2002. (Cited on page 5.)

J. F. Kane. Effects of rare codon clusters on high-level expression of heterologous proteins in escherichia coil. *Current Opinion in Biotechnology*, 6, 1995. (Cited on page 52.)

D. Kebede and G. Ahlgren. Optimum growth conditions and light utilization efficiency of spirulina platensis (equals arthrospira fusiformis) (cyanophyta) from lake chitu, ethiopia. *Hydrobioligia*, 332, 1996. (Cited on page 4.)

V. Khosla. Where will biofuels and biomass feedstocks come from? White paper, 2008. (Cited on page 4.)

E. Kimura. Metabolic engineering of glutamate production. *Advances in Biochemical Engineering/Biotechnology*, 79, 2003. (Cited on page 16.)

P. W. King, M. C. Posewitz, M. L. Ghirardi, and M. Seibert. Functional studies of [fefe] hydrogenase maturation in an escherichia coli biosynthetic system. *Journal Of Bacteriology*, 188, 2006. (Cited on page 52.)

R. D. Knight, S. J. Freeland, and L. F. Landweber. A simple model based on mutation and selection explains trends in codon and amino-acid usage and gc composition within and across genomes. *Genome Biology*, 2, 2001. (Cited on page 52.)

T. Knight. Idempotent vector design for standard assembly of biobricks. *Dspace*, 2003. (Cited on page 23.)

C. Madden, M. D. Vaughn, I. Díez-Pérez, K. A. Brown, P. W. King, D. Gust, A. L. Moore, and T. A. Moore. Catalytic turnover of [fefe]-hydrogenase based on single-molecule imaging. *Journal of the American Chemical Society*, 134, 2012. (Cited on page 7.)

A. Melis, L. Zhang, M. Forestier, M. L. Ghirardi, and M. Seibert. Sustained photobiological hydrogen gas production upon reversible inactivation of oxygen evolution in the green alga chlamydomonas reinhardtii. *American Society of Plant Physiologists*, 122, 2000. (Cited on page 8.)

S. Merchant, S. Prochnik, O. Vallon, E. Harris, S. Karpowicz, G. Witman, A. Terry, A. Salamov, L. Fritz-Laylin, L. Maréchal-Drouard, W. Marshall, L. Qu, D. Nelson, A. Sanderfoot, M. Spalding, V. Kapitonov, Q. Ren, P. Ferris, E. Lindquist, H. Shapiro, S. Lucas, J. Grimwood, J. Schmutz, P. Cardol, H. Cerutti, G. Chanfreau, C. Chen, V. Cognat, M. Croft, R. Dent, S. Dutcher, E. Fernández, H. Fukuzawa, D. González-Ballester, D. González-Halphen, A. Hallmann, M. Hanikenne, M. Hippler, W. Inwood, K. Jabbari, M. Kalanon, R. Kuras, P. Lefebvre, S. Lemaire, A. Lobanov, M. Lohr, A. Manuell, I. Meier, L. Mets, M. Mittag, T. Mittelmeier, J. Moroney, J. Moseley, C. Napoli, A. Nedelcu, K. Niyogi, S. Novoselov, I. Paulsen, G. Pazour, S. Purton, J. Ral, D. Riaño-Pachón, W. Riekhof, L. Rymarquis, M. Schroda, D. Stern, J. Umen, R. Willows, N. Wilson, S. Zimmer, J. Allmer, J. Balk, K. Bisova, C. Chen, M. Elias, K. Gendler, C. Hauser, M. Lamb, H. Ledford, J. Long, J. Minagawa, M. Page, J. Pan, W. Pootakham, S. Roje, A. Rose, E. Stahlberg, A. Terauchi, P. Yang, S. Ball, C. Bowler, C. Dieckmann, V. Gladyshev, P. Green, R. Jorgensen, S. Mayfield, B. Mueller-Roeber, S. Rajamani, R. Sayre, P. Brokstein, I. Dubchak, D. Goodstein, L. Hornick, Y. Huang, J. Jhaveri, Y. Luo, D. Martínez, W. Ngau, B. Otillar, A. Poliakov, A. Porter, L. Szajkowski, G. Werner, K. Zhou, I. Grigoriev, D. Rokhsar, and A. Grossman. The chlamydomonas genome reveals the evolution of key animal and plant functions. *Science*, 318, 2007. (Cited on page 58.)

J. Meyer. [fefe] hydrogenases and their evolution: A genomic perspective. *Cellular and Molecular Life Sciences*, 64, 2007. (Cited on page 7.)

F. Mignone, C. Gissi, S. Liuni, and G. Pesole. Untranslated regions of mrnas. *Genome Biology*, 3, 2002. (Cited on page 51.)

