Winter 12-15-2014

Metabolic Engineering of Cyanobacteria for Photosynthetic Production of Drop-In Liquid Fuels

Bertram Michael Berla
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

School of Engineering and Applied Sciences
Energy, Environmental, and Chemical Engineering

Dissertation Examination Committee:

Himadri B. Pakrasi, Chair
Yinjie Tang, Co-Chair
Robert Blankenship
Gautam Dantas
Cynthia Lo
Fuzhong Zhang

Metabolic Engineering of Cyanobacteria for Photosynthetic Production of Drop-In Liquid Fuels

by

Bertram M. Berla

A dissertation presented to the
Graduate School of Arts & Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

December 2014
St. Louis, Missouri

Table of Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 6 does not contain any figures.

Appendix Chapter 1

Appendix Chapter 2

Appendix Chapter 3

Appendix Chapter 4

List of Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 6 does not contain any tables.

Appendix Chapter 1

Appendix Chapter 1 does not contain any tables.

Appendix Chapter 2

Appendix Chapter 3

Appendix Chapter 4

Acknowledgements

I consider myself very fortunate to have been mentored by a number of incredible teachers during my education. My scientific education began with my dad, an endlessly curious person who never met an uninteresting question, and rarely stops trying to learn the answers. Talking while driving around in the car as a youngster is where I got my first taste for science. My mom has also always been an incredible supporter of my curiosity, and the curiosity of other kids like me, through her work in founding the Gifted Resource Council, where I attended summer camps and weekend classes to learn about space, biology, ancient history, and many other things. My first exposure to doing science in a real-life lab was afforded to me by my friend and childhood neighbor, Dr. Alan Permutt, MD. Alan was a renowned diabetes researcher at Washington University who unfortunately passed away two years ago. His encouragement and example, however, made a big impact on my decision to pursue a career in science. Although this may seem unusual to many, the teachers from my high school years who had the biggest impact on me as a scientist were both in the humanities. Pat Noland, who taught me AP English, was the first teacher who forced me (often kicking and screaming) to explain in detail exactly how a particular piece of evidence supported a conclusion I wanted to draw in an essay. I consider this intellectual rigor to be the most valuable skill I've ever learned. Kurt Knoedelseder, my theater teacher, taught me about the hard work required and the satisfaction achieved from putting all of oneself into a big project, and only considering the work done when the final product was truly excellent. I hope that I have done the lessons I learned from these teachers justice in this dissertation.

During my undergraduate career, my research mentor, Prof. Evan DeLucia, deserves the credit for really getting me hooked on science. Evan gave me the freedom to develop a project that interested and engaged me, and provided me with incredible resources to pursue my curiosity. For that, I am forever grateful. Prof. Susan Fahrbach was also instrumental in encouraging and helping me to get involved with undergraduate research through the Undergraduate Mentoring in Environmental Biology fellowship program. In the years after I finished my bachelor's degree, Dr. Daniel Schachtman and Dr. Leslie Hicks at the Donald Danforth Plant Science Center were great mentors to me, and gave me the opportunity to learn valuable skills, including writing and publishing my first real scientific papers.

During my graduate school career, I got to work in several excellent labs. Prof. Pratim Biswas gave me the opportunity to work in his lab the summer before beginning as a student at WUSTL, and this was my first experience working with cyanobacteria. Prof. Yinjie Tang also hosted me for a valuable rotation project where I learned about the ^{13}C-assisted metabolic analysis techniques that are used in much of this dissertation. Yinjie has continued to be a great friend and mentor during my graduate school career as my co-advisor and committee co-chair. Finally, I settled in the lab of Prof. Himadri Pakrasi, my graduate advisor and committee chair. Himadri's lab has been a wonderful place to work for me. Himadri has given me the incredible gift of time and freedom to pursue scientific questions that have ignited my curiosity, even when they aren't exactly the questions I originally set out to answer. Beyond this, Himadri has connected me with a wealth of resources and collaborations that enabled me to turn my curiousity into interesting science. For that, I am eternally grateful. In Himadri's lab, Dr. Abhay Singh, Dr. Anindita Bandyopadhyay, and (now) Dr. Jeff Cameron have especially served as mentors to me, who taught me so much of the technical knowledge I needed to do this work. I

also probably learned more than I taught through my interactions with Andrew Ng, a very talented undergraduate student who I know has a bright future in science. In addition, the rest of the lab members and classmates who have been my colleagues during these last 6 years are a huge reason for my reaching this milestone, and I owe you all a debt of gratitude. I especially want to thank my long-time collaborators from Penn State University, Dr. Rajib Saha, Tom Mueller, and Prof. Costas Maranas, who are co-authors of several chapters of this dissertation.

One of the things I have learned (and that I think all graduate students know by the end) is that one of the hardest parts of graduate school is dealing with constant "failures" and learning to bounce back emotionally from experiments that often don't work. For that emotional resilience, I primarily have to thank the never-ending support of my family and friends. My wife Jenny, daugher Evan, and of course my dog Monster are the biggest reasons I've always believed I could finish this project. In addition to their support, my mother Bev Berla, my father Eric, and my sister Rachel Cruse have always provided me with all the support and love I could ever want on this journey. I've even got to add on to that with my wonder in-laws, Bob and Jill Waxler. Thank you all so much for your love and support. In addition, my second family of close friends have helped me to smile, stay grounded, and keep working.

Finally, I would like to acknowledge the contribution of my committee members, including my committee chair Prof. Himadri Pakrasi, co-chair Prof. Yinjie Tang, and Profs. Bob Blankenship, Gautam Dantas, Cynthia Lo, and Fuzhong Zhang. Thank you for generously giving of your time and guidance to help me through this process. Additionally, I would like to thank Prof. Tuan-Hua David Ho, who served on my committee before his retirement from Washington University. I would also like to thank Profs. Milorad Dudukovic and Tae Seok Moon, whom I

consider as honorary committee members for their valuable insight and guidance during my graduate school career.

Specifically, I want to acknowledge those who have contributed directly to creating all of the content in this dissertation. Complete citations for all previously published work are given in my *curriculum vitae*. Chapter 1 originally appeared in *Frontiers in Microbiology* and is adapted and reproduced here with permission from Nature Publishing Group. I especially want to thank my co-authors Dr. Rajib Saha, Cheryl Immethun, Prof. Tae Seok Moon, Prof. Costas Maranas, and Prof. Himadri Pakrasi. Development of the original manuscript was supported by funding from the Office of Science (BER), U. S. Department of Energy to Drs. Himadri B. Pakrasi and Costas D. Maranas, grant DE-SC0006870, and a National Science Foundation Graduate Research Fellowship to Cheryl M. Immethun. My own role in this work was that I wrote several sections of the paper, and edited the others.

My co-authors for chapter 2 are Dr. Rajib Saha, Prof. Costas Maranas, and Prof. Himadri Pakrasi. We are grateful to Dr. Jeff Cameron and Prof. Toivo Kallas for insightful discussions in the development of this work. This work was supported by funding from the Office of Science (BER), U. S. Department of Energy, grant DE-SC0006870, to Drs. Himadri B. Pakrasi and Costas D. Maranas. This study also made use of the NIH/NIGMS Biomedical Mass Spectrometry Resource at Washington University in St. Louis, MO, which is supported by National Institutes of Health/National Institute of General Medical Sciences Grant #8P41GM103422. I would like to thank Jan Crowley at that facility for her help, as well. My own role in this work was that I conceived of and conducted the experiments described (except the *in silico* work, which was performed by RS), and wrote the paper.

Chapter 3 was co-authored by Dr. Rajib Saha, myself, Alex Verseput, Prof. Himadri

Pakrasi, and Prof. Costas Maranas. This work originally appeared in the journal *PLoS ONE* under a Creative Commons Open-Access License. We acknowledge Thanura Elvitigala and Lawrence Page for contributions during the amino acid and pigment measurements. We also thank Anthony P. Burgard and Alireza Zomorrodi for discussions during the model reconstruction process. This study was supported by a grant from the United States Department of Energy, Biological and Environmental Research (DOE-BER), grant DE-SC0006870. My role in this work was to make the measurements of biomass composition described, as well as working with RS and AV to reconstruct the *in silico* model. I also wrote some sections of the paper and edited the paper.

Chapter 4 is in part reproduced from an article co-authored by myself and Prof. Himadri Pakrasi that originally appeared in *Applied and Environmental Microbiology*. Those sections appears here with permission from the American Society of Microbiology. Other portions of this chapter are from a manuscript in preparation by Andrew Ng, myself, and Prof. Himadri Pakrasi. We thank Dr. Thanura Elvitigala for valuable advice and other members of the Pakrasi lab for collegial discussions. We would also like to thank Chris Sawyer at the Genome Technology Access Center (GTAC) at Washington University for assistance with qPCR data collection and analysis. The research published in *AEM* was supported by funding from the Office of Science (BER), U. S. Department of Energy, grant DE-SC0006870, and from the Consortium for Clean Coal Utilization at Washington University. My own role in this work was that I conducted the experiments related to the microarrays and plasmid copy number. I assisted AN in performing the other experiments. I wrote the *AEM* paper and the portions of the unpublished work that appear here.

Chapter 5 includes unpublished work conducted, analyzed, and written by myself. This work was supported by funding from the Office of Science (BER), U. S. Department of Energy, grant DE-SC0006870, and from the Consortium for Clean Coal Utilization at Washington University.

Appendix Chapter 1 originally appeared in *The Biotechnology Journal* and is reproduced here with permission from the publisher, John Wiley and Sons. My co-authors were Le You, Lian He, Prof. Himadri Pakrasi, and Prof. Yinjie Tang. This research was funded by an NSF Career Grant (MCB0954016) to Yinjie Tang. It was also supported by grants from the DOE Office of Science (BER), grant DE-SC0006870, and the Consortium for Clean Coal Utilization to Himadri Pakrasi. We thank James Ballard and Katrina Leyden for their writing assistance. Along with LY and YJT, I designed these experiments, and LY and I co-wrote this paper.

Appendix chapter 2 originally appeared in *BMC Systems Biology* and is reproduced here with permission from the publishers. This works authors were Thomas J Mueller, myself, Prof. Himadri Pakrasi, and Prof. Costas Maranas. This work was supported by funding from the Office of Science (BER), U.S. Department of Energy to Drs. Costas D. Maranas and Himadri B. Pakrasi, grant DE-SC0006870. I assisted TM in completing this work and in writing and editing the paper.

Appendix Chapter 3 originally appeared in *The Journal of Visualized Experiments* and is reproduced here with permission from the publishers. This works authors were Le You, Dr. Lawrence Page, Prof. Xueyang Feng, myself, Prof. Himadri Pakrasi, and Prof. Yinjie Tang. This study was supported by an NSF Career Grant (MCB0954016) and a DOE Bioenergy Research Grant (DEFG0208ER64694). I assisted LY in editing this manuscript and demonstrated several aspects of the method in the accompanying video.

Appendix chapter 4 originally appeared in *Microbiology* and is reproduced here with permission from the publishers. This work's authors were Prof. Xueyang Feng, Dr. Anindita Bandyopadhyay, myself, Dr. Lawrence Page, Dr. Bing Wu, Prof. Himadri Pakrasi, and Prof. Yinjie Tang. This study was supported in part by a US Department of Energy (DOE) bioenergy research grant (DEFG0208ER64694) and in part by a grant from the National Science Foundation (MCB0954016). The authors are also grateful to Jing Jiang, Jeff Cameron and Dr Jana Stöckel for their kind help with measurements and discussions. My own role in this work was that I assisted XY and AB in making measurements and edited the manuscript.

Bert Berla

Washington University in St. Louis

December 2014

Dedicated to my Dad,

whose endless curiosity inspired me to set off on this journey,

and whose life-long personal growth inspired me to finish it.

And to my Mom,

my first and best teacher who always believed in and cared for me,

whether or not she had any idea what I was talking about.

ABSTRACT OF THE DISSERTATION

Metabolic Engineering of Cyanobacteria for Photosynthetic Production
of Drop-In Liquid Fuels
by
Bertram M. Berla
Doctor of Philosophy in Energy, Environmental, and Chemical Engineering
Washington University in St. Louis, 2014
Professor Himadri B. Pakrasi, Chair
Professor Yinjie Tang, Co-Chair

Cyanobacteria are oxygenic phototrophs with great potential as hosts for renewable fuel and chemical production. They grow very quickly (compared with plants) and can use sunlight for energy and CO_2 as a carbon source (unlike yeast or *E. coli*). While cyanobacteria have been engineered to make many chemicals that are native and non-native parts of their metabolism, this work is concerned with the production of heptadecane in *Synechocystis* sp. PCC 6803. Heptadecane is in a class of natural products produced by all cyanobacteria, but in quantities insufficient for industrialization. Towards this future goal, we have built enabling systems for the overproduction of fuels and chemicals in *Synechocystis* 6803 and cyanoabacteria generally. These tools include plasmid vectors for the overproduction of heterologous proteins and genome-scale metabolic models that can predict strategies for metabolite overproduction. We have shown that the vectors we developed are helpful in controlling the level and timing of heterologous protein expression using a fluorescent reporter, and have made progress towards heptadecane overproduction. During this process, we have also found that heptadecane is crucial for cold tolerance and modulates cyclic electron flow in photosynthesis. In addition to measuring this phenotype *in vivo*, we have analyzed it *in silico* using our genome-scale metabolic model and have gained insight into the role of cyclic electron flow in photosynthesis generally.

Chapter 1

Synthetic Biology of Cyanobacteria:

Unique Challenges and Opportunities

1.1. Introduction

Cyanobacteria have garnered a great deal of attention recently as biofuel-producing organisms. Their key advantage over other bacteria is their ability to use photosynthesis to capture energy from sunlight and convert CO_2 into products of interest. As compared with eukaryotic algae and plants, cyanobacteria are much easier to manipulate genetically and grow much faster, with doubling times as low as 2 hours. They have been engineered to produce a wide and ever-expanding range of products including fatty acids, long-chain alcohols, alkanes, ethylene, polyhydroxybutyrate, 2,3-Butanediol, ethanol, and hydrogen. These processes have been reviewed recently (Gronenberg *et al.* 2013; Nozzi *et al.* 2013) and will not be covered in great detail in this dissertation. The highest titers of ethanol (5.5 g/L) (Gao *et al.* 2012), 2,3-butanediol (2.4 g/L) (Oliver *et al.* 2013) and lactic acid (2.2 g/L) (Varman *et al.* 2013) have been produced, but these product yields remain far lower than could be commercially viable and are produced far more slowly than using heterotrophic bacteria or yeast. On the other hand, an ever-expanding array of products have been produced in cyanobacteria including volatile products that could ease separation from cultures such as ethylene (Guerrero *et al.* 2012), isobutyraldehyde (Atsumi *et al.* 2009), hydrogen (Berto *et al.* 2011) and isoprene (Bentley *et al.* 2014). In addition, some more unique metabolites have been produced in *Synechocystis* sp. PCC 6803. These include the plant secondary metabolite p-coumaric acid (Xue *et al.* 2014) the isoprenoid squalene (Englund *et al.* 2014) and diesel-range alkanes (Wang *et al.* 2013). While these products have been produced only at much lower titers they are in some ways unique to the green lineage and thus may be preferable targets for production in cyanobacteria.

In this dissertation we will look towards how the techniques of the emerging field of synthetic biology and a deeper understanding of cyanobacterial biology might bear fruit in

improving the output specifically of diesel-range alkanes. Due to the low price of these commodity fuels, it is critical to maximize the productivity of engineered strains to make them economically competitive. I believe that the tools of synthetic biology can help with this challenge and that a detailed understanding of any system one is attempting to engineer is critical for success. Specifically, this initial chapter will cover systems, parts, and methods of analysis for synthetic biology systems. Synthetic biology requires a well-characterized host or 'chassis' strain that can be genetically manipulated with ease and predictability. Ideally, the host should grow quickly and tolerate a range of environmental conditions. The host should be simple to cultivate using readily available laboratory equipment and inexpensive growth media. Simple, rapid, and high-throughput techniques should be available for procedures like DNA/RNA isolation, metabolomics, and proteomics. To achieve modular, 'plug-and-play' modification of the host strain, its metabolism and regulatory systems must be well-characterized under a wide variety of relevant conditions. Since cyanobacterial biofuel production processes will need to use sunlight as an energy source to be economically and environmentally useful, the day/night cycle will be particularly relevant; The intermittent nature of this energy source will be a key engineering challenge. We will discuss which cyanobacterial chassis have been used and their relative merits and unique traits. Ultimately, the hope is that one of these strains might be developed to become a 'green *E. coli*' for which a wide variety of genetic parts and systems are available for easy modification. Next, we will discuss the critical issue of how gene expression can be controlled in cyanobacteria. Compared with other systems, there are few examples of simple and effective controllable promoters in cyanobacteria. We will also discuss methods for analysis of gene expression using light-emitting reporters and for global analysis of metabolism using either constraint-based modeling or measurement of ^{13}C labeling. Examples of many of

these techniques' use in *Synechocystis* sp. PCC 6803 make up the remainder of this thesis beyond this chapter and are mentioned at the end of this chapter (see "This Work").

1.2. Genetic Modification of Cyanobacteria

Several strains of cyanobacteria are known which are readily amenable to genetic modification (See table 1.1). Such modifications can be performed either *in cis* (through choromosome editing) or *in trans* (through plasmid addition) and synthetic biology experiments have used both approaches. We discuss advantages and disadvantages of each approach, as well as recent technical developments below. While even the best cyanobacterial model systems are still far from being a 'green *E. coli*', many tools are already available and more are being developed. The future holds great promise for this field.

1.2.1. Genetic Modification *in Cis*: Chromosome Editing

Cis genetic modification is the most common approach in cyanobacterial synthetic biology. This approach takes advantage of the capability of many cyanobacterial strains for natural transformation and homologous recombination (see table 1.1) to create insertion, deletion, or replacement mutations in cyanobacterial chromosomes. Traditionally, strains have been transformed with selectable markers linked to any sequence of interest and flanked by sequences homologous to any non-essential sequence on the chromosome (See figure 1.1).

This strategy allows the creation of targeted mutations to the chromosome, but sometimes raises concerns about segregation in polyploid strains. However, once segregated, such mutations can be stable over long time periods even in the absence of selective pressure from added antibiotics (Liu *et al.* 2011; Wang *et al.* 2013). While such stability is desirable, systems that

create major metabolic demand, by for example redirecting flux into biofuel-producing pathways, will face greater selective pressures for mutation or loss of heterologous genes.

Recently, several methods have been developed that allow the creation of markerless mutations in cyanobacterial chromosomes (figure 1.1b). Two of these methods operate on a similar principle: First, a conditionally toxic gene is linked to an antibiotic resistance cassette and then inserted into the chromosome, with selection for antibiotic-resistant mutants. Next, a second transformation is carried out in which the resistance cassette and toxin gene are deleted, and markerless mutants are selected which have lost the toxic gene. A cassette containing an internal repeated region can also undergo a second recombination event under negative selection without a second transformation (Viola *et al.* 2014). This principle has been used in cyanobacteria with the *B. subtilis* levansucrase synthase gene *sacB*, which confers sucrose sensitivity (Lagarde *et al.* 2000) as well as with *E. coli mazF*, a general protein synthesis inhibitor expressed under a nickel-inducible promoter (Cheah *et al.* 2013) and *acsA* from *Synechococcus* sp. PCC 7002, an acyltransferase that makes the strain sensitive to acrylic acid (Begemann *et al.* 2013). These latter systems have advantages for cyanobacterial strains that are naturally sucrose-sensitive. Either method allows the reuse of a single selectable marker for making multiple successive changes to the chromosome. In addition to these methods, a third system operates on a similar principle - a cyanobacterial strain that is streptomycin resistant due to a mutation in the *rps12* gene can be made streptomycin-sensitive by expressing a second heterologous copy of wild type *rps12* linked to a kanamycin (or other antibiotic) resistance cassette as well as any sequence of interest. Streptomycin-resistant, kanamycin-sensitive markerless mutants can be recovered in a second transformation (Takahama *et al.* 2004). Although this method can also be used to make successive markerless mutants, it requires a background strain that is streptomycin-resistant due

to an altered ribosome. Thus, it may not be an ideal method for synthetic biology studies that seek to draw conclusions about translation in wild-type systems. For the ability to transfer any translated genetic part or parts involved in translation (such as ribosome binding sites) to other strains, this mutation could be problematic. A possible advantage of this system is that both selections are positive selections, whereas the *sacB, mazF,* or *acsA* systems require a negative selection in their second transformation. Care must be taken to ensure that resistance is due to loss, as opposed to mutation, of the counter-selectable marker. Recombinase-based systems including Cre-LoxP (in *Anabaena sp.* PCC7120, (Zhang *et al.* 2007)) or FLP/FRT (in *Synechocystis sp.* PCC6803 and *Synechococcus elongatus* PCC7942, (Tan *et al.* 2013)) have also been used to engineer mutants that lack a selectable marker. However, these methods leave a scar sequence, meaning that the final chromosomal sequence is not completely user-specifiable and also that multiple mutations using this technique in the same cell line may potentially lead to undesirable crossover events or other unexpected results.

Until recently, it has been difficult to create mutants at high throughput in cyanobacterial strains, as transposon-based methods developed for use in other strains can work poorly in cyanobacterial hosts. However, libraries can be created in other strains and subsequently transferred to a cyanobacterial host via homologous recombination. A Tn7-based library containing ~10,000 lines was recently created to screen for strains with increased polyhydroxybutyrate (PHB) production (Tyo *et al.* 2009) and a similar approach has been taken for finding mutants in circadian clock function in *Synechococcus* 7942 (Holtman *et al.* 2005) and later extended to include insertions into nearly 90% of open reading frames in that strain (Chen *et al.* 2012). Chromosomal DNA fragments were first cloned into a plasmid library in *E. coli* and then the library was mutagenized with Tn7 before homologous recombination back into the

cyanobacterial host strain. This could be an especially valuable approach for validating genome-scale models of cyanobacterial metabolism (see chapters 3-4 and appendix chapters 1-4).

1.2.2 Genetic Modification *in Trans*: Foreign Plasmids

Although transgene expression *in cis* is the most common approach in cyanobacterial research, genes are also routinely expressed in cyanobacteria *in trans* (Huang *et al.* 2010; Landry *et al.* 2012; Huang *et al.* 2013; Taton *et al.* 2015). In synthetic biology and metabolic engineering of other prokaryotes, this is by far the more common approach, and has led to such standardized approaches as "Bio-Brick" assembly in which standardized genetic 'parts' such as promoters, ribosome binding sites, genes, and terminators can be readily swapped in and out of standard plasmids (http://partsregistry.org). This move towards standardization of genetic parts is a critical aim for synthetic biology, independent of the chassis organism or method of transformation. However, a limited number of plasmids are available for expression in cyanobacterial hosts. Plasmid assembly for expression *in cis* or *in trans* in cyanobacterial hosts has generally been performed in *E. coli* because of the longer growth times that would be associated with assembling vectors in cyanobacterial hosts (figure 1.2a). This requires broad host range plasmids. However, with the rise of *in vitro* assembly methods such as SLIC (Li *et al.* 2007), Gibson assembly (Gibson *et al.* 2009), CPEC (Quan *et al.* 2009), fusion PCR (Szewczyk *et al.* 2007), and Golden Gate (Engler *et al.* 2011), this limitation may become less important over time (figure 1.2b). These next-generation cloning methods have been reviewed elsewhere (Hilson *et al.* 2012) and will not be covered here. Fusion PCR has been used to construct linear DNA fragments for homologous recombination in cyanobacterial chromosomes (Nagarajan *et al.* 2011), but to our knowledge replicative vectors for cyanobacteria have so far not been constructed without the use of a 'helper' heterotrophic strain. Techniques for *in vivo* assembly of

plasmids that have been developed for yeast (Shao *et al.* 2009) may be adaptable to cyanobacteria because of their facility for homologous recombination (figure 1.2c). Such an improvement could greatly speed up the process of making cyanobacterial mutant strains, either for modification *in cis* or *in trans*. The major technical challenge for such an approach is that the long time after transformation required to isolate cyanobacterial mutants (typically 1 week or more) means it is critical to have high-fidelity assembly methods to avoid a time-consuming screening process.

Although shuttle vectors do exist for cyanobacteria, there has been little characterization of their copy numbers in cyanobacterial hosts, and the lack of replicative vectors with varied copy numbers limits the valuable ability to control the expression level of heterologous genes by selecting their copy number (Jones *et al.* 2000; Dunlop *et al.* 2011). Plasmids derived from RSF1010 appear to have a copy number of 10-30 (or ~1-3 per chromosome) in *Synechocystis sp.* PCC 6803 (Ng *et al.* 2000; Huang *et al.* 2010), but copy numbers of other broad host-range plasmids have not been quantified to date. Endogenous plasmids of cyanobacteria have also been used as target sites for expression of heterologous genes in *Synechococcus sp.* PCC 7002 (Xu *et al.* 2011) and in *Synechocystis* sp. PCC 6803 (This work, see chapter 2). This strain harbors several endogenous plasmids whose copy numbers range from ~1-8 per chromosome, with an approximate chromosome copy number of 6 per cell. *Synechocystis* sp. PCC 6803 also has plasmids whose copy numbers span a similar range (from ~0.4-8 per chromosome (Berla *et al.* 2012)). The origins of replication from these plasmids constitute a source of genetic parts that could be used to generate cyanobacterial expression plasmids having a range of copy numbers, and which could potentially be modified to create higher or lower-copy plasmids that are compatible with existing plasmids in various cyanobacterial systems. The range of shuttle

vectors that have been used in cyanobacterial hosts has been recently reviewed (Wang *et al.* 2012). While many tools are available for genetic modification of these biotechnologically promising strains, opportunities abound to develop new and improved tools that will allow research to proceed faster.

1.3 Unique Challenges of the Cyanobacterial Lifestyle

Organisms that survive using sunlight as a primary nutrient face unique challenges. These must be better understood and addressed to fulfill the biotechnological promise of cyanobacteria through synthetic biology.

1.3.1 Life in a Diurnal Environment

A primary goal of synthetic biology in cyanobacteria is to use photosynthesis to convert CO_2 into higher-value products such as biofuels and chemical precursors. To make such a process economically and environmentally feasible will require using sunlight as a primary energy source. While some cyanobacteria are facultative heterotrophs, their key advantage over obligate heterotrophic bacteria is photosynthesis. Unlike heterotrophic growth environments where carbon and energy sources can be provided more uniformly both in space and time, sunlight will only be available during the day and will be attenuated as it passes through the culture. Under certain conditions, cultures may be able to take advantage of a 'flashing light effect' to integrate spatially uneven illumination by storing chemical energy when in bright light near the reactor surface and using that energy to conduct biochemistry during time spent in the dark away from the reactor surface (Sforza *et al.* 2012; He *et al.* 2015). This ability will depend on light intensity, mixing rates, reactor geometry, and likely other factors. Certain diazotrophic

cyanobacteria can even use daylight to continue growth during the night. *Cyanothece sp.* ATCC 51142 and several other strains (Taniuchi *et al.* 2008; Latysheva *et al.* 2012; Pfreundt *et al.* 2012) are unicellular diazotrophic cyanobacteria that perform photosynthesis and accumulate glycogen during the day, and then during the night break down their glycogen reserves to supply energy for nitrogen fixation. Thus, these strains spread out the energy available from sunlight over a 24-hour period. This process involves a genome-wide oscillation in transcription, with more than 30% of genes oscillating in expression between day and night (Stockel *et al.* 2008). To take full advantage of sunlight, synthetic systems must be created that are capable of responding appropriately to this challenging dynamic environment. It has recently been shown that biofuel-producing strains that dynamically tune the expression of heterologous pathways in response to their own intracellular conditions produce more biofuel and exhibit greater stability of heterologous pathways (Zhang *et al.* 2012). As challenging as the design of such a system was for batch heterotrophic cultures, it will be even more challenging in production environments that include a diurnal light cycle.

While not all strains exhibit as complete a physiological change between day and night as *Cyanothece* 51142, all cyanobacteria do have a circadian clock that adapts them to their autotrophic lifestyle. In non-diazotrophic strains, a large percentage of the genome remains under circadian control (Beck *et al.* 2014). The cyanobacterial circadian clock is anchored by master regulators KaiA, KaiB, and KaiC, which act by cyclically phosphorylating and dephosphorylating each other (Akiyama 2012). While the circadian rhythm can be reconstituted *in vitro* using the three Kai proteins in the presence of ATP (Nakajima *et al.* 2005), the accurate maintenance of this clock *in vivo* depends on proper protein turnover (Holtman *et al.* 2005), on codon selection in the *kaiBC* transcript (Xu *et al.* 2013), on transcriptional feedback (Teng *et al.*

2013), and on the controlled response of the entire program of cellular transcription to the output of the KaiABC oscillator. While disturbing rhythmicity can lead to strains that grow better under constant light, the circadian clock is adaptive for strains living in a dynamic environment (Woelfle *et al.* 2004; Xu *et al.* 2013). Therefore, integrating synthetic gene circuits such as biofuel production processes into the circadian rhythm of cyanobacterial hosts will likely lead to both improved production and improved strain stability in outdoor production environments.

1.3.2. Redirecting Carbon Flux by Decoupling Growth from Production

While redirecting carbon flux is a challenge in all metabolic engineering efforts, it has been suggested that stringent control of fixed carbon partitioning among central metabolic pathways poses a major limitation to chemical production especially in photosynthetic organisms (Melis 2013). During the growth phase, it may be true that carbon partitioning is tightly controlled by any number of mechanisms including metabolite channeling or simply high demand for metabolic intermediates. However, biofuel production during non-growth phases (Atsumi *et al.* 2009; Liu *et al.* 2011; Varman *et al.* 2013; Wang *et al.* 2013) demonstrates that under appropriate conditions, cyanobacterial hosts can produce biofuel compounds with higher selectivity, since biofuel can be produced by metabolically active cells even in the absence of growth. Enhancing their productivity in this phase is a major opportunity for cyanobacterial synthetic biologists to overcome these limits on carbon partitioning. Capturing this opportunity will require designing complete metabolic circuits that remain highly active during stationary phase.

1.3.3 RNA-Based Regulation

Recently, regulation of gene expression through RNA mechanisms has received great attention across bacterial clades (Selinger *et al.* 2000; Sharma *et al.* 2010; Mitschke *et al.* 2011).

While these mechanisms of regulation may be important in all bacteria, their prominence is perhaps the greatest in the cyanobacteria and may help these diurnal organisms adapt to their highly dynamic environment: in a recent dRNA-seq study, many of the most highly expressed RNAs belonged to families of non-coding RNAs which are present in nearly all sequenced cyanobacteria, but not in any other organisms (Gierga *et al.* 2009; Mitschke *et al.* 2011) and these transcripts were shown to be important in regulating the dirunal rhythm of this strain (Beck *et al.* 2014). While their high expression in *Synechocystis* 6803 suggests functional importance for non-coding RNAs, few have clearly elucidated functions to date. *syr1* overexpression has been shown to lead to a severe growth defect in *Synechocystis* 6803 (Mitschke *et al.* 2011). Another small RNA, *isiR*, has a critical function in stress response in *Synechocystis* 6803. *isiR* binds to the mRNA (*isiA*) for the iron-stress inducible protein, which when translated, forms a ring around trimers of photosystem I, preventing their activity and thus oxidative stress in the absence of sufficient iron (Duhring *et al.* 2006). The binding of *isiR* to *isiA* appears to result in rapid degradation. This particular arrangement allows a very rapid and emphatic response to iron repletion in cyanobacteria, since a large pool of *isiA* transcripts can be quickly silenced and marked for degradation by transcription of the antisense *isiR*. Of particular relevance to this dissertation is that while the clusters for alkane biosynthesis in cyanobacteria initially appeared to be polycistronic, all of the strains so far examined actually produce only monocistronic transcripts. Many of these strains also include a small non-coding RNA immediately upstream of *ado* whose function is so far unknown (Klahn *et al.* 2014). Although little is so far known about the generality of this type of regulation, the dynamics of this response might also be effective to use for synthetic systems in cyanobacteria that live in the presence of light as an intermittently available but critical nutrient.

While non-coding RNA has received a lot of recent attention, two-component systems make up the most widely studied family of environmental response regulators in cyanobacteria. Many of these systems have known functions in response to diverse environmental stimuli such as nitrogen, phosphorous, CO_2, temperature, salt, and light intensity and quality (Ashby *et al.* 2006; Montgomery 2007). Many of the most widely-used systems in the construction of synthetic biological devices (such as the *ara* and *lux* clusters) use 2-component systems, and even combine 2-component systems with non-coding RNA to control system dynamics (Waters *et al.* 2006). As synthetic biology advances into the construction of more and more complex systems, there will be a growing need to understand and use all of the different mechanisms available for control of gene expression and enzyme activity in cyanobacteria.

1.4. Parts for Cyanobacterial Synthetic Biology

While cyanobacteria are promising organisms for biotechnology, synthetic biology tools for these organisms lag behind what has been developed for *E. coli* and yeast (Heidorn *et al.* 2011; Markley *et al.* 2014). Furthermore, synthetic biology tools developed in *E. coli* or yeast often do not function as designed in cyanobacteria (Huang *et al.* 2010). Here, we discuss inducible promoters and reporters in cyanobacteria, and cultivation systems that will allow their testing at increased throughput. Refining such systems will make cyanobacterial synthetic biology more user-friendly, a central goal for developing the 'green *E. coli*.'

1.4.1. Inducible Promoters

Creation of synthetic biology systems that predictably respond to a specific signal often depends upon inducible promoters for transcriptional control. An ideal inducible promoter will

have the following properties: (1) It will not be activated in the absence of inducer. (2) It will produce a predictable response to a given concentration of inducer or repressor. This response may be digital (i.e., on/off) or graded change with different concentrations of inducer/repressor. (3) The inducer at saturating concentrations should have no harmful effect on the host organism. (4) The inducer should be cheap and stable under the growth conditions of the host. Finally, (5) the inducible system should act orthogonally to the host cell's transcriptional program. Ideal transcriptional repressors should not bind to native promoters and if non-native transcriptional machinery is used (such as T7 RNA polymerase) it should not initiate transcription from native promoters. Promoters must perform as ideally as possible in order to be used in the construction of more complex genetic circuits (Moon *et al.* 2012).

Many common inducible promoters in cyanobacteria respond to transition metals. These have often been the basis of metal detection systems (Erbe *et al.* 1996; Boyanapalli *et al.* 2007; Peca *et al.* 2007; Peca *et al.* 2008; Blasi *et al.* 2012). Cyanobacteria balance metal intake for the organisms' needs against potential oxidative stress and protein denaturation (Michel *et al.* 2001; Peca *et al.* 2008) via tightly regulated systems. As shown in Table 1.2, cyanobacteria's metal-responsive promoters frequently show greater than 100-fold dynamic range. For example, the promoter for the *Synechocystis* sp. PCC 6803 gene, *coaA*, was induced 500 fold by 6 μM Co^{2+}(Guerrero *et al.* 2012), and P_{smt} from *Synechococcus elongatus* PCC 7942 was induced 300-fold by 2 μM Zn^{2+} (Erbe *et al.* 1996). The most responsive cyanobacterial promoters reported were P_{nrsB} from *Synechocystis* sp. PCC 6803, responding 1000-fold to 0.5 μM Ni^{2+}(Peca *et al.* 2007), and P_{isiAB} also from *Synechocystis* sp. PCC 6803, repressed 5000-fold by 30 μM Fe^{3+} following depletion (Kunert *et al.* 2003).

While the sensitivity of these promoters to low concentrations of ions may seem like an advantage, in practice it can make them difficult to use. Glassware must be thoroughly cleaned according to special protocols to remove trace metals and cells often have to be starved for extended periods, inducing stress responses, to use such inducible systems. Additionally, promoters endogenous to a chassis strain are woven into a complex, incompletely understood regulatory system. In this system, promoters are activated by multiple inducers, such as P*coaT* (Co^{2+} and Zn^{2+}) and P*ziaA* (Cd^{2+} and Zn^{2+}), both from *Synechocystis* sp. PCC 6803 and inducers can also activate multiple promoters, such as Cd^{2+} inducing P*ziaA* and P*isiA* (Blasi *et al.* 2012). Thus, these promoters fall short according to criteria 2, 3, and 5 described above.

While few good choices have so far been available for inducible promoters in cyanobacteria, it will be helpful to understand the differences in the cellular machinery of *E. coli* and cyanobacteria in order to adapt existing systems for use in a cyanobacterial 'green *E. coli*'. First, RNA polymerase (RNAP) is structurally different between *E. coli* and cyanobacteria. In cyanobacteria the β' subunit of the RNAP holoenzyme is split into two parts, as opposed to one in most eubacteria, creating a different DNA binding domain (Imamura *et al.* 2009). Being photosynthetic, circadian, and sometimes nitrogen-fixing, cyanobacteria also employ three sets of interconnected σ factors that are different than those used by *E. coli* (Imamura *et al.* 2009). Guererro *et al*. (2012) looked at the variation in the -35 and -10 regions of $P_{A1lacO-1}$ and P_{trc}. P_{trc} is not inducible in *Synechocystis* sp PCC 6803 and had the "standard" bacterial structure in these regions while $P_{A1lacO-1}$, which produced an eight fold response to IPTG in the same host, had a different structure in both regions. They postulated that *Synechocystis* 6803's sigma factors had different selectivity for these two regions. In fact, by systematically altering the bases between -10 and the transcription start site, a library of TetR-regulated promoters with improved

inducibility were created in *Synechocystis* sp. strain ATCC27184 (a glucose-tolerant derivative of *Synechocystis* 6803). The best performing promoter induced a 290-fold change in response to 1 ug/ml aTc (Huang *et al.* 2013). This work demonstrates the improvements that can be seen when modifying parts to work in a particular chassis. However, the light-sensitivity of the inducer aTc required the use of special growth lights that may have had other effects on photoautotrophic metabolism. Another system that suffers from similar limitations has recently been developed that is inducible by green light (Abe *et al.* 2014). Although this system is useful in laboratory studies, filtering light to remove particular wavelengths at an industrial scale would be impractical. Further studies that follow in this vein of using well-characterized sythetic biology parts and modifying them to function optimally in a particular cyanobacterial chassis are likely to bear fruit.