Y. Miura, S. Ohta, M. Mano, and K. Miyamoto. Isolation and characterization of a unicellular marine green alga exhibiting high activity in dark hydrogen production. *Agricultural and Biological Chemistry*, 50, 1986. (Cited on page 9.)

Y. Miura, C. Saitoh, S. Matsuoka, and K. Miyamoto. Stably sustained hydrogen production with high molar yield through a combination of a marine green alga and a photosynthetic bacterium. *Bioscience, Biotechnology, and Biochemistry*, 56, 1992. (Cited on page 9.)

A. Moore. Biofuels are dead: long live biofuels(?) - part one. *New Biotechnology*, 25, 2008. (Cited on page 4.)

D. W. Mulder, E. M. Shepard, J. E. Meuser, N. Joshi, P. W. King, M. C. Posewitz, J. B. Broderick, , and J. W. Peters. Insights into [fefe]-hydrogenase structure, mechanism, and maturation. *Structure*, 19, 2011. (Cited on page 7.)

Y. Nakamura, T. Gojobori, and T. Ikemura. Codon usage tabulated from the international dna sequence databases: status for the year 2000. *Nucl Acids Res*, 28, 2000. (Cited on pages 82 and 83.)

M. Nakao, S. Okamoto, M. Kohara, T. Fujishiro, T. Fujisawa, S. Sato, S. Tabata, T. Kaneko, and Y. Nakamura. Cyanobase: the cyanobacteria genome database update 2010. *Nucleic Acids Research*, 38, 2010. (Cited on pages 56 and 80.)

Phusion High-Fidelity PCR Kit. NEB, 2015a. (Cited on page 22.)

Quick Ligation Protocol. NEB, 2015b. (Cited on page 23.)

K. Okonechnikov, O. Golosova, M. Fursov, and the UGENE team. Unipro ugene: a unified bioinformatics toolkit. *Bioinformatics*, 28, 2012. (Cited on page 61.)

W. J. Oswald. High rate ponds in waste disposal. *Developments in industrial microbiology*, 4, 1963. (Cited on page 4.)

A. R. Poteete, C. Rosadini, and C. St. Pierre. Gentamicin and other cassettes for chromosomal gene replacement in escherichia coli. *BioTechniques*, 41, 2006. (Cited on page 26.)

T. L. Richard. Challenges in scaling up biofuels infrastructure. *Science*, 329, 2010. (Cited on page 4.)

A. Ruffing, T. J. Jensen, L. M. Strickland, S. Meserole, and D. Tallant. Systems-level synthetic biology for advanced biofuel production. Technical Report SAND2015-2319 579481, Sandia National Laboratories (SNL-NM), Albuquerque, NM (United States), 2015. (Cited on pages 46, 47, and 65.)

A. M. Ruffing. Improved free fatty acid production in cyanobacteria with synechococcus sp. pcc 7002 as host. *Frontiers in Bioengineering and Biotechnology*, 2, 2014. (Cited on pages 10 and 16.)

T. Sakamoto and D. A. Byant. Synergistic effect of high-light and low temperature on cell growth of the delta12 fatty acid desaturase mutant in synechococcus sp. pcc 7002. *Photosynth Res*, 72, 2002. (Cited on page 9.)

C. Schäfer, B. Friedrich, and O. Lenz. Novel, oxygen-insensitive group 5 [nife]-hydrogenase in ralstonia eutropha. *Applied and Environmental Microbiology*, 79, 2013. (Cited on page 7.)

N. C. Shaner, R. E. Campbell, P. A. Steinbach, B. N. G. Giepmans, A. E. Palmer, and R. Y. Tsien. Improved monomeric red, orange and yellow fluorescent proteins derived from discosoma sp. red fluorescent protein. *Nature Biotechnology*, 22, 2004. (Cited on page 19.)

S. F. Shih and G. J. Gascho. Water requirement for sugarcane production. *Transactions of the ASABE*, 23, 1980. (Cited on page 4.)

S. Shima and R. K. Thauer. A third type of hydrogenase catalyzing h2 activation. *The Chemical Record*, 7, 2006. (Cited on page 6.)

K. Y. Show, D. J. Lee, J. H. Tay, C. Y. Lin, and J. S. Chang. Biohydrogen production: Current perspectives and the way forward. *International Journal of Hydrogen Energy*, 37(20):15616 – 15631, 2012. ISSN 0360-3199. The 2011 Asian Bio-Hydrogen and Biorefinery Symposium (2011ABBS). (Cited on pages 5 and 6.)

A. Singha, P. S. Nigamb, and J. D. Murphya. Renewable fuels from algae: An answer to debatable land based fuels. *Bioresource Technology*, 102, 2010. (Cited on page 5.)