Recent progress has been made in creating more functional IPTG-inducible sysytems for cyanobacteria. By varying the inter-operator spacing within a library of P*trc*2O promoters (Camsund *et al.* 2014), a promoter that showed 13-fold induction by IPTG in *Synechocystis* sp. PCC 6803. In *Synechococcus* sp. PCC 7002, a library of hybrid promoters including elements of P*cpcB* from *Synechocytis* 6803 and lac operators showed nearly 50-fold induction (Markley *et al.* 2014). The lack of inducibility seen in lac-derived promoters in cyanobacteria could also be a function of inadequate transport of IPTG into cells. Concentrations of IPTG above 1 mM have been shown to induce lac-derived promoters in organisms without an active lactose permease, like many cyanobacteria. By introducing an active lactose permease into *Pseudomonas fluorescens,* inducibility was boosted 5 times at 0.1 mM IPTG (Hansen *et al.* 1998). Evolving the Lac repressor for improved inducibility is another strategy. Gene expression improved ten times with 1 μM IPTG through rounds of error prone PCR and DNA shuffling (Satya Lakshmi *et al.*

2009). Strength of expression and inducibility may also vary between different cyanobacterial strains. IPTG caused as much as a 36-fold response using the *trc* promoter in *Synechococcus elongatus* PCC 7942, but little or no response in *Synechocystis* sp. PCC 6803 (See Table 1.2). Phylogenetic analysis of σ factors from six different cyanobacterial strains, including *Synechocystis* sp. PCC 6803, showed *S. elongatus* 7942 to be distinctive. *S. elongatus* 7942 has σ factors that are unique to marine cyanobacteria as well as a group 3 σ factor similar to those from the heterocyst-forming *Anabaena* sp. PCC 7120 (Imamura *et al.* 2009). Understanding these strain-specific differences will enhance the synthetic biologist's ability to design promoters with ideal characteristics in their chassis of choice. This relates to the ability to take up inducers as well as the optimal characteristics of inducers (as in the light-sensitivity of aTc) as described above.

In addition to transcriptional control, recent progress has been made in inducible expression in cyanobacteria using translational control. A theophylline-responsive riboswitch at the 5' end of a transcript adopts a conformation that makes its ribosome binding site inaccessible in the absence of theophylline. In the presence of the inducer, the riboswitch's tail binds the inducer, freeing the ribosome binding site and enabling translation (Topp *et al.* 2007). These parts have been proposed as excellent choices for inducible expression in new systems because the interaction between inducer and transcript is direct and does not depend on other parts such as sigma factors, nor is it likely to have off-target interactions (Topp *et al.* 2010). This system has been adapted for use in cyanobacteria with some initial success in regulating the expression of reporter proteins (Nakahira *et al.* 2013; Ma *et al.* 2014). This system seems to succeed well on all of the conditions for an ideal inducible promoter mentioned above, and with further development could be an ideal system.

1.4.2. Reporters

Characterization of synthetic biological circuits depends on a reporting method to track the expression, interaction and position of proteins. Preferably the reporter should be detected without destruction of the organisms or additional inputs. Bacterial luciferase and fluorescent proteins are the most common non-invasive reporters. The *lux* operon is frequently used for reporting in cyanobacteria (Michel *et al.* 2001; Mackey *et al.* 2007; Peca *et al.* 2008) and is well suited for real time reporting of gene expression due to the short half-life of the relevant enzymes (Ghim *et al.* 2010). The superior brightness of fluorescent proteins makes them more ideal for subcellular localization via microscopy or for cell-sorting methods. Fluorescent proteins are produced in an array of colors and also do not require additional substrates. Their use in cyanobacteria is somewhat complicated by the fluorescence of the organism's photosynthetic pigments, but Cerulean, GFPmut3B (a mutant of green fluorescent protein) and EYFP (enhanced yellow fluorescent protein) have all been used successfully in cyanobacteria as reporters of gene expression (Huang *et al.* 2010; Heidorn *et al.* 2011; Landry *et al.* 2012; Huang *et al.* 2013).

Bacterial luciferase luminesces upon oxidation of reduced flavin mononucleotide (Meighen 1993). Fluorescent proteins also require oxygen to correctly fold and fluoresce (Hansen *et al.* 2001). The light-dark cycle of nitrogen-fixing cyanobacteria provides temporal separation of the oxygen-sensitive nitrogenase from oxygen-evolving photosynthesis (Golden *et al.* 1997). During the dark cycle, respiration reduces intra-cellular oxygen levels so that nitrogenase can function. Therefore, neither bacterial luciferase nor traditional fluorescent proteins can likely be used to study cyanobacteria in their dark cycle or to report on synthetic biology systems that operate in these oxygen-depleted conditions. Using blue light photoreceptors from *Bacillus subtilis* and *Pseudomonas putida*, oxygen-independent flavin

mononucleotide-binding florescent proteins have been devised (Drepper *et al.* 2007). With an excitation wavelength of 450 nm and an emission wavelength of 495 nm, they should perform well in cyanobacteria, although no data supporting this has been published yet. Functionality of these new fluorescent proteins was also improved by replacing a phenylalanine suspected of quenching with serine or threonine, resulting in a doubling of the brightness (Mukherjee *et al.* 2012). This expanding variety of easily readable reporter systems will be extremely valuable for cyanobacterial synthetic biology.

1.4.3. Cultivation Systems

To date, most synthetic biology and metabolic engineering work in cyanobacteria has been performed using simple, low-tech cultivation methods such as shake flasks or bubbling tubes grown under standard fluorescent light sources. Often, laboratory incubators have simply been retrofitted by the addition of fluorescent light sources available in home improvement stores. However, as light and CO_2 are major nutrients for cyanobacteria, it is critical to properly standardize the inputs of these resources to reliably characterize biological parts. It is also critical to increase the throughput of cyanobacterial growth systems to be able to screen the large numbers of variants that can be generated by combinatorial methods, as is routinely performed by growing heterotrophic bacterial cultures in 96-well plate format. Growth of cyanobacteria in 6-well plates can be routinely performed in our lab and by others (Huang *et al.* 2013) along with 24-well plates (Simkovsky *et al.* 2012), but growth in 96-well plates is poor, limiting assay throughput and requiring more space in lighted chambers under consistent illumination, which is often a limitation. Simple, low-cost systems to reproducibly grow many cyanobacterial cultures in parallel are necessary.

1.5. Genome-Scale Modeling and Fluxomics of Cyanobacteria

A primary aim of cyanobacterial synthetic biology is the production of particular metabolites as biofuels or platform chemicals. As such, better understanding the metabolic phenotypes of wild-type and synthetic strains is a critical aim. While cyanobacterial metabolomics have been recently reviewed (Schwarz *et al.* 2013), here we describe recent progress in genome-scale modeling and fluxomics of cyanobacteria. These approaches can help guide the creation of synthetic strains with desirable metabolic phenotypes such as biofuel overproduction via *in silico* prediction or *in vivo* measurement of metabolic fluxes (See figure 1.3). Specific to cyanobacterial systems, we highlight a number of challenges including complexity of modeling the photosynthetic metabolism and performing flux balance analysis, poor annotations of important metabolic pathways, and unavailability of *in vivo* gene essentiality information for most cyanobacteria. Finally, we focus on recent advancements in this area.

1.5.1. Challenges

Incorporating Photoautotrophy into Metabolic Models

Flux balance analysis (FBA) is a tool to make quantitative *in silico* predictions about metabolism (Fell *et al.* 1986; Savinell *et al.* 1992; Varma *et al.* 1993; Orth *et al.* 2010). An FBA model incorporates the stoichiometry of all genome-encoded metabolic reactions and assumes steady-state growth, such as during exponential phase. This assumption leads to a model that consists of a system of algebraic equations stating that the rate of producing any given metabolite is equal to the rate of consuming that metabolite. A solution to this system of equations is a possible answer to the question "what are all the metabolic fluxes in this system?" Since there are usually more reactions than metabolites, this system of equations is underdetermined and has many possible solutions. Therefore, one has to pick a solution that satisfies a biological

objective, such as maximal growth, energy production, or byproduct formation (Varma *et al.* 1994) or specify the ranges of each flux for which that objective is satisfied (Zomorrodi *et al.* 2012). For this purpose, a model will also include upper and lower bounds of fluxes that constrain the model to produce physically and biologically reasonable solutions.

Success of FBA greatly depends on the quality of the metabolic network reconstruction as well as the availability of regulatory constraints under a given environmental or growth condition. For instance, constraints can be added that disable or limit fluxes due to known regulatory constraints or substrate availability (Zomorrodi *et al.* 2012). For cyanobacteria, the major challenges to develop a genome-scale metabolic model and subsequently perform FBA are the same ones faced by these organisms in their diurnal environment: how to incorporate light and how to differentiate light and dark metabolisms. Although it has been nearly a decade since publication of the first study applying flux balance analysis to cyanobacteria, it is only recently that models have incorporated complete descriptions of the light reactions of photosynthesis (Nogales *et al.* 2012; Saha *et al.* 2012; Vu *et al.* 2012; Knoop *et al.* 2013). In so doing, these authors were able to highlight the critical importance of alternate electron flow pathways to growth under diverse environmental conditions, and to identify differences in metabolism during carbon-limited and light-limited growth. Additionally, this work helped to resoluve conflicts about the alleged existence of a glyoxylate shunt in cyanobacteria (Knoop *et al.* 2013). However, debate remains among photosynthesis researchers about the exact form of the light reactions (Heyes *et al.* 2005; Kopecna *et al.* 2013). This uncertainty about the exact stoichiometry of metabolism is a challenge for the predictive power of FBA in photosynthetic systems. While FBA requires the assumption of a pseudo-steady state, all cyanobacteria must alternate between day and night metabolisms during a diurnal cycle. A recent model (Saha *et al.* 2012) of

Cyanothece sp. ATCC 51142 utilizes proteomic data to model the diurnal rhythm of this strain, which fixes carbon during the day and nitrogen during the night (see section 1.3.1 above).

Incomplete Genome Annotation

Genome-scale models are built starting with an annotated genome sequence (see figure 1.3), which allows prediction of which metabolic reactions are available in a given strain. However, genome annotation is constantly evolving, and open questions remain about important metabolic reactions in cyanobacteria.

The understanding of several key pathways in cyanobacteria has been recently revised. Zhang and Bryant (Zhang *et al.* 2011) identified enzymes from *Synechococcus* 7002 that can complete the TCA cycle *in vitro* and have homologues in most cyanobacterial species, which were previously thought to possess an incomplete TCA cycle. Based on this information, *Synechocystis* 6803 model *i*Syn731 (Saha *et al.* 2012) allows for a complete TCA cycle including these reactions. However, using flux variability analysis (Mahadevan *et al.* 2002; Mahadevan *et al.* 2003) it was determined that this alternate pathway is not essential for maximal biomass production (unpublished results). Recently, experiments using ^{13}C metabolic tracers, detailed in appendix chapter 1 of this work (You *et al.* 2014), have provided direct, *in vivo* evidence for the activity of this pathway, but found that activity to be quite low under normal growth conditions.

Fatty acid metabolism in cyanobacteria has unique properties that have been recently uncovered due to increased interest in these pathways for biofuel production. Both *Synechocystis* sp. PCC 6803 and *Synechococcus elongatus* sp. PCC 7942 contain a single candidate gene annotated for fatty acid activation. While in both organisms the gene is annotated as acyl-CoA synthetase, it shows only acyl-ACP synthetase activity instead (Kaczmarzyk *et al.* 2010). Further

analysis also shows the importance of acyl-ACP synthetase in enabling the transfer of fatty acids across the membrane (von Berlepsch *et al.* 2012). Quinone synthesis is another pathway with conflicting annotations. Cyanobacteria contain neither ubiquinone nor menaquinone (Collins *et al.* 1981). Despite the lack of ubiquinone within cyanobacteria, a number of cyanobacterial genomes contain homologs for six *E. coli* genes involved in ubiquinone biosynthesis (Sakuragi 2004). Given these homologous genes it is probable that plastoquinone, a quinone molecule participating in the electron transport chain, is produced in cyanobacteria using a pathway very similar to that of ubiquinone production in proteobacteria. Other recent work (Wu *et al.* 2010) showed that *Cyanothece* 51142 contains an alternative pathway for isoleucine biosynthesis. Threonine ammonia-lyase, catalyzing the conversion of threonine to 2-ketobutyrate, is absent in *Cyanothece* 51142. Instead, this organism uses a citramalate pathway with pyruvate and acetyl-CoA as precursors for isoleucine synthesis. An intermediate in this pathway, namely ketobutyrate, can be converted to higher alcohols (propanol and butanol) via this non-fermentative alcohol production pathway. These active areas of research will help to better define cyanobacterial metabolism and allow the generation of models that can more accurately predict cellular phenotypes. While newer fluxomics techniques can yield powerful results in well-characterized strains, developing a 'green *E. coli*' will also require expanded knowledge of biochemistry that to date can only come from older methods of single gene or single protein analysis.

Fewer Mutant Resources to Test Model Accuracy

The quality or accuracy of any genome-scale metabolic model can be tested by contrasting the *in silico* growth phenotype with available experimental data on the viability of single or multiple gene knockouts (Thiele *et al.* 2010). Any discrepancies between model

predictions and observed results can aid in model refinement (Kumar *et al.* 2009). For model strains besides cyanobacteria, concerted efforts to create complete mutant libraries have led to improvements in metabolic modeling. To the best of our knowledge, extensive *in vivo* gene essentiality data are available only for *Synechocystis* 6803 among the cyanobacteria in the CyanoMutants database (Nakamura *et al.* 1999; Nakao *et al.* 2010), but only for ~119 genes, compared with 731 genes associated with metabolic reactions in a recent genome-scale model (Saha *et al.* 2012). Thus, only a small subset of the model predictions on gene essentiality can be evaluated using available data for *Synechocystis* 6803, and the proportion is much less for any other strain. While a genome-wide library of knockout mutants has been created in *Synechococcus* 7942 (Chen *et al.* 2012) segregation (and thus essentiality) has only been checked for a small selection of these mutants and is not available in any large-scale public database to date. Unavailability of such mutant information limits model validation and in turn hurts the value of computational predictions from FBA. Efforts to create complete mutant libraries in model cyanobacterial strains would improve the fidelity of genome-scale metabolic models, leading to testable hypotheses about how to alter metabolism for metabolite overproduction.

1.5.2. Recent Advances

Detailed Genome-Scale Models

Genome-scale models contain detailed Gene-Protein-Reaction associations, a stoichiometric representation of all possible reactions occurring in an organism, and a set of appropriate regulatory constraints on each reaction flux. They are differentiated from more basic FBA models simply by their completeness – they span all or nearly all of the metabolic reactions encoded in a genome. Thus, these models can have greater predictive value than those of only central metabolism. *Cyanothece* 51142 is one of the most potently diazotrophic unicellular

cyanobacteria characterized and the first diazotrophic cyanobacterium to be completely sequenced (Welsh *et al.* 2008). The first genome-scale model for *Cyanothece* 51142, *i*Cce806, is recently developed (Vu *et al.* 2012), while another more recent genome-scale model *i*Cyt773 contains an additional 266 unique reactions spanning pathways such as lipid, pigment and alkane biosynthesis (Saha *et al.* 2012). *i*Cyt773 also models diurnal metabolism by including flux regulation based on available day/night protein expression data (Stockel *et al.* 2011) and developing separate (light/dark) biomass equations. These models greatly enhance the ability to make computational predictions about this unique and promising diazotrophic organism.

Since *Synechocystis* 6803 is a model cyanobacterial strain, it has long been the target for modeling of photosynthetic central metabolism (Yang *et al.* 2002; Shastri *et al.* 2005). More recent models (Knoop *et al.* 2010; Montagud *et al.* 2011) analyze growth under different conditions and detect bottlenecks and gene knock-out candidates to enhance metabolite production (e.g., ethanol, succinate, and hydrogen). In addition, some such predictions using *iSyn*731 are discussed in chapter 4 of this work. A recent model represents the photosynthetic apparatus in detail, detects alternate flow pathways of electrons and also pinpoints photosynthetic robustness during photoautotrophic metabolism (Nogales *et al.* 2012). *i*Syn731, the latest of all *Synechocystis* 6803 models, integrates all recent developments and supplements them with improved metabolic capability and additional literature evidence. As many as 322 unique reactions are introduced in *i*Syn731 including reactions distributed in pathways such as heptadecane and fatty acid biosynthesis (Saha *et al.* 2012). Furthermore, *i*Syn731 is the first model for which both gene essentiality data (Nakamura *et al.* 1999) and MFA flux data (Young *et al.* 2011) are utilized to assess the predictive quality. This model has also recently been used to study the effect of cyclic electron flow on the growth of *Synechocystis* 6803 (Chapter 5 of this

work). Additionally, genome-scale modeling has been extended to include another model cyanobacterium, *Synechococcus* sp. PCC 7002 (Hamilton *et al.* 2012). Other model strains highlighted in table 1.1 have not yet had genome-scale models generated for their metabolism. Thus, stoichiometric models are emerging as a valuable tool for use across model cyanobacterial systems.

^{13}C MFA Analysis

While *in silico* models are great tools for generating hypotheses on how to use synthetic biology interventions to alter metabolism, they need to be complemented by fluxomics methods that allow *in vivo* measurement of metabolic fluxes to assess these interventions. Such a suite of tools allows the closure of the design-build-test engineering cycle in synthetic biology. To this end, Young et al. (2011) have developed a method to measure fluxes in autotrophic metabolism via dynamic isotope labeling measurements. In this approach, cultures are fed with a step-change from naturally labeled bicarbonate to $NaH^{13}CO_3$ and the labeling patterns of metabolic intermediates are followed over a time-coures to determine relative rates of metabolic flux. Previous studies (Yang *et al.* 2002) have also assessed metabolic fluxes under mixotrophic growth conditions, using a pseudo-steady-state approach in which cells are fed with ^{13}C labeled glucose and metabolic fluxes are inferred from labeling patterns of proteinogenic amino acids. These studies have been extremely useful in identifying fluxes that exist *in vivo*, but have previously been regarded as wasteful or futile cycles, such as the oxidative pentose phosphate pathway and RuBP oxygenation. Comparisons between flux measurements (Young *et al.* 2011) and flux predictions (Saha *et al.* 2012) for *Synechocystis* 6803 have revealed the necessity of additional regulatory information for accurate *in silico* predictions of phenotype. ^{13}C-MFA coupled with simpler tracer experiments also helped to delineate the role of a recently identified

alternative TCA cycle in cyanobacteria, as detailed in appendix chapter 1 of this work (You *et al.* 2014). These modeling and fluxomics efforts have resulted in deeper understanding of the metabolic capabilities of the modeled strains and of cyanobacteria in general.

1.6 Conclusions

Cyanobacterial synthetic biology offers great promise for enhancing efforts to produce biofuels and chemicals in photoautotrophic hosts. While several cyanobacterial chassis strains have been used in synthetic biology efforts, the tools for their manipulation and analysis need greater development to unlock this potential and develop a 'green *E. coli*'. Metabolic modeling is a complementary tool that can help guide the creation of synthetic strains with desirable phenotypes. By developing the tools for strain manipulation and control, synthetic biologists can unlock a bright future for the biotechnological use of abundant light and CO_2.

1.7 This Work

This initial chapter has outlined how the tools of systems and synthetic biology might in general contribute to metabolic engineering of cyanobacteria for biofuel production. The following chapters give examples of that as applied to production of heptadecane in *Synechocystis* sp. PCC 6803. In particular, chapter 2 details my attempts to characterize the physiological role of heptadecane as a metabolite in *Synechocystis* sp. PCC 6803 by analyzing a knockout mutant that does not produce heptadecane. The work in chapter 2 makes use a genome-scale metabolic model to contextualize our observation that an alkane-free mutant displays increased cyclic electron flow, especially at low temperature where the mutant grows poorly.

Chapter 3 details the construction and validation of that genome-scale metabolic model. Chapter 4 covers my efforts to engineer cyanobacteria for overproduction of heterologous genes specifically at stationary phase by first analyzing microarray and plasmid copy number data to find genes and replicons whose expression is specific to stationary phase. Subsequently, we used the promoters of those genes and those replicons as genetic parts to construct a synthetic biology system for protein overexpression at stationary phase. Chapter 5 covers my attempts to overproduce alkanes in *Synechocystis* sp. PCC 6803 via metabolic engineering and modifications of growth conditions. This chapter also contains background on the production of alkanes by cyanobacteria.

The appendix chapters to this dissertation are a number of other studies of which I have been a co-author during my time at Washington University. These studies are in the areas of *in silico* metabolic modeling and the measurement of metabolic fluxes using ^{13}C-labeling methods in cyanobacteria. Appendix chapter 1 provides direct evidence for the existence of a complete TCA cycle in *Synehocystis* sp. PCC 6803 via a pathway that is common to most cyanobacteria. However, this pathway is different from the better known TCA cycle pathway in heterotrophs and appears to primarily act as a bifurcated pathway as opposed to a cycle. Appendix chapter 4 details the analysis of the metabolism of *Cyanothece* sp. ATCC 51142 using ^{13}C tracers in continuous light. Appendix chapter 3 gives a general method for the experiments in appendix chapters 1 and 4 with an accompanying video that provides a user-friendly explanation of these detailed procedures. Finally, appendix chapter 2 details a rapid method for the construction of genome-scale models of novel sequenced cyanobacteria.

1.8 References

Abe, K., Miyake, K., Nakamura, M., Kojima, K., Ferri, S., Ikebukuro, K. and Sode, K. (2014). "Engineering of a green-light inducible gene expression system in Synechocystis sp. PCC6803." *Microb Biotechnol* **7**(2): 177-183.

Akiyama, S. (2012). "Structural and dynamic aspects of protein clocks: how can they be so slow and stable?" *Cell Mol Life Sci* **69**(13): 2147-2160.

Ashby, M. K. and Houmard, J. (2006). "Cyanobacterial two-component proteins: structure, diversity, distribution, and evolution." *Microbiol Mol Biol Rev* **70**(2): 472-509.

Atsumi, S., Higashide, W. and Liao, J. C. (2009). "Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde." *Nat Biotechnol* **27**(12): 1177-1180.

Beck, C., Hertel, S., Rediger, A., Lehmann, R., Wiegard, A., Kolsch, A., Heilmann, B., Georg, J., Hess, W. R. and Axmann, I. M. (2014). "Daily expression pattern of protein-encoding genes and small noncoding RNAs in synechocystis sp. strain PCC 6803." *Appl Environ Microbiol* **80**(17): 5195-5206.

Begemann, M. B., Zess, E. K., Walters, E. M., Schmitt, E. F., Markley, A. L. and Pfleger, B. F. (2013). "An organic acid based counter selection system for cyanobacteria." *PLoS One* **8**(10): e76594.

Bentley, F. K., Zurbriggen, A. and Melis, A. (2014). "Heterologous expression of the mevalonic acid pathway in cyanobacteria enhances endogenous carbon partitioning to isoprene." *Mol Plant* **7**(1): 71-86.

Berla, B. M. and Pakrasi, H. B. (2012). "Upregulation of plasmid genes during stationary phase in Synechocystis sp. strain PCC 6803, a cyanobacterium." *Appl Environ Microbiol* **78**(15): 5448-5451.

Berto, P., D'Adamo, S., Bergantino, E., Vallese, F., Giacometti, G. M. and Costantini, P. (2011). "The cyanobacterium Synechocystis sp. PCC 6803 is able to express an active [FeFe]-hydrogenase without additional maturation proteins." *Biochem Biophys Res Commun* **405**(4): 678-683.

Blasi, B., Peca, L., Vass, I. and Kos, P. B. (2012). "Characterization of stress responses of heavy metal and metalloid inducible promoters in synechocystis PCC6803." *J Microbiol Biotechnol* **22**(2): 166-169.

Boyanapalli, R., Bullerjahn, G. S., Pohl, C., Croot, P. L., Boyd, P. W. and McKay, R. M. (2007). "Luminescent whole-cell cyanobacterial bioreporter for measuring Fe availability in diverse marine environments." *Appl Environ Microbiol* **73**(3): 1019-1024.

Camsund, D., Heidorn, T. and Lindblad, P. (2014). "Design and analysis of LacI-repressed promoters and DNA-looping in a cyanobacterium." *J Biol Eng* **8**(1): 4.

Cheah, Y. E., Albers, S. C. and Peebles, C. A. (2013). "A novel counter-selection method for markerless genetic modification in Synechocystis sp. PCC 6803." *Biotechnol Prog* **29**(1): 23-30.

Chen, Y., Holtman, C. K., Taton, A. and Golden, S. S. (2012). "Functional Analysis of the *Synechococcus elongatus* PCC 7942 Genome." *Functional Genomics and Evolution of Photosynthetic Systems*. R. Burnap and W. Vermaas, Springer. **33:** 119-137.

Collins, M. D. and Jones, D. (1981). "Distribution of Isoprenoid Quinone Structural Types in Bacteria and Their Taxonomic Implications." *Microbiological Reviews* **45**(2): 316-354.

Drepper, T., Eggert, T., Circolone, F., Heck, A., Krauss, U., Guterl, J. K., Wendorff, M., Losi, A., Gartner, W. and Jaeger, K. E. (2007). "Reporter proteins for in vivo fluorescence without oxygen." *Nat Biotechnol* **25**(4): 443-445.

Duhring, U., Axmann, I. M., Hess, W. R. and Wilde, A. (2006). "An internal antisense RNA regulates expression of the photosynthesis gene isiA." *Proc Natl Acad Sci U S A* **103**(18): 7054-7058.

Dunlop, M. J., Dossani, Z. Y., Szmidt, H. L., Chu, H. C., Lee, T. S., Keasling, J. D., Hadi, M. Z. and Mukhopadhyay, A. (2011). "Engineering microbial biofuel tolerance and export using efflux pumps." *Mol Syst Biol* **7**: 487.

Engler, C. and Marillonnet, S. (2011). "Generation of families of construct variants using golden gate shuffling." *Methods Mol Biol* **729**: 167-181.

Englund, E., Pattanaik, B., Ubhayasekera, S. J., Stensjo, K., Bergquist, J. and Lindberg, P. (2014). "Production of squalene in Synechocystis sp. PCC 6803." *PLoS One* **9**(3): e90270.

Erbe, J. L., Adams, A. C., Taylor, K. B. and Hall, L. M. (1996). "Cyanobacteria carrying an smt-lux transcriptional fusion as biosensors for the detection of heavy metal cations." *J Ind Microbiol* **17**(2): 80-83.

Fell, D. A. and Small, J. R. (1986). "Fat synthesis in adipose tissue. An examination of stoichiometric constraints." *Biochem J* **238**(3): 781-786.

Gao, Z., Zhao, H., Li, Z., Tan, X. and Lu, X. (2012). "Photosynthetic production of ethanol from carbon dioxide in genetically engineered cyanobacteria." *Energy and Environmental Science* **5**: 9857-9865.

Ghim, C. M., Lee, S. K., Takayama, S. and Mitchell, R. J. (2010). "The art of reporter proteins in science: past, present and future applications." *BMB Rep* **43**(7): 451-460.

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 3rd and Smith, H. O. (2009). "Enzymatic assembly of DNA molecules up to several hundred kilobases." *Nat Methods* **6**(5): 343-345.

Gierga, G., Voss, B. and Hess, W. R. (2009). "The Yfr2 ncRNA family, a group of abundant RNA molecules widely conserved in cyanobacteria." *RNA Biol* **6**(3): 222-227.

Golden, S. S., Ishiura, M., Johnson, C. H. and Kondo, T. (1997). "Cyanobacterial Circadian Rhythms." *Annu Rev Plant Physiol Plant Mol Biol* **48**: 327-354.

Gronenberg, L. S., Marcheschi, R. J. and Liao, J. C. (2013). "Next generation biofuel engineering in prokaryotes." *Curr Opin Chem Biol*.

Guerrero, F., Carbonell, V., Cossu, M., Correddu, D. and Jones, P. R. (2012). "Ethylene synthesis and regulated expression of recombinant protein in Synechocystis sp. PCC 6803." *PLoS One* **7**(11): e50470.

Hamilton, J. J. and Reed, J. L. (2012). "Identification of Functional Differences in Metabolic Networks Using Comparative Genomics and Constraint-Based Models." *PLoS One* **7**(4).

Hansen, L. H., Knudsen, S. and Sorensen, S. J. (1998). "The effect of the lacY gene on the induction of IPTG inducible promoters, studied in Escherichia coli and Pseudomonas fluorescens." *Curr Microbiol* **36**(6): 341-347.

Hansen, M. C., Palmer, R. J., Jr., Udsen, C., White, D. C. and Molin, S. (2001). "Assessment of GFP fluorescence in cells of Streptococcus gordonii under conditions of low pH and low oxygen concentration." *Microbiology* **147**(Pt 5): 1383-1391.

He, L., Wu, G. and Tang, Y. J. (2015). "Simulating cyanobacterial phenotypes in photobioreactors by coupling flux balance analysis, kinetics and hydrodynamics." *Submitted to Journal of Applied Microbiology and Biotechnology*.

Heidorn, T., Camsund, D., Huang, H. H., Lindberg, P., Oliveira, P., Stensjo, K. and Lindblad, P. (2011). "Synthetic biology in cyanobacteria engineering and analyzing novel functions." *Methods Enzymol* **497**: 539-579.

Heyes, D. J. and Hunter, C. N. (2005). "Making light work of enzyme catalysis: protochlorophyllide oxidoreductase." *Trends in Biochemical Sciences* **30**(11): 642-649.

Hilson, N., Rosengarten, R. and Keasling, J. (2012). "j5 DNA Assembly Design Automation Software." *ACS Synthetic Biology* **1**(1): 14-21.

Holtman, C., Chen, Y., Sandoval, P., Gonzales, A., Nalty, M., Thomas, T., Youderian, P. and Golden, S. (2005). "High-Throughput Functional Analysis of the Synechococcus elongatus PCC 7942 Genome." *DNA Research* **12**: 103-115.

Huang, H. H., Camsund, D., Lindblad, P. and Heidorn, T. (2010). "Design and characterization of molecular tools for a Synthetic Biology approach towards developing cyanobacterial biotechnology." *Nucleic Acids Res* **38**(8): 2577-2593.

Huang, H. H. and Lindblad, P. (2013). "Wide-dynamic-range promoters engineered for cyanobacteria." *J Biol Eng* **7**(1): 10.

Imamura, S. and Asayama, M. (2009). "Sigma factors for cyanobacterial transcription." *Gene Regul Syst Bio* **3**: 65-87.

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

Kaczmarzyk, D. and Fulda, M. (2010). "Fatty acid activation in cyanobacteria mediated by acyl-acyl carrier protein synthetase enables fatty acid recycling." *Plant Physiol* **152**(3): 1598-1610.

Klahn, S., Baumgartner, D., Pfreundt, U., Voigt, K., Schon, V., Steglich, C. and Hess, W. R. (2014). "Alkane Biosynthesis Genes in Cyanobacteria and Their Transcriptional Organization." *Front Bioeng Biotechnol* **2**: 24.

Knoop, H., Grundel, M., Zilliges, Y., Lehmann, R., Hoffmann, S., Lockau, W. and Steuer, R. (2013). "Flux balance analysis of cyanobacterial metabolism: the metabolic network of Synechocystis sp. PCC 6803." *PLoS Comput Biol* **9**(6): e1003081.

Knoop, H., Zilliges, Y., Lockau, W. and Steuer, R. (2010). "The Metabolic Network of Synechocystis sp. PCC 6803: Systemic Properties of Autotrophic Growth." *Plant Physiology* **154**(1): 410-422.

Kopecna, J., Sobotka, R. and Komenda, J. (2013). "Inhibition of chlorophyll biosynthesis at the protochlorophyllide reduction step results in the parallel depletion of Photosystem I and Photosystem II in the cyanobacterium Synechocystis PCC 6803." *Planta* **237**(2): 497-508.

Kumar, V. S. and Maranas, C. D. (2009). "GrowMatch: an automated method for reconciling in silico/in vivo growth predictions." *PLoS Comput Biol* **5**(3): e1000308.

Kunert, A., Vinnemeier, J., Erdmann, N. and Hagemann, M. (2003). "Repression by Fur is not the main mechanism controlling the iron-inducible isiAB operon in the cyanobacterium Synechocystis sp. PCC 6803." *FEMS Microbiol Lett* **227**(2): 255-262.

Lagarde, D., Beuf, L. and Vermaas, W. (2000). "Increased Production of Zeaxanthin and Other Pigments by Application of Genetic Engineering Techniques to Synechocystis sp. Strain PCC 6803." *Applied and Environmental Microbiology* **66**(1): 64-72.

Landry, B., Stockel, J. and Pakrasi, H. (2012). "Use of Degradation Tags to Control Protein Levels in the Cyanobacterium *Synechocystis* sp. Strain PCC 6803." *Applied and Environmental Microbiology* **70**(8): 2833-2835.

Latysheva, N., Junker, V. L., Palmer, W. J., Codd, G. A. and Barker, D. (2012). "The evolution of nitrogen fixation in cyanobacteria." *Bioinformatics* **28**(5): 603-606.

Li, M. Z. and Elledge, S. J. (2007). "Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC." *Nat Methods* **4**(3): 251-256.

Liu, X., Sheng, J. and Curtiss, R., 3rd (2011). "Fatty acid production in genetically modified cyanobacteria." *Proc Natl Acad Sci U S A* **108**(17): 6899-6904.

Ma, A. T., Schmidt, C. M. and Golden, J. W. (2014). "Regulation of gene expression in diverse cyanobacterial species by using theophylline-responsive riboswitches." *Appl Environ Microbiol* **80**(21): 6704-6713.

Mackey, S. R., Ditty, J. L., Clerico, E. M. and Golden, S. S. (2007). "Detection of rhythmic bioluminescence from luciferase reporters in cyanobacteria." *Methods Mol Biol* **362**: 115-129.

Mahadevan, R., Edwards, J. S. and Doyle, F. J. (2002). "Dynamic flux balance analysis of diauxic growth in Escherichia coli." *Biophysical Journal* **83**(3): 1331-1340.

Mahadevan, R. and Schilling, C. H. (2003). "The effects of alternate optimal solutions in constraint-based genome-scale metabolic models." *Metabolic Engineering* **5**: 264-276.

Markley, A. L., Begemann, M. B., Clarke, R. E., Gordon, G. C. and Pfleger, B. F. (2014). "Synthetic Biology Toolbox for Controlling Gene Expression in the Cyanobacterium Synechococcus sp. strain PCC 7002." *ACS Synth Biol*.

Meighen, E. A. (1993). "Bacterial bioluminescence: organization, regulation, and application of the lux genes." *FASEB J* **7**(11): 1016-1022.

Melis, A. (2013). "Carbon partitioning in photosynthesis." *Curr Opin Chem Biol*.

Michel, K. P., Pistorius, E. K. and Golden, S. S. (2001). "Unusual regulatory elements for iron deficiency induction of the idiA gene of Synechococcus elongatus PCC 7942." *J Bacteriol* **183**(17): 5015-5024.

Mitschke, J., Georg, J., Scholz, I., Sharma, C. M., Dienst, D., Bantscheff, J., Voss, B., Steglich, C., Wilde, A., Vogel, J. and Hess, W. R. (2011). "An experimentally anchored map of transcriptional start sites in the model cyanobacterium Synechocystis sp. PCC6803." *Proc Natl Acad Sci U S A* **108**(5): 2124-2129.

Montagud, A., Zelezniak, A., Navarro, E., de Cordoba, P., Urchueguia, J. F. and Patil, K. R. (2011). "Flux coupling and transcriptional regulation within the metabolic network of the photosynthetic bacterium Synechocystis sp PCC6803." *Biotechnology Journal* **6**(3): 330-342.

Montgomery, B. L. (2007). "Sensing the light: photoreceptive systems and signal transduction in cyanobacteria." *Mol Microbiol* **64**(1): 16-27.

Moon, T. S., Lou, C., Tamsir, A., Stanton, B. C. and Voigt, C. A. (2012). "Genetic programs constructed from layered logic gates in single cells." *Nature* **491**(7423): 249-253.

Mukherjee, A., Weyant, K. B., Walker, J. and Schroeder, C. M. (2012). "Directed evolution of bright mutants of an oxygen-independent flavin-binding fluorescent protein from Pseudomonas putida." *J Biol Eng* **6**(1): 20.

Nagarajan, A., Winter, R., Eaton-Rye, J. and Burnap, R. (2011). "A synthetic DNA and fusion PCR approach to the ectopic expression of high levels of the D1 protein of photosystem II in Synechocystis sp. PCC 6803." *J Photochem Photobiol B* **104**(1-2): 212-219.

Nakahira, Y., Ogawa, A., Asano, H., Oyama, T. and Tozawa, Y. (2013). "Theophylline-dependent riboswitch as a novel genetic tool for strict regulation of protein expression in Cyanobacterium Synechococcus elongatus PCC 7942." *Plant Cell Physiol* **54**(10): 1724-1735.

Nakajima, M., Imai, K., Ito, H., Nishiwaki, T., Murayama, Y., Iwasaki, H., Oyama, T. and Kondo, T. (2005). "Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation in vitro." *Science* **308**(5720): 414-415.

Nakamura, Y., Kaneko, T., Miyajima, N. and Tabata, S. (1999). "Extension of CyanoBase. CyanoMutants: repository of mutant information on Synechocystis sp. strain PCC6803." *Nucleic Acids Res* **27**(1): 66-68.

Nakao, M., Okamoto, S., Kohara, M., Fujishiro, T., Fujisawa, T., Sato, S., Tabata, S., Kaneko, T. and Nakamura, Y. (2010). "CyanoBase: the cyanobacteria genome database update 2010." *Nucleic Acids Res* **38**(Database issue): D379-381.

Ng, W. O., Zentella, R., Wang, Y., Taylor, J. S. and Pakrasi, H. B. (2000). "PhrA, the major photoreactivating factor in the cyanobacterium Synechocystis sp. strain PCC 6803 codes for a cyclobutane-pyrimidine-dimer-specific DNA photolyase." *Arch Microbiol* **173**(5-6): 412-417.

Nogales, J., Gudmundsson, S., Knight, E. M., Palsson, B. O. and Thiele, I. (2012). "Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis." *Proc Natl Acad Sci U S A* **109**(7): 2678-2683.

Nozzi, N. E., Oliver, J. W. and Atsumi, S. (2013). "Cyanobacteria as a Platform for Biofuel Production." *Front Bioeng Biotechnol* **1**: 7.

Oliver, J. W., Machado, I. M., Yoneda, H. and Atsumi, S. (2013). "Cyanobacterial conversion of carbon dioxide to 2,3-butanediol." *Proc Natl Acad Sci U S A* **110**(4): 1249-1254.

Orth, J. D., Thiele, I. and Palsson, B. O. (2010). "What is flux balance analysis?" *Nature Biotechnology* **28**(3): 245-248.

Peca, L., Kos, P. B., Mate, Z., Farsang, A. and Vass, I. (2008). "Construction of bioluminescent cyanobacterial reporter strains for detection of nickel, cobalt and zinc." *FEMS Microbiol Lett* **289**(2): 258-264.

Peca, L., Kos, P. B. and Vass, I. (2007). "Characterization of the activity of heavy metal-responsive promoters in the cyanobacterium Synechocystis PCC 6803." *Acta Biol Hung* **58 Suppl**: 11-22.

Pfreundt, U., Stal, L. J., Voss, B. and Hess, W. R. (2012). "Dinitrogen fixation in a unicellular chlorophyll d-containing cyanobacterium." *ISME J* **6**(7): 1367-1377.

Quan, J. and Tian, J. (2009). "Circular polymerase extension cloning of complex gene libraries and pathways." *PLoS One* **4**(7): e6441.

Saha, R., Verseput, A. T., Berla, B. M., Mueller, T. J., Pakrasi, H. B. and Maranas, C. D. (2012). "Reconstruction and comparison of the metabolic potential of cyanobacteria Cyanothece sp. ATCC 51142 and Synechocystis sp. PCC 6803." *PLoS One* **7**(10): e48285.

Sakuragi, Y. (2004). *Studies of Quinones in Cyanobacteria*. Doctor of Philosophy, The Pennsylvania State University.

Satya Lakshmi, O. and Rao, N. M. (2009). "Evolving Lac repressor for enhanced inducibility." *Protein Eng Des Sel* **22**(2): 53-58.

Savinell, J. M. and Palsson, B. O. (1992). "Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism." *J Theor Biol* **154**(4): 421-454.

Schwarz, D., Orf, I., Kopka, J. and Hagemann, M. (2013). "Recent Applications of Metabolomics Toward Cyanobacteria." *Metabolites* **3**(1): 72-100.

Selinger, D. W., Cheung, K. J., Mei, R., Johansson, E. M., Richmond, C. S., Blattner, F. R., Lockhart, D. J. and Church, G. M. (2000). "RNA expression analysis using a 30 base pair resolution Escherichia coli genome array." *Nat Biotechnol* **18**(12): 1262-1268.

Sforza, E., Simionato, D., Giacometti, G. M., Bertucco, A. and Morosinotto, T. (2012). "Adjusted light and dark cycles can optimize photosynthetic efficiency in algae growing in photobioreactors." *PLoS One* **7**(6): e38975.

Shao, Z. and Zhao, H. (2009). "DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways." *Nucleic Acids Res* **37**(2): e16.

Sharma, C. M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiss, S., Sittka, A., Chabas, S., Reiche, K., Hackermuller, J., Reinhardt, R., Stadler, P. F. and Vogel, J. (2010). "The primary transcriptome of the major human pathogen Helicobacter pylori." *Nature* **464**(7286): 250-255.

Shastri, A. A. and Morgan, J. A. (2005). "Flux balance analysis of photoautotrophic metabolism." *Biotechnology Progress* **21**(6): 1617-1626.

Simkovsky, R., Daniels, E. F., Tang, K., Huynh, S. C., Golden, S. S. and Brahamsha, B. (2012). "Impairment of O-antigen production confers resistance to grazing in a model amoeba-cyanobacterium predator-prey system." *Proc Natl Acad Sci U S A* **109**(41): 16678-16683.

Stockel, J., Jacobs, J. M., Elvitigala, T. R., Liberton, M., Welsh, E. A., Polpitiya, A. D., Gritsenko, M. A., Nicora, C. D., Koppenaal, D. W., Smith, R. D. and Pakrasi, H. B. (2011). "Diurnal rhythms result in significant changes in the cellular protein complement in the cyanobacterium Cyanothece 51142." *PLoS One* **6**(2): e16680.

Stockel, J., Welsh, E. A., Liberton, M., Kunnvakkam, R., Aurora, R. and Pakrasi, H. B. (2008). "Global transcriptomic analysis of Cyanothece 51142 reveals robust diurnal oscillation of central metabolic processes." *Proc Natl Acad Sci U S A* **105**(16): 6156-6161.

Szewczyk, E., Nayak, T., Oakley, C. E., Edgerton, H., Xiong, Y., Taheri-Talesh, N., Osmani, S. A. and Oakley, B. R. (2007). "Fusion PCR and gene targeting in Aspergillus nidulans." *Nat Protoc* **1**(6): 3111-3120.

Takahama, K., Matsuoka, M., Nagahama, K. and Ogawa, T. (2004). "High-Frequency Gene Replacement in Cyanobacteria Using a Heterologous rps12 Gene." *Plant Cell Physiology* **45**(3): 333-339.

Tan, X., Liang, F., Cai, K. and Lu, X. (2013). "Application of the FLP/FRT recombination system in cyanobacteria for construction of markerless mutants." *Appl Microbiol Biotechnol*.

Taniuchi, Y., Yoshikawa, S., Maeda, S., Omata, T. and Ohki, K. (2008). "Diazotrophy under continuous light in a marine unicellular diazotrophic cyanobacterium, Gloeothece sp. 68DGA." *Microbiology* **154**(Pt 7): 1859-1865.

Taton, A., Unglaub, F., Wright, N. E., Zeng, W. Y., Paz-Yepes, J., Brahamsha, B., Palenik, B., Peterson, T. C., Haerizadeh, F., Golden, S. S. and Golden, J. W. (2015). "Broad-host-range vector system for synthetic biology and biotechnology in cyanobacteria." *Nucleic Acids Res* **42**(17): e136.

Teng, S. W., Mukherji, S., Moffitt, J. R., de Buyl, S. and O'Shea, E. K. (2013). "Robust circadian oscillations in growing cyanobacteria require transcriptional feedback." *Science* **340**(6133): 737-740.

Thiele, I. and Palsson, B. O. (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nat Protoc* **5**(1): 93-121.

Topp, S. and Gallivan, J. P. (2007). "Guiding bacteria with small molecules and RNA." *J Am Chem Soc* **129**(21): 6807-6811.

Topp, S., Reynoso, C. M., Seeliger, J. C., Goldlust, I. S., Desai, S. K., Murat, D., Shen, A., Puri, A. W., Komeili, A., Bertozzi, C. R., Scott, J. R. and Gallivan, J. P. (2010). "Synthetic riboswitches that induce gene expression in diverse bacterial species." *Appl Environ Microbiol* **76**(23): 7881-7884.

Tyo, K. E., Jin, Y. S., Espinoza, F. A. and Stephanopoulos, G. (2009). "Identification of gene disruptions for increased poly-3-hydroxybutyrate accumulation in Synechocystis PCC 6803." *Biotechnol Prog* **25**(5): 1236-1243.

Varma, A., Boesch, B. W. and Palsson, B. O. (1993). "Stoichiometric Interpretation of Escherichia-Coli Glucose Catabolism under Various Oxygenation Rates." *Applied and Environmental Microbiology* **59**(8): 2465-2473.

Varma, A. and Palsson, B. O. (1994). "Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110." *Appl Environ Microbiol* **60**(10): 3724-3731.

Varman, A. M., Yu, Y., You, L. and Tang, Y. J. (2013). "Photoautotrophic production of D-lactic acid in an engineered cyanobacterium." *Microb Cell Fact* **12**: 117.

Viola, S., Ruhle, T. and Leister, D. (2014). "A single vector-based strategy for marker-less gene replacement in Synechocystis sp. PCC 6803." *Microb Cell Fact* **13**: 4.

von Berlepsch, S., Kunz, H. H., Brodesser, S., Fink, P., Marin, K., Flugge, U. I. and Gierth, M. (2012). "The acyl-acyl carrier protein synthetase from Synechocystis sp. PCC 6803 mediates fatty acid import." *Plant Physiol* **159**(2): 606-617.

Vu, T. T., Stolyar, S. M., Pinchuk, G. E., Hill, E. A., Kucek, L. A., Brown, R. N., Lipton, M. S., Osterman, A., Fredrickson, J. K., Konopka, A. E., Beliaev, A. S. and Reed, J. L. (2012). "Genome-scale modeling of light-driven reductant partitioning and carbon fluxes in diazotrophic unicellular cyanobacterium Cyanothece sp. ATCC 51142." *PLoS Comput Biol* **8**(4): e1002460.

Wang, B., Pugh, S., Nielsen, D. R., Zhang, W. and Meldrum, D. R. (2013). "Engineering cyanobacteria for photosynthetic production of 3-hydroxybutyrate directly from CO." *Metab Eng* **16C**: 68-77.

Wang, B., Wang, J., Zhang, W. and Meldrum, D. R. (2012). "Application of synthetic biology in cyanobacteria and algae." *Front Microbiol* **3**: 344.

Wang, W., Liu, X. and Lu, X. (2013). "Engineering cyanobacteria to improve photosynthetic production of alka(e)nes." *Biotechnol Biofuels* **6**(1): 69.

Waters, C. M. and Bassler, B. L. (2006). "The Vibrio harveyi quorum-sensing system uses shared regulatory components to discriminate between multiple autoinducers." *Genes Dev* **20**(19): 2754-2767.

Welsh, E. A., Liberton, M., Stockel, J., Loh, T., Elvitigala, T., Wang, C., Wollam, A., Fulton, R. S., Clifton, S. W., Jacobs, J. M., Aurora, R., Ghosh, B. K., Sherman, L. A., Smith, R. D., Wilson, R. K. and Pakrasi, H. B. (2008). "The genome of Cyanothece 51142, a unicellular diazotrophic cyanobacterium important in the marine nitrogen cycle." *Proc Natl Acad Sci U S A* **105**(39): 15094-15099.

Woelfle, M. A., Ouyang, Y., Phanvijhitsiri, K. and Johnson, C. H. (2004). "The adaptive value of circadian clocks: an experimental assessment in cyanobacteria." *Curr Biol* **14**(16): 1481-1486.

Wu, B., Zhang, B. C., Feng, X. Y., Rubens, J. R., Huang, R., Hicks, L. M., Pakrasi, H. B. and Tang, Y. J. J. (2010). "Alternative isoleucine synthesis pathway in cyanobacterial species." *Microbiology-Sgm* **156**: 596-602.

Xu, Y., Alvey, R. M., Byrne, P. O., Graham, J. E., Shen, G. and Bryant, D. A. (2011). "Expression of genes in cyanobacteria: adaptation of endogenous plasmids as platforms for high-level gene expression in Synechococcus sp. PCC 7002." *Methods Mol Biol* **684**: 273-293.

Xu, Y., Ma, P., Shah, P., Rokas, A., Liu, Y. and Johnson, C. H. (2013). "Non-optimal codon usage is a mechanism to achieve circadian clock conditionality." *Nature* **495**(7439): 116-120.

Xue, Y., Zhang, Y., Cheng, D., Daddy, S. and He, Q. (2014). "Genetically engineering Synechocystis sp. Pasteur Culture Collection 6803 for the sustainable production of the plant secondary metabolite p-coumaric acid." *Proc Natl Acad Sci U S A* **111**(26): 9449-9454.

Yang, C., Hua, Q. and Shimizu, K. (2002). "Metabolic flux analysis in Synechocystis using isotope distribution from 13C-labeled glucose." *Metab Eng* **4**(3): 202-216.

You, L., Berla, B., He, L., Pakrasi, H. B. and Tang, Y. J. (2014). "13C-MFA delineates the photomixotrophic metabolism of Synechocystis sp. PCC 6803 under light- and carbon-sufficient conditions." *Biotechnol J* **9**(5): 684-692.

Young, J. D., Shastri, A. A., Stephanopoulos, G. and Morgan, J. A. (2011). "Mapping photoautotrophic metabolism with isotopically nonstationary (13)C flux analysis." *Metabolic Engineering* **13**(6): 656-665.

Zhang, F., Carothers, J. M. and Keasling, J. D. (2012). "Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids." *Nat Biotechnol* **30**(4): 354-359.

Zhang, S. Y. and Bryant, D. A. (2011). "The Tricarboxylic Acid Cycle in Cyanobacteria." *Science* **334**(6062): 1551-1553.

Zhang, Y., Pu, H., Wang, Q., Cheng, S., Zhao, W., Zhang, Y. and Zhao, J. (2007). "PII is important in regulation of nitrogen metabolism but not required for heterocyst formation in the Cyanobacterium Anabaena sp. PCC 7120." *J Biol Chem* **282**(46): 33641-33648.

Zomorrodi, A. R., Suthers, P. F., Ranganathan, S. and Maranas, C. D. (2012). "Mathematical optimization applications in metabolic networks." *Metab Eng* **14**(6): 672-686.

Table 1.1: Model strains of cyanobacteria for synthetic biology.

Strain	Genetic Methods	Ideal Growth Temp (C)	Doubling Time (hours)	Metabolism	Genome-Scale model(s)?	Selected Reference
Synechocystis sp. PCC 6803	conjugation, natural transformation, Tn5 mutagenesis, fusion PCR	30	6-12	mixotrophic, autotrophic	Yes	(Heidorn *et al.* 2011)
Notes:	Extensive systems biology datasets are available					
Synechococcus elongatus PCC 7942	conjugation, natural transformation, Tn5 mutagenesis	38	12-24	autotrophic, mixotrophic*	No	(Chen *et al.* 2012)
Notes:	A model strain for the study of circadian clocks					
Synechococcus sp. PCC 7002	conjugation, natural transformation	38	3.5	mixotrophic, autotrophic	Yes	(Xu *et al.* 2011; Markley, *et al.* 2014)
Notes:	Among the fastest-growing strains known					
Anabaena variabilis PCC 7120	conjugation, natural transformation	30	>24	mixotrophic, autotrophic	No	(Zhang *et al.* 2007)
Notes:	Nitrogen-fixing, Filamentous					
Leptolyngbya sp. Strain BL0902	conjugation, Tn5 mutagenesis	30	~20	autotrophic	No	(Taton *et al.* 2012)
Notes:	Filamentous, Grows well in outdoor photo-bioreactors in a broad range of conditions					

* This strain has been engineered to utilize several sugars, although the wild-type strain does not (McEwen *et al.* 2013)

Table 1.2: Inducible promoters used in cyanobacterial hosts.

Promoter	Source	Inducer/ Repressor & Concentration	Expression Host	Expressed Gene	Dynamic Range (fold-change unless specified)	Measure of Expression	Reference
A1lacO-1	*E. coli*	Inducer IPTG 1 mM	*Synechocystis* sp. PCC 6803	Gene encoding EFE from *Pseudomonas syringae*	8	170 nl ethylene / (ml h)	Guerrero *et al.* 2012
arsB	*Synechocystis* sp. PCC 6803	Inducer AsO2- 720 µM	*Synechocystis* sp. PCC 6803	*arsB*	100	RT-PCR	Blasi *et al.* 2012
coa	*Synechocystis* sp. PCC 6803	Inducer Co2+ 6 µM	*Synechocystis* sp. PCC 6803	gene encoding EFE from *Pseudomonas syringae*	500	48 nl ethylene / (ml h)	Guerrero *et al.* 2012
coat	*Synechocystis* sp. PCC 6803	Inducer Co2+ 1 µM	*Synechocystis* sp. PCC 6803	*coaT*	10	RT-PCR	Blasi *et al.* 2012
coat	*Synechocystis* sp. PCC 6803	Inducer Zn2+ 4 µM	*Synechocystis* sp. PCC 6803	*coaT*	8	RT-PCR	Blasi *et al.* 2012
coat	*Synechocystis* sp. PCC 6803	Inducer Co2+ 6.4 µM	*Synechocystis* sp. PCC 6803	*coaR+luxAB*	70	70 (relative luminescence)	Peca *et al.* 2008
coat	*Synechocystis* sp. PCC 6803	Inducer Zn2+ 3.2 µM	*Synechocystis* sp. PCC 6803	*coaR+luxAB*	25	25 (relative luminescence)	Peca *et al.* 2008
coat	*Synechocystis* sp. PCC 6803	Inducer Co2+ 3 µM	*Synechocystis* sp. PCC 6803	*coaT*	10	RT-PCR	Peca *et al.* 2007
coat	*Synechocystis* sp. PCC 6803	Inducer Zn2+ 5 µM	*Synechocystis* sp. PCC 6803	*coaT*	10	RT-PCR	Peca *et al.* 2007
idiA	*Synechococcus elongatus* PCC 7942	Repressor Fe2+ 0.043mM	*Synechococcus elongatus* PCC 7942	*luxAB*	170	Luminescence 5.3×10^6 cpm	Michel *et al.* 2001

Table 1.2 (continued): Inducible promoters used in cyanobacterial hosts.

Promoter	Source	Inducer/ Repressor & Concentration	Expression Host	Expressed Gene	Dynamic Range (fold-change unless specified)	Measure of Expression	Reference
isiAB	*Synechocystis* sp. PCC 6803	Repressor Fe3+ 30 µM	*Synechocystis* sp. PCC 6803	*isiAB+gfp*	5000	From 5000 GFP RFU	Kunert *et al.* 2003
isiAB	*Synechococcus* sp. strain PCC 7002	Repressor Fe3+ 100 nM	*Synechococcus* sp. strain PCC 7002	*luxAB from Vibrio harveyi*	2	From 1.2 x 10^8 RLU $cell^{-1}$ s^{-1}	Boyanapalli *et al.* 2007
LlacO1[1]	*E. coli*	Inducer IPTG 1mM	*Synechococcus elongatus* PCC7942	*alsS (B. subtilis), alsD (A. hydrophila),* and *adh (C. beijerinckii)*	1.6	Rel. activity of sADH and ALS	Oliver *et al.* 2013
nir	*Anabaena* sp. PCC 7120	Inducer/Repressor NO3-/NH4+ 5.9 mM/10.0 mM	*Anabaena* sp. PCC 7120	*nir*	Qualified, but not quantified	250 mg labeled proteins for NMR/L	Desplancq *et al.* 2005
nirA	*Synechococcus elongatus* PCC 7942	Inducer/Repressor NO3-/NH4+ 15.0 mM/3.75 mM	*Synechococcus elongatus* PCC 7942	*cmpABCD*	5	260 nmol HCO3-/mg Chl	Omata *et al.* 1999
nirA	*Synechococcus elongatus* PCC 7942	Inducer/Repressor NO3-/NH4+ 17.6 mM/17.6 mM	*Synechocystis* sp. PCC 6803	gene encoding p-hydroxy-phenylpyruvate dioxygenase from *A. thaliana*	25	250 ng tocopherol / mg dcw	Qi *et al.* 2005
nrsB	*Synechocystis* sp. PCC 6803	Inducer Ni2+ 5 µM	*Synechocystis* sp. PCC 6803	*nrsB*	400	RT-PCR	Blasi *et al.* 2012
nrsB	*Synechocystis* sp. PCC 6803	Inducer Co2+ 3 µM	*Synechocystis* sp. PCC 6803	*nrsB*	10	RT-PCR	Peca *et al.* 2007
nrsBACD	*Synechocystis* sp. PCC 6803	Inducer Ni2+ 6.4 µM	*Synechocystis* sp. PCC 6803	*nrsR+luxAB*	50	50 (relative luminescence)	Peca *et al.* 2008
nrsB	*Synechocystis* sp. PCC 6803	Inducer Ni2+ 0.5 µM	*Synechocystis* sp. PCC 6803	*nrsB*	1000	RT-PCR	Peca *et al.* 2007

Table 1.2 (continued): Inducible promoters used in cyanobacterial hosts.

Promoter	Source	Inducer/ Repressor & Concentration	Expression Host	Expressed Gene	Dynamic Range (fold-change unless specified)	Measure of Expression	Reference
petE	*Synechocystis* sp. PCC 6803	Inducer Cu2+ 0.3μM	*Anabaena* sp. PCC 7120	*hetN* (prevents heterocyst formation)	Qualified, but not quantified	0% heterocysts from 10% uninduced	Callahan *et al.* 2001
petE	*Synechocystis* sp. PCC 6803	Inducer Cu2+ 0.5 μM	*Synechocystis* sp. PCC 6803	gene encoding EFE from *Pseudomonas syringae*	5	28 nl ethylene / ml h	Guerrero *et al.* 2012
petE	*Synechocystis* sp. PCC 6803	Inducer Cu2+ 3μM	*Anabaena* sp. PCC 7120	*hetP*	4.5	8% heterocyst frequency	Higa *et al.* 2010
psbA1	*Anabaena* sp. PCC 7120	Inducer light 30 μEm-2s-1	*Anabaena* sp. PCC 7120	*hetR* from *E. coli*	Qualified, but not quantified	17% heterocyst frequency	Chaurasia *et al.* 2011
psbA2	*Synechocystis* sp. PCC6803	Inducer light 500 μE m^{-2} s^{-1}	*Synechocystis* sp. PCC6803	*ispS* from *Pueraria montana* (kudzu)	Qualified, but not quantified	~50 mg isoprene per g DCW per day	Lindberg *et al.* 2010
psbA2[2]	*Synechocystis* sp. PCC 6803	Inducer light 50 μE m^{-2} s^{-1}	*Synechocystis* sp. PCC 6803	*hydA1* from *Chlamydomonas reinhardtii*	Qualified, but not quantified	130 nmol H_2 mg Chl^{-1} min^{-1}	Berto *et al.* 2011
smt	*Synechococcus elongatus* PCC 7942	Inducer Zn2+ 2 μM	*Synechococcus elongatus* PCC 7942	*luxCDABE* from *Vibrio fisheri*	300	325,000 cps luminescence	Erbe *et al.* 1996
smt	*Synechococcus elongatus* PCC 7002	Inducer Zn2+ 2 μM	*Synechocystis* sp. PCC 6803	gene encoding EFE from *Pseudomonas syringae*	2	2 nl ethylene / ml h	Guerrero *et al.* 2012
tetR[3]	*E. coli*	Inducer aTc 1000 ng/ml	*Synechocystis* sp. strain ATCC27184	EYFP	290	over 10,000 emission/cell (a.u.)	Huang *et al.* 2013
trc	*E. coli*	Inducer IPTG 1mM	*Synechococcus elongatus* PCC 7942	*uidA* from *E. coli* (β-Glucuronidase)	36	340 nmol MU min^{-1} mg $protein^{-1}$	Geerts *et al.* 1995

Table 1.2 (continued): Inducible promoters used in cyanobacterial hosts.

Promoter	Source	Inducer/ Repressor & Concentration	Expression Host	Expressed Gene	Dynamic Range (fold-change unless specified)	Measure of Expression	Reference
trc[4]	*E. coli*	Inducer IPTG 1 mM	*Synechocystis* sp. PCC 6803	gene encoding EFE from *Pseudomonas syringae*	No significant difference	170 nl ethylene/ml h	Guerrero *et al.* 2012
trc10	*E. coli*	Inducer IPTG 2 mM	*Synechocystis* sp. PCC 6803	gene encoding GFPmut3B	1.6	101 (relative to P*lacI* activity)	Huang *et al.* 2010
trc20	*E. coli*	Inducer IPTG 2 mM	*Synechocystis* sp. PCC 6803	gene encoding GFPmut3B	4	12 (relative to P*lacI* activity)	Huang *et al.* 2010
trp-lac	*E. coli*	Inducer IPTG 100 µM	*Synechococcus elongatus* PCC 7942	*invA* and *glf* genes from *Zymomonas mobilis*	160 for fructose, 30 for glucose	160 µM fructose + 30 µM glucose	Niederholt-meyer *et al.* 2010
ziaA[5]	*Synechocystis* sp. PCC 6803	Inducer Zn2+ 3.5 µM	*Synechocystis* sp. PCC 6803	*hydA1* from *Chlamydomonas reinhardtii*	Qualified, but not quantified	109 nmol H2 mg Chl-1 min-1	Berto *et al.* 2011
ziaA	*Synechocystis* sp. PCC 6803	Inducer Cd2+ 2 µM	*Synechocystis* sp. PCC 6803	*ziaA*	10	RT-PCR	Blasi *et al.* 2012
ziaA	*Synechocystis* sp. PCC 6803	Inducer Zn2+ 4 µM	*Synechocystis* sp. PCC 6803	*ziaA*	40	RT-PCR	Blasi *et al.* 2012
ziaA	*Synechocystis* sp. PCC 6803	Inducer Zn2+ 5 µM	*Synechocystis* sp. PCC 6803	*ziaA*	40	RT-PCR	Peca *et al.* 2007
Theophylline-activated riboswitch[6]	Synthetic	Inducer Theophylline 2 mM	*Synechococcus elongatus* PCC 7942	*luxAB*	190.4	Luminescence	Nakahira *et al.* 2013
Hybrid Ptrc + Theophylline-activated riboswitch[6]	*E. coli,* Synthetic	Inducer Theophylline/IPTG 2 mM/1 mM	*Synechococcus elongatus* PCC 7942 (and others)	*yfp*	31	Fluorescence	Ma *et al.* 2014

Table 1.2 (continued): Inducible promoters used in cyanobacterial hosts.

Promoter	Source	Inducer/ Repressor & Concentration	Expression Host	Expressed Gene	Dynamic Range (fold-change unless specified)	Measure of Expression	Reference
cpcG2	*Synechocystis* sp. PCC 6803	Inducer Green light 20 μE m^{-2} s^{-1}	*Synechocystis* sp. PCC 6803	*gfp*	~5	Fluorescence	Abe *et al.* 2014
cpt hybrid	*Synechocystis* sp. PCC 6803 + *E. coli*	Inducer IPTG 5 mM	*Synechococcus* sp. PCC 7002	*yfp*	48	Fluorescence	Markley *et al.* 2014
*trc*2O	*E. coli / synthetic*	Inducer IPTG 1 mM	*Synechocystis* sp. PCC 6803	*yfp*	13	Fluorescence	Camsund *et al.* 2014

[1] – Leaky production of 2,3-butanediol, no IPTG and 1 mM IPTG similar

[2] – In the presence of 5 μM DCMU, which inhibits PSII-dependent oxygen evolution

[3] – Grown in the dark on 5 mM glucose

[4] – P_{lac} variants had differential expression early in growth phase but dynamic range was reduced as growth proceeded.

[5] – In the presence of 5 μM DCMU, which inhibits the PSII-dependent oxygen evolution

[6] – While not a promoter, the theophylline-dependent riboswitch holds promise for translational expression control in diverse hosts.

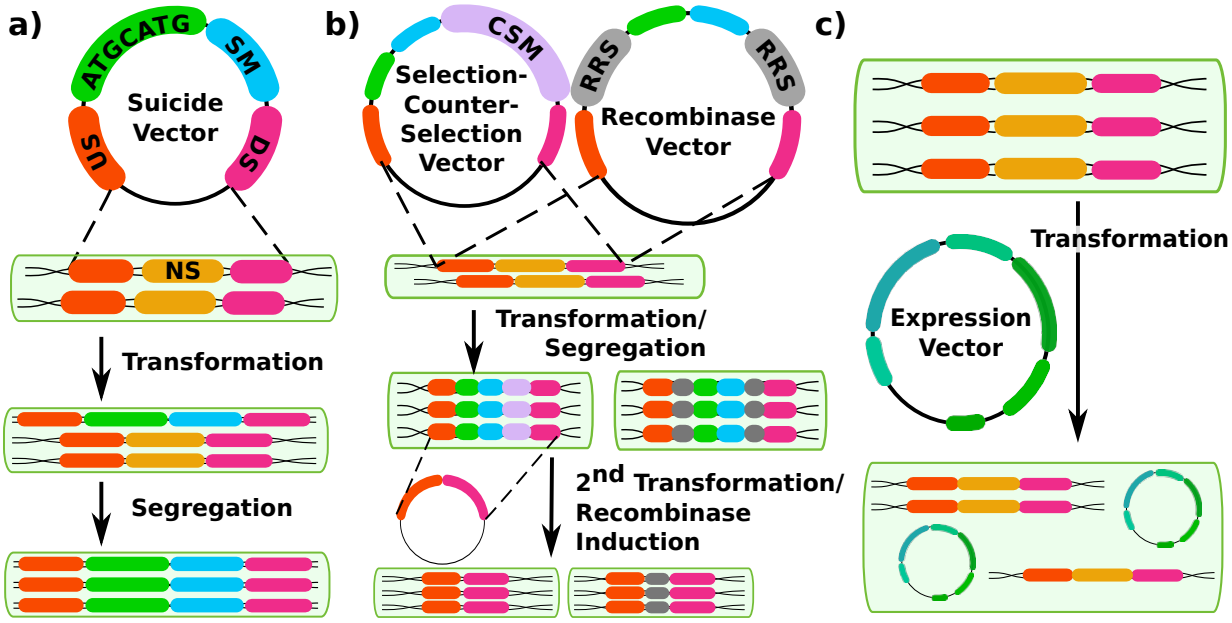


Figure 1.1: **Different methods for constructing cyanobacterial mutants.** (a) shows the traditional method using double homologous recombination to insert a suicide vector into the genome at a neutral site (NS, gold) with upstream (US, orange) and downstream (DS, magenta) flanking regions in the vector. The insert contains an arbitrary sequence of interest (ATGCATG, green) and a selectable marker (SM, blue). (b) shows 2 methods of creating markerless mutants, either by selection-counterselection or by using a recombinase system such as FLP/FRT, The counter-selection method's first step is the same as for the method in panel a, except that the insert also contains a counter-selectable marker (CSM, purple) such as *sacB*. A second transformation is performed to create a markerless mutant. Alternatively, the insert can contain recombinase recognition sites (RRS, gray) that are controlled by an inducible recombinase at a second (or the same) site in the genome. While it erases the selectable marker, this method does leave a scar sequence behind. (c) shows genetic modification in *trans* via expression plasmids.

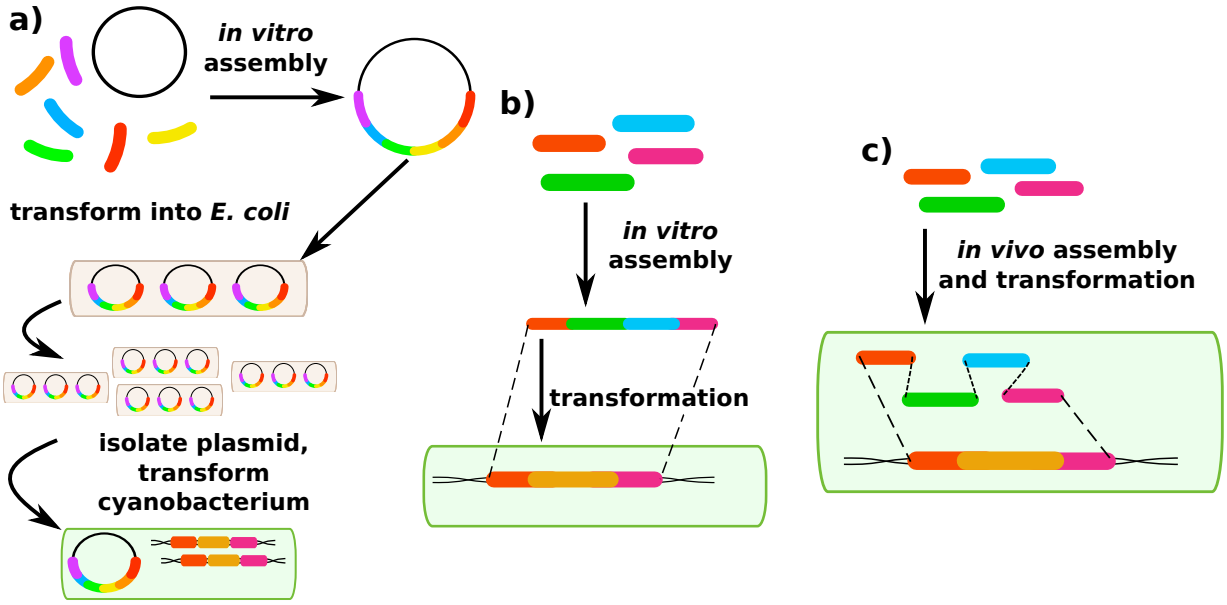


Figure 1.2: **DNA assembly methods.** Traditionally in cyanobacterial synthetic biology, plasmids are assembled *in vitro* and then propagated in *E. coli* before being transformed into cyanobacteria (a). More recently, methods have been developed for *in vitro* assembly and direct transformation via fusion PCR (b). Recently, a method has been developed for *in vivo* plasmid assembly via homologous recombination in yeast which may also be applicable in certain cyanobacterial strains.

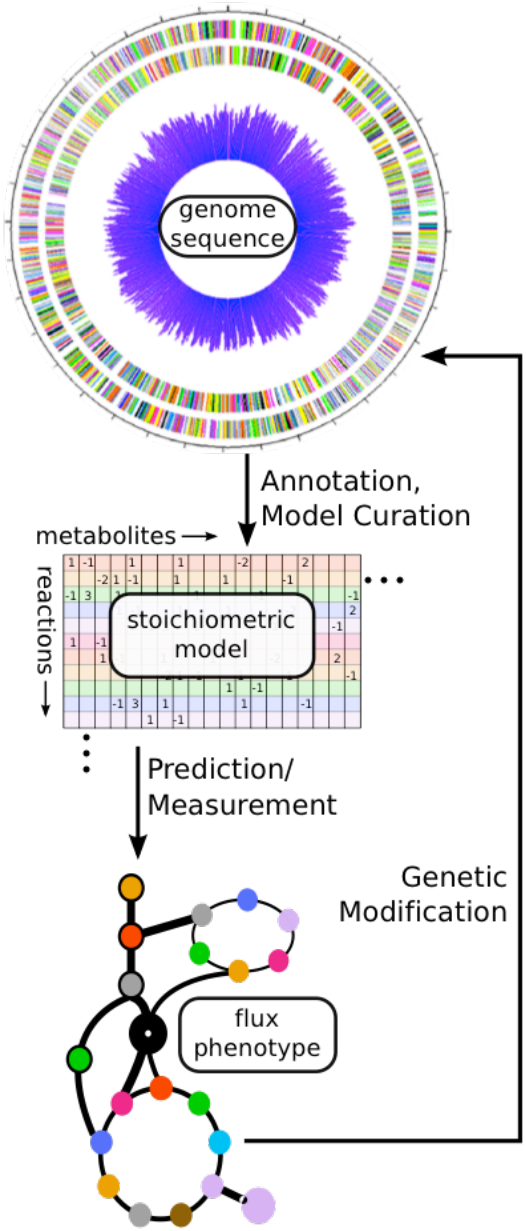


Figure 1.3: **Using fluxomics and genome-scale models to link genotype to metabolic phenotype.** From an annotated genome sequence, a stoichiometric model of metabolism can be constructed. That model can be solved via either prediction of an optimal flux phenotype (FBA) or measurement of actual flux phenotype (^{13}C-MFA). These results can help suggest modifications for altering the phenotype of the cell in a desired manner. In this way, a synthetic biologist can design new strains, build them using genetic modification methods, and test their phenotypes before designing new modifications in an iterative fashion.

Chapter 2

Cyanobacterial Alkanes Promote Growth in Cold Stress and Modulate Cyclic Photophosphorylation

2.1. Introduction

Cyanobacteria are the most ancient group of oxygenic photosynthetic organisms. These bacteria evolved the first thylakoid membranes, the structures on which they, as well as algae and plants, split water to produce O_2 and transform solar into chemical energy. In cyanobacteria, these membranes universally include alkanes and/or alkenes of 15-19 carbons. Recently, two pathways for production of these metabolites have been discovered (Schirmer *et al.* 2010; Mendez-Perez *et al.* 2011; Warui *et al.* 2011; Pandelia *et al.* 2013). Although these hydrocarbons were identified nearly 50 years ago (Han *et al.* 1968; Winters *et al.* 1969) and are produced at molar concentrations similar to chlorophyll *a*, little is known about their physiological role.

2.1.1. Cyanobacterial Alkanes

Medium-chain hydrocarbons are produced by a number of different organisms, including insects (Reed *et al.* 1994), birds (Cheesbrough *et al.* 1988), plants (Aarts *et al.* 1995), algae (Ladygina *et al.* 2006), and cyanobacteria and fulfill a wide variety of roles from waterproofing to energy storage. However, several factors distinguish cyanobacterial alkane biosynthesis in terms of its biological and biotechnological interest. For one, cyanobacteria produce large amounts of alkane, as much as 0.25% by weight in the wild-type strain (Coates *et al.* 2014), and most cyanobacteria do so using a small, soluble enzyme. This alkane abundance is similar on a molar basis to chlorophyll *a* and indicates that under appropriate conditions, the enzyme must operate quite efficiently. Although early studies on cyanobacterial ADO (aldehyde deformylating oxygenase) indicated that it operated extremely slowly or unstably, this may have been due to early confusion about the stoichiometry of the reaction and its mechanism. For example, while some studies reported that activity could only be detected under anaerobic conditions, it later became clear that the enzyme uses O_2 as a substrate to oxidize the aldehyde carbon and convert it

to formate (Das *et al.* 2014). This abundance of cyanobacterial alkanes and their potential usefulness as diesel fuels makes them promising targets for biofuel production. However, despite this promise, limited progress has been made so far in engineering either cyanobacteria to overproduce alkanes or other organisms to overproduce them using cyanobacterial enzymes (See chapter 4 of this work for further discussion).

From a biological perspective, the function of hydrocarbon-producing pathways in cyanobacteria has until recently been unknown. However, emerging evidence points to the importance of these pathways in cyanobacteria. There are two known pathways for cyanobacterial alka(e)ne biosynthesis, which either elongate and then decarboxylate fatty acids using a polyketide synthase-like complex (PKS-type), or reduce and then deformylate fatty acids using a pair of soluble enzymes (ADO-type). In general, the PKS-type pathway produces an α-olefin, while the ADO-type pathway produces linear, saturated species, although methylated species have been found in organisms carrying either pathway. One or the other of these pathways, but never both, is present in all fully sequenced cyanobacteria (Coates *et al.* 2014; Klahn *et al.* 2014). At the time of writing of this thesis, a small number of draft genomes exist in which no genes for alkane or alkene production have yet been found, but the fact that these are draft genomes combined with the universality of these genes in finished cyanobacterial genomes suggests that this absence is due to incomplete assembly or annotation. The ADO pathway is much more common – it has been identified in 164 strains as opposed to 17 for the PKS-type pathway (Coates *et al.* 2014; Klahn *et al.* 2014). The universality and abundance of cyanobacterial alka(e)nes indicates that they must have some important biological function.

2.1.2. The Cyanobacterial Thylakoid Membrane

In this study, we have investigated the function of the heptadecane produced by *Synechocystis* sp. PCC 6803 (hereafter *Synechocystis* 6803). This model strain harbors the ADO-type pathway and is easily amenable to genetic manipulation. It was the first photosynthetic organism to have its genome completely sequenced (Kaneko *et al.* 1996) and is a common model system for studies on photosynthesis as well as synthetic biology and metabolic engineering (Berla *et al.* 2013). Although efforts have been made to overproduce heptadecane as a biofuel molecule, they have met with limited success (Howard *et al.* 2013; Kaiser *et al.* 2013; Wang *et al.* 2013). This resistance to overproduction of a promising biofuel highlights the gap that still exists between the dream of a plug-and-play microbial cell factory and the reality of a dynamically regulated, intricately interconnected, and often still mysterious cellular environment. Therefore, it is important to identify the mechanisms underpinning overproduction resistance to break down these barriers.