M. R. Soboleski, J. Oaks, and W. P. Halford. Green fluorescent protein is a quantitative reporter of gene expression in individual eukaryotic cells. *FASEB Journal*, 19, 2005. (Cited on page 19.)

R. Sodoyer. Expression systems for the production of recombinant pharmaceuticals. *Biodrugs*, 18, 2004. (Cited on pages 15, 51, and 52.)

R. Y. Stanier, R. Kunisawa, M. Mandel, and G. Cohen-Bazire. Purification and properties of unicellular bluegreen algae (order chroococcales). *Bacteriological Reviews*, 35, 1971. (Cited on page 34.)

S. E. Stevens and R. D. Porter. Transformation in agmenellum quadruplicatum. *Proceedings of the National Academy of Sciences of the United States of America*, 77, 1980. (Cited on page 9.)

H. Teeling, B. M. Fuchs, D. Becher, C. Klockow, A. Gardebrecht, C. M. Bennke, M. Kassabgy, S. Huang, A. J. Mann, J. Waldmann, M. Weber, A. Klindworth, A. Otto, J. Lange, J. Bernhardt, C. Reinsch, M. Hecker, J. Peplies, F. D. Bockelmann, U. Callies, G. Gerdts, A. Wichels, K. H. Wiltshire, F. O. Glöckner, T. Schweder, and

R. Amann. Substrate-controlled succession of marine bacterio-plankton populations induced by a phytoplankton bloom. *Science*, 4, 2012. (Cited on page 10.)

R. Y. Tsien. The green fluorescent protein. *Annu. Rev. Biochem*, 67, 1998. (Cited on page 17.)

P. M. Vignais and B. Billoud. Occurrence, classification, and biological function of hydrogenases: An overview. *Chemical Reviews*, 107, 2007. (Cited on pages 6, 7, and 8.)

G. S. Waldo, B. M. Standish, J. Berendzen, and T. C. Terwilliger. Rapid protein-folding assay using green fluorescent protein. *Nature Biotechnology*, 17, 1999. (Cited on page 19.)

M. S. Wetz and P. A. Wheeler. Response of bacteria to simulated upwelling phytoplankton blooms. *Marine Ecology Progress Series*, 272, 2004. (Cited on page 10.)

WHO. Ambient (outdoor) air quality and health, March 2014. (Cited on page 3.)

K. Wiebe, A. Croppenstedt, T. Raney, J. Skoet, M. Zurek, the FAO Agricultural, D. E. D. of FAO, J. Tschirley, and M. Cluff. *The state of food and agriculture 2008*. Food and Agriculture Organization of the United Nations, 2008. (Cited on pages 3 and 4.)

R. H. Wijffels and M. J. Barbosa. An outlook on microalgal biofuels. *Science*, 13, 2013. (Cited on pages 4, 16, and 74.)

M. Winkler, J. Esselborn, and T. Happe. Molecular basis of [fefe]-hydrogenase function: An insight into the complex interplay between protein and catalytic cofactor. *Biochimica et Biophysica Acta*, 1827, 2013. (Cited on page 7.)

Y. Xu, R. M. Alvey, P. O. Byrne, J. E. Graham, G. Shen, and D. A. Bryant. Expression of genes in cyanobacteria: Adaptation of endogenous plasmids as platforms for high-level gene expression in synechococcus sp. pcc 7002. In R. Carpentier, editor, *Photosynthesis Research Protocols*, volume 684 of *Methods in Molecular Biology*, pages 273–293. Humana Press, 2011. ISBN 978-1-60761-924-6. (Cited on pages 16, 17, 46, 47, and 65.)

S. Yoshihara, X. Geng, S. Okamoto, K. Yura, T. Murata, M. Go, M. Ohmori1, and M. Ikeuchi. Mutational analysis of genes involved in pilus structure, motilityand transformation competency in the unicellular motile cyanobacteriumsynechocystis sp. pcc6803. *Plant and Cell Physiology*, 42, 2001. (Cited on page 47.)

Z. Zhang, S. Schwartz, L. Wagner, and W. Miller. A greedy algorithm for aligning dna sequences. *Journal of Computational Biology*, 7, 2000. (Cited on pages 84 and 85.)

J. Zhou, H. Zhang, H. Meng, Y. Zhu, G. Bao, Y. Zhang, Y. Li, and Y. Ma. Discovery of a super-strong promoter enables efficient production of heterologous proteins in cyanobacteria. *Scientific Reports*, 4, 2014. (Cited on pages 17 and 43.)

G. Zittelli, V. Tomasello, E. Pinzani, and M. Tredici. Outdoor cultivation of arthrospira platensis during autumn and winter in temperate climates. *Journal of Applied Phycology*, 8, 1996. (Cited on page 4.)

DECLARATION

Put your declaration here.

Trondheim, Norway, May 2015

Christopher Wright Williams