The cyanobacterial thylakoid membrane is unique among biomembranes because it houses both oxygen-evolving photosynthesis and a full complement of respiratory enzymes (Schubert *et al.* 1995; Ohkawa *et al.* 2000; Cooley *et al.* 2001). This diversity of functions poses both a challenge and an opportunity – it allows the organism to adapt to a huge range of conditions and maintain balance in diverse metabolic pathways. However, this membrane system must also maintain both its physical composition and its activity across a wide range of environmental conditions. Any deviations in the redox poise of electron transport components can lead to metabolic imbalance and oxidative damage. Since photosynthetic organisms have a limited ability to not absorb light energy when the sun is out, their photosynthetic pathways must remain prepared to carry flux. While heterotrophs can slow their uptake of nutrients during

stress, autotrophs have little ability to avoid the sun, and can not save sunlight for later when conditions are better – they must 'make hay (or some other biomass) while the sun shines'. One particularly challenging and well-studied environmental condition for cyanobacteria is cold stress. Cyanobacteria modify their membranes in response to cold stress by synthesizing unsaturated lipids that remain fluid (Wada *et al.* 1990; Mikami *et al.* 2002; Ludwig *et al.* 2012). Recovery from photoinhibition depends on maintaining the fluidity of the membrane through fatty acid modification (Gombos *et al.* 1994). It has also been shown that cold-stress limits the ability of the cyanobacterium, *Synechococcus* sp. PCC 7002, to utilize nitrate, and requires urea as a reduced nitrogen source for optimal growth (Sakamoto *et al.* 1998; Sakamoto *et al.* 2002).

Figure 2.1 gives an overview of the main components of the photosynthetic machinery housed in the thylakoid membrane. This additional membrane system exists inside the cytoplasm of nearly all cyanobacterial strains, often occupying most of its volume (Liberton *et al.* 2011). These components are responsible for capturing solar energy in the forms of ATP and NADPH to power carbon fixation as well as the rest of cellular metabolism. It is critical that these energy sources are produced so as to match their consumption. A number of pathways allow the cell to strike this balance while also maintaining the redox poise of all electron transfer components (Allen 2002; Nogales *et al.* 2012). Successful forward electron transfer depends critically on maintenance of redox poise for all components, with deviations leading to unintended reactions and oxidative stress. There are two primary pathways for photosynthetic energy production. In the linear electron transport pathway, electrons travel from water to NADP(H). They are first excited by light at photosystem II (PSII) where water is split and O_2 is evolved. These excited electrons are then transported by plastoquinone (PQ) inside the thylakoid membrane to the cytochrome b_6f complex. Next, they are transported by soluble acceptors such as plastocyanin

in the thylakoid lumen to PSI, where they are again excited by light before donation to final acceptors in the cytoplasm, including NADP(H), nitrate, and others. Along the way, various steps in the pathway are coupled to transport of protons into the thylakoid lumen to power ATP synthesis by an F_1F_0 ATP synthase. This ATP synthesis requires 14 protons to generate 3 ATP, unlike those found in most heterotrophs, which require only 12 protons (Seelert *et al.* 2000). The second pathway highlighted in Figure 2.1 is a cyclic pathway, in which electrons from PSI are donated back to the PQ pool. While several alternative cyclic routes have been proposed, the pathway with the highest quantum yield involves NDH-1 donating electrons from NADPH to the PQ pool, since this pathway pumps additional protons as compared with others (Battchikova *et al.* 2011; Kramer *et al.* 2011). When electrons are recycled in this pathway, no NADPH but more ATP is produced. Thus, it has been suggested that cyclic electron transport pathways are critical for achieving the appropriate balance of ATP and NADPH to power CO_2 fixation (Allen 2002; Kramer *et al.* 2011; Nogales *et al.* 2012). However, these electron transport pathways must also power other cellular processes such as nitrogen assimilation, macromolecule synthesis, and the carbon-concentrating mechanism. In addition to the high-yield NDH pathway, cyanobacteria also include other forms of NDH-1 specialized for roles in the CO_2-concentrating mechanism (Price 2011) as well as succinate dehydrogenase (Cooley *et al.* 2001) that can participate in cyclic electron transport around PSI. Pseudo-cyclic pathways involving PSII and PSI can also supply extra ATP while reducing oxygen instead of $NADP^+$ (Schubert *et al.* 1995; Howitt *et al.* 1998; Nogales *et al.* 2012; Vu *et al.* 2012). Table 2.1 gives an overview of the quantum efficiency of different alternative electron flow pathways in *Synechocystis* 6803 for ATP and NADPH production. Because of its prominent role as a model system for photosynthesis studies, far more

is known about such pathways in *Synechocystis* 6803 as compared with any other cyanobacterium.

Until recently, no physiological role of the ubiquitous alkanes and alkenes found in cyanobacterial membranes had been identified. It was recently found that the α-olefins produced by the PKS-type pathway in *Synechococcus* sp. PCC 7002 play a role in cold-tolerance of that strain (Mendez-Perez *et al.* 2014). The strain produces a mono- and a diunsaturated olefin, and the diunsaturated species was found to accumulate at low temperature and to be essential for growth at low temperature (22 C compared to that organism's optimum of 38 C). In the present work, we find a similar phenotype for heptadecane in *Synechocystis* sp. PCC 6803. We show that membrane alkanes support optimal photosynthesis at low temperatures. Beyond this growth phenotype, we show that a strain that does not produce alkanes relies more heavily on cyclic electron transport, especially at low temperatures. We examine this result in the context of a genome-scale metabolic model that we generated previously for this strain (Saha *et al.* 2012). We used Flux Balance Analysis (FBA) (Orth *et al.* 2010) to explore the role of this pathway in helping cyanobacteria respond to environmental stress. We argue that alkanes are critical metabolites at low temperature because they help to maintain the balance of critical photosynthetic activities in the thylakoid membrane. We hypothesize that in the absence of alkanes, excess cyclic electron transport is required to maintain redox poise and that this excess leads to the observed slow growth of the mutant strain.

2.2. Materials and Methods

2.2.1. Mutant Construction

Plasmid pSL2192 was constructed via SLIC (Li *et al.* 2012) and restriction-based cloning. Initially, the fragments immediately upstream and downstream of the *Synechocystis* 6803 genes *sll0208* and *sll0209* encoding aldehyde deformylating oxygenase and fatty acyl-ACP reductase, respectively, were cloned into the pUC118 backbone on either side of a kanR cassette from pUC4K via SLIC. Primers were designed using j5 software (Hillson *et al.* 2012) and synthesized by IDT (Coralville, IA). Homologous sequences added to assembly fragments at the 5' ends of primers are in lowercase in table 2.2 below, while sequences binding to template DNA are in uppercase. Subsequently, this kanamycin cassette was replaced by a larger HincII fragment of the same plasmid.

Wild-type *Synechocystis* 6803 was transformed with this plasmid via natural transformation and transformants were isolated on BG11 media containing 20 μg/mL of kanamycin. Once colonies appeared, they were restreaked onto fresh media containing 40 μg/mL of kanamycin and the expected insertion site was confirmed via colony PCR. Mutant segregation was confirmed by the absence of any detectable PCR fragment originating from the *sll0208* or *sll0209* loci as shown in figure 2.2.

2.2.2. Culture Conditions

Cultures were grown in shake flasks at 200 rpm with BG-11 media under 30 μE of white light. For growth experiments, cultures were pre-incubated at their respective growth temperature for 48 hours before being diluted to OD_{730} = 0.05 (~10^7 cells/mL) in fresh BG-11 media. Cell growth was monitored by measuring OD_{730} daily using a BioTek plate reader (Biotek, Winooski, VT). For heptadecane analysis, samples were taken after 8 days of growth.

For redox kinetics experiments, cultures were grown for 5 days at 30 C, then resuspended in fresh media before measurement at either 30 C or 20 C.

2.2.3. Extraction and Analysis of Alkanes

2 mL of culture was pelleted by centrifugation and combined with 1 mL of ethyl acetate and 0.5 mL of 0.1 mm glass beads. Cells were lysed by bead beating for 3 cycles of 1 minute, with 5 minutes rest between cycles. Glass beads and debris were pelleted by centrifugation for 10 minutes at 16,000 x *g* then the upper ethyl acetate layer was removed for analysis. Chlorophyll *a* was determined on a DW-2000 spectrophotometer according to the formula [chl *a*] (μg/mL) = $(16.29*A_{665}) - (8.24 \times A_{652})$ (Lichtenthaler 1987). Alkanes were determined on an Agilent 6890 GC-MS fitted with a 12 meter DB5-MS column as previously (Schirmer *et al.* 2010) and quantified by comparison with an n-heptadecane standard (Sigma-Aldrich, St. Louis, MO).

2.2.4. Photophysiology Experiments

For analysis of P_{700} redox kinetics, cultures were grown as above for 5 days, then diluted 2-fold in fresh media and grown for 24 hours at either 20 C or 30 C. Cells were harvested by centrifugation and resuspended in fresh media to a chlorophyll concentration of 10 μg/mL. Samples were then maintained in the light at their growth temperature until ready for analysis (within several hours). Before each measurement, any required inhibitors were added and then the sample was dark-adapted for 2 minutes. A sub-saturating (130 $\mu E/m^2/sec$) pulse of actinic light from an orange LED source was used to illuminate the sample for 5 seconds. During that pulse and for 10 seconds of recovery afterwards, the redox state of P_{700} was monitored by absorption at 705 nm.

2.2.5. Flux Balance Analysis

Flux balance analysis (FBA)(Varma *et al.* 1994) was carried out on our previously developed *Synechocystis i*Syn731 model (Saha *et al.* 2012) to evaluate maximum biomass production in photosynthetic condition under various ratios of NDH-1 to PSI activity. The flux distribution for each of these states was inferred using FBA:

Maximize $v_{biomass}$

Subject to

$$\sum_{j=1}^{m} S_{ij} v_j = 0 \;\forall\; \mathrm{i} \in 1,\ldots..,\mathrm{n} \tag{2.1}$$

$$v_{j,\min} \leq v_j \leq v_{j,\max} \;\forall\; \mathrm{j} \in 1,\ldots..,\mathrm{m} \tag{2.2}$$

$$v_{ATPm} \geq 10 \tag{2.3}$$

$$v_{H_2CO_3\_uptake} + v_{CO_2\_uptake} \leq 100 \tag{2.4}$$

$$v_{PSIphoton\_uptake} + v_{PSIIphoton\_uptake} \leq 1000 \tag{2.5}$$

$$v_{NDH} = \mathrm{n}\, v_{PSI} \tag{2.6}$$

$$v_{PSI} = v_{PSI\_2} \tag{2.7}$$

$$v_{Cytox} \leq \varepsilon \tag{2.8}$$

Here, S_{ij} is the stoichiometric coefficient of metabolite i in reaction j and v_j is the flux value of reaction j. Parameters $v_{j,min}$ and $v_{j,max}$ denote the minimum and maximum allowable fluxes for reaction j, respectively. $v_{Biomass}$, v_{ATPm}, v_{H2CO3_uptake}, v_{CO2_uptake}, $v_{PSIphoton\_uptake}$, $v_{PSIIphoton\_uptake}$, v_{NDH}, v_{PSI} and v_{PSI_2} represent the flux of biomass formation, ATP maintenance, bicarbonate, carbon-di-oxide, PSI photon and PSII photon uptake, NAD(P)H dehydrogenase, PSI (involving reduced/oxidized plastocyanin) and PSI_2 (involving ferro/ferri-cytochrome)

reactions, respectively. And, n is the ratio of NDH-1 activity to PSI activity that is varied in the range of 0.01 and 2 with an increment of 0.01 (equivalent to a recycle rate between 0 and 100%). Based on the modeling practice (Saha *et al.* 2012; Vu *et al.* 2012), as shown in equations (2.4) and (2.5), photosynthetic conditions in *Synechocystis* sp. PCC 6803 are represented via setting a ratio of photon uptake (in the form PSI and PSII photon) to carbon uptake (in the form of CO_2 and H_2CO_3) to 1000:100. Also the ATP maintenance requirement of the cell is set to be 10 mmole/g-DW-h and activity of both the PSI reactions is assumed to be equal (Saha *et al.* 2012). In order to clarify the role of the recycle ratio on the fitness of the organism, the major pseudo-cyclic electron flows involving cytochrome oxidase are downregulated using the constraint as described in equation (8). Here, the value of the parameter **ε** was set to 0.01.

Once we had the values of optimal biomass for various ratios of NDH to PSI activity (as discussed above) from the *Synechocystis i*Syn731 model, flux variability analysis (Kumar *et al.* 2011) for the reactions (which participate in electron transport chain in thylakoid lumen and photosynthesis) was performed based on the following formulation:

Maximize/Minimize v_j

Subject to equations (2.1) – (2.8) as well as the additional constraint:

$$v_{Biomass} \geq v_{Biomass}^{\min} \tag{2.9}$$

Here, $v_{Biomass}^{\min}$ is the minimum level of biomass production. In this case we fixed it to be the optimal value obtained under a specific NADH to PSI activity ratio in photosynthetic condition for the *Synechocystis i*Syn731 model.

CPLEX solver (version 12.4, IBM ILOG) was used in the GAMS (version 24.4.4, GAMS Development Corporation) environment for solving the aforementioned optimization models. All

computations were carried out on Intel Xeon E5450 Quad-Core 3.0 GHz and Intel Xeon X5675 Six-Core 3.06 GH that are part of the lionxj and lionxf clusters (Intel Xeon E and X type processors and 128 and 128 GB memory, respectively) of High Performance Computing Group of The Pennsylvania State University.

2.3. Results

2.3.1. Mutant Construction

We constructed a plasmid vector, pNOalk (Figure 2.2A), to replace the ADO-type pathway for heptadecane biosynthesis in the naturally competent cyanobacterium *Synechocystis* sp. PCC 6803 with a kanamycin resistance cassette via double homologous recombination. After approximately six months and many rounds of patching on kanamycin-containing BG-11 media (20-40 μg/mL), we confirmed that no wild-type gene copies remained in the mutant strain via PCR (Figure 2.2B). *Synechocystis* 6803 has a high and flexible genome copy number and genes that confer a fitness advantage can be maintained in a merodiploid state when replaced by selective markers. In our experience most mutant strains segregate within 2-3 patchings, but even after several months we had to screen several mutant lines to identify one that was fully segregated. We also confirmed that the strain did not produce detectable heptadecane via GC-MS and regularly reconfirmed this throughout our experiments.

2.3.2. Growth, Heptadecane Production, and Fatty Acids Analysis

Synechocystis 6803 grows optimally at 30 C (Tasaka *et al.* 1996). Although the NOalk strain grew at nearly the same rate as the wild-type at 30 C, its growth was slower than the wild-type at 25 C and severely hampered at 20 C (Figure 2.3A). Growth of the wild-type strain was

only slightly slower at 20 C than at 30 C. At these lower temperatures, the wild-type strain also produced approximately twice as much heptadecane as when grown at 30 C (Figure 2.3B). This increased production of heptadecane at low temperature and poor growth in its absence indicate that heptadecane plays an important role in cold tolerance. No significant differences were observed between the fatty acids profile of the wild type and mutant strains (Figure 2.3C). In both strains, 16:0 fatty acids were by far the most abundant, accounting for more than 70% of fatty acids. All other individual species accounted for less than 10% of the total fatty acids.

2.3.3. Photosynthetic Analysis

To investigate why the NOalk mutant grows poorly at low temperature, we analyzed the kinetics of PSI reaction center oxidation/reduction using a JTS-10 spectrophotometer (BioLogic, Grenoble, France). Photo-oxidized PSI reaction centers (P_{700}^{+}) have much lower absorption of red light than the reduced form (P_{700}), so a decrease in absorbance at 705 nm is correlated with increased oxidation of P_{700}. This redox cycle occurs each time PSI absorbs a photon, and significant net oxidation of P_{700} can occur in the light, especially in the presence of photosynthetic inhibitors. Probing the kinetics of P_{700} reduction and oxidation is a powerful method for studying photosynthesis and especially cyclic electron transport (Joliot *et al.* 2005; Marathe *et al.* 2012). We suspended exponentially growing WT and NOalk cells in fresh BG-11 media to a chlorophyll concentration of 10 μg/mL. We exposed dark-adapted cells to a 5 second pulse of actinic light and measured the oxidation of P_{700}, then switched off the light and measured its re-reduction in the dark over 10 seconds (See figure 2.4). We measured these redox kinetics at both 30 C and 20 C and in the presence of the inhibitors DCMU (1,6-dichloromethyl urea) and DBMIB (2,5-dibromo-3-methyl-6-isopropyl benzoquinone). While DCMU blocks the activity of PSII and thus linear electron transport, it allows cyclic electron transport around PSI

to continue. DBMIB is a quinone analogue that inhibits cytochrome b_6f and thus disables both linear and cyclic electron transport (Trebst 2007), as well as respiratory pathways involving cytochrome b_6f (Ohkawa *et al.* 2000; Cooley *et al.* 2001).

Figures 2.4A and 2.4C show the effects of these inhibitors at 30 C, while 2.4B and 2.4D show the kinetics at 20 C. Figures 2.4C and 2.4D show the details of the re-reduction of P_{700}^+ in the dark for 2.4A and 2.4B, respectively. In the absence of inhibitors, the redox kinetics of P_{700} are nearly the same in the WT and NOalk strains, with the oxidation being slightly faster and reduction slightly slower in the mutant. However, in the presence of inhibitors, the kinetics of reduction diverge. The wild-type strain shows a greater change in oxidation at steady-state than the mutant in the presence of either DCMU or DBMIB (fig. 2.4A and 2.4B). To account for this difference and allow easy comparison among conditions, the kinetics of P_{700}^+ re-reduction in the dark (fig. 2.4C and 2.4D) have been normalized to the maximal oxidation observed within each measurement.

While the re-reduction rates are similar for the WT and NOalk in the absence of inhibitors, the mutant is less sensitive to DCMU as seen by its faster re-reduction with this inhibitor. This difference indicates a lesser reliance on PSII activity to re-reduce P_{700}^+ in the mutant. Therefore, the mutant must be supplying more electrons to P_{700}^+ via cyclic electron transport or respiratory pathways, or else expending fewer electrons via quinol oxidases (See figure 2.1). Assuming that CEF is the major contributor to P_{700}^+ re-reduction, we have used these data to calculate the percentage contribution of cyclic vs. linear electron transport to P_{700}^+ re-reduction (Marathe *et al.* 2012) (figure 2.5). The mutant strain at either 20 or 30 C has a greater reliance on cyclic electron transport. However, this difference becomes larger at 20 C, with the cyclic process accounting for nearly 20% of electron flow to P_{700}^+. This would indicate that the

cells are recycling approximately 1 electron in 5 in the mutant at 20 C, as opposed to one electron in 11 in the wild-type at 20 C and 1 in 17 in the wild-type at 30 C.

Another notable feature of the data presented in figure 2.4 is the greater sensitivity of the mutant to DBMIB. At low temperature, the re-reduction half time for the mutant approaches 1 second, which is consistent with back-reaction from ferredoxin, or potentially with the transfer of electrons from NADPH to P_{700}^{+} by any number of non-physiological routes. However, for the wild-type at either temperature, the re-reduction is 1.5-2 times faster.

2.3.4. Flux Balance Modeling of Linear and Cyclic Electron Transport

We earlier developed a genome-scale model of photoautotrophic metabolism in *Synechocystis* 6803, *i*Syn731 (Saha *et al.* 2012). This model includes detailed descriptions of both linear and cyclic electron transport pathways, as well as alternative electron flow pathways, such as cytochrome oxidase, Mehler reaction, and photorespiration. In this study, we have used the model for flux balance analysis (FBA) (Orth *et al.* 2010) to predict the distribution of intracellular fluxes associated with a maximal growth rate under various rates of cyclic electron flow. Using similar techniques, it was recently found that a range of electron flow pathways allow for robust maintenance of redox poise under a range of environmental conditions in multiple cyanobacterial strains (Nogales *et al.* 2012; Vu *et al.* 2012). We chose to analyze the effect of varied CEF rates because of the observed effect of the mutation on cyclic electron transport. We found that biomass production is optimal at a recycle rate of 25% (Figure 2.6) and decreases approximately linearly both above and below that value albeit with different slopes. At lower cyclic electron flow rates, ATP production is limiting, while at high recycle rates NADPH production becomes limiting. However, the specific recycle rate associated with maximal growth is dependent on environmental conditions such as light intensity and spectral quality (PSI and

PSII have different absorption spectra), nitrogen source, and the presence of reduced carbon sources. Presumably owing to the many redundant pathways (see table 2.1) in the cell that allow for modulation of the ATP:NADPH ratio, growth was predicted to be possible at any recycle rate under the conditions tested. However, this complementation among pathways may only be possible *in silico* due to kinetic or redox poise limitations that exist *in vivo* (Vu *et al.* 2012) and are not captured in current models.

A closer observation in the *in silico* flux ranges via flux variability analysis (Mahadevan *et al.* 2003) reveals that with the increase of the recycle ratio, the model predicts expanding ranges of ATP and NADPH production. Based on these results, the model is predicting that the cell is activating energy inefficient reactions and/or pathways in above-optimal recycle ratios to consume whichever energy source is produced in excess for a particular solution. Because the appropriate ATP:NADPH balance for biomass production must be maintained, these inefficient pathways decrease the quantum efficiency of biomass production.

2.4. Discussion

2.4.1. Alkane Production at Low Temperature

Similarly to the alkane produced by *Synechocystis* 6803, α-olefins produced by *Synechococcus* 7002 were found to be essential for growth at low temperature. While this strain grows optimally at 38 C, a knockout mutant producing no alkanes did not grow at 22 C (Mendez-Perez *et al.* 2014). *Synechococcus* 7002 produces both a 19:1 and a 19:2 α-olefin. The level of the 19:2 olefin increased at 22 C, although the total alkene pool decreased. This finding is consistent with the classical paradigm of membrane unsaturation as a response to cold (Wada

et al. 1991). In contrast, we found that the unsaturated hydrocarbon heptadecane accumulated in response to cold. This difference may help to explain why the two pathways for alka(e)ne production are never found to occur together in the same strain (Coates *et al.* 2014), and suggests that their mechanisms of action may be distinct from each other despite the similar phenotypes of these mutants.

2.4.2. P_{700} Redox Kinetics

Synechocystis 6803 exhibits increased cyclic electron flow at low temperature, especially in the NOalk mutant strain. Cyclic electron flow is known to serve diverse roles in autotrophs. Recent evidence suggests that cyclic electron transport in *Chlamydomonas* is regulated directly by the redox state of the cell and acts to prevent over-reduction of the stroma (Takahashi *et al.* 2013). CEF is also crucial for acclimation of *Arabidopsis thaliana* to fluctuating light (Suorsa *et al.* 2012). Additionally, CEF can help to maintain the balance of ATP:NADPH required for CO_2 fixation (Shikanai 2014). Since the thylakoid membrane ATP synthase requires 14 protons to produce 3 ATP, linear electron flow can not provide for the 3:2 ratio of ATP:NADPH to power CO_2 fixation by the Calvin Cycle (Allen 2002) or for the higher ratios required by the whole of metabolism (Nogales *et al.* 2012; Saha *et al.* 2012; Vu *et al.* 2012). For this reason alone, it has been suggested that PSI would have to recycle approximately 1 electron in 5 to the PQ pool or expend an even greater proportion in pseudo-cyclic electron flow in green algae. The story in cyanobacteria is a bit different. Because of the proton-pumping NDH-1 in the cyanobacterial thylakoid membrane (Kramer *et al.* 2011), only about 1 electron in 9 would need to be recycled to provide for CO_2 fixation, although the optimal number depends on the exact mix of nutrients consumed for biomass production. Different nitrogen and carbon sources will require different inputs of ATP and NADPH for biomass synthesis, as discussed below in the context of our FBA

analysis. Additionally, because cyanobacterial thylakoid membranes include a cytochrome c terminal oxidase (CtaI/II) (Schubert *et al.* 1995), pseudo-cyclic electron flow around PSII can operate with the same quantum yield of ATP as CEF around PSI, yielding 0.86 ATP per photon (See table 2.1). Thus, the highly redundant cyanobacterial thylakoid membrane provides many options for the cell to balance ATP and NADPH production under different light quality and quantity, along with the redox state of various electron carriers. However, it appears that the higher rates of cyclic flow in the absence of alkanes lead to a reduction in this photosynthetic flexibility, perhaps because of the need to maintain the redox poise of certain electron transport components. It remains unclear what exactly the source of this limitation might be. One possibility is that the lack of alkanes reduces membrane fluidity at low temperature, leading to some loss of photosynthetic activity or restriction in electron transport, such as the intramembrane trafficking of reducing equivalents by plastoquinone. In turn, this metabolic inflexibility leads to non-ideal thylakoid reductant partitioning. While the mutant displayed a similar effect on its redox kinetics at either ideal or low temperature, the mutant only grew more slowly at low temperature. It seems that with alkanes present, cellular metabolism was able to accommodate the challenge of maintaining a near-maximal growth rate at 25 or 20 C. However without alkanes these low temperatures exceeded the capacity of the cell to adapt (See figure 2.7). Taken together, these data suggest that enhanced CEF in the mutant contributes to maintaining redox poise, but restricts the flexibility of the cell to adapt to changing environmental conditions.

The slow re-reduction of P_{700}^{+} in the NOalk mutant with DBMIB might be explained in several ways. The re-reduction of P_{700}^{+} in the presence of DBMIB proceeds primarily by charge recombination within PSI in green algae (Alric 2010) and it is likely that the same holds true for

cyanobacteria. With a relatively oxidized NADP(H) pool, such charge recombination would proceed more slowly. Another possibility is that formate dehydrogenase contributes to the re-reduction of P_{700}^{+} via cytochrome c553 under these conditions. Because alkane production from fatty acids is a source of formate in the cell, this pathway might be slowed in the mutant. While it is unlikely that this pathway is a major contributor to P_{700}^{+} reduction during normal growth, it might have an effect during low temperature and poisoning with DBMIB.

2.4.3. Flux Balance Modeling

The FBA derived optimal recycle rate agrees with our observations of cyclic electron flow. We measured a recycle rate of approximately 1 electron in 5 in the mutant at 20 C compared to an optimal recycle ratio of 1 electron in 4 via FBA. While the measured recycle rate for optimal growth was a bit lower (1 electron in 17 in WT cells at 30 C), it is expected that experimental measurements would be underestimates of the actual rate and that modeling results would overestimate the ideal rate. Both of these discrepancies are caused by the potential activity of cytochrome oxidase. *In vivo*, this activity would lead us to underestimate CEF rates by intercepting some of the electrons donated to cytochrome b_6f from reaching PSI. Others have attempted to block cytochrome oxidase activity using potassium cyanide in combination with DCMU for estimation of cyclic electron flow in cyanobacteria, but this inhibitor has given inconsistent results. In some cases, it leads to faster re-reduction of P_{700}^{+} (Bernat *et al.* 2011; Marathe *et al.* 2012), while in other cases the re-reduction is slower (Ivanov *et al.* 2000; Marathe *et al.* 2012) indicating that this inhibitor has non-target effects. Another possible reason for an underestimate of CEF is that the process is slowed in the presence of DCMU because of a relatively oxidized thylakoid lumen in the mutant. Transfer from PSI electron acceptors to the PQ pool has been regarded as the rate-limiting step in CEF, so an oxidized pool of PSI acceptors

could slow the kinetics of CEF (Maxwell *et al.* 1976; Alric 2010). *In silico*, by restricting the activity of cytochrome oxidase (and other alternative electron flow pathways) to explore the effect of the dominant cyclic pathway via NDH-1 (Ohkawa *et al.* 2000), we eliminated a potential source of ATP and thus required the cyclic pathway to carry a higher flux for ATP production. It should also be noted that an exact match between the simulated environmental conditions in our model and those experienced in our growth studies is difficult to assess, and as we have shown, environmental conditions have a significant impact on cyclic electron flow.

In future work, FBA will serve as a valuable tool for studying the effects of different environmental conditions on cyanobacteria in general and on cyclic electron transport specifically, as we believe this process is a key tool that allows photosynthetic organisms to acclimate to a dynamic environment. We have demonstrated here that FBA can serve to analyze phenotypes beyond what is explicitly included in the model. Our model contains no information related to the effects of temperature or the regulatory role of heptadecane, but by inputting our observations of those effects (increased CEF) into the model, we identified a likely secondary outcome of activation of inefficient metabolic cycles for growth. These metabolic reactions would have been impossible to directly measure using available technologies. Thus, an *in silico* approach helped us to understand the mechanistic role of alkanes in a unique way.

2.5. Conclusions

The low temperature of 20 C used in this study is well within the normal daily range that might be expected in a temperate climate where mid-day temperatures reached or exceeded the optimum of 30 C for *Synechocystis* 6803. At these temperatures, alkanes appear to play a role in

modulating reductant balance and maintaining the redox balance of the photosynthetic electron transport chain, as evidenced by increased cyclic electron flow in their absence. A photosynthetic strain adapted to living at a wide range of temperatures would gain a growth advantage from being able to use the light that is available in cooler morning and evening hours. Redox balance must be tightly controlled by cyanobacteria to avoid redox stress (Schuurmans *et al.* 2014), including across the full range of environmental conditions in the organism's habitat. It is not unexpected that these hydrocarbons, which are both universal and unique metabolites to the cyanobacterial phylum, should be involved in the challenge of living with light as a primary energy source. While plants and algae face this same challenge, their chloroplasts' thylakoid membranes are adapted to a very different set of challenges than those in cyanobacteria, which house both photosynthesis and respiration. These membranes also include different types of hydrocarbons, such as sterols. Thus, this paper puts forth and provides evidence for the functional role for this recently-characterized and biotechnologically important class of cyanobacterial metabolites.

2.6. References

Aarts, M. G., Keijzer, C. J., Stiekema, W. J. and Pereira, A. (1995). "Molecular characterization of the CER1 gene of arabidopsis involved in epicuticular wax biosynthesis and pollen fertility." *Plant Cell* **7**(12): 2115-2127.

Allen, J. (2002). "Photosynthesis of ATP-electrons, proton pumps, rotors, and poise." *Cell* **110**(3): 273-276.

Alric, J. (2010). "Cyclic electron flow around photosystem I in unicellular green algae." *Photosynth Res* **106**(1-2): 47-56.

Battchikova, N., Wei, L., Du, L., Bersanini, L., Aro, E. M. and Ma, W. (2011). "Identification of novel Ssl0352 protein (NdhS), essential for efficient operation of cyclic electron transport around photosystem I, in NADPH:plastoquinone oxidoreductase (NDH-1) complexes of Synechocystis sp. PCC 6803." *J Biol Chem* **286**(42): 36992-37001.

Berla, B. M., Saha, R., Immethun, C. M., Maranas, C. D., Moon, T. S. and Pakrasi, H. B. (2013). "Synthetic biology of cyanobacteria: unique challenges and opportunities." *Front Microbiol* **4**: 246.

Bernat, G., Appel, J., Ogawa, T. and Rogner, M. (2011). "Distinct roles of multiple NDH-1 complexes in the cyanobacterial electron transport network as revealed by kinetic analysis of P700+ reduction in various Ndh-deficient mutants of Synechocystis sp. strain PCC6803." *J Bacteriol* **193**(1): 292-295.

Cheesbrough, T. M. and Kolattukudy, P. E. (1988). "Microsomal preparation from an animal tissue catalyzes release of carbon monoxide from a fatty aldehyde to generate an alkane." *J Biol Chem* **263**(6): 2738-2743.

Coates, R. C., Podell, S., Korobeynikov, A., Lapidus, A., Pevzner, P., Sherman, D. H., Allen, E. E., Gerwick, L. and Gerwick, W. H. (2014). "Characterization of cyanobacterial hydrocarbon composition and distribution of biosynthetic pathways." *PLoS One* **9**(1): e85140.

Cooley, J. W. and Vermaas, W. F. (2001). "Succinate dehydrogenase and other respiratory pathways in thylakoid membranes of Synechocystis sp. strain PCC 6803: capacity comparisons and physiological function." *J Bacteriol* **183**(14): 4251-4258.

Das, D., Ellington, B., Paul, B. and Marsh, E. N. (2014). "Mechanistic insights from reaction of alpha-oxiranyl-aldehydes with cyanobacterial aldehyde deformylating oxygenase." *ACS Chem Biol* **9**(2): 570-577.

Gombos, Z., Wada, H. and Murata, N. (1994). "The recovery of photosynthesis from low-temperature photoinhibition is accelerated by the unsaturation of membrane lipids: a mechanism of chilling tolerance." *Proc Natl Acad Sci U S A* **91**(19): 8787-8791.

Han, J., McCarthy, E. D., Hoeven, W. V., Calvin, M. and Bradley, W. H. (1968). "Organic geochemical studies, ii. A preliminary report on the distribution of aliphatic hydrocarbons in algae, in bacteria, and in a recent lake sediment." *Proc Natl Acad Sci U S A* **59**(1): 29-33.

Hillson, N. J., Rosengarten, R. D. and Keasling, J. D. (2012). "j5 DNA assembly design automation software." *ACS Synth Biol* **1**(1): 14-21.

Howard, T. P., Middelhaufe, S., Moore, K., Edner, C., Kolak, D. M., Taylor, G. N., Parker, D. A., Lee, R., Smirnoff, N., Aves, S. J. and Love, J. (2013). "Synthesis of customized

petroleum-replica fuel molecules by targeted modification of free fatty acid pools in Escherichia coli." *Proc Natl Acad Sci U S A* **110**(19): 7636-7641.
Howitt, C. A. and Vermaas, W. F. (1998). "Quinol and cytochrome oxidases in the cyanobacterium Synechocystis sp. PCC 6803." *Biochemistry* **37**(51): 17944-17951.
Ivanov, A. G., Park, Y. I., Miskiewicz, E., Raven, J. A., Huner, N. P. and Oquist, G. (2000). "Iron stress restricts photosynthetic intersystem electron transport in Synechococcus sp. PCC 7942." *FEBS Lett* **485**(2-3): 173-177.
Joliot, P. and Joliot, A. (2005). "Quantification of cyclic and linear flows in plants." *Proc Natl Acad Sci U S A* **102**(13): 4913-4918.
Kaiser, B. K., Carleton, M., Hickman, J. W., Miller, C., Lawson, D., Budde, M., Warrener, P., Paredes, A., Mullapudi, S., Navarro, P., Cross, F. and Roberts, J. M. (2013). "Fatty aldehydes in cyanobacteria are a metabolically flexible precursor for a diversity of biofuel products." *PLoS One* **8**(3): e58307.
Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M. and Tabata, S. (1996). "Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions." *DNA Res* **3**(3): 109-136.
Klahn, S., Baumgartner, D., Pfreundt, U., Voigt, K., Schon, V., Steglich, C. and Hess, W. R. (2014). "Alkane Biosynthesis Genes in Cyanobacteria and Their Transcriptional Organization." *Front Bioeng Biotechnol* **2**: 24.
Kramer, D. M. and Evans, J. R. (2011). "The importance of energy balance in improving photosynthetic productivity." *Plant Physiol* **155**(1): 70-78.
Kumar, V. S., Ferry, J. G. and Maranas, C. D. (2011). "Metabolic reconstruction of the archaeon methanogen Methanosarcina Acetivorans." *Bmc Systems Biology* **5**.
Ladygina, N., Dedyuknina, E. and Vainshtein, M. (2006). "A review of microbial synthesis of hydrocarbons." *Process Biochemistry* **41**(5): 1001-1014.
Li, M. Z. and Elledge, S. J. (2012). "SLIC: a method for sequence- and ligation-independent cloning." *Methods Mol Biol* **852**: 51-59.
Liberton, M., Austin, J. R., 2nd, Berg, R. H. and Pakrasi, H. B. (2011). "Insights into the complex 3-D architecture of thylakoid membranes in unicellular cyanobacterium Cyanothece sp. ATCC 51142." *Plant Signal Behav* **6**(4): 566-569.
Lichtenthaler, H. K. (1987). "Chlorophylls and carotenoids: Pigments of photosynthetic biomembranes " *Methods in Enzymology* **148**: 350-382.
Ludwig, M. and Bryant, D. A. (2012). "Synechococcus sp. Strain PCC 7002 Transcriptome: Acclimation to Temperature, Salinity, Oxidative Stress, and Mixotrophic Growth Conditions." *Front Microbiol* **3**: 354.
Mahadevan, R. and Schilling, C. H. (2003). "The effects of alternate optimal solutions in constraint-based genome-scale metabolic models." *Metab Eng* **5**(4): 264-276.
Marathe, A. and Kallas, T. (2012). *Cyclic Electron Transfer Pathways in Syncechococcus sp. PCC 7002 Cyanobacteria During Photosynthesis At High Light Intensity*. Master of Science - Biology, The University of Wisconsin Oshkosh.

Maxwell, P. C. and Biggins, J. (1976). "Role of cyclic electron transport in photosynthesis as measured by the photoinduced turnover of P700 in vivo." *Biochemistry* **15**(18): 3975-3981.

Mendez-Perez, D., Begemann, M. B. and Pfleger, B. F. (2011). "Modular synthase-encoding gene involved in alpha-olefin biosynthesis in Synechococcus sp. strain PCC 7002." *Appl Environ Microbiol* **77**(12): 4264-4267.

Mendez-Perez, D., Herman, N. A. and Pfleger, B. F. (2014). "A desaturase gene involved in formation of 1,14-nonadecadiene in Synechococcus sp. strain PCC 7002." *Appl Environ Microbiol*.

Mikami, K., Kanesaki, Y., Suzuki, I. and Murata, N. (2002). "The histidine kinase Hik33 perceives osmotic stress and cold stress in Synechocystis sp PCC 6803." *Mol Microbiol* **46**(4): 905-915.

Nogales, J., Gudmundsson, S., Knight, E. M., Palsson, B. O. and Thiele, I. (2012). "Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis." *Proc Natl Acad Sci U S A* **109**(7): 2678-2683.

Ohkawa, H., Pakrasi, H. B. and Ogawa, T. (2000). "Two types of functionally distinct NAD(P)H dehydrogenases in Synechocystis sp. strain PCC6803." *J Biol Chem* **275**(41): 31630-31634.

Orth, J. D., Thiele, I. and Palsson, B. O. (2010). "What is flux balance analysis?" *Nat Biotechnol* **28**(3): 245-248.

Pandelia, M. E., Li, N., Norgaard, H., Warui, D. M., Rajakovich, L. J., Chang, W. C., Booker, S. J., Krebs, C. and Bollinger, J. M., Jr. (2013). "Substrate-triggered addition of dioxygen to the diferrous cofactor of aldehyde-deformylating oxygenase to form a diferric-peroxide intermediate." *J Am Chem Soc* **135**(42): 15801-15812.

Price, G. D. (2011). "Inorganic carbon transporters of the cyanobacterial CO2 concentrating mechanism." *Photosynth Res* **109**(1-3): 47-57.

Reed, J. R., Vanderwel, D., Choi, S., Pomonis, J. G., Reitz, R. C. and Blomquist, G. J. (1994). "Unusual mechanism of hydrocarbon formation in the housefly: cytochrome P450 converts aldehyde to the sex pheromone component (Z)-9-tricosene and CO2." *Proc Natl Acad Sci U S A* **91**(21): 10000-10004.

Saha, R., Verseput, A. T., Berla, B. M., Mueller, T. J., Pakrasi, H. B. and Maranas, C. D. (2012). "Reconstruction and comparison of the metabolic potential of cyanobacteria Cyanothece sp. ATCC 51142 and Synechocystis sp. PCC 6803." *PLoS One* **7**(10): e48285.

Sakamoto, T. and Bryant, D. A. (1998). "Growth at low temperature causes nitrogen limitation in the cyanobacterium Synechococcus sp. PCC 7002." *Arch Microbiol* **169**(1): 10-19.

Sakamoto, T. and Bryant, D. A. (2002). "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**(3): 231-242.

Schirmer, A., Rude, M. A., Li, X., Popova, E. and del Cardayre, S. B. (2010). "Microbial biosynthesis of alkanes." *Science* **329**(5991): 559-562.

Schubert, H., Matthijs, H. C. P. and Mur, L. R. (1995). "*In Vivo* Assay of P700 Redox Changes in the Cyanobacterium *Fremyella Diplosiphon* and the Role of Cytochrome-C-Oxidase in Regulation of Photosynthetic Electron-Transfer." *Photosynthetica* **31**(4): 517-527.

Schuurmans, R. M., Schuurmans, J. M., Bekker, M., Kromkamp, J. C., Matthijs, H. C. P. and Hellingwerf, K. J. (2014). "The Redox Potential of the Plastoquinone Pool of the

Cyanobacterium Synechocystis Species Strain PCC 6803 Is under Strict Homeostatic Control." *Plant Physiology* **165**(1): 463-475.
Seelert, H., Poetsch, A., Dencher, N. A., Engel, A., Stahlberg, H. and Muller, D. J. (2000). "Structural biology. Proton-powered turbine of a plant motor." *Nature* **405**(6785): 418-419.
Shikanai, T. (2014). "Central role of cyclic electron transport around photosystem I in the regulation of photosynthesis." *Curr Opin Biotechnol* **26**: 25-30.
Suorsa, M., Jarvi, S., Grieco, M., Nurmi, M., Pietrzykowska, M., Rantala, M., Kangasjarvi, S., Paakkarinen, V., Tikkanen, M., Jansson, S. and Aro, E. M. (2012). "PROTON GRADIENT REGULATION5 is essential for proper acclimation of Arabidopsis photosystem I to naturally and artificially fluctuating light conditions." *Plant Cell* **24**(7): 2934-2948.
Takahashi, H., Clowez, S., Wollman, F. A., Vallon, O. and Rappaport, F. (2013). "Cyclic electron flow is redox-controlled but independent of state transition." *Nat Commun* **4**: 1954.
Tasaka, Y., Gombos, Z., Nishiyama, Y., Mohanty, P., Ohba, T., Ohki, K. and Murata, N. (1996). "Targeted mutagenesis of acyl-lipid desaturases in Synechocystis: evidence for the important roles of polyunsaturated membrane lipids in growth, respiration and photosynthesis." *EMBO J* **15**(23): 6416-6425.
Trebst, A. (2007). "Inhibitors in the functional dissection of the photosynthetic electron transport system." *Photosynth Res* **92**(2): 217-224.
Varma, A. and Palsson, B. O. (1994). "Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use." *Nature Biotechnology* **12**: 994-998.
Vu, T. T., Stolyar, S. M., Pinchuk, G. E., Hill, E. A., Kucek, L. A., Brown, R. N., Lipton, M. S., Osterman, A., Fredrickson, J. K., Konopka, A. E., Beliaev, A. S. and Reed, J. L. (2012). "Genome-scale modeling of light-driven reductant partitioning and carbon fluxes in diazotrophic unicellular cyanobacterium Cyanothece sp. ATCC 51142." *PLoS Comput Biol* **8**(4): e1002460.
Wada, H. and Murata, N. (1990). "Temperature-Induced Changes in the Fatty Acid Composition of the Cyanobacterium, Synechocystis PCC6803." *Plant Physiol* **92**(4): 1062-1069.
Wada, H. and Murata, N. (1991). "[The cyanobacterial membrane architecture in relation to the low-temperature tolerance]." *Tanpakushitsu Kakusan Koso* **36**(9): 1604-1610.
Wang, W., Liu, X. and Lu, X. (2013). "Engineering cyanobacteria to improve photosynthetic production of alka(e)nes." *Biotechnol Biofuels* **6**(1): 69.
Warui, D. M., Li, N., Norgaard, H., Krebs, C., Bollinger, J. M., Jr. and Booker, S. J. (2011). "Detection of formate, rather than carbon monoxide, as the stoichiometric coproduct in conversion of fatty aldehydes to alkanes by a cyanobacterial aldehyde decarbonylase." *J Am Chem Soc* **133**(10): 3316-3319.
Winters, K., Parker, P. L. and Van Baalen, C. (1969). "Hydrocarbons of blue-green algae: geochemical signfficance." *Science* **163**(3866): 467-468.

Table 2.1: Alternative electron flow pathways and their quantum yields of ATP and NADPH. Cyclic electron flow pathways around PSI are highlighted in green, and pseudocyclic pathways involving PSII are highlighted in blue.

Pathway	**PSII:PSI activity ratio**	**ATP quantum yield (hv^{-1})**	**NADPH quantum yield (hv^{-1})**
Linear electron flow	1:1	.32	.25
Cyclic electron flow (NDH-1)	0:1	.86	0
Cyclic electron flow (FQR/SDH/NDH-2/NO3 reductase)	0:1	.43	0
Cytochrome oxidase	1:0	.86	0
Mehler reaction/Hydrogenase	1:1	.35	0

Table 2.2: Oligonucleotides used in this study. Sequences binding to template DNA are in capital letters, while sequences added to the 5' end for assembly via SLIC (see methods) are in lowercase.

Primer name	Sequence
sll0208_US_forward	aacagctatgaccatgattacgaatt**CAAAATCTCCGGTCGGGTAAC**
sll0208_US_reverse	gaatatggctcat**AGGGGCGTTGGACTCCTG**
MCS_sll0209_DS_forward	gctcggatccggtaccgtcgactctaga**GCCGACAGGATAGGGCGTG**
sll0209_DS_reverse	tgtaaaacgacggccagtgccaagct**GACAAAAGTGAATGGATGCCCG**
kanR_forward	gtccaacgcccct**ATGAGCCATATTCAACGGGAAAC**
kanR_MCS_reverse	tagagtcgacggtaccggatccgagctc**TTAGAAAAACTCATCGAGCATCAAATG**
A_f	**CGCCCAAGCGGTTGCTGAAGA**
A_r	**GCGCCACAAACCGCTACCGT**
B_f	**AGATGGCGGAACTCTTGCCGGA**
B_r	**ATACCTTGGCGTCCCCCTGCA**
C_f	**TCAGCTACGGCGAAGCCCTCA**
C_r	**CCTAAAGAGCTACTAAAGGGC**

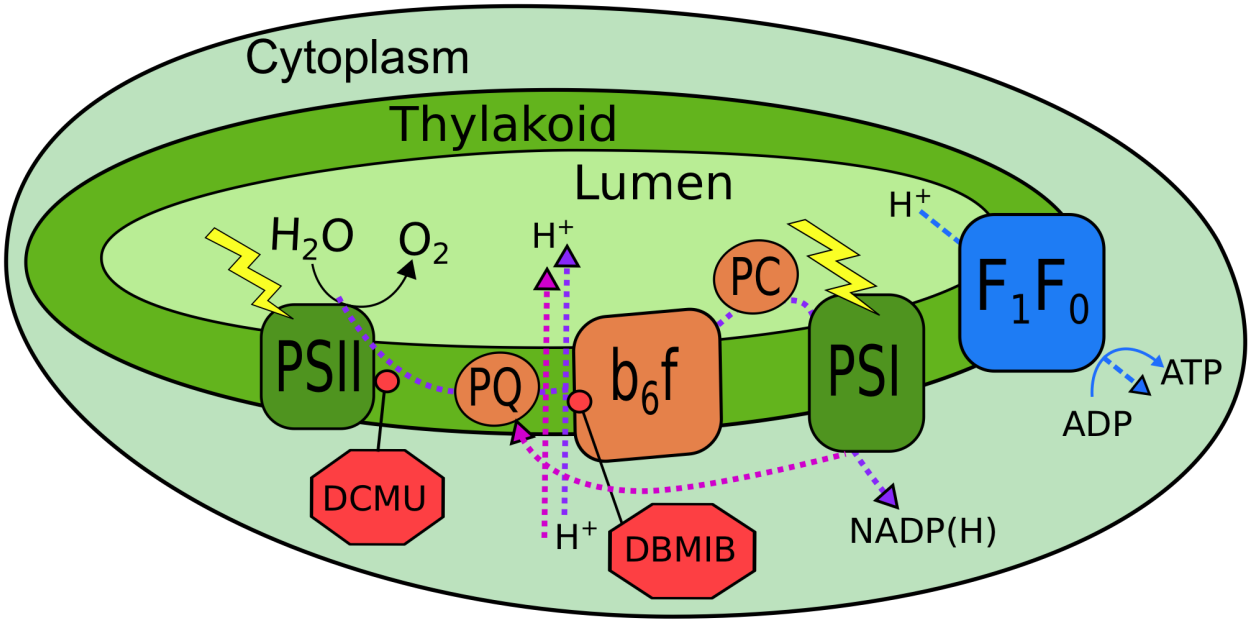


Figure 2.1: Cartoon of cyanobacterial photosynthetic electron transport pathways. Indicated are photosystems I (PSI/P_{700}) and II (PSII). In the linear electron transport pathway (dotted violet line), light is first absorbed by PSII, then excited electrons are transported inside the membrane by plastoquinone (PQ) to cytochrome b_6f, then through the thylakoid lumen by the soluble carrier plastocyanin (PC) to PSI. At PSI, electrons are excited by light a second time and then reduce $NADP^+$. Along the way, protons are transported inside the lumen to power ATP synthesis via the F_1F_0 ATP synthase. In the cyclic pathway (dotted magenta line), electrons are re-donated from PSI to the PQ pool. Thus, the cyclic pathway produces ATP at the expense of NADPH. Inhibitors used in this study and their sites of inhibition are also indicated in red octagons. DCMU blocks electron transfer from PSII to PQ and DBMIB prevents oxidation of plastoquinone by the cytochrome b_6f complex.

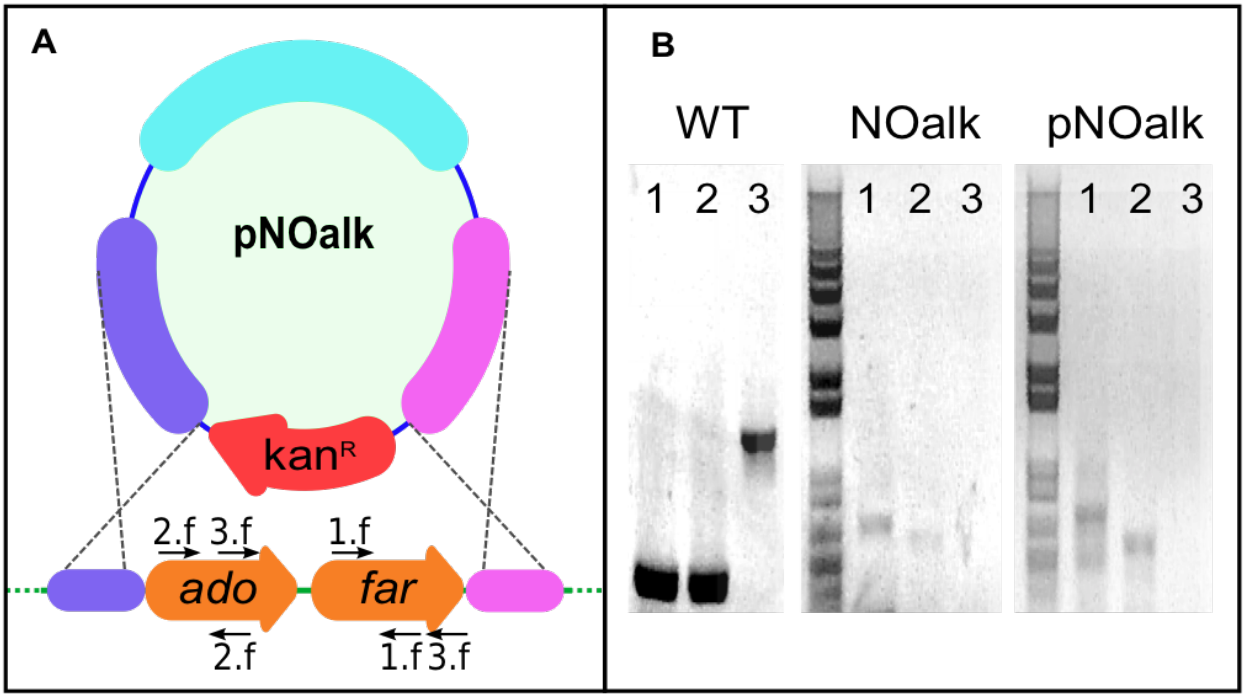


Figure 2.2: Knockout mutant construction strategy (A) and confirmation by PCR (B). We constructed a plasmid, pNOalk, containing sequences flanking the genes for the ADO-type heptadecane biosynthesis pathway in *Synechocystis* 6803 around a kanamycin resistance cassette (A). We confirmed the absence of these genes from the mutant strain via PCR with 3 different primer sets (B) using genomic DNA from the wild-type strain or the mutant strain (NOalk), or the plasmid pNOalk as template. Binding sites of the three primer sets (1,2,3) on the wild-type chromosome are indicated in panel a.

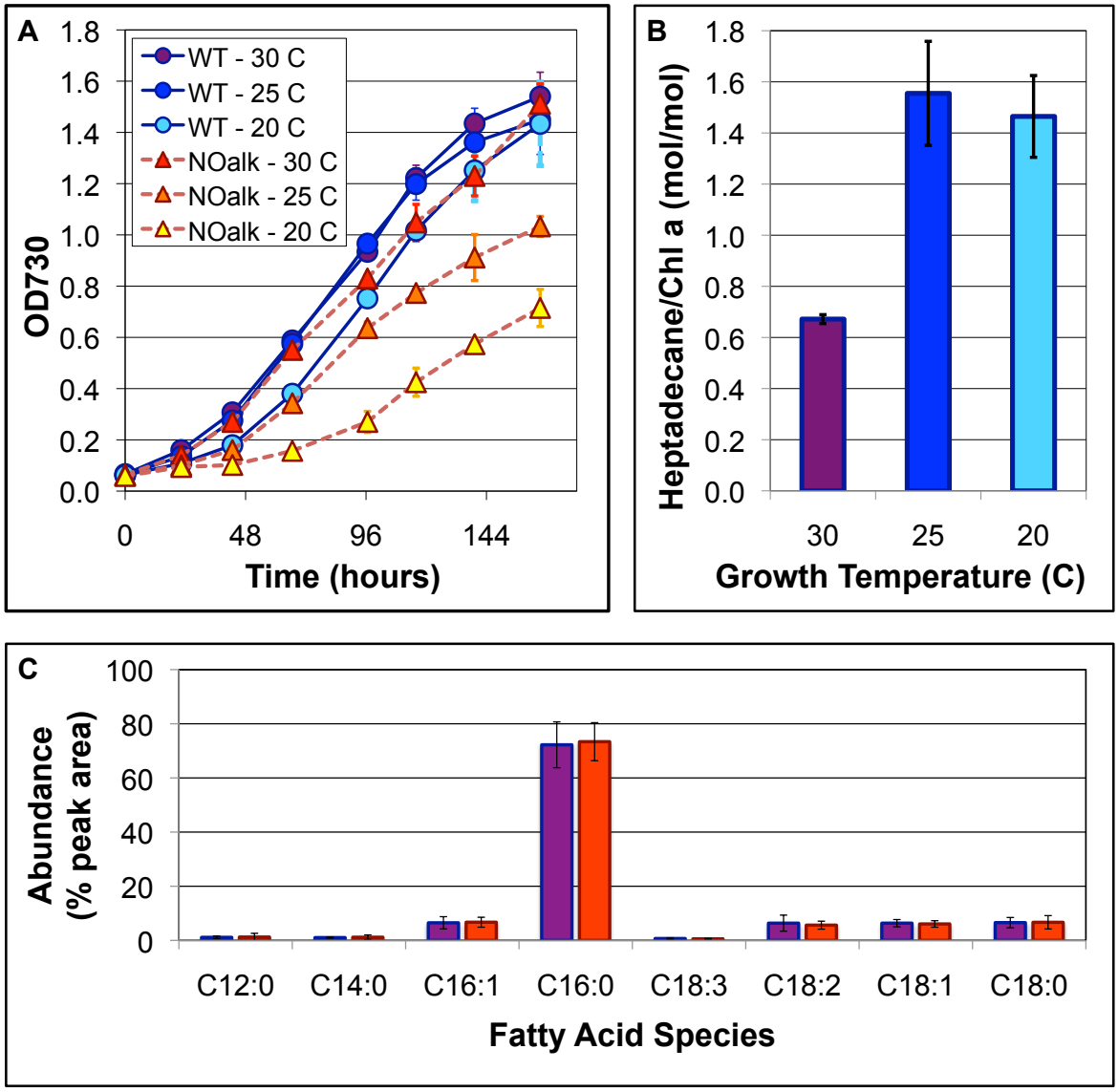


Figure 2.3: Growth and alkane production of WT and noALK strains at various temperatures. (A) Cultures were grown in shake flasks at 200 rpm with BG-11 media under 30 µE of white light. Cultures were pre-incubated at their respective growth temperature for 48 hours before being diluted to OD_{730} = 0.05 (~10^7 cells/mL) in fresh BG-11 media. Cell growth was monitored by measuring OD_{730} daily on a BioTek (Winooski, VT) plate reader. (B) After 8 days, aliquots were taken and extracted with ethyl acetate and glass beads in a bead-beater. Heptadecane was measured via GC-MS and normalized to chlorophyll a concentration. Error bars are ± SD for n=3 for both A and B. Where error bars are not seen, the error is smaller than the symbol shown.

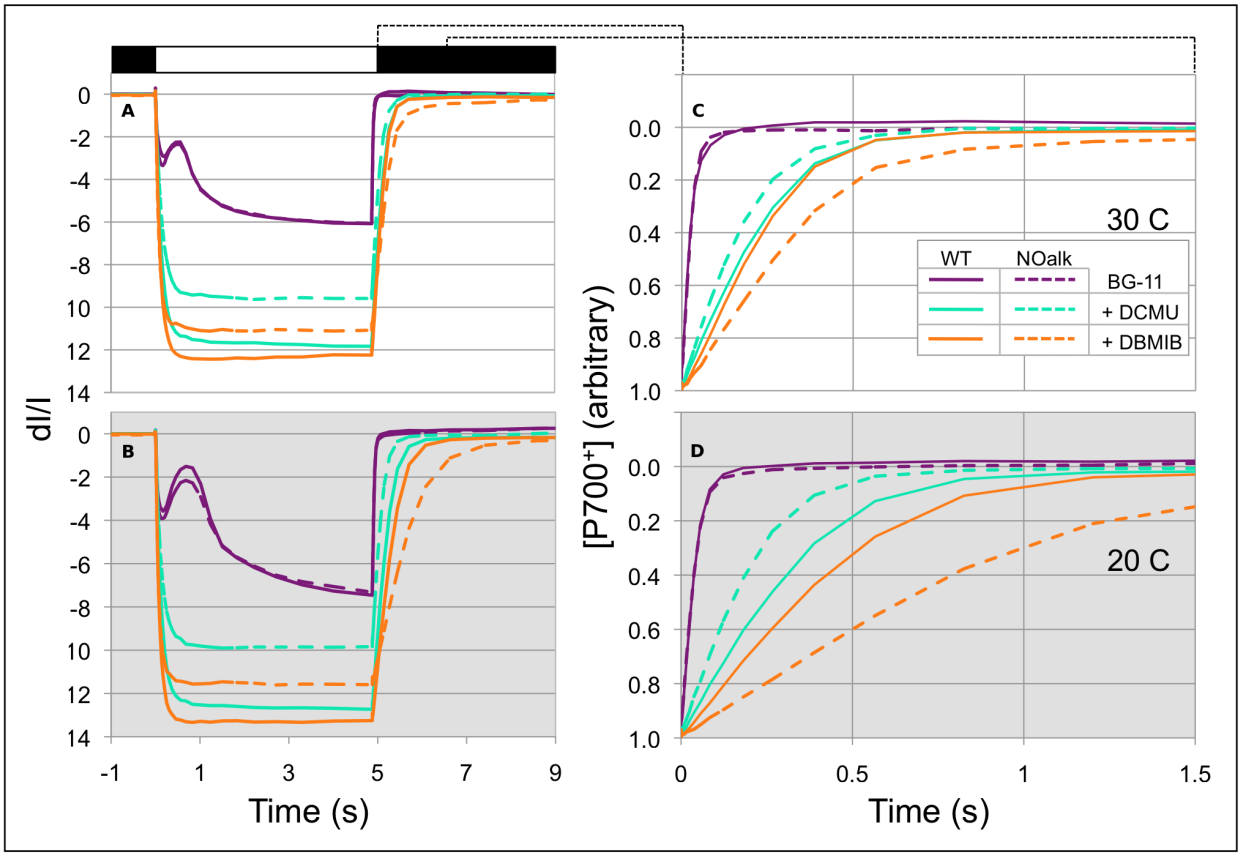


Figure 2.4: P_{700} redox kinetics for WT and noALK strains at 20 and 30 C. Using a JTS-10 spectrophotometer, cell suspensions were dark-adapted and then exposed to a pulse of orange actinic light (to excite both PSII and PSI) for 5 seconds. The actinic light was then turned off (indicated by black/white bar above panel A). During this time-course, measuring flashes of 705 nm light probed the redox state of the P_{700} reaction center. Data were collected from cells that had been grown at 30 C, then measured at 30 C (panel A,C) or shifted to 20 C before measurement (panel B,D). Panels C and D show the details of re-reduction of P_{700}^{+} in the dark for the experiments in panels A and B, respectively. Inhibitors of linear (10 µM DCMU, which inhibits transfer from PSII to the quinone pool), and linear + cyclic (1 µM DBMIB, which blocks cytochrome b6) were added to assess the mechanism of growth inhibition for the noALK strain at low temperature. Each trace is an average of 3 independent experiments.

A

Strain	Inhibitor	re-reduction $t_{1/2}$ (ms)	
		30 C	20 C
WT	None	12.1 ± 0.9	27.7 ± 0.3
	DCMU	170.3 ± 10.2	239.6 ± 18.2
	DBMIB	187.0 ± 6.2	333.3 ± 34.7
	%CEF	7.2 ± 0.6	11.6 ± 0.3
NOalk	None	13.4 ± 0.4	30.7 ± 1.4
	DCMU	132.4 ± 8.8	149.7 ± 3.4
	DBMIB	263.9 ± 13.9	631.1 ± 33.0
	%CEF	10.2 ± 0.5	20.5 ± 0.8

B

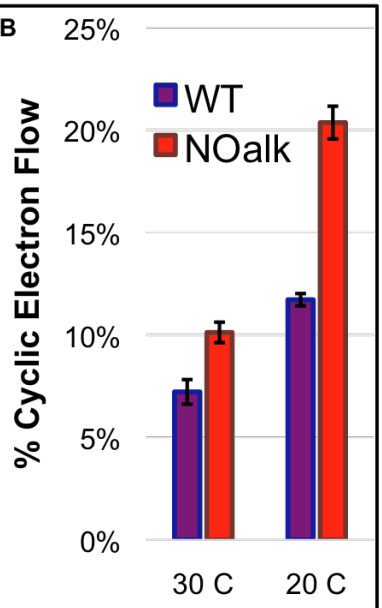


Figure 2.5: Strain NOalk uses a higher ratio of cyclic:linear electron transport. Panel A gives the half-times for re-reduction of P_{700}^{+} in the dark, calculated from the traces shown in figure 2.3C and 2.3D. Panel B shows the percentage of electron flow to P_{700}^{+} that is cyclic, calculated from the data in panel A.

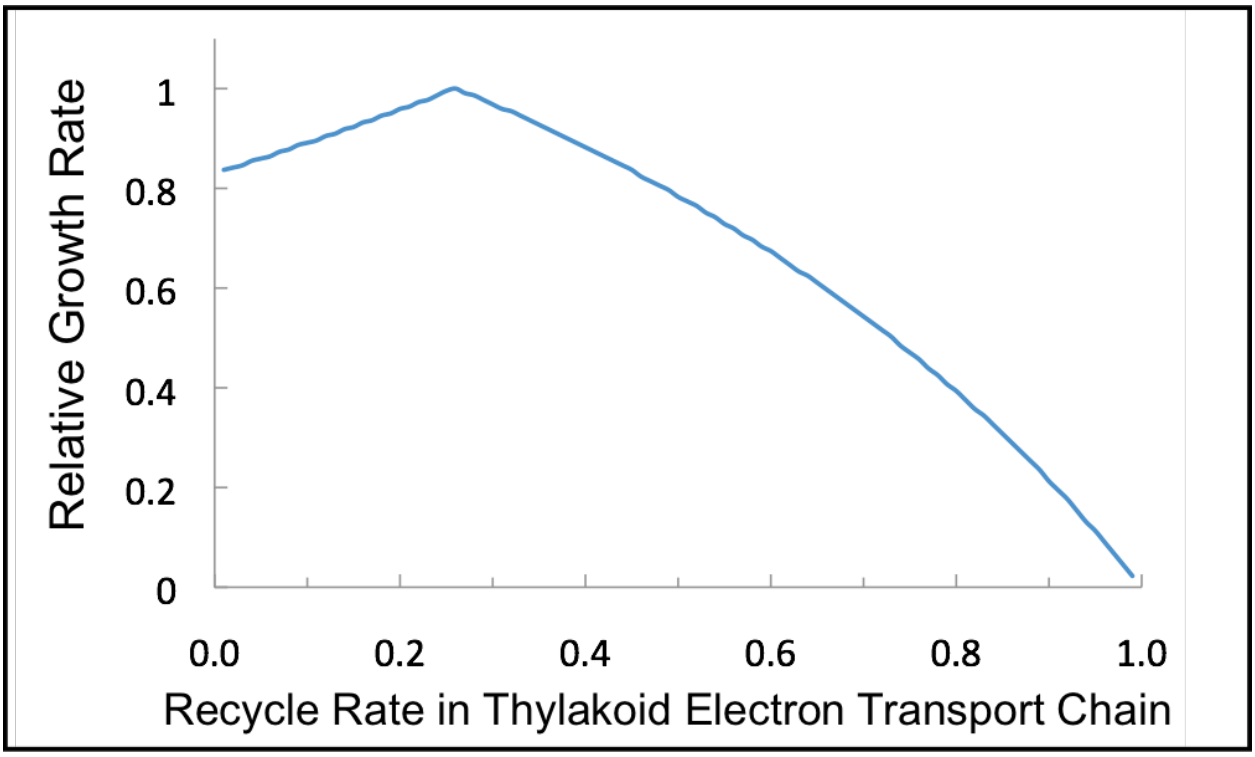


Figure 2.6: The simulated effect of cyclic electron transport on growth rate using *i*Syn731. We modeled the effect of varying the recycle rate of electrons from PSI to the PQ pool on light-limited growth of *Synechocystis* 6803. Optimal biomass yield was achieved at a recycle rate of 0.25. Biomass production remained possible at any recycle rate between 0.00 and 0.99. The recycle rate was defined as the NDH-1 catalyzed electron flux from NADPH to plastoquinone divided by the electron flux into PSI. These simulations were carried out with alternative electron flow pathways (including succinate dehydrogenase and cytochrome oxidase, see table 2.1) restricted to minimal flux. See methods section for further details.

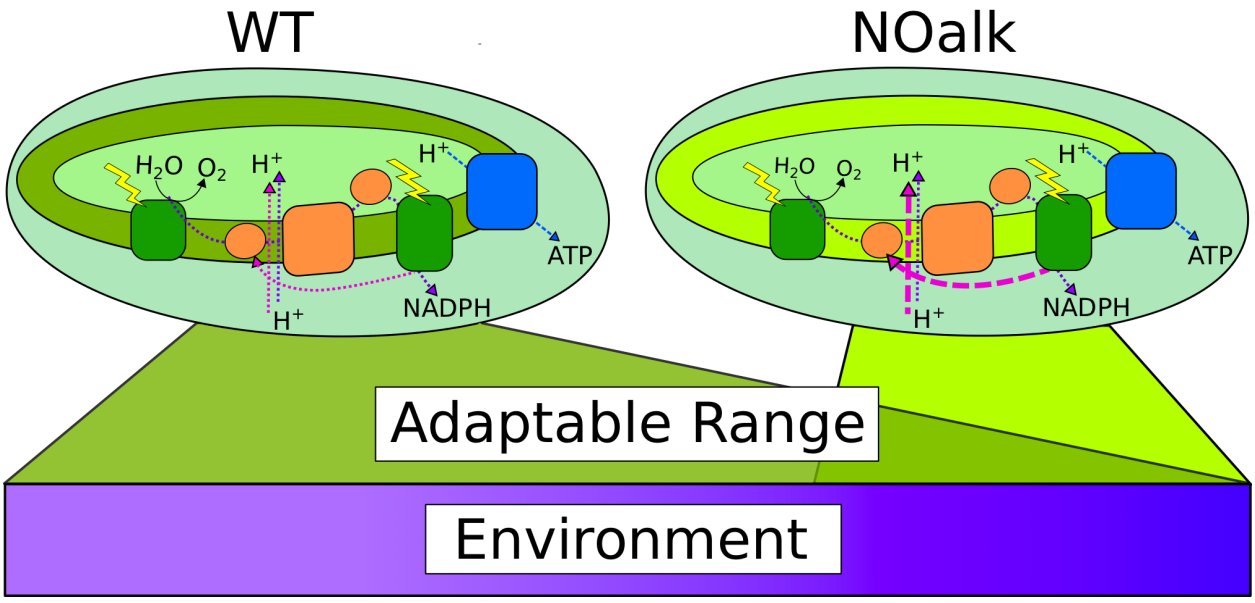


Figure 2.7: Alkanes impact the adaptability of cyanobacteria to environmental conditions. The lack of alkanes constrains the thylakoid electron transport chain to a higher recycle rate of electrons from photosystem I to the plastoquinone pool. This inflexibility in reductant partitioning leads to a narrower range of environmental conditions (in particular temperature) in which the strain can grow optimally

Chapter 3

Reconstruction and Comparison of the Metabolic Potential of Cyanobacteria *Cyanothece* sp. ATCC 51142 and *Synechocystis* sp. PCC 6803

3.1. Introduction

Cyanobacteria are primary producers in aquatic environments and contribute significantly to biological carbon sequestration, O_2 production and the nitrogen cycle (Bryant *et al.* 2006; Popa *et al.* 2007; Moisander *et al.* 2010). Their inherent photosynthetic capability and ease in genetic modifications are two significant advantages over other microbes in the industrial production of valuable bioproducts. In contrast to other microbial production processes requiring regionally limited cellulosic feedstocks, cyanobacteria only need CO_2, sunlight, water and a few mineral nutrients to grow. The short life cycle and transformability of cyanobacteria combined with a detailed understanding of their biochemical pathways are significant advantages of cyanobacteria as efficient platforms for harvesting solar energy and producing bio-products such as short chain alcohols, hydrogen and alkanes (Ducat *et al.* 2011).

The genus *Cyanothece* includes unicellular, diazotrophic cyanobacteria. *Cyanothece* sp. ATCC 51142 (hereafter *Cyanothece* 51142) is one of the most potent diazotrophs characterized and the first to be completely sequenced (Welsh *et al.* 2008). This organism can fix atmospheric nitrogen at rates higher than many filamentous cyanobacteria and also accommodates the biochemically incompatible processes of photosynthesis and nitrogen fixation within the same cell by temporally separating them (Zehr 2005). *Synechocystis* sp. PCC 6803 (hereafter *Synechocystis* 6803), the first photosynthetic organism with a completely sequenced genome (Kaneko *et al.* 1996), is probably the most extensively studied model organism for photosynthetic processes (Knoop *et al.* 2010). It is also closely related to *Cyanothece* 51142 and shares many characteristics with all *Cyanothece*. The genome of *Cyanothece* 51142 is about 35% larger than that of *Synechocystis* 6803 mostly due to the presence of genes related to nitrogen fixation and temporal regulation (Welsh *et al.* 2008). *Synechocystis* 6803 has been the subject of

many targeted genetic manipulations (e.g., expression of heterologous gene products) as a photo-biological platform for the production of valuable chemicals such as poly-beta-hydroxybutyrate, isoprene, hydrogen and other biofuels (Wu *et al.* 2001; McHugh 2005; Turner *et al.* 2008; Liu *et al.* 2009; Navarro *et al.* 2009; Bandyopadhyay *et al.* 2010; Knoop *et al.* 2010; Lindberg *et al.* 2010; Min *et al.* 2010). However, genetic tools for *Cyanothece* 51142 are still lacking thus hampering its use as a bio-production strain even though it has many attractive native pathways. For example, *Cyanothece* 51142 can produce (in small amounts) pentadecane and other hydrocarbons while containing a novel (though incomplete) non-fermentative pathway for producing butanol (Schirmer *et al.* 2010; Wu *et al.* 2010).

A breakthrough in solar biofuel production will require following one of two strategies: 1) obtaining photosynthetic strains that naturally have high-throughput pathways analogous to those in known biofuel producers, or 2) creating cellular environments conducive for heterologous enzyme function. Despite its attractive capabilities including nitrogen fixation and H_2 production (Bandyopadhyay *et al.* 2010), unfortunately genetic tools are not currently available to efficiently test engineering interventions directly for *Cyanothece* 51142. Therefore, a promising path forward may be to use *Synechocystis* 6803 as a "proxy" (for which a comprehensive genetic toolkit is available) and subsequently transfer knowledge gained during experimentation with *Synechocystis* 6803 to *Cyanothece* 51142. This requires high quality metabolic models for both organisms. Comprehensive genome-wide metabolic reconstructions include the complete inventory of metabolic transformations of a given cyanobacterial system. Comparison of the metabolic capabilities of *Cyanothece* 51142 and *Synechocystis* 6803 derived from their corresponding genome-scale models will provide valuable insights into their niche biological functions and also open up new avenues for economical biofuel production.

As discussed in Chapter 1.5 of this work, Genome-scale models (GSM) contain gene to protein to reaction associations (GPRs) along with a stoichiometric representation of all possible biotransformations known to occur in an organism combined with a set of appropriate regulatory constraints on each reaction flux (Reed *et al.* 2006; Puchalka *et al.* 2008). By defining the global metabolic space and flux distribution potential, GSMs can assess allowable cellular phenotypes under specific environmental conditions (Reed *et al.* 2006; Puchalka *et al.* 2008). The first genome-scale model for *Cyanothece* 51142 was recently published (Vu *et al.* 2012). The authors addressed the complexity of the electron transport chain (ETC) and explored further the specific roles of photosystem I (PSI) and photosystem II (PSII). In contrast, *Synechocystis* 6803 has been the target for metabolic model reconstruction for quite some time (Yang *et al.* 2002; Shastri *et al.* 2005; Hong *et al.* 2007; Fu 2009; Knoop *et al.* 2010; Montagud *et al.* 2010; Montagud *et al.* 2011; Nogales *et al.* 2012). Most of these earlier efforts for *Synechocystis* 6803 focused on only central metabolism (Yang *et al.* 2002; Shastri *et al.* 2005; Hong *et al.* 2007). Knoop *et al.* (Knoop *et al.* 2010) and Montagud *et al.* (Montagud *et al.* 2010; Montagud *et al.* 2011) developed genome-scale models for *Synechocystis* 6803, analyzed growth under different conditions, identified gene knock-out candidates for enhanced succinate production and performed flux coupling analysis to detect potential bottlenecks in ethanol and hydrogen production. A more recent model describes in detail the photosynthetic apparatus, identifies alternate electron flow pathways and highlights the high photosynthetic robustness of *Synechocystis* 6803 during photoautotrophic metabolism (Nogales *et al.* 2012). (Knoop *et al.* 2013) extended their earlier efforts to address recent developments in the central metabolism of cyanobacteria, including detection of a novel complete TCA cycle and ongoing debates about the presence of a glyoxylate shunt in this clade. All these efforts have brought about an improved

understanding of the metabolic capabilities of *Synechocystis* 6803 and cyanobacteria in general.

This chapter introduces high-quality genome-scale models for *Cyanothece* 51142, *i*Cyt773, and *Synechocystis* 6803, *i*Syn731, (as shown in Table 3.1) that integrate recent developments (Nogales *et al.* 2012; Vu *et al.* 2012), supplements them with additional literature evidence and highlights their similarities and differences. The detailed descriptions of the model are not included in this work, but can be found in the supplementary materials of the original article on which it is based (Saha *et al.* 2012). As many as 322 unique reactions are introduced in the *Synechocystis* *i*Syn731 model and 266 in *Cyanothece* *i*Cyt773. New pathways include, among many, a TCA bypass (Zhang *et al.* 2011), heptadecane biosynthesis (Schirmer *et al.* 2010) and detailed fatty acid biosynthesis in *i*Syn731 and comprehensive lipid and pigment biosynthesis and pentadecane biosynthesis (Schirmer *et al.* 2010) in *i*Cyt773. For the first time, not only extensive gene essentiality data (Nakamura *et al.* 1999) is used to assess the quality of the developed model (i.e., *i*Syn731) but also the allowable model metabolic phenotypes are contrasted against MFA flux data (Young *et al.* 2011). This comparison is extremely valuable because it validates the model by comparing its predictions to *in vivo* measurements of metabolic fluxes. The diurnal rhythm of *Cyanothece* metabolism is modeled for the first time via developing separate (light/dark) biomass equations and regulating metabolic fluxes based on available protein expression data over light and dark phases (Stockel *et al.* 2011).

3.2. Materials and Methods

3.2.1. Measurement of Biomass Precursors

Growth Conditions

Wild-type *Synechocystis* 6803 and *Cyanothece* 51142 were grown for several days from an initial OD_{730} of ~0.05 to ~0.4. *Synechocystis* 6803 was grown in BG-11 medium (Allen 1968) and *Cyanothece* 51142 in ASP2 medium (Reddy *et al.* 1993) with (+N) or without (-N) nitrate. All cultures were grown in shake flasks with continuous illumination of ~100 μmol photons/m^2/sec provided from cool white fluorescent tubes. *Synechocystis* was maintained at 30°C and *Cyanothece* at 25°C. For *Synechocystis,* the illumination was constant and doubling time was ~24 hours. *Cyanothece* alternated between 12 hours of light and 12 hours of darkness, with a doubling time of ~48 hours.

Pigments

1 mL of cells of both *Synechocystis* 6803 and *Cyanothece* 51142 (from light and dark phases) was pelleted and extracted twice with 5 mL 80% aqueous acetone and the extracts pooled. Spectra of this extract and of a sample of whole cells were taken on a DW2000 spectrophotometer (Olis, GA, USA) against 80% acetone or BG-11 media as a reference. Chlorophyll a contents were calculated as reported (Porra *et al.* 1989) from the acetone extract. Total carotenoid concentrations were also calculated from the acetone extract according to a published method (Lichtenthaler 1987). The relative amounts of different carotenoids included in the biomass equation were estimated according to known ratios (Steiger *et al.* 1999). Concentrations of phycocyanin were estimated from the spectra of intact cells (Arnon *et al.* 1974). All measurements were taken in triplicate.

Amino Acids

Total protein contents were measured using a Pierce BCA Assay kit. Amino acid proportions were determined according to published shotgun proteomics data for both *Cyanothece* 51142 and *Synechocystis* 6803 across a range of conditions (Stoeckel *et al.* 2008) according to the following procedure: From peptide-level data, each mass spectral observation of a peptide was taken as an instance of a particular protein. The amino acid composition of each protein was taken from data in Cyanobase (http://*genome.kazusa.or.jp/cyanobase)* and thus the 'proteome' was taken to include all of the proteins whose peptides were observed in our data set, in proportion according to how often their peptides were observed. Amino acid frequencies were averaged across the proteome by a weighting factor of number of observations divided by the number of amino acids in the protein, similar to RPKM normalization for next-gen sequencing (Mortazavi 2008).

Other Cellular Components

The compositions of other cellular components of *Synechocystis* 6803 and *Cyanothece* 51142 were estimated based on values in the literature. DNA and RNA contents for *Synechocystis* 6803 were reported by Shastri and Morgan (Shastri *et al.* 2005). The remaining biomass components of *Synechocystis* 6803 (i.e., lipid, soluble pool and inorganic ions) were extracted from the measurements carried out by Nogales et al.(Nogales *et al.* 2012). For *Cyanothece* 51142, biochemical compositions of macromolecules such as lipids, RNA, DNA and soluble pool were extracted from the measurements reported by Vu et al. (Vu *et al.* 2012).

3.2.2. Model Simulations

Flux balance analysis (FBA) (Varma *et al.* 1994) was employed in both the model validation and model testing phases. *Cyanothece i*Cyt773 and *Synechocystis i*Syn731 models were evaluated in terms of biomass production under several scenarios: light and dark phases, heterotrophic and mixotrophic conditions. Flux distributions for each one of these states were inferred using FBA:

Maximize $v_{biomass}$

Subject to

$$\sum_{j=1}^{M} S_{ij} v_j = 0,\ \forall i \in 1,...,N \tag{3.1}$$

$$v_{j,min} \leq v_j \leq v_{j,max},\ \forall j \in 1,...,M \tag{3.2}$$

Here, S_{ij} is the stoichiometric coefficient of metabolite i in reaction j and v_j is the flux value of reaction j. Parameters $v_{j,min}$ and $v_{j,max}$ denote the minimum and maximum allowable fluxes for reaction j, respectively. Light and dark phases in *Cyanothece* 51142 are represented via modifying the minimum or maximum allowable fluxes with the following constraints, respectively:

$$v_{Glytr} = 0 \text{ and } v_{Glyctr} = 0 \tag{3.3}$$

$$v_{CO_2tr} = 0,\ v_{Glytr} = 0,\ v_{light} = 0,\ v_{cf} = 0 \tag{3.4}$$

Here, $v_{Biomass}$ is the flux of biomass reaction and v_{Glytr}, v_{Glyctr} and v_{CO2tr} are the fluxes of glycerol, glycogen and carbon dioxide transport reactions and v_{light} and v_{cf} are the fluxes of light reactions and carbon fixation reactions. For light phase, constraint (3.3) was included in the linear model, whereas for dark phase constraint (3.4) was included.

Once the *Synechocystis i*Syn731 model was validated, it was further tested for *in silico* gene essentiality. The following constraint(s) was included individually in the linear model to represent any mutant:

$$v_{\text{mutant}} = 0 \tag{3.5}$$

Here, v_{mutant} represents flux of reaction(s) associated with any genetic mutation.

Flux variability analysis (Kumar *et al.* 2011) for the reactions (for which photoautotrophic ^{13}C MFA measurements (Young *et al.* 2011) were available) was performed based on the following formulation:

Maximize/Minimize v_j

Subject to

$$\sum_{j=1}^{m} S_{ij} v_j = 0 \ \forall \ \text{i} \in 1,\ldots\ldots, \text{n} \tag{3.6}$$

$$v_{j,\min} \leq v_j \leq v_{j,\max} \ \forall \ \text{j} \in 1, \ldots\ldots, \text{m} \tag{3.7}$$

$$v_{Biomass} \geq v_{min}^{Biomass} \tag{3.8}$$

Here, $v_{min}^{Biomass}$ is the minimum level of biomass production. In this case we fixed it to be the optimal value obtained under light condition for the *Synechocystis i*Syn731 model.

CPLEX solver (version 12.1, IBM ILOG) was used in the GAMS (version 23.3.3, GAMS Development Corporation) environment for implementing GapFind and GapFill (Satish Kumar *et al.* 2007) and solving the aforementioned optimization models. All computations were carried out on Intel Xeon E5450 Quad-Core 3.0 GH and Intel Xeon E5472 Quad-Core 3.0 GH processors that are the part of the lionxj cluster (Intel Xeon E type processors and 96 GB memory) of High Performance Computing Group of The Pennsylvania State University.

3.3. Results and Discussion

3.3.1. Model Components

Biomass Composition and the Diurnal Cycle

The biomass equation approximates the dry biomass composition by draining all building blocks or precursor molecules in their physiologically relevant ratios. Most of the earlier genome-scale modeling efforts (Fu 2009; Knoop *et al.* 2010; Montagud *et al.* 2010) of *Synechocystis* 6803 contain approximate biomass equations completely or partially adopted from other species without direct measurements. This can adversely affect the accuracy of maximum biomass yield calculations, gene essentiality predictions and knockouts for overproduction.

Biomass composition for *Synechocystis i*Syn731 and *Cyanothece i*Cyt773 models were generated by defining all essential cellular biomass content values by experimental measurement or collection from existing literature (*see 'Materials and Methods' for details)*. Macromolecules present in both cyanobacteria such as protein, carbohydrates, lipids, DNA, RNA, pigments, soluble pool and inorganic ions were assigned to their corresponding metabolic precursors (e.g., L-glycine, glucose, 16C-lipid, ATP, dGTP, beta-carotene, coenzyme A and potassium respectively, *see Supplementary file S3 for the complete list of biomass components*). Based on the experimental measurements of precursor molecules needed to form a gram of the biomass, stoichiometric coefficients were assigned. For *Synechocystis* 6803 we measured compositions of proteins and pigments and extracted compositions of the remaining biomass macromolecules from the model by Nogales *et al.* (Nogales *et al.* 2012). Thereby we developed biomass equations for three different conditions: photoautotrophic, mixotrophic and heterotrophic (see Supplementary file S3). Experimental measurements (described in the Materials and Methods section and also in Supplementary File S3) showed that biomass composition (i.e., mainly

pigments) varies for *Cyanothece* 51142 between light and dark conditions and nitrogen supplementation. Since pigments such as chlorophyll, carotenoids and phycocyanobilin play important roles in photosynthetic processes their quantities are consequently higher under light conditions. In the presence of light *Cyanothece* 51142 uses photosynthesis to store solar energy in the form of carbohydrates (i.e., glycogen), while in dark it expends that energy to fix nitrogen. Surprisingly, no significant change was measured in the carbohydrate pool between light and dark phases due to infinitesimal contribution of photosynthetically stored carbohydrates to total carbohydrate content in the biomass of *Cyanothece* 51142. Aggregate quantities of the remaining biomass macromolecules for *Cyanothece* 51142 such as lipids, RNA, DNA and soluble pool were extracted from the most recent *Cyanothece* 51142 model by Vu *et al*. (2012) to develop biomass equations for light and dark phases *(*see Supplementary file S3).

An earlier characterization study for *Cyanothece* 51142 revealed that 113 proteins are expressed in higher abundance in the light phase while 137 are expressed in higher abundance in dark conditions (Stockel *et al*. 2011). The constructed model spans 26 light-specific proteins, associated with 36 reactions mainly involved in fatty acid, pigment, and amino acid metabolism and 11 dark-specific proteins accounting for 16 reactions from glycolysis, purine, pyrimidine, pyruvate, and amino acid metabolism *(*see Supplementary files S4). Separate biomass equations as well as two regulatory structures for the model were derived in order to represent diurnal metabolic differences for *Cyanothece* 51142 (see Supplementary File S4). In contrast, diurnal differences observed in *Synechocystis* 6803 (Kucho *et al*. 2005) are less pronounced (i.e., observed for only 54 genes) and less well functionally annotated (i.e., 32 genes with 'unassigned' functions). When compared to existing biomass equations of *Synechocystis* 6803 (Knoop *et al*. 2010; Montagud *et al*. 2010) we found significantly lower values for the percent

weight contribution of proteins towards the biomass pool (i.e., 52% for *Synechocystis* 6803 and 53% for *Cyanothece* 51142 vs. 84% and 66%, respectively). The new protein biomass contribution is in better agreement with the previously reported value of 55% for *Cyanothece* 51142 (Tredici *et al.* 1986).

Identification and Correction of Network Gaps

Upon ensuring biomass formation, GapFind (Satish Kumar *et al.* 2007) was applied to assess network connectivity and blocked metabolites. By applying Gapfill (Satish Kumar *et al.* 2007), putative reconnection hypotheses were identified for blocked metabolites. Only the suggested modifications that were independently corroborated using literature sources and also did not lead to the introduction of thermodynamically infeasible cycles were included in the model. For *Synechocystis i*Syn731 model, GapFind identified 207 blocked metabolites. Note that there exist 125 blocked metabolites in the *i*JN678 model (Nogales *et al.* 2012). GapFill identified unblocking hypotheses for 138 blocked metabolites. However, 88 of them led to the generation of infeasible thermodynamic cycles and thus were excluded. For only 5 blocked metabolites corroborating evidence for reconnection was obtained by adding 10 reactions (i.e., 2 metabolic, 4 transport and 4 exchange reactions). The added metabolic reactions have unknown gene associations (see Supplementary file S1 for detailed information) while all 4 added transport reactions involve passive diffusion and thus are not associated with any specific gene(s) or protein(s). Ultimately, the 45 remaining blocked metabolites with reconnection mechanisms suggested by GapFill (but unconfirmed) along with 69 blocked metabolites with no reconnection hypotheses were retained in the model *i*Syn731, while metabolites such as ubiquinone, a potential alternate substrate for succinate dehydrogenase (Nogales *et al.* 2012), was excluded from *i*Syn731.

For the *Cyanothece* *i*Cyt773 model, 74 blocked metabolites were found after applying GapFind. Note that there are 66 blocked metabolites in *i*Cce806 (Vu *et al.* 2012). Two exchange reactions were added to allow the uptake of glucose and thyaminose ensuring biomass production under heterotrophic or mixotrophic conditions. Four blocked metabolites directly adopted from *i*Cce806 (during the draft model creation phase) were linked to five reactions with spurious gene associations and thus both metabolites and reactions were removed from *i*Cyt773. GapFill suggested re-connection mechanisms for 52 blocked metaboloites (out of a total of 70). However, for 12 blocked metabolites the re-connection model modifications led to the creation of thermodynamically infeasible cycles and thus were discarded. Corroborating evidence for the reconnection of 30 blocked metabolites was identified through the addition of 19 GapFill suggested reactions (i.e., 8 metabolic, 7 transport and 4 exchange reactions). Of the eight added metabolic reactions we found direct literature evidence for five, homology-based evidence for one while two reactions are spontaneous (see Supplementary file S2 for detailed information). All seven added transport reactions are through passive diffusion and thus are not connected with any specific gene(s) or protein(s). Ten remaining blocked metabolites for which GapFill suggested reconnection hypotheses (along with 22 with no reconnection hypotheses) were left blocked in *i*Cyt773 as no information to corroborate the GapFill suggested changes was found in the published literature and databases. For example, biotin is produced in *Cyanothece* 51142; however, there is no literature evidence to support the presence of the initial step of the primary production pathway (i.e., conversion of pimeloyl-CoA from pimelate) and the intermediate step (i.e., biotransformation of 7,8-diamino-nonanoate from 8-amino-7-oxononanoate). This indicates that *Cyanothece* 51142 may utilize a currently unknown pathway for producing biotin. The six other blocked metabolites are involved in the nonfermentative alcohol production pathway (as

explained in model comparison section) known to be incomplete in *Cyanothece* 51142. Table 3.2 summarizes the results related to connectivity restoration of *Synechocystis* *i*Syn731 and *Cyanothece* *i*Cyt773 models.

GPR Associations and Elemental and Charge Balancing

GPR associations connect genotype to phenotype by linking gene(s) (G) that code for the protein(s) (P) that catalyze a particular reaction (R). They are important to trace correctly as they provide the means to target at the gene level any change in the network desired at the reaction level. This is critical because genes may catalyze multiple reactions in multiple pathways. Many earlier models for *Synechocystis* 6803 do not provide in detail complex GPR associations, rather list only gene(s) and enzyme(s) involved in a specific reaction (Knoop *et al.* 2010; Montagud *et al.* 2010; Montagud *et al.* 2011). For both *i*Cyt773 and *i*Syn731 models, we included comprehensive GPR associations (see Table 3.1 for detailed information). All four intracellular compartments (i.e., periplasm, cytosol, thylakoid lumen and carboxysome) were assumed to have the same pH (7.2) and subsequently, metabolites were assigned appropriate protonation states corresponding to this pH and each reaction was elementally and charge balanced.

Under high light intensity in photoautotrophic conditions, *Cyanothece* *i*Cyt773 model produces 0.026 mole biomass/mole carbon fixed whereas *Synechocystis* *i*Syn731 yields 0.021 mole biomass/mole carbon fixed. These yields are almost identical to the ones calculated using the most recent models of *Cyanothece* 51142 (Vu *et al.* 2012) and *Synechocystis* 6803 (Nogales *et al.* 2012). Experimental measurements of biomass yields are in the same order of magnitude with model predictions for the two organisms, 0.072 (Reddy *et al.* 1993; Vu *et al.* 2012) and 0.082 mole biomass/mole carbon fixed (Bentley *et al.* 2012), respectively.

3.3.2. Comparing Flux Predictions of *i*Syn731 Against Experimental Measurements

We superimposed photoautotrophic flux measurements (Young *et al.* 2011) for *Synechocystis* 6803 onto *i*Syn731 model predictions to assess if the measurements are consistent with the model and whether the biomass maximization assumption correctly apportions fluxes to the metabolic network. For each reaction that was assigned a flux we calculated the flux-range under the maximum biomass assumption. Table 3.3 and Figure 3.1 summarize the obtained results for a basis of 100 millimole of CO_2 plus H_2CO_3 uptake (Young *et al.* 2011). In seven (out of thirty one) cases the measured flux is fully contained within the model predicted ranges obtained upon maximizing biomass formation implying model consistency with MFA measurements. In contrast, under the maximum biomass assumption for thirteen fluxes the ranges underestimate and for four fluxes the ranges overestimate the experimentally deduced flux ranges while for seven fluxes the model derived flux ranges partially overlap with the experimental ones.

Perhaps the most informative discrepancy is for the CO_2 fixing RuBisCO (RBC) reaction, which has a measured flux range of (123.00 to 132.00) vs. the model-calculated range of (102.49 to 106.33). In both cases the increased RBC flux (in comparison to the basis of 100 millimole of CO_2 plus H_2CO_3 uptake) is needed to counteract the carbon loss due to the CO_2 releasing reactions such as isocitrate dehydrogenase (ICD) and pyruvate dehydrogenase (PDH). We find that flux ranges, under the maximum biomass production assumption, of reactions such as glucose 6-phosphate dehydrogenase (G6PD), 6-phosphogluconolactonase (6PGL) and phosphogluconate dehydrogenase (6PGD) in oxidative pentose phosphate (OPP) pathway are negligible (0.00 to 0.03). In contrast, the experimentally derived range for OPP is from 12 to 21. This is approximately equal to the difference between the model-predicted vs. experimentally

deduced RBC reaction range implying the persistence of OPP flux even under the photoautotrophic condition (Young *et al.* 2011) despite the presence of a more efficient NADPH production route through photosynthesis as predicted by the model (under max biomass). The high values Young et al. (2011) obtained for the OPP fluxes were surprising as OPP is not a very efficient route for cyanobacteria to generate reducing power. This may reflect some inherent biological constraint that is not captured by the optimality assumption. For example, photosynthetic NADPH production may be limited by the need to maintain the redox poise of the electron transport chain (See chapter 5 of this work). Although this strategy may not be optimal from a flux balance perspective, the need to avoid oxidative damage may explain the use of this apparently inefficient pathway.

Model predicted lower flux ranges for RBC are propagated to seven other reactions in the Calvin cycle (i.e., phosphoglycerate kinase (PGK), glyceraldehyde-3-phosphate dehydrogenase (13PDG), triose-phosphate isomerase (TPI), transketolase (TKT1), ribose-5-phosphate isomerase (RPI), ribulose 5-phosphate 3-epimerase (RPE) and phosphoribulokinase (PRK). The remaining six reaction fluxes with lower model predicted fluxes compared to measurements (Young *et al.* 2011) are all in the TCA cycle (i.e., citrate synthase (CS), aconitase (ACONT), isocitrate dehydrogenase (ICD), succinate dehydrogenase (SUCD) and malic enzyme (ME1 and ME2) reactions). Even under the max biomass assumption, SUCD is not required to carry any flux due to the presence of other succinate dehydrogenases (as part of respiratory chain) in the *i*Syn731 model. This is in keeping with more recent ^{13}C-MFA data showing that flux through the cyanobacterial TCA bypass is minimal during normal growth (You *et al.* 2014). Furthermore, in contrast with experimental observations, under the maximum biomass assumption, the model predicts no flux through the malic enzyme (ME) reactions presumably because it is a less

energy-efficient route (i.e., phosphoenolpyruvate → oxaloacetate → malate → pyruvate) for pyruvate generation than the pyruvate kinase (PYK) reaction (Young *et al.* 2011).

There are nine reactions with experimentally derived ranges completely subsumed within the ones derived under the maximum biomass assumption. Five of them are in the Calvin cycle: fructose-bisphosphate aldolase (FBA), fructose-bisphosphatase (FBP), Sedoheptulose 1,7-bisphosphate D-glyceraldehyde-3-phosphate-lyase (SBGPL), sedoheptulose-bisphosphatase (SBP) and bidirectional transaldolase (TAL). The first four reactions are essential with experimentally deduced flux ranges of (53.00 to 66.00) for FBA and FBP and (29.00 to 43.00) for SBGPL and SBP. In contrast, the calculated flux ranges (-0.08 to 73.17) for FBA and SBGPL and (0.00 to 73.17) for FBP and SBP imply that they are *in silico* non-essential. As depicted in Figure 3.1, these reactions are involved in the production of sedoheptulose 7-phosphate (S7P) from fructose 1,6-bisphosphate (FDP). An alternative production route for S7P is afforded in the model through the bidirectional transaldolase (TAL) reaction from fructose 6-phosphate (F6P) alluding to an explanation for the wider flux ranges derived using the model. Experimental and model predicted flux ranges for TAL are (-6.00 to 9.00) and (-35.93 and 37.32), respectively. Upon restricting the TAL flux ranges in the calculations to the ones found experimentally, the flux variability analysis shrinks the flux ranges for FBA and FBP to (28.22 to 43.27) and (28.22 to 43.33) and for SBGPL and SBP to (29.82 to 44.87) and (29.82 to 44.87), respectively which are very close to the experimentally measured ranges. This is indicative that in addition to the maximization of biomass formation, additional restrictions (e.g., photosynthetic efficiency and relative selectivity of RuBISCO for carboxylation over oxidation) limit the range of fluxes that the aforementioned glycolytic fluxes may span *in vivo*. Note that the presence of experimentally

measured fluxes is important to test the model and the adopted maximization principle. We were fortunate in this case to have access to such data as for most organisms they are absent.

Phosphoglycerate mutase (PGM) and enolase (ENO) reactions have very similar model derived and experimentally obtained flux ranges. Model-predicted flux values of the remaining two reactions, pyruvate kinase (PYK) and pyruvate dehydrogenase (PDH), could reach as low as zero due to the metabolic flexibility that the *i*Syn731 model possesses by having alternate enzymes with different cofactor specificities. The max biomass flux range of fumarase (FUM) is found to be (-7.26 to 1.49), compared to the experimentally measured (1.70 to 2.00). Therefore, it appears that under the photoautotrophic condition, the forward direction is kinetically favorable. By restricting the reaction to be irreversible the model predicted a FUM flux range of (0.00 to 1.49) which is close to the experimentally derived one (see Figure 3.1). However, contrary to MFA measurements these reactions (FUM and ME) are dispensable for *in silico* biomass production.

3.3.3. *i*Syn731 Model Testing Using *in Vivo* Gene Essentiality Data

The quality of model *i*Syn731 for Synechocystis 6803 was tested using experimental data on the viability (or lack thereof) of single gene knockouts. We used the CyanoMutants database (Nakamura *et al.* 1999; Nakao *et al.* 2010) that includes *in vivo* gene essentiality data for 119 genes (i.e., 19 essential and 100 nonessential) with metabolic functions in *i*Syn731 model. Cases that were flagged with incomplete segregation in the database were omitted in *i*Syn731 model comparisons. We examined the feasibility of biomass production for the model *i*Syn731 by comparing the maximum biomass formation upon imposing the gene knockout with the maximum theoretical yield of the *wild-type* organism. A threshold of 10% of the maximum theoretical yield was used as a cutoff (Kumar *et al.* 2009). Comparisons between *in vivo* and *in*

silico results led to four possible outcomes, as previously delineated by Kumar *et al*., GG, GNG, NGG and NGNG (Kumar *et al.* 2009). Initially, the model correctly predicted 18 out of 19 essential genes (i.e., 18 NGNG and 1 GNG) and 74 out of 100 non-essential genes (i.e., 73 GG and 27 NGG). We next explored the causes of these discrepancies and attempted to mitigate them whenever possible.

The single GNG case corresponds to mutant Δ*chlA*$_I$ exhibiting no growth under aerobic conditions (Minamizaki *et al.* 2008). The ChlA$_{\text{I}}$ system is a Mg-protoporphyrin IX monomethylester (MPE) cyclase system that is responsible for forming the isocyclic ring (E-ring) in chlorophylls under aerobic conditions. The model allowed for the BchE and ChlA$_{\text{II}}$ systems (alternate cyclase systems) to complement for the loss of the ChlA$_{\text{I}}$ system leading to an *in silico* viable mutant. However, the same literature source suggested that both BchE and ChlA$_{\text{II}}$ systems are unlikely to be active under aerobic conditions and thus rescue mutant Δ*chlA*$_I$. This prompted the introduction of a regulatory restriction in *i*Syn731 model where only ChlA$_{\text{I}}$ reactions were active under aerobic conditions as MPE while ChlA$_{\text{II}}$ and BchE system reactions were deactivated. Using these regulatory restrictions resolves the single GNG inconsistency.

Twenty (out of 27) NGG cases were associated with Photosystem I (PSI), Photosystem II (PSII) and other photosynthesis reactions. While reconstructing the model, we assumed that all genes involved in photosynthetic reaction system were essential to the functioning of the overall system. Published literature (Jansson *et al.* 1987; Chitnis *et al.* 1989; Burnap *et al.* 1991; Nakamoto 1995; Shen *et al.* 1997) suggests that genes involved in photosynthetic reactions form complex interdependencies. We used NCBI COBALT multiple sequence alignment tool (Papadopoulos *et al.* 2007) to construct a phylogenetic tree of the genes associated with each photosystem along with BLASTp searches to identify putative complementation relationships

between genes to explain the inconsistencies between the predicted *in silico* and *in vivo* growth. Genes deemed homologous (i.e., lie adjacent in the phylogenetic tree) were linked with "OR" GPR relations implying that the loss of one gene can be complemented by the other. Seven out of twenty NGG cases (i.e., *psaD, psaI* and *psbA2* for PSI and PSII and *cpcC2, cpcC1, cpcD,* and *apcD* for other photosynthesis reactions) were resolved by modifying the corresponding GPR using an OR relation (Chitnis *et al.* 1989; Chitnis *et al.* 1989; Delorimier *et al.* 1990; Nakamoto 1995; Ughy *et al.* 2004; Jallet *et al.* 2012). However, no phylogenetically adjacent or related (or homologous) genes were found for the remaining 13 NGG cases (*psaE, psbD2, psbO, psbU, psbV, psb28, psbX, psb27, petE, cpcA, cpcB, apcE, apcF*) (Chitnis *et al.* 1989; Burnap *et al.* 1991; Shen *et al.* 1993; Shen *et al.* 1995; Manna *et al.* 1997; Shen *et al.* 1997; Shen *et al.* 1998; Jallet *et al.* 2012). For these cases, the genes were deemed nonessential to the functioning of the reactions in question (i.e., photosynthesis reactions) and thereby the corresponding GPRs were modified to show an OR relation between each of these genes and an '*unknown gene'*, similar to what was previously performed in the refinement of the *i*MM904 model (Mo *et al.* 2009) (see Supplementary File S5 for detailed information). This procedure incorporates the fact that while many of these gene products participate in some way in an optimal an robust photosynthesis, they are not always required for growth.

The remaining seven NGG cases are associated with a variety of metabolic functions. One such case is the Δ*modBC* mutant corresponding to the sole ABC molybdate transporter in the model. Literature evidence (Zahalak *et al.* 2004) revealed that a related cyanobacterium, *Anabaena variabilis* ATCC 29413, could continue to grow despite the loss of its molybdate ABC transporter due to the presence of another low affinity molybdate transporter or an inducible sulfate transport system that can serve as a low affinity molybdate transporter when required. We

found the same gene coding for the sulfate transporter in *A. variabilis* (*cysA*) in the *i*Syn731 model allowing the resolution of the discrepancy by adding a *cysA*-linked alternate molybdate transporter. Another NGG case is mutant Δ*crtO* that cannot produce echinenone (a biomass component) in *i*Syn731 with no effect on observed growth. Therefore, it appears that *i*Syn731 cannot capture the flexibility of *Synechocystis* 6803 metabolism (Fernandez-Gonzalez *et al.* 1997) when echinenone production is restricted. The remaining five NGG cases are spread across many metabolic pathways. The Δ*ctaA* mutant eliminates the copper ABC transporter without affecting growth, which alludes to the existence of another unknown mode of copper uptake not present in *i*Syn731 (Tottey *et al.* 2001; Tottey *et al.* 2002). The Δ*menG* mutant eliminates a reaction for the production of phylloquinone while mutant Δ*ppd* affects the production of homogentisate, a precursor for both tocopherols and plastoquinone. Finally, the Δ*vte3* mutant affects the production of both plastoquinone and α-tocopherol (Cheng *et al.* 2003) and the viable Δ*ccmA* mutant restricts the production of chorismate (a precursor to aromatic amino acids) and also restricts carboxysome formation (Ogawa *et al.* 1994; Dahnhardt *et al.* 2002; Sakuragi *et al.* 2002). These six inconsistencies between the model predictions and growth data imply that the cyanobacterium can co-opt another metabolic process to (partially) complement for the gene loss. Unlike the case of the Δ*modBC* mutant, we have found no plausible mechanism for the six remaining mutants.

After resolving the discrepancies, as described above, *i*Syn731 correctly predicted all 19 essential genes (i.e., 19 NGNG and 0 GNG) and 94 (out of 100) non-essential genes (i.e., 94 GG and 6 NGG). Figure 3.2 shows our results and comparisons against two other available *Synechocystis* 6803 models by (Knoop *et al.* 2010) and (Nogales *et al.* 2012). We used the CyanoMutants database (Nakamura *et al.* 1999) to identify 114 genes (i.e., 19 essential and 95

nonessential) having metabolic functions in the *i*JN678 model by (Nogales *et al.* 2012). Out of 114 genes the *i*JN678 model correctly predicted 18 essential genes (i.e., 18 NGNG and 1 GNG) and 69 non-essential genes (i.e., 69 GG and 26 NGG). The model by Knoop *et al.* was tested for 51 mutants but we found that only 43 (i.e., 7 essential and 36 non-essential) of them were reported to have complete segregation (Nakamura *et al.* 1999). Of these 43, Knoop *et al.*'s model correctly predicted 5 essential genes (i.e., 5 NGNG and 2 GNG) and 32 nonessential genes (i.e., 32 GG and 4 NGG). The specificity and sensitivity of each of these three models were also calculated and displayed at the bottom of Figure 3.2. Although an updated version of the model from (Knoop *et al.* 2013) was released in between the original publication of this article and its revision for this thesis, we did not re-test gene essentiality or any other parameters of that model.

All 114 genes tested for *i*JN678 were also present in the *i*Syn731 model. 26 NGG and one GNG cases present in *i*JN678 model correspond to NGG and GNG cases that were either fixed or still present in *i*Syn731 as discussed before. Lethal mutant Δ*ppa* is correctly predicted as NGNG in *i*Syn731 but deemed GNG in (Knoop *et al.* 2010) model. This was because *ppa* in *i*Syn731 codes for the degradation of both triphosphate into diphosphate and diphosphate to phosphate. Only the latter activity is linked to *ppa* in Knoop et al's model. Out of 4 NGG cases, two involve Δ*cmpA* and Δ*cmpB* mutants. Both these genes are involved in the ABC transporter system for bicarbonate from periplasm to cytosol. *i*Syn731 avoids this inconsistency as it contains an alternate sodium and bicarbonate co-transport system.

3.3.4. Model Comparisons

Synechocystis 6803 Model Comparisons

The *i*Syn731 model integrates the description in the photosystems of the model presented by Nogales *et al*. and adds additional detail. One notable difference is that *i*Syn731 uses a

separate photon for each reaction center (i.e., PSI and PSII) as they are optimized for different ranges of wavelength (Taiz *et al.* 2002), whereas *i*JN678 uses a single photon shared by both photosystem reactions. As many as 322 new reactions (see Figure 3.3A), are added in *i*Syn731 distributed across many pathways. Most of the additions are in the lipid and fatty acid metabolism to support the synthesis of measured fatty acids and lipids present in the biomass equation. This list includes myristic acid (14-carbon saturated fatty acid) and lauric acid (12-carbon saturated fatty acid). *i*JN678 contained four reactions exhibiting unbounded flux (i.e., two duplicate glycine cleavage reactions and two duplicate leucine transaminase reactions). They form a thermodynamically infeasible cycle (see Figure 3.4A for leucine transaminase reactions) that was resolved in *i*Syn731 by eliminating redundant functions. In addition, the glycine cleavage system was recast in detail by abstracting the separate action of the four enzymes (named the T-, P-, L-, and H-proteins) that ultimately catalyze the demethylamination of glycine.

*i*Syn731 improves upon *i*JN678 by eliminating lumped reactions whenever a multi-step description is available and expands the range of functions carried out with alternate cofactors. As many as twelve reactions with an enoyl-[acyl-carrier-protein] reductase function were linked with not only NADP but also with the more rare NAD cofactor specificity. Another important difference between *i*Syn731 and *i*JN678 is the cellular location of CO_2 fixation by RuBisCO. Cyanobacteria possess a micro-compartment (Badger *et al.* 2003; Yeates *et al.* 2008) called the carboxysome encapsulating RuBisCO and carbonic anhydrase (CA). *i*Syn731 adds carboxysome as a cellular compartment and also all necessary transport reactions. Recently, Zhang and Bryant (2011) hypothesized the existence of a functional TCA cycle in most cyanobacterial species using a 2-ketoglutarate to succinate bypassing step. *i*Syn731 allows for a complete TCA cycle using the bypassing step. In addition, *i*Syn731 contains an intact heptadecane biosynthesis

pathway as recently described (Schirmer *et al.* 2010) unlike earlier *Synechocystis* 6803 models (Fu 2009; Knoop *et al.* 2010; Montagud *et al.* 2010; Nogales *et al.* 2012) (see Figure 3.5A for distribution of unique reactions in *i*Syn731).

Cyanothece 51142 Model Comparisons

The *i*Cyt773 model for *Cyanothece* 51142 improves upon the *i*Cce806 model (Vu *et al.* 2012). *i*Cyt773 segregates reactions into the periplasm, thylakoid lumen, carboxysome, and cytoplasm compartments thus introducing an additional 60 transport reactions compared to *i*Cce806. Unlike *i*Cce806, *i*Cyt773 does not track macromolecule synthesis for DNA, RNA, and proteins to maintain consistency with the *Synechocystis* 6803 model. This difference accounts for 69 genes present in *i*Cce806 but absent from *i*Cyt773. *i*Cce806 contained 15 reactions which formed five cycles that could carry unbounded metabolic flux (i.e., thermodynamically infeasible cycles). All these cycles were eliminated by restricting reaction directionality and eliminating reactions that were linear combinations of others coded by the same gene (see Figure 3.4B).

*i*Cyt773 contains 43 unique genes and 266 unique reactions (including transport and alternate cofactor utilizing reactions) as shown in Figure 3.3B. Figure 3.5B depicts the distribution of the new reactions across different pathways. Most of the additions are found in lipid and pigment biosynthesis pathways. The *i*Cyt773 model captures in detail the lipid biosynthesis pathway composed of 73 reactions and links as many as 28 biomass precursor lipids (e.g., sulfoquinovosyldiacylglycerols, monogalactosyldiacylglycerols, digalactosyldiacyl-glycerols, and phosphatidylglycerols) directly to the biomass equation. The porphyrin and chlorophyll metabolism and carotenoid biosynthesis pathways were updated to include 24 reactions for the production of accessory pigments such as echinenone, an accessory pigment, and (3Z)-phycocyanobilin, a phycobilin. Accessory pigments donate electrons to chlorophyll

rather than directly to photosynthesis. Phycobilins are adapted for many wavelengths not absorbed by chlorophyll thus broadening the spectrum useful for photosynthesis. The variety of pigments in cyanobacteria is well documented (Glazer 1977; Paerl 1984; Poutanen *et al.* 2001) providing so far untapped avenues for engineering increased efficiency in photosynthesis and control of electron transfer processes in biological systems. Another new function in *i*Cyt773 is L-Aspartate Oxidase. L-Aspartate Oxidase allows the deamination of aspartate, forming oxaloacetate a key TCA-cycle metabolite and ammonia. The impact of this addition to *i*Cyt773 is not evident under the photoautotrophic condition but becomes relevant for growth in a medium containing aspartate. *i*Cyt773 also uniquely supports the synthesis of pentadecane as documented by (Schirmer *et al.* 2010) and contains an (almost) complete non-fermentative citramalate pathway as suggested by (Wu *et al.* 2010).

A number of lumped reactions in *i*Cce806 were recast in detail. For example, pyruvate dehydrogenase (PDH) is a three-enzyme complex that carries out the biotransformation of pyruvate to acetyl-CoA in three steps using five separate cofactors (i.e., TPP, CoA, FAD, lipoate, and NAD). Similar detail was used for lumped steps in the metabolism of glycine, histidine, and serine. All additions to the list of reactions in *i*Cyt773 were corroborated using genome annotations (Welsh *et al.* 2008) or published literature (Collins *et al.* 1981; Min *et al.* 2010; Schirmer *et al.* 2010; Wu *et al.* 2010) with the exception of ten enzymes, whose function in the lipid and pigment biosynthesis pathways was required for biomass production.

A shift in biomass composition was observed under light, dark, and nitrate supplemented (light and dark) conditions. These differences were captured in four separate biomass descriptions present in *i*Cyt773. In addition, we used data from (Stockel *et al.* 2011) on the diurnal oscillations for approximately 20% of proteins in *Cyanothece* 51142 to identify

regulatory reaction shutdowns in our metabolic model. Supplementary File S4 lists the reactions that were inactivated under light and dark conditions, respectively. As expected, the nitrogenase genes *cce_0559* and *cce_0560*, known to be active in the absence of light, exhibited low spectral counts under light conditions. In contrast, photosystem II gene *cce_1526*, showed no spectral count under dark conditions. Unexpectedly, the data suggested that the Mehler reactions associated gene (*cce_2580*), known to be active in *Synechocystis* 6803 (Allahverdiyeva *et al.* 2011) and expected to be active in *Cyanothece* 51142, exhibited lower expression in light than in dark conditions.

iSyn731 and iCyt773 Models Comparison

Figure 3.3C illustrates the total number of common and unique reactions and metabolites between *i*Syn731 and *i*Cyt773 models. The *Cyanothece* 51142 genome (Welsh *et al.* 2008; Bandyopadhyay *et al.* 2011) is 1.5 times larger than that of *Synechocystis* 6803 (Kaneko *et al.* 1996), nevertheless *i*Cyt773 is smaller than *i*Syn731 due to differences in the level of detail of annotation and biochemical characterization. As many as 670 reactions and 596 metabolites are shared by both models corresponding to 47% and 63% of the total reactome and metabolome, respectively (see Figure 3.3C). The higher degree of conservation of metabolites (as opposed to reactions) across the two cyanobacteria suggests that lifestyle adaptations tend to usher new enzymatic activities that most of the time make use of the same metabolite pool without introducing new metabolites. There are 486 reactions that are unique to *i*Syn731 with no counterpart in *i*Cyt773. These reactions are not preferentially allotted to specific pathways. Instead they are spread over tens of different pathways. Primary metabolism reactions dispersed throughout fatty acid biosynthesis, lipid metabolism, oxidative phosphorylation, purine and pyrimidine metabolism, transport and exchange reactions account for 295 reactions. Secondary

metabolism including chlorophyll and cyanophycin metabolism, folate, terpenoid, phenylpropanoid and flavonoid biosynthesis accounts for the remaining 191 *i*Syn731-specific reactions. Interestingly, the 276 iCyt773-specific reactions span the same set of diverse pathways implying that the two organisms have adopted unique/divergent biosynthetic capabilities for similar metabolic needs. Fifty-eight span primary metabolism pathways such as purine and pyrimidine metabolism, fatty acid and lipid biosynthesis, amino acid biosynthesis. The remaining 218 reactions describe secondary metabolism such as terpenoid biosynthesis, chlorophyll and cyanophycin biosynthesis, plastoquinone and phyloquinone biosynthesis (see Supplementary File S6 for detail information). The much larger set of unique *i*Syn731-specific reactions compared to *i*Cyt773 reflect more complete genome annotation and biochemical characterization rather than augmented metabolic versatility.

A number of distinct differences in metabolism between the two organisms have been accounted for in the two models. For example, *i*Cyt773 does not have the enzyme threonine ammonia-lyase, which catalyzes the conversion of threonine to 2-ketobutyrate and as a consequence lacks the traditional route for isoleucine synthesis. Instead it employs part of the alternative citramalate pathway for isoleucine synthesis with pyruvate and acetyl-CoA as precursors. Follow up literature queries revealed the existence of this alternative pathway in *Cyanothece* 51142 (Wu *et al.* 2010). Ketobutyrate, an intermediate in the citramalate pathway, can be readily converted to higher alcohols, such as propanol and butanol, via a non-fermentative alcohol production pathway. Using the *i*Cyt773 model, we determined that only 2-ketoacid decarboxylase is missing from these three-step processes. In contrast, *i*Syn731 was found to have only the traditional route for isoleucine production with the citramalate pathway completely absent (see Figure 3.6A). In another example, the fermentative 1-butanol pathway is known to be

incomplete in both organisms. By querying the developed models we can pinpoint exactly which steps are absent. Specifically, the conversion between 3-hydroxybutanoyl-CoA and butanal is missing in both models. In addition to higher alcohols, higher alkanes (C13 and above) are important biofuel molecules as the main constituents of diesel and jet fuel (Schirmer *et al.* 2010). Recently reported (Schirmer *et al.* 2010) novel genes involved in the biosynthesis of alkanes in several cyanobacterial strains were incorporated in the models. Metabolic differences in *Cyanothece* 51142 and *Synechocystis* 6803 lead to the production of pentadecane in *Cyanothece* 51142 and heptadecane in *Synechocystis* 6803, presumably according to the selectivity of the respective fattyl acyl-ACP reductases (see Figure 3.6B).

Model *i*Cyt773, in contrast to *i*Syn731, does not have a complete urea cycle as it lacks the enzyme L-arginine aminohydrolase catalyzing the production of urea from L-arginine. Literature sources (Quintero *et al.* 2000; Bandyopadhyay *et al.* 2011) support this finding and explain the absence of a functional urea cycle as a consequence of the nitrogen-fixation ability of *Cyanothece* 51142 (Solomon *et al.* 2010; Tripp *et al.* 2010). Because *Cyanothece* 51142 can fix nitrogen directly from the atmosphere and produce ammonium via the enzyme nitrogenase, genes corresponding to the activity of L-arginine aminohydrolase and urease (for breaking down urea) become redundant. In addition to nitrogen metabolism, *i*Cyt773 and *i*Syn731 models reveal marked differences in anaerobic metabolic capabilities. Unlike *i*Syn731, *i*Cyt773 includes an L-lactate dehydrogenase activity that enables the complete fermentative lactate production pathway. On the other hand, *i*Syn731 contains the anaerobic chlorophyll biosynthetic pathway using enzyme protoporphyrin IX cyclase (BchE) that is absent in *i*Cyt773. Other differences in metabolism include lipid and fatty acid synthesis, fructose-6-phosphate shunt and nitrogen fixation. Model *i*Syn731 traces the location of the double bond for unsaturated fatty acid

synthesis pathways, as two separate isomers of unsaturated C_{18} fatty acids are part of the biomass description. *i*Cyt773 allows for the shunting of fructose-6-phosphate into erythrose-4-phosphate along with acetate and ATP using the fructose-6-phosphate phosphoketolase activity. Finally, both *i*Syn731 and *i*Cyt773 contain multiple hydrogenases allowing both to produce hydrogen. However, only the latter has a nitrogenase activity that can fix nitrogen while simultaneously producing hydrogen.

3.3.5. Using *i*Syn731 and *i*Cyt773 to Estimate Production Yields

We tested the recently developed models *i*Syn731 and *i*Cyt773 by comparing the predicted maximum theoretical product yields with experimentally measured values for two very different metabolic products: isoprene and hydrogen. Isoprene, a volatile hydrocarbon and potential feedstock for biofuel, is mostly produced in plants under heat stress (Lindberg *et al.* 2010). Cyanobacteria offer promising production alternatives as they can grow to high densities in bioreactors and produce isoprene directly from photosynthesis intermediates (Lindberg *et al.* 2010). *Synechocystis* 6803 has all but one gene (encoding isoprene synthase) in the methyl-erythritol-4-phosphate (MEP) pathway for isoprene synthesis from dimethylallyl phosphate (DMAPP). Upon cloning the isoprene synthase from kudzu vine (*Pueraria montana*) into *Synechocystis* 6803 isoprene production was demonstrated using sunlight and atmospheric CO_2 of 4.3×10^{-4} mole isoprene/mole carbon fixed (Connor *et al.* 2010). We calculated the maximum isoprene yield using *i*Syn731 to be 3.63×10^{-5} mole isoprene/ mole carbon fixed upon adding the isoprene synthase activity to the model and simulating the conditions described in (Bentley *et al.* 2012) under maximum biomass production. Similar isoprene yields were obtained with *i*JN678 (Nogales *et al.* 2012) while earlier models of *Synechocystis* 6803 (Fu 2009; Knoop *et al.* 2010; Montagud *et al.* 2010; Montagud *et al.* 2011) lack the MEP pathway (partially or completely)

and thus do not support isoprene production. The underestimation of the experimentally observed isoprene yield by the model predicted maximum yield may be due to sub-optimal growth of the production strain, differences in the list of measured biomass components, missing isoprene-relevant reactions from the model or more likely a combination of the above factors.

Both *Cyanothece* 51142 and *Synechocystis* 6803 produce hydrogen by utilizing nitrogenase and hydrogenase activities, respectively (Bandyopadhyay *et al.* 2010). Under subjective dark conditions (Bandyopadhyay *et al.* 2010) whereby (i) stored glycogen acts as a carbon source, (ii) photosynthesis harnesses light energy, and (iii) nitrogenase activity is not restricted, hydrogen production yield for *Cyanothece* 51142 was measured at 49.67 mole/mole glycogen consumed. Simulating the same conditions in *i*Cyt773 and *i*Cce806 (Vu *et al.* 2012) leads to maximum theoretical yields for hydrogen production of 48.43 mole/mole glycogen and 102.4 mole/mole glycogen, respectively. The entire amount of hydrogen produced in *i*Cyt773 is due to the nitrogenase activity. In contrast, the predicted doubling of the maximum hydrogen yield in *i*Cce806 is due to the utilization of the reverse direction of two hydrogen dehydrogenase reactions without any nitrogenase activity. Utilization of the nitrogenase reaction requires the use and recycling of more ATP than simply running the dehydrogenase reactions in reverse. However, it has been reported that hydrogen production in *Cyanothece* 51142 is primarily mediated by the nitrogenase enzyme (Bandyopadhyay *et al.* 2010) in the dark phase. This lends support to the irreversibility of the dehydrogenase reactions (under dark condition) as present in the *i*Cyt773 model. Experimental results for *Synechocystis* 6803 support up to 4.24 mole/mole glycogen consumed (Antal *et al.* 2005; Bandyopadhyay *et al.* 2010) of hydrogen production. *i*Syn731 predicts a maximum hydrogen theoretical yield of 2.28 mole/mole glycogen consumed while *i*JN678 yields a value of 2.00 mole/mole glycogen consumed. Again the factors outlined

for isoprene production may explain the lower theoretical yields predicted by the two models. The small difference between the model predicted yields is due to the presence of one step lumped biotransformation between isocitrate and oxoglutarate via isocitrate dehydrogenase in *i*JN678. *i*Syn731 describes this biotransformation in two steps (isocitrate → oxalosuccinate → oxoglutarate) (Muro-Pastor *et al.* 1996) generating an additional NADPH and subsequently more hydrogen via the hydrogenase reaction.

3.4. Conclusion

In this paper, we expanded upon existing models to develop two genome-scale metabolic models, *Synechocystis i*Syn731 and *Cyanothece i*Cyt773, for cyanobacterial metabolism by integrating all available knowledge available from public databases and published literature. All metabolite and reaction naming conventions are consistent between the two models allowing for direct comparisons. Systematic gap filling analyses led to the bridging of a number of network gaps in the two models and the elimination of orphan metabolites. Two separate biomass equations as well as two different versions of *Cyanothece i*Cyt773 models were developed for light and dark phases to represent diurnal regulation. The development of two separate models for *Cyanothece* 51142 (i.e., light and dark) provides the two "end-points" for the future development of dynamic metabolic models capturing the temporal evolution (Stoeckel *et al.* 2008; Colijn *et al.* 2009; Jensen *et al.* 2011; Stockel *et al.* 2011; Landry *et al.* 2013) of fluxes during the transition phases DFBA (Mahadevan *et al.* 2003). Comparisons against available ^{13}C MFA measurements for *Synechocystis* 6803 (Young *et al.* 2011) revealed that the *i*Syn731 model upon biomass maximization yields flux ranges that are generally consistent with experimental

data. Discrepancies between the two identify metabolic nodes where regulatory constraints are needed to recapitulate physiological behavior. The ability of *i*Syn731 to predict the fate of single gene knock-outs was further improved (specificity of 0.94 and sensitivity of 1.00) by reconciling *in silico* growth predictions with *in vivo* gene essentiality data (Nakamura *et al.* 1999). Similar analyses could also be carried out for *Cyanothece i*Cyt773 model once such flux measurements and *in vivo* gene essentiality data become available.

It is becoming widely accepted that focusing on a single pathway at a time without quantitatively assessing the system-wide implications of genetic manipulations may be responsible for suboptimal production levels of various biofuels and products. By accounting for both primary and some secondary metabolic pathways, the *Cyanothece i*Cyt773 model can be used to explore *in silico* the effect of genetic modifications aimed at increased production of useful biofuel molecules. By taking full inventory of *Cyanothece* 51142 metabolism (as abstracted in *i*Cyt773), and applying available strain optimization techniques (Kim *et al.* 2010; Ranganathan *et al.* 2010) optimal gene modifications could be pursued for a variety of targets in coordination with experimental techniques. In particular, the availability of a microaerobic environment in *Cyanothece* 51142 at certain times during the diurnal cycle can be exploited for the expression of novel pathways that are not usually found in oxygenic cyanobacterial strains that largely maintain an aerobic environment. However, the use of *Cyanothece* 51142 as a bio-production platform is currently hampered by the inability to efficiently carry out genetic modifications.

By systematically cataloguing the shared (and unique) metabolic content in *i*Syn731 and *i*Cyt773, successful genetic interventions assessed experimentally for *Synechocystis* 6803 can be "translated" to *Cyanothece* 51142. For example, it has been reported (Tan *et al.* 2011; Gao *et al.*

2012) that overproduction of fatty alcohols can be achieved in *Synechocystis* 6803 upon cloning a fatty acyl-CoA reductase (*far*) from Jojoba (*Simmondsia chinensis*) and the over-expression of gene *slr1609* coding for an acyl-ACP synthetase. By using models *i*Syn731 and *i*Cyt773 we can infer that in addition to cloning *far* from Jojoba, over-expression of gene *cce_1133* coding for a native acyl-ACP synthetase would be needed to bring about the same overproduction in *Cyanothece* 51142.

As *Synechcocystis* 6803 is also a popular host for metabolic engineering efforts, *i*Syn731 can also offer insights in how to improve production in this strain. In chapter 4 of this work, we describe attempts to use *i*Syn731 and the OptForce algorithm to enhance production of heptadecane.

3.5. Supplemental Data

The following materials are available in the online version of the original article on which this chapter is based (Saha *et al.* 2012):

1. **Supplemental File S1.** *Synechocystis* *i*Syn731 model along with established GPR, metabolite, gene and protein information.
2. **Supplemental File S2.** *Cyanothece* *i*Cyt773 model along with established GPR, metabolite, gene and protein information.
3. **Supplemental File S3.** Biomass component measurements and stoichiometry of biomass equation.
4. **Supplemental File S4.** Reactions with diurnal activation/inactivation.

5. **Supplemental File S5.** Comparison of *in silico* vs. *in vivo* gene essentiality results for *i*Syn731 and modifications made in GPR associations.
6. **Supplemental File S6.** Comparison between *Synechocystis i*Syn731 and *i*Cyt773 models in terms of genes, proteins, reactions and metabolites.
7. **Supplemental File S7.** SBML file of *Synechocystis i*Syn731 model.
8. **Supplemental File S8.** SBML file of *Cyanothece* 51142 *i*Cyt773 model.

3.6. References

Allahverdiyeva, Y., Ermakova, M., Eisenhut, M., Zhang, P., Richaud, P., Hagemann, M., Cournac, L. and Aro, E. M. (2011). "Interplay between flavodiiron proteins and photorespiration in Synechocystis sp. PCC 6803." *J Biol Chem* **286**(27): 24007-24014.

Allen, M. M. (1968). "Simple Conditions for Growth of Unicellular Blue-Green Algae on Plates." *Journal of Phycology* **4**(1): 1-&.

Antal, T. K. and Lindblad, P. (2005). "Production of H2 by sulphur-deprived cells of the unicellular cyanobacteria Gloeocapsa alpicola and Synechocystis sp. PCC 6803 during dark incubation with methane or at various extracellular pH." *J Appl Microbiol* **98**(1): 114-120.

Arnon, D. I., Mcswain, B. D., Tsujimot.Hy and Wada, K. (1974). "Photochemical Activity and Components of Membrane Preparations from Blue-Green-Algae .1. Coexistence of 2 Photosystems in Relation to Chlorophyll Alpha and Removal of Phycocyanin." *Biochimica Et Biophysica Acta* **357**(2): 231-245.

Badger, M. R. and Price, G. D. (2003). "CO2 concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution." *J Exp Bot* **54**(383): 609-622.

Bandyopadhyay, A., Elvitigala, T., Welsh, E., Stockel, J., Liberton, M., Min, H., Sherman, L. A. and Pakrasi, H. B. (2011). "Novel metabolic attributes of the genus cyanothece, comprising a group of unicellular nitrogen-fixing Cyanothece." *Mbio* **2**(5).

Bandyopadhyay, A., Stockel, J., Min, H., Sherman, L. A. and Pakrasi, H. B. (2010). "High rates of photobiological H2 production by a cyanobacterium under aerobic conditions." *Nat Commun* **1**: 139.

Bentley, F. K. and Melis, A. (2012). "Diffusion-based process for carbon dioxide uptake and isoprene emission in gaseous/aqueous two-phase photobioreactors by photosynthetic microorganisms." *Biotechnol Bioeng* **109**(1): 100-109.

Bryant, D. A. and Frigaard, N. U. (2006). "Prokaryotic photosynthesis and phototrophy illuminated." *Trends Microbiol* **14**(11): 488-496.

Burnap, R. L. and Sherman, L. A. (1991). "Deletion Mutagenesis in Synechocystis Sp Pcc6803 Indicates That the Mn-Stabilizing Protein of Photosystem-Ii Is Not Essential for O2 Evolution." *Biochemistry* **30**(2): 440-446.

Cheng, Z., Sattler, S., Maeda, H., Sakuragi, Y., Bryant, D. A. and DellaPenna, D. (2003). "Highly divergent methyltransferases catalyze a conserved reaction in tocopherol and plastoquinone synthesis in cyanobacteria and photosynthetic eukaryotes." *Plant Cell* **15**(10): 2343-2356.

Chitnis, P. R., Reilly, P. A., Miedel, M. C. and Nelson, N. (1989). "Structure and Targeted Mutagenesis of the Gene Encoding 8-Kda Subunit of Photosystem-I from the Cyanobacterium Synechocystis Sp Pcc-6803." *Journal of Biological Chemistry* **264**(31): 18374-18380.

Chitnis, P. R., Reilly, P. A. and Nelson, N. (1989). "Insertional Inactivation of the Gene Encoding Subunit-Ii of Photosystem-I from the Cyanobacterium Synechocystis Sp Pcc-6803." *Journal of Biological Chemistry* **264**(31): 18381-18385.

Colijn, C., Brandes, A., Zucker, J., Lun, D. S., Weiner, B., Farhat, M. R., Cheng, T. Y., Moody, D. B., Murray, M. and Galagan, J. E. (2009). "Interpreting Expression Data with

Metabolic Flux Models: Predicting Mycobacterium tuberculosis Mycolic Acid Production." *Plos Computational Biology* **5**(8).
Collins, M. D. and Jones, D. (1981). "Distribution of isoprenoid quinone structural types in bacteria and their taxonomic implication." *Microbiol Rev* **45**(2): 316-354.
Connor, M. R. and Atsumi, S. (2010). "Synthetic Biology Guides Biofuel Production." *Journal of Biomedicine and Biotechnology* **2010**(541698).
Dahnhardt, D., Falk, J., Appel, J., van der Kooij, T. A., Schulz-Friedrich, R. and Krupinska, K. (2002). "The hydroxyphenylpyruvate dioxygenase from Synechocystis sp. PCC 6803 is not required for plastoquinone biosynthesis." *FEBS Lett* **523**(1-3): 177-181.
Delorimier, R., Bryant, D. A. and Stevens, S. E. (1990). "Genetic-Analysis of a 9 Kda Phycocyanin-Associated Linker Polypeptide." *Biochimica Et Biophysica Acta* **1019**(1): 29-41.
Ducat, D. C., Way, J. C. and Silver, P. A. (2011). "Engineering cyanobacteria to generate high-value products." *Trends Biotechnol* **29**(2): 95-103.
Fernandez-Gonzalez, B., Sandmann, G. and Vioque, A. (1997). "A new type of asymmetrically acting beta-carotene ketolase is required for the synthesis of echinenone in the cyanobacterium Synechocystis sp. PCC 6803." *J Biol Chem* **272**(15): 9728-9733.
Fu, P. C. (2009). "Genome-scale modeling of Synechocystis sp PCC 6803 and prediction of pathway insertion." *Journal of Chemical Technology and Biotechnology* **84**(4): 473-483.
Gao, Q., Wang, W., Zhao, H. and Lu, X. (2012). "Effects of fatty acid activation on photosynthetic production of fatty acid-based biofuels in Synechocystis sp. PCC6803." *Biotechnol Biofuels* **5**(1): 17.
Glazer, A. N. (1977). "Structure and molecular organization of the photosynthetic accessory pigments of cyanobacteria and red algae." *Mol Cell Biochem* **18**(2-3): 125-140.
Hong, S. J. and Lee, C. G. (2007). "Evaluation of central metabolism based on a genomic database of Synechocystis PCC6803." *Biotechnology and Bioprocess Engineering* **12**(2): 165-173.
Jallet, D., Gwizdala, M. and Kirilovsky, D. (2012). "ApcD, ApcF and ApcE are not required for the Orange Carotenoid Protein related phycobilisome fluorescence quenching in the cyanobacterium Synechocystis PCC 6803." *Biochim Biophys Acta* **1817**(8): 1418-1427.
Jansson, C., Debus, R. J., Osiewacz, H. D., Gurevitz, M. and McIntosh, L. (1987). "Construction of an Obligate Photoheterotrophic Mutant of the Cyanobacterium Synechocystis 6803 : Inactivation of the psbA Gene Family." *Plant Physiol* **85**(4): 1021-1025.
Jensen, P. A., Lutz, K. A. and Papin, J. A. (2011). "TIGER: Toolbox for integrating genome-scale metabolic models, expression data, and transcriptional regulatory networks." *BMC Systems Biology* **5**(147).
Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M. and Tabata, S. (1996). "Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions." *DNA Res* **3**(3): 109-136.
Kim, J. and Reed, J. L. (2010). "OptORF: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains." *BMC Systems Biology* **4**(53).

Knoop, H., Grundel, M., Zilliges, Y., Lehmann, R., Hoffmann, S., Lockau, W. and Steuer, R. (2013). "Flux balance analysis of cyanobacterial metabolism: the metabolic network of Synechocystis sp. PCC 6803." *PLoS Comput Biol* **9**(6): e1003081.

Knoop, H., Zilliges, Y., Lockau, W. and Steuer, R. (2010). "The Metabolic Network of Synechocystis sp. PCC 6803: Systemic Properties of Autotrophic Growth." *Plant Physiology* **154**(1): 410-422.

Kucho, K., Okamoto, K., Tsuchiya, Y., Nomura, S., Nango, M., Kanehisa, M. and Ishiura, M. (2005). "Global analysis of circadian expression in the cyanobacterium Synechocystis sp. strain PCC 6803." *J Bacteriol* **187**(6): 2190-2199.

Kumar, V. S., Ferry, J. G. and Maranas, C. D. (2011). "Metabolic reconstruction of the archaeon methanogen Methanosarcina Acetivorans." *Bmc Systems Biology* **5**(28).

Kumar, V. S. and Maranas, C. D. (2009). "GrowMatch: an automated method for reconciling in silico/in vivo growth predictions." *PLoS Comput Biol* **5**(3): e1000308.

Landry, B. P., Stockel, J. and Pakrasi, H. B. (2013). "Use of degradation tags to control protein levels in the Cyanobacterium Synechocystis sp. Strain PCC 6803." *Appl Environ Microbiol* **79**(8): 2833-2835.

Lichtenthaler, H. K. (1987). "Chlorophylls and Carotenoids - Pigments of Photosynthetic Biomembranes." *Methods in Enzymology* **148**: 350-382.

Lindberg, P., Park, S. and Melis, A. (2010). "Engineering a platform for photosynthetic isoprene production in cyanobacteria, using Synechocystis as the model organism." *Metabolic Engineering* **12**(1): 70-79.

Liu, X. Y. and Curtiss, R. (2009). "Nickel-inducible lysis system in Synechocystis sp PCC 6803." *Proceedings of the National Academy of Sciences of the United States of America* **106**(51): 21550-21554.

Mahadevan, R., Edwards, J. and Doyle, F. (2003). "Dynamic Flux Analysis of diauxic growth in *Escherichia coli*." *Biophysical Journal* **83**: 1331-1340.

Manna, P. and Vermaas, W. (1997). "Lumenal proteins involved in respiratory electron transport in the cyanobacterium Synechocystis sp. PCC6803." *Plant Molecular Biology* **35**(4): 407-416.

McHugh, K. (2005). *Hydrogen production methods*. Virginia, MPR Associates, Inc.

Min, H. and Sherman, L. A. (2010). "Hydrogen production by the unicellular, diazotrophic cyanobacterium Cyanothece sp. strain ATCC 51142 under conditions of continuous light." *Appl Environ Microbiol* **76**(13): 4293-4301.

Minamizaki, K., Mizoguchi, T., Goto, T., Tamiaki, H. and Fujita, Y. (2008). "Identification of two homologous genes, chlAI and chlAII, that are differentially involved in isocyclic ring formation of chlorophyll a in the cyanobacterium Synechocystis sp. PCC 6803." *J Biol Chem* **283**(5): 2684-2692.

Mo, M. L., Palsson, B. O. and Herrgard, M. J. (2009). "Connecting extracellular metabolomic measurements to intracellular flux states in yeast." *BMC Syst Biol* **3**: 37.

Moisander, P. H., Beinart, R. A., Hewson, I., White, A. E., Johnson, K. S., Carlson, C. A., Montoya, J. P. and Zehr, J. P. (2010). "Unicellular cyanobacterial distributions broaden the oceanic N2 fixation domain." *Science* **327**(5972): 1512-1514.

Montagud, A., Navarro, E., de Cordoba, P. F., Urchueguia, J. F. and Patil, K. R. (2010). "Reconstruction and analysis of genome-scale metabolic model of a photosynthetic bacterium." *Bmc Systems Biology* **4**: -.

Montagud, A., Zelezniak, A., Navarro, E., de Cordoba, P., Urchueguia, J. F. and Patil, K. R. (2011). "Flux coupling and transcriptional regulation within the metabolic network of the photosynthetic bacterium Synechocystis sp PCC6803." *Biotechnology Journal* **6**(3): 330-342.

Mortazavi, A. W., BA; McCue, K; Schaeffer, L; Wold, B (2008). "Mapping and quantifying mammalian transcripts by RNA-Seq." *Nature Methods* **5**(7): 621-628.

Muro-Pastor, M. I., Reyes, J. C. and Florencio, F. J. (1996). "The NADP+-isocitrate dehydrogenase gene (icd) is nitrogen regulated in cyanobacteria." *J Bacteriol* **178**(14): 4070-4076.

Nakamoto, H. (1995). "Targeted inactivation of the gene psaI encoding a subunit of photosystem I of the cyanobacterium Synechocystis sp PCC 6803." *Plant and Cell Physiology* **36**(8): 1579-1587.

Nakamura, Y., Kaneko, T., Miyajima, N. and Tabata, S. (1999). "Extension of CyanoBase. CyanoMutants: repository of mutant information on Synechocystis sp. strain PCC6803." *Nucleic Acids Res* **27**(1): 66-68.

Nakao, M., Okamoto, S., Kohara, M., Fujishiro, T., Fujisawa, T., Sato, S., Tabata, S., Kaneko, T. and Nakamura, Y. (2010). "CyanoBase: the cyanobacteria genome database update 2010." *Nucleic Acids Res* **38**(Database issue): D379-381.

Navarro, E., Montagud, A., de Cordoba, P. F. and Urchueguia, J. F. (2009). "Metabolic flux analysis of the hydrogen production potential in Synechocystis sp PCC6803." *International Journal of Hydrogen Energy* **34**(21): 8828-8838.

Nogales, J., Gudmundsson, S., Knight, E. M., Palsson, B. O. and Thiele, I. (2012). "Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis." *Proc Natl Acad Sci U S A* **109**(7): 2678-2683.

Ogawa, T., Marco, E. and Orus, M. I. (1994). "A gene (ccmA) required for carboxysome formation in the cyanobacterium Synechocystis sp. strain PCC6803." *J Bacteriol* **176**(8): 2374-2378.

Paerl, H. W. (1984). "Cyanobacterial Carotenoids - Their Roles in Maintaining Optimal Photosynthetic Production among Aquatic Bloom Forming Genera." *Oecologia* **61**(2): 143-149.

Papadopoulos, J. S. and Agarwala, R. (2007). "COBALT: constraint-based alignment tool for multiple protein sequences." *Bioinformatics* **23**(9): 1073-1079.

Popa, R., Weber, P. K., Pett-Ridge, J., Finzi, J. A., Fallon, S. J., Hutcheon, I. D., Nealson, K. H. and Capone, D. G. (2007). "Carbon and nitrogen fixation and metabolite exchange in and between individual cells of Anabaena oscillarioides." *Isme Journal* **1**(4): 354-360.

Porra, R. J., Thompson, W. A. and Kriedemann, P. E. (1989). "Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls a and b extracted with four different solvents: verification of the concentration of chlorophyll standards by atomic absorption spectroscopy." *Biochim Biophys Acta* **975**: 384-394.

Poutanen, E. L. and Nikkila, K. (2001). "Carotenoid pigments as tracers of cyanobacterial blooms in recent and postglacial sediments of the Baltic Sea." *Ambio* **30**(4-5): 179-183.

Puchalka, J., Oberhardt, M. A., Godinho, M., Bielecka, A., Regenhardt, D., Timmis, K. N., Papin, J. A. and Martins dos Santos, V. A. (2008). "Genome-scale reconstruction and analysis of the Pseudomonas putida KT2440 metabolic network facilitates applications in biotechnology." *PLoS Comput Biol* **4**(10): e1000210.

Quintero, M. J., Muro-Pastor, A. M., Herrero, A. and Flores, E. (2000). "Arginine catabolism in the cyanobacterium Synechocystis sp. Strain PCC 6803 involves the urea cycle and arginase pathway." *J Bacteriol* **182**(4): 1008-1015.

Ranganathan, S., Suthers, P. F. and Maranas, C. D. (2010). "OptForce: An Optimization Procedure for Identifying All Genetic Manipulations Leading to Targeted Overproductions." *Plos Computational Biology* **6**(4).

Reddy, K. J., Haskell, J. B., Sherman, D. M. and Sherman, L. A. (1993). "Unicellular, Aerobic Nitrogen-Fixing Cyanobacteria of the Genus Cyanothece." *Journal of Bacteriology* **175**(5): 1284-1292.

Reddy, K. J., Haskell, J. B., Sherman, D. M. and Sherman, L. A. (1993). "Unicellular, aerobic nitrogen-fixing cyanobacteria of the genus Cyanothece." *J Bacteriol* **175**(5): 1284-1292.

Reed, J. L., Patel, T. R., Chen, K. H., Joyce, A. R., Applebee, M. K., Herring, C. D., Bui, O. T., Knight, E. M., Fong, S. S. and Palsson, B. O. (2006). "Systems approach to refining genome annotation." *Proc Natl Acad Sci U S A* **103**(46): 17480-17484.

Saha, R., Verseput, A. T., Berla, B. M., Mueller, T. J., Pakrasi, H. B. and Maranas, C. D. (2012). "Reconstruction and comparison of the metabolic potential of cyanobacteria Cyanothece sp. ATCC 51142 and Synechocystis sp. PCC 6803." *PLoS One* **7**(10): e48285.

Sakuragi, Y., Zybailov, B., Shen, G., Jones, A. D., Chitnis, P. R., van der Est, A., Bittl, R., Zech, S., Stehlik, D., Golbeck, J. H. and Bryant, D. A. (2002). "Insertional inactivation of the menG gene, encoding 2-phytyl-1,4-naphthoquinone methyltransferase of Synechocystis sp. PCC 6803, results in the incorporation of 2-phytyl-1,4-naphthoquinone into the A(1) site and alteration of the equilibrium constant between A(1) and F(X) in photosystem I." *Biochemistry* **41**(1): 394-405.

Satish Kumar, V., Dasika, M. S. and Maranas, C. D. (2007). "Optimization based automated curation of metabolic reconstructions." *BMC Bioinformatics* **8**: 212.

Schirmer, A., Rude, M. A., Li, X. Z., Popova, E. and del Cardayre, S. B. (2010). "Microbial Biosynthesis of Alkanes." *Science* **329**(5991): 559-562.

Shastri, A. A. and Morgan, J. A. (2005). "Flux balance analysis of photoautotrophic metabolism." *Biotechnology Progress* **21**(6): 1617-1626.

Shen, G. Z., Boussiba, S. and Vermaas, W. F. J. (1993). "Synechocystis Sp Pcc-6803 Strains Lacking Photosystem-I and Phycobilisome Function." *Plant Cell* **5**(12): 1853-1863.

Shen, J. R., Ikeuchi, M. and Inoue, Y. (1997). "Analysis of the psbU gene encoding the 12-kDa extrinsic protein of photosystem II and studies on its role by deletion mutagenesis in Synechocystis sp. PCC 6803." *Journal of Biological Chemistry* **272**(28): 17821-17826.

Shen, J. R., Qian, M., Inoue, Y. and Burnap, R. L. (1998). "Functional characterization of Synechocystis sp. PCC 6803 Delta psbU and Delta psbV mutants reveals important roles of cytochrome c-550 in cyanobacterial oxygen evolution." *Biochemistry* **37**(6): 1551-1558.

Shen, J. R., Vermaas, W. and Inoue, Y. (1995). "The Role of Cytochrome C-550 as Studied through Reverse Genetics and Mutant Characterization in Synechocystis Sp Pcc-6803." *Journal of Biological Chemistry* **270**(12): 6901-6907.

Solomon, C. M., Collier, J. L., Berg, G. M. and Glibert, P. M. (2010). "Role of urea in microbial metabolism in aquatic systems: a biochemical and molecular review." *Aquatic Microbial Ecology* **59**(1): 67-88.

Steiger, S., Schafer, L. and Sandmann, G. (1999). "High-light-dependent upregulation of carotenoids and their antioxidative properties in the cyanobacterium Synechocystis PCC 6803." *Journal of Photochemistry and Photobiology B-Biology* **52**(1-3): 14-18.

Stockel, J., Jacobs, J. M., Elvitigala, T. R., Liberton, M., Welsh, E. A., Polpitiya, A. D., Gritsenko, M. A., Nicora, C. D., Koppenaal, D. W., Smith, R. D. and Pakrasi, H. B. (2011). "Diurnal rhythms result in significant changes in the cellular protein complement in the cyanobacterium Cyanothece 51142." *PLoS One* **6**(2): e16680.

Stoeckel, J., Welsh, E. A., Liberton, M., Kunnvakkam, R., Aurora, R. and Pakrasi, H. B. (2008). "Global transcriptomic analysis of Cyanothece 51142 reveals robust diurnal oscillation of central metabolic processes." *Proceedings of the National Academy of Sciences of the United States of America* **105**(16): 6156-6161.

Taiz, L. and Zeiger, E. (2002). *Plant Physilogy*. Massachusetts, Sinauer Associates, Inc., Publishers.

Tan, X., Yao, L., Gao, Q., Wang, W., Qi, F. and Lu, X. (2011). "Photosynthesis driven conversion of carbon dioxide to fatty alcohols and hydrocarbons in cyanobacteria." *Metab Eng* **13**(2): 169-176.

Tottey, S., Rich, P. R., Rondet, S. A. and Robinson, N. J. (2001). "Two Menkes-type atpases supply copper for photosynthesis in Synechocystis PCC 6803." *J Biol Chem* **276**(23): 19999-20004.

Tottey, S., Rondet, S. A., Borrelly, G. P., Robinson, P. J., Rich, P. R. and Robinson, N. J. (2002). "A copper metallochaperone for photosynthesis and respiration reveals metal-specific targets, interaction with an importer, and alternative sites for copper acquisition." *J Biol Chem* **277**(7): 5490-5497.

Tredici, M. R., Margheri, M. C., Philippis, R. D., Materass, R., Bocci, F. and Tomaselli (1986). "Conversion of solar energy into the energy of biomass by culture of marine cyanobacteria." *Proceedings of the 1986 International Congress on Renewable Energy Sources* **1**: 191-199.

Tripp, H. J., Bench, S. R., Turk, K. A., Foster, R. A., Desany, B. A., Niazi, F., Affourtit, J. P. and Zehr, J. P. (2010). "Metabolic streamlining in an open-ocean nitrogen-fixing cyanobacterium." *Nature* **464**(7285): 90-94.

Turner, J., Sverdrup, G., Mann, M. K., Maness, P. C., Kroposki, B., Ghirardi, M., Evans, R. J. and Blake, D. (2008). "Renewable hydrogen production." *International Journal of Energy Research* **32**(5): 379-407.

Ughy, B. and Ajlani, G. (2004). "Phycobilisome rod mutants in Synechocystis sp strain PCC6803." *Microbiology-Sgm* **150**: 4147-4156.

Varma, A. and Palsson, B. O. (1994). "Metabolic Flux Balancing - Basic Concepts, Scientific and Practical Use." *Bio-Technology* **12**(10): 994-998.

Vu, T. T., Stolyar, S. M., Pinchuk, G. E., Hill, E. A., Kucek, L. A., Brown, R. N., Lipton, M. S., Osterman, A., Fredrickson, J. K., Konopka, A. E., Beliaev, A. S. and Reed, J. L. (2012). "Genome-scale modeling of light-driven reductant partitioning and carbon fluxes in diazotrophic unicellular cyanobacterium Cyanothece sp. ATCC 51142." *PLoS Comput Biol* **8**(4): e1002460.

Welsh, E. A., Liberton, M., Stockel, J., Loh, T., Elvitigala, T., Wang, C., Wollam, A., Fulton, R. S., Clifton, S. W., Jacobs, J. M., Aurora, R., Ghosh, B. K., Sherman, L. A., Smith, R. D., Wilson, R. K. and Pakrasi, H. B. (2008). "The genome of Cyanothece 51142, a

unicellular diazotrophic cyanobacterium important in the marine nitrogen cycle." *Proc Natl Acad Sci U S A* **105**(39): 15094-15099.

Wu, B., Zhang, B. C., Feng, X. Y., Rubens, J. R., Huang, R., Hicks, L. M., Pakrasi, H. B. and Tang, Y. J. J. (2010). "Alternative isoleucine synthesis pathway in cyanobacterial species." *Microbiology-Sgm* **156**: 596-602.

Wu, G. F., Shen, Z. Y. and Wu, Q. Y. (2001). "Possibility to improve the cyanobacterial poly-beta-hydroxybutyrate biosynthesis level." *Journal of Chemical Engineering of Japan* **34**(9): 1187-1190.

Yang, C., Hua, Q. and Shimizu, K. (2002). "Metabolic flux analysis in Synechocystis using isotope distribution from 13C-labeled glucose." *Metab Eng* **4**(3): 202-216.

Yeates, T. O., Kerfeld, C. A., Heinhorst, S., Cannon, G. C. and Shively, J. M. (2008). "Protein-based organelles in bacteria: carboxysomes and related microcompartments." *Nat Rev Microbiol* **6**(9): 681-691.

You, L., Berla, B., He, L., Pakrasi, H. B. and Tang, Y. J. (2014). "13C-MFA delineates the photomixotrophic metabolism of Synechocystis sp. PCC 6803 under light- and carbon-sufficient conditions." *Biotechnol J* **9**(5): 684-692.

Young, J. D., Shastri, A. A., Stephanopoulos, G. and Morgan, J. A. (2011). "Mapping photoautotrophic metabolism with isotopically nonstationary (13)C flux analysis." *Metabolic Engineering* **13**(6): 656-665.

Zahalak, M., Pratte, B., Werth, K. J. and Thiel, T. (2004). "Molybdate transport and its effect on nitrogen utilization in the cyanobacterium Anabaena variabilis ATCC 29413." *Molecular Microbiology* **51**(2): 539-549.

Zehr, J. P., Church, M.J., and Moisander, P.H. (2005). "Diversity, distribution and biogeochemical significance of nitrogen-fixing microorganisms in anoxic and suboxic ocean environments." *NATO Series book on past and present water column anoxia*, Springer**:** 337-369.

Zhang, S. Y. and Bryant, D. A. (2011). "The Tricarboxylic Acid Cycle in Cyanobacteria." *Science* **334**(6062): 1551-1553.

Table 3.1: *Synechocystis* 6803 *i*Syn731 and *Cyanothece* 51142 *i*Cyt773 model statistics

	***Synechocystis 6803* *i*Syn731 model**	***Cyanothece 51142* *i*Cyt773 model**
Included genes	731	773
Proteins	511	465
Single functional proteins	348	336
Multifunctional proteins	91	83
Isozymes	4	1
Multimeric proteins	32	22
Others[a]	36	23
Reactions	1,156	946
Metabolic reactions	972	761
Transport reactions	127	128
GPR associations		
Gene associated (metabolic/transport)	827	686
Spontaneous[b]	180	158
Nongene associated (metabolic/transport)	59	16
No protein associated	90	86
Exchange reactions	57	57
Metabolites[c]	996	811
Cytosolic	862	675
Carboxisomic	8	8
Thylakoidic	10	9
Periplasmic	59	62
Extracellular	57	57

[a] - Others include proteins involve in complex relationships, e.g. multiple proteins act as protein complex which is one of the isozymes for any specific reaction.

[b] - Spontaneous reactions are those without any enzyme as well as gene association.

[c] - Metabolites represent total number of metabolites with considering their compartmental specificity.

Table 3.2: Summary of connectivity restoration in *Synechocystis* 6803 *i*Syn731 and *Cyanothece* 51142 *i*Cyt773 models

	***Synechocystis 6803 i*Syn731**	***Cyanothece 51142 i*Cyt773**
Number of blocked metabolites	207	74
Number of metabolites with GapFill (Satish Kumar *et al.* 2007) suggested reconnection strategies	138	52
Number of metabolites whose reconnection forms a cycle	88	12
Number of metabolites with validated reconnection mechanisms	5	30
Number of added reactions to the model	10	19

Table 3.3: ^{13}C MFA flux measurements (Young *et al*. 2011) vs. model flux predictions.

Reaction	Flux measurements by Young *et al.*, (2011)		Flux ranges predicted by *i*JN678 model *(With max biomass)*		Flux ranges predicted by *i*Syn731 model *(With max biomass)*	
	95% LB	95% UB	LB	UB	LB	UB
RBC	123.00	132.00	109.02	109.10	102.49	106.33
PGK	219.00	237.00	187.11	187.25	182.70	182.92
13PDG	219.00	237.00	187.11	196.36	182.70	201.96
GAPDH	90.00	99.00	74.98	75.07	73.40	73.50
FBA	53.00	66.00	-0.17	74.85	-0.08	73.17
FBP	53.00	66.00	0.00	74.85	0.00	73.17
PGI	15.00	24.00	0.68	0.73	0.82	0.84
G6PD	12.00	21.00	0.00	0.05	0.00	0.03
6PGL	12.00	21.00	0.00	0.05	0.00	0.03
6PGD	12.00	21.00	0.00	0.05	0.00	0.03
PRK	123.00	132.00	109.02	109.10	106.24	106.32
SBGPL	29.00	43.00	-0.17	74.85	-0.08	73.17
SBP	29.00	43.00	0.00	74.85	0.00	73.17
TAL	-6.00	9.00	-36.74	38.28	-35.93	37.32
TKT1	37.20	37.50	36.57	36.60	36.66	36.79
RPI	35.40	35.70	35.18	35.21	35.82	35.86
TKT2	35.40	35.70	37.25	37.28	36.18	36.23
RPE	75.50	76.20	73.83	73.88	72.01	72.10
PGM	22.90	23.60	26.83	26.95	25.92	29.79
ENO	23.40	23.80	26.84	26.95	25.92	29.79
PYK	7.90	11.10	0.00	13.88	0.00	16.72
PDH	11.50	12.00	0.00	8.97	0.00	13.46
CS	3.00	3.40	2.15	2.21	1.35	1.37
ACONT	3.00	3.40	2.15	2.21	1.35	1.37
ICD	3.00	3.00	2.15	2.21	1.32	1.37
SUCD	0.00	0.40	0.00	0.00	0.00	0.00
FUM	1.70	2.00	-5.44	1.55	-7.26	1.49
MDH	1.90	5.20	5.35	5.61	7.15	7.32
ME1	3.70	6.90	0.00	0.17	0.00	0.08
ME2	3.70	6.90	-	-	0.00	0.08
PPC	9.90	13.20	11.74	11.98	12.25	12.37

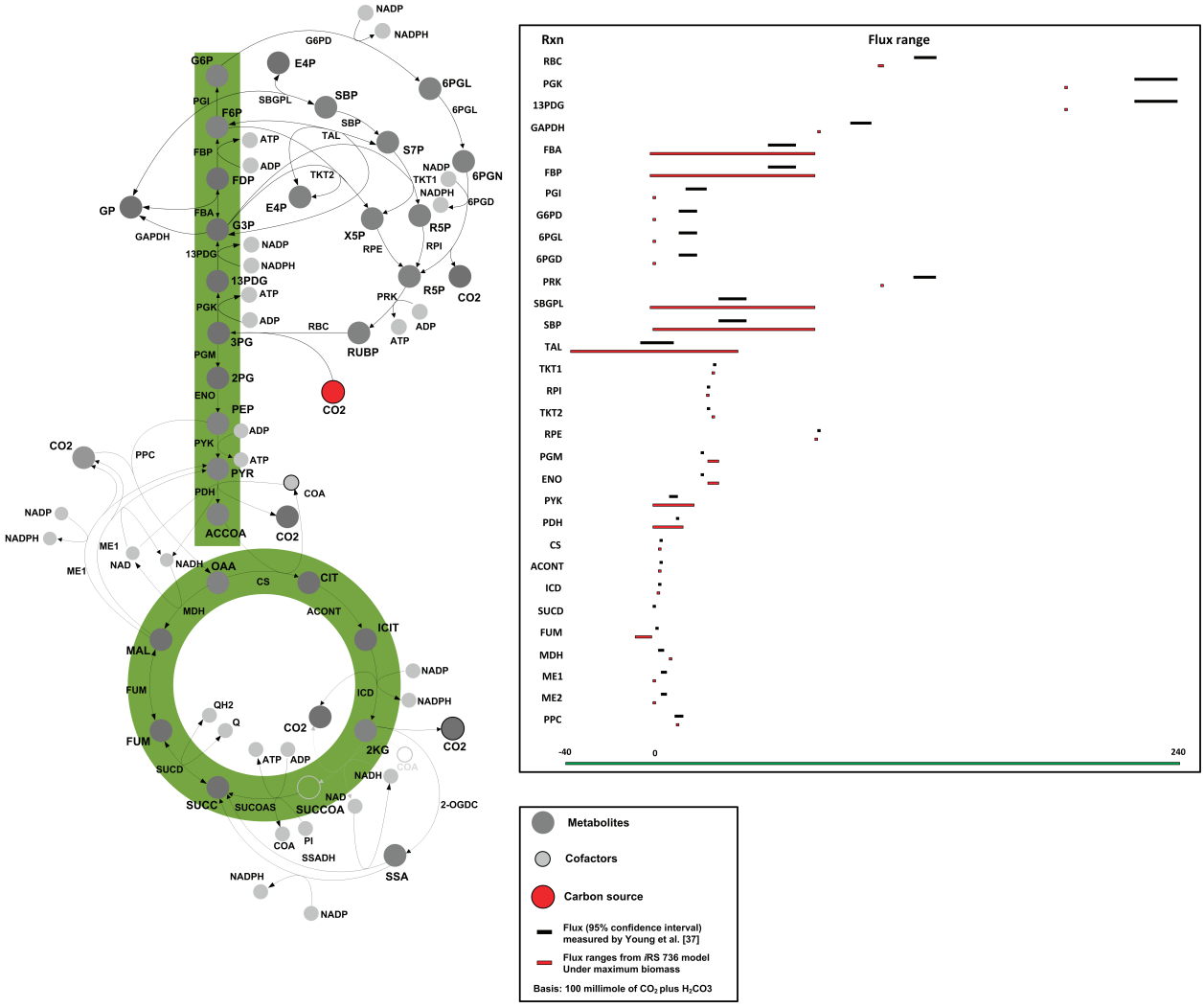


Figure 3.1: Comparison of model derived and experimentally measured flux ranges for *Synechocystis* sp. PCC 6803 under the maximum biomass condition. The basis is 100 millimole of CO_2 plus H_2CO_3. Experimental measurements are from Young *et al*. (2011)

A)

Knoop *et al.* [12]

In Silico Results	*In Vivo* Results	
	Growth	No Growth
Growth	32/36	2/7
No Growth	4/36	5/7

***i*JN678 [32]**

In Silico Results	*In Vivo* Results	
	Growth	No Growth
Growth	69/95	1/19
No Growth	26/95	18/19

***i*Syn731**

In Silico Results	*In Vivo* Results	
	Growth	No Growth
Growth	94/100	0/19
No Growth	6/100	19/19

B)

Reconstruction	Knoop *et al.* [12]	*i*JN678 [32]	*i*Syn731 [this study]
Specificity	0.88	0.73	0.94
Sensitivity	0.71	0.95	1.00

$$\text{Specificity} = \frac{GG}{GG + NGG}$$

$$\text{Sensitivity} = \frac{NGNG}{NGNG + GNG}$$

Figure 3.2: Comparison of gene essentiality/viability data with predictions by a number of *Synechocystis* 6803 models. (A) Tabulated growth (G) or non-growth (NG) predictions and experimental data. The first number denotes the number of GG, GNG, NGG and NGNG combinations whereas the second number signifies the number of experimentally observed lethal or viable mutants. (B) Specificity and sensitivity of all three models. Note that GG denotes both *in silico* and *in vivo* growth, NGG represents no growth *in silico* but *in vivo* growth. NGNG implies no growth for either *in silico* or *in vivo*, whereas GNG marks growth *in silico* but no growth *in vivo*.

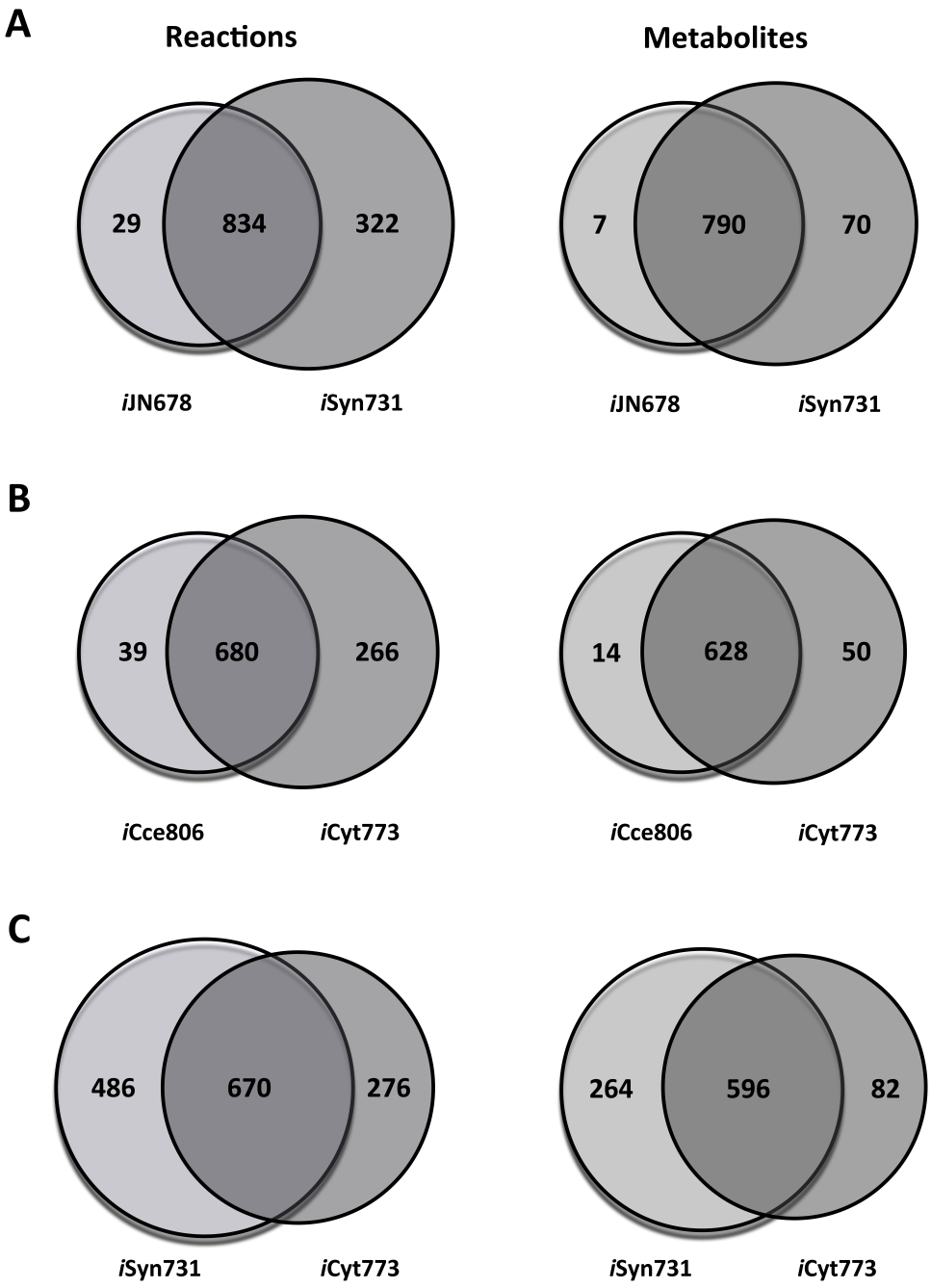


Figure 3.3: Venn diagram depicting (common and unique) reactions and metabolites between models. (A) *i*JN678 (Nogales *et al*. 2012) and *i*Syn731, (B) *i*Cce806 (Vu *et al*. 2012) and *i*Cyt773, and (C) *i*Syn731 and *i*Cyt773 models.

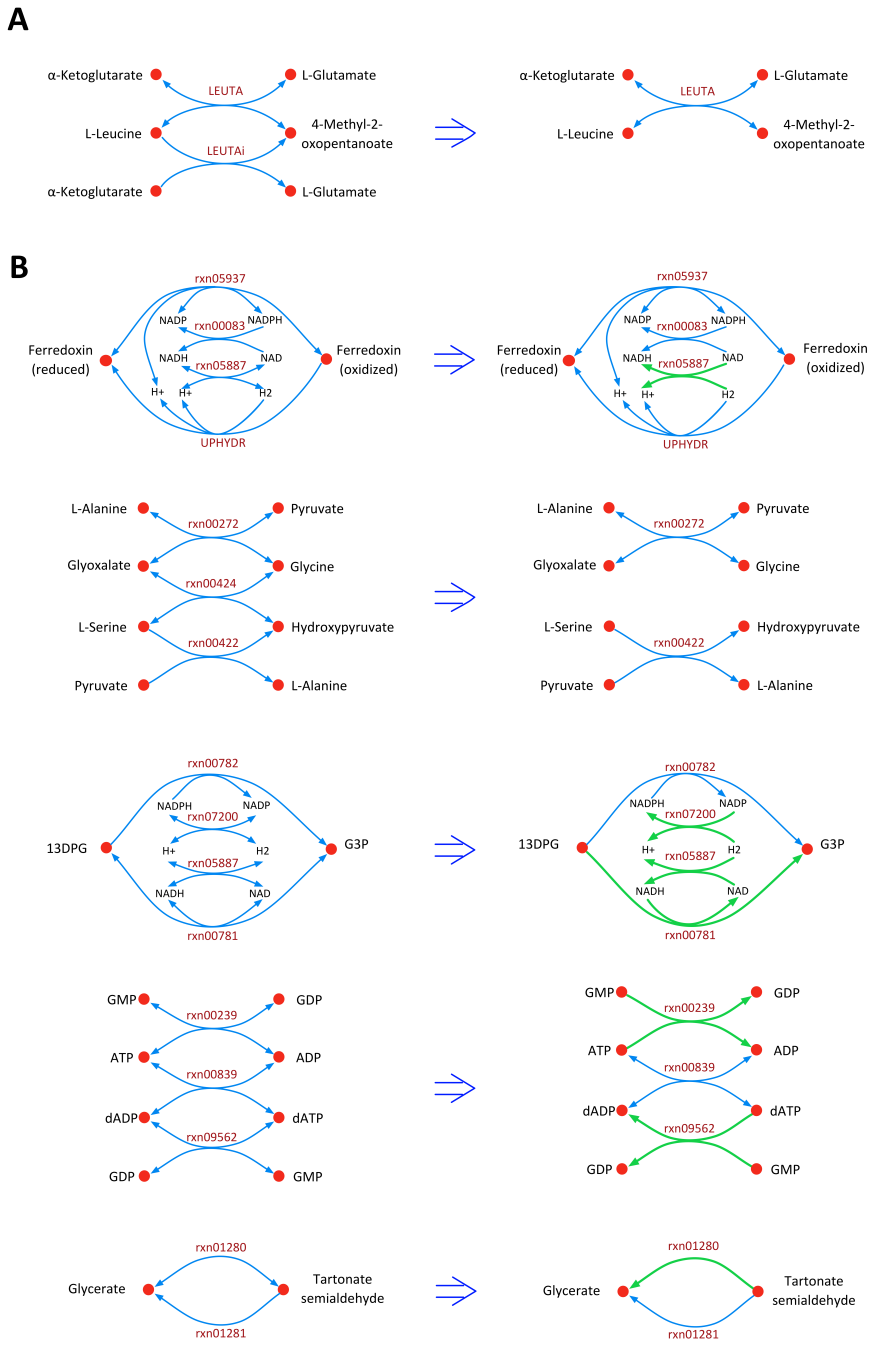


Figure 3.4: Schematics that illustrate the thermodynamically infeasible cycles and subsequent resolution strategies. (A) Cycles present in *i*JN678 (Nogales *et al*. 2012) and (B) Cycles present in *i*Cce805 (Vu *et al*. 2012). Blue colored lines represent the original reaction directionality whereas green ones denote modified directionality to eliminate cycles.

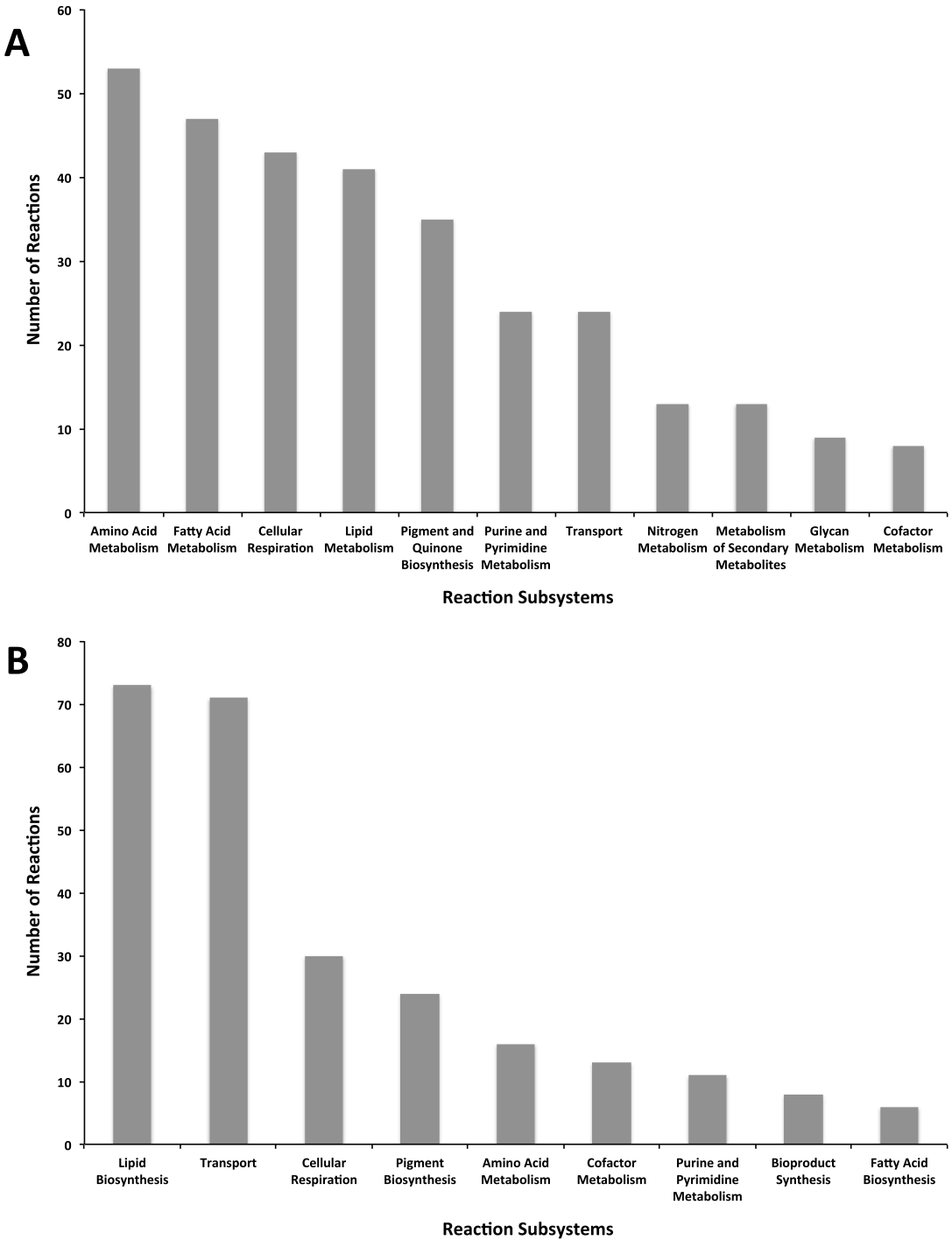


Figure 3.5: List of added reactions across pathways. (A) *i*Syn731 compared to *i*JN678 (Nogales *et al*. 2012), and (B) *i*Cyt773 compared to *i*Cce806 (Vu *et al.* 2012).

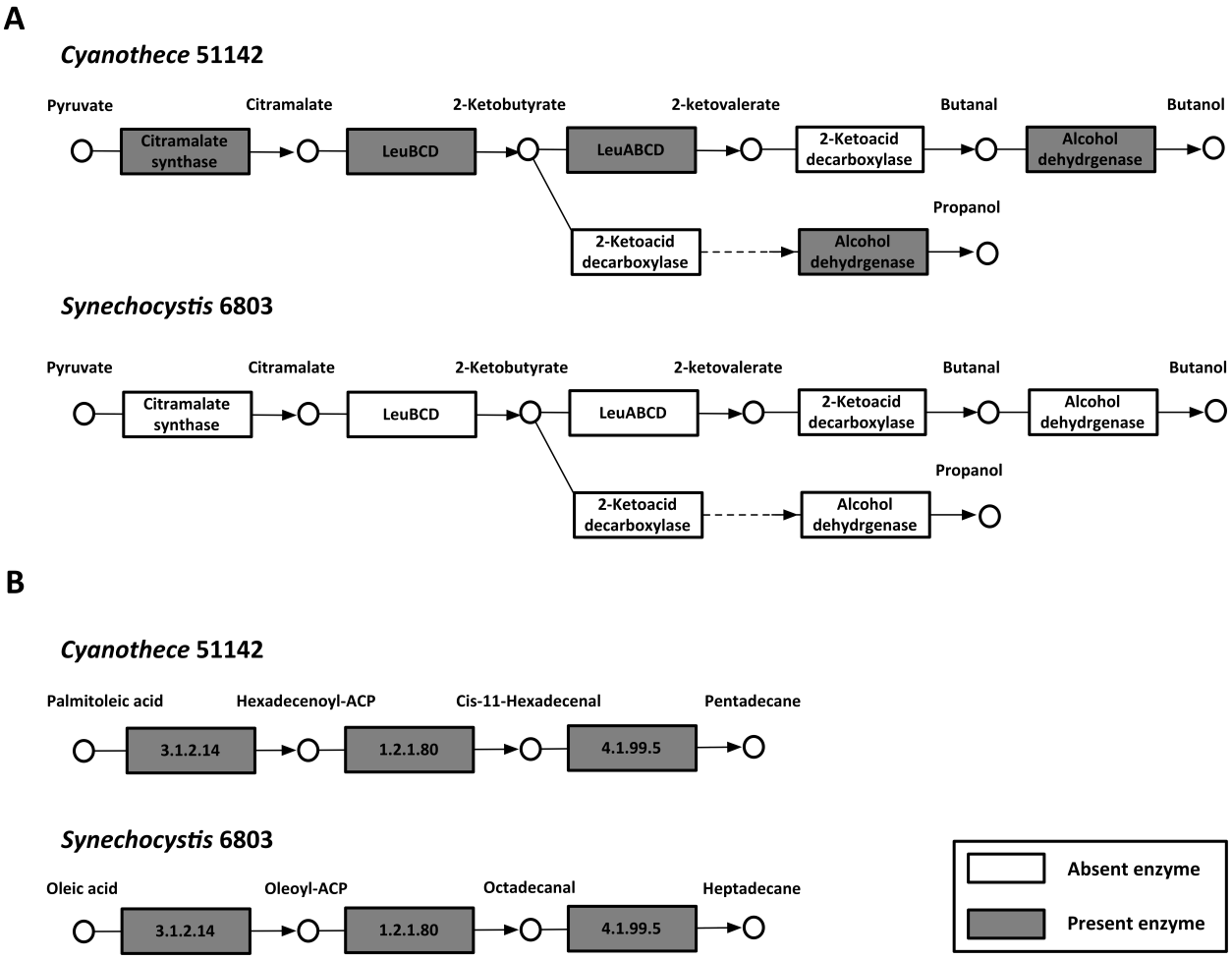


Figure 3.6: Examples of pathways that differ between the two cyanobacteria. (A) Nonfermentative alcohol production pathway highlighting the present and absent enzymes in *Cyanothece* 51142 and *Synechocystis* 6803, and (B) Alkane biosynthesis pathways in *Cyanothece* 51142 and *Synechocystis* 6803.

Chapter 4

Neutral sites on endogenous plasmids in *Synechocystis* sp. PCC 6803 enable increased protein expression and are composable with strong promoters.

4.1. Introduction

4.1.1. Synthetic Biology Tools in Cyanobacteria

As discussed in detail in chapter 1 of this dissertation, synthetic biology tools to enable metabolic engineering in cyanobacteria, though under active development, still limit progress in this area. While a range of strong promoters have been identified including variants of P*trc* (Camsund *et al.* 2014), P*cpcB* (Zhou *et al.* 2011; Markley *et al.* 2014), and others, promoter controllability is still an obstacle. While chemical inducers are a common laboratory strategy for controlling activity of heterologous pathways, autonomously controlled pathways that respond to the environment can reduce the operating costs of bioprocesses, simplify their operation, and ultimately lead to higher product yield (Zhang *et al.* 2012). Therefore, we set out to create a system for high-level protein expression in *Synechocystis* 6803 that is specific to stationary phase. By first allowing biomass to accumulate and then inducing product formation, competition between growth and product formation can be reduced and stability of production strains can be improved.

This process is likely to require the use of several well-characterized synthetic biology parts working together. For this reason, the composability of parts, meaning their ability to function together in ways that are predictable based on how they work alone, is a key property. This chapter will describe and analyze the creation of parts and systems for high-level expression of proteins specifically during stationary phase.

4.1.2. Stationary Phase

Bacterial cultures enter stationary phase when either nutrient limitation or build-up of growth byproducts ceases cell division. However, this does not necessarily imply

that cells become metabolically inactive. Well after the growth period had ended, a strain of *Synechococcus elongatus* PCC 7942 engineered to produce isobutyraldehyde continued to make this biofuel precursor (Atsumi *et al.* 2009). Artificial 'leaves' have been constructed from *Rhodopseudomonas palustris* cells embedded in latex that can produce H_2 photoheterotrophically for over 5 months without cell growth (Gosse *et al.* 2010). P*fic* in *Escherichia coli* was recently used to produce a high titer of a bacteriotoxin at stationary phase without any inducer, and without detectable growth-limiting toxin during exponential phase (Cao *et al.* 2011). Separating growth from production of biofuels has been identified as a key strategy for developing economically viable photosynthetic biofuel processes (Melis 2013).

In the present study we have performed microarray studies to identify genes in *Synechocystis* sp. PCC 6803 whose expression continues after a culture has stopped growing. We propose that the 5' UTRs of these genes might contain promoters that would be useful in synthetic biology applications for expression of heterologous proteins during stationary phase. Interestingly, the genes whose expression is highest after growth ceases are almost all genes of unknown function, and many are located on the small plasmids of the *Synechocystis* 6803 genome. We have identified neutral sites on both the chromosome of *Synechocystis* 6803 and on a small plasmid. We tested the expression of a reporter protein (EYFP) from these sites using a strong promoter and found that expression was higher during stationary phase.

4.2 Analysis of Expression at Stationary Phase

4.2.1. Cultures and Microarray Analysis

To identify genes active during stationary phase in *Synechocystis* 6803, we grew replicate cultures in BG11 medium bubbled with air + 5% CO_2 (autotrophic) or with air + 5 mM glucose (mixotrophic). Temperature was maintained at 30°C and light intensity at 100 μE m^{-2} s^{-1} from cool white fluorescent lamps. Cell growth was monitored by measurement of OD_{730} on a BioTek μQuant plate reader (BioTek, Vermont, USA). Cultures were sampled for microarrays in exponential phase and twice during stationary phase (see figure 1), and analyzed as described (Singh *et al.* 2008). Briefly, 2 replicate microarrays were analyzed for each of 2 replicate cultures for each nutritional condition. Data were LOWESS-normalized in the MATLAB Bioinformatics Toolbox. Normalized probe intensities were grouped by genes and t-tested to determine significant up- or down-regulation ($p < 0.05$). The complete table of genes analyzed by these microarrays is available online as supplementary material to this dissertation.

4.2.2. Stationary Phase Promoter Score (SPPS)

To quantify the activity of potential promoters at stationary phase, we calculated a 'Stationary Phase Promoter Score' (SPPS) for each open reading frame (ORF) in our microarray experiment, which is given by equation 4.1.

$$\text{SPPS} = \log_2(\text{fold-change}) + \log_2(\text{normalized expression}) \quad (4.1)$$

The fold-changes between expoential and stationary phase were averaged across the two time points in stationary phase sampled and nutritional conditions. Normalized expression is the mean LOWESS-normalized intensity of all microarray spots corresponding to a gene divided by

the mean normalized intensity for all genes. This normliazed expression calculation was at stationary phase only.

Many of the genes with the highest SPPS are located on endogenous plasmids, especially pSysA, pCA2.4, and pCC5.2 (Table 4.1). The upstream sequences of those ORFs are given in Table 4.2. The genome of *Synechocystis* 6803 includes 1 circular chromosome of 3.57 Mb, 4 larger plasmids of 44 to 120 kb (pSysA, pSysG, pSysM, pSysX), and 3 smaller plasmids of 2.4 to 5.2 kb (pCA2.4, pCB2.4, pCC5.2) (Yang *et al.* 1993; Yang *et al.* 1994; Kaneko *et al.* 1996; Xu *et al.* 1997; Kaneko *et al.* 2003). Although the plasmids of *Synechocystis* 6803 have received limited study, plasmid-borne genes are required for glucose tolerance (Kahlon *et al.* 2006) and encode a 2-component system responsive to low-oxygen (Summerfield *et al.* 2011). The 3 smaller plasmids contain only 10 ORFs, and *repA* on pCA2.4 is the only one with an annotated function (Nakao *et al.* 2010). Most genes on pSysA, pSysG, and pSysM were up-regulated during stationary phase in either nutritional condition. Under mixotrophic conditions, nearly all genes on the smaller plasmids (12/14) were also up-regulated (Table 4.3).

In terms of function, our results agree with previous studies of the exponential to linear growth transition in *Synechocystis* 6803 (Foster *et al.* 2007) and *E. coli* (Haddadin *et al.* 2005), which found that photosynthesis (in *Synechocystis* 6803) and energy production processes (in both strains) were down-regulated. However, the largest category of regulated genes in our study is that of unknown and hypothetical genes. Despite their unknown functions, the promoters of these genes are expected to serve as useful parts in synthetic biology studies (Shimada *et al.* 2004; Miksch *et al.* 2005). Attempts to harness these promoters as parts for synthetic biology are discussed later in this chapter.

4.2.3. Plasmid Copy Numbers

Because plasmid copy numbers often increase during stationary phase, we were interested to test whether this phenomenon might explain the observed up-regulation of plasmid genes (Table 4.4). We measured plasmid copy numbers per chromosome via qPCR (Lee *et al.* 2006). Briefly, we designed qPCR primer sets targeting each of the 8 replicons in the *Synechocystis* 6803 genome. These primer sequences are given in Table 4.4. We extracted genomic DNA from the same cultures sampled for our microarray experiments via phenol-chloroform extraction. For each replicon, we produced and purified a PCR product using the qPCR primers, and constructed standard curves of threshold cycle number vs. amount of template DNA to determine the efficiency of each PCR reaction. In the same qPCR run, we determined the plasmid copy numbers in each growth stage and nutritional condition using the chromosome as a reference. These data were corrected for the efficiency of each PCR reaction and the analysis was performed on 2 separate days. The data shown in Table 4.4 are the average of those 2 separate experiments.

The 3 smaller plasmids have higher copy numbers in the range of ~3 to 7 at stationary phase under autotrophic conditions, and at both exponential and stationary phase under mixotrophic growth conditions. The copy numbers of the 4 larger plasmids range from ~0.3 to 1.2 per chromosome, and vary less with growth phase. Copy numbers of pSysA, pSysM, and pSysX are about twice as high during mixotrophic growth as during autotrophic growth, but only slightly higher for pSysG.

Copy numbers of pSysA, pCA2.4, and pCC5.2, the plasmids containing the highest-SPPS genes, did not increase at stationary phase in all nutritional conditions, indicating that expression levels of such genes are controlled both at the gene dosage and transcriptional levels. For

synthetic biology applications, the flexibility afforded by a range of available gene copy numbers and promoter specificities will serve as a benefit, since higher-copy plasmids have been associated with growth deficits, lower productivity and lower inducibility (Jones *et al.* 2000). High copy plasmids from *E. coli* have been modified for use in *Synechocystis* 6803 (Huang *et al.* 2010) and have copy numbers between ~1 (Marraccini *et al.* 1993) and ~3 (Ng *et al.* 2000) per chromosome (10-30 per cell). These plasmids can be maintained with antibiotics, in contrast to endogenous cyanobacterial plasmids that have higher copy numbers and can be modified to contain heterologous genes and maintained based on essential sequences they carry (Xu *et al.* 2011).

We have identified genes up-regulated during the transition to stationary phase under various nutritional conditions in *Synechocystis* 6803. These genes are mostly encoded on plasmids, whose copy numbers range between ~0.4 and 7 per chromosome. The transcriptional behavior of these genes' promoters may make them useful for synthetic biology in applications where expression is desired only at stationary phase, to maximize production while not interfering with biomass accumulation during the growth phase. The higher copy numbers of these plasmids relative to the chromosome may also make them useful insertion sites for high expression of heterologous genes.

4.3. High-Copy Plasmids for Heterologous Gene Expression

To test the utility of the small plasmids of *Synechocystis* 6803 as neutral sites for cloning, we created reporter strains that express EYFP from a site on one of those plasmids. We identified a neutral site (NSP1) on plasmid pCC5.2 by inspection of the annotated plasmid sequence for a

region that did not contain any predicted genes. Figure 4.2A shows our strategy for this work. Figure 4.2B shows a map of this small plasmid with the targeted neutral site identified. In addition, using recently published RNA-seq data, we identified 2 neutral sites on the chromosome of *Synechocystis* 6803 from which no expression appears to occur under reference conditions (Mitschke *et al.* 2011). These regions are also indicated in Figure 4.2B. We constructed neutral site-targeting vectors via CPEC containing ~600 bp upstream and downstream of each neutral site flanking a P*trc_eyfp*_Km^R cassette (Huang *et al.* 2010). We created kanamycin-resistant mutants of *Synechocystis* 6803 expressing the EYFP cassette from each of these three locations, as well as a 4^{th} carrying the broad host range vector pPMQAK1 from which the expression cassette was derived. After growing starter cultures for 3 days, we re-diluted cultures in fresh media to OD_{730} of 0.02. Three independent replicates of each culture were then transferred to 12 well plates for growth at 30 C under ~50 μE m^{-2} s^{-1} of light. Every 24 hours, samples were taken and EYFP fluorescence was measured using a BioTek Synergy Mx plate reader at excitation/emission wavelengths of 485/528 nm. All fluorescence measurements were normalized to OD_{730}. The results of this experiment are shown in figure 4.3. The cassette placed in NSP1 gives much higher expression than the cassette placed in NSC1 or NSC2, which are virtually identical to each other. At day 2, the ratio of NSP1 expression to NSC1 is 8.2, and this value increases to 14.5 after 8 days of culture. This 1.8-fold increase agrees reasonably well with the 4-fold increase we observed earlier in plasmid copy number at stationary phase under autotrophic conditions (Table 4.4). The EYFP cassette expressed from pPMQAK1 gave intermediate levels of expression, which did not change relative to the chromosome during the period examined.

Figure 4.3C shows results of PCR to confirm the insertion of the EYFP cassette at its target site in either NSC1, NSC2, or NSP1. Even after extensive restreaking on kanamycin-containing media, the mutation in NSC1 did not segregate. However, as the lower panel of figure 4.3B shows, the mutation was still maintained after growth for one month in the absence of antibiotic, albeit at a seemingly lower allele frequency. A similar lack of segregation was observed for a mutant created in the pAQ1 small plasmid of *Synechococcus* sp. PCC 7002 (Begemann *et al.* 2013). One explanation of this observation is that the increased copy number of the plasmid makes segregation less likely and thus it will require more time. This explanation assumes that segregation is essentially a random process – after enough cell divisions, a population of daughter cells will emerge that no longer contains any wild type plasmids. For a replicon with a higher copy number, this is less likely to occur in a short time. It is also possible that NSC1 is not, in fact a neutral site, and that some essential transcript originates from this locus. Further experiments will be needed to resolve this question.

Finally, we sought to show that our neutral sites could be composed predictably with different promoters. In addition to the mutants described above, we constructed mutants with P*trc*1O replaced by P*cpcB*$_{560}$ (Zhou *et al.* 2011), P*psbA2*, P*slr9003*, and P*pSysA_116*. The latter two promoters are those defined as having the highest SPPS in this study (see table 4.1). They originate both from replicons that increased their abundance at stationary phase (*slr9003* is on pCC5.2) and those that did not (pSysA). We conducted similar growth curve experiments as for the P*trc*1O strains described above with these new strains (see figure 4.4). Surprisingly, P*psbA2*, P*slr9003*, and P*pSysA_116* gave barely detectable fluorescence in our experiments, so those data are not shown here. P*cpcB*$_{560}$ has previously been shown to be an extremely strong promoter. CpcB is one of the most abundant proteins in *Synechocystis* 6803 and this promoter can lead to

expression of a heterologous protein as among the most abundant in the cell. We observed that this promoter was 2-4 times stronger than P*trc*1O during exponential phase, but later in growth the expression from these two promoters converged to very similar levels. In our microarray experiments, we observed that *cpcB* itself was down-regulated by ~60% at stationary phase in autotrophic conditions. If P*trc*1O itself is not regulated by the transition to stationary phase, then this down-regulation of P*cpcB* would nicely explain our observations. Thus, we find that novel combinations of promoters, genes, and expression sites can show composability in *Synechcocystis* 6803. This property will aid in the future construction of novel synthetic biological systems.

4.4. Conclusions and Future Directions

We have shown that small plasmids in *Synechocystis* 6803 can be used for expression of heterologous proteins, and that these proteins will be expressed more highly at stationary phase if appropriate promoters are chosen. We have also identified promoters that might be useful as parts for increasing the specificity of expression at that phase of culture. However, of the two putative stationary phase-active promoters that we tested, neither one gave significant expression. Given the increasing availability of RNA-seq data for *Synechocystis* 6803, it should be possible to better understand the transcriptional units on the plasmids of *Synechocystis* 6803 and use that detailed knowledge to locate more active promoters with desired growth-phase specific expression. By composing these promoters with neutral sites that become more abundant at stationary phase, a functional auto-inducing system that first grows a dense culture and then uses available CO_2 and sunlight to make a product can be constructedm, with a higher fold of

induction that observed here (1.8-fold). By responding to the culture environment, this system would relieve the need to add chemical inducers of transcription or translation.

4.5. Supplementary Material

Supplementary file 1 (.xls) included with the online version of this dissertation includes a full table of the microarray data described in this chapter.

4.6. References

Atsumi, S., Higashide, W. and Liao, J. C. (2009). "Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde." *Nat Biotechnol* **27**(12): 1177-1180.

Begemann, M. B., Zess, E. K., Walters, E. M., Schmitt, E. F., Markley, A. L. and Pfleger, B. F. (2013). "An organic acid based counter selection system for cyanobacteria." *PLoS One* **8**(10): e76594.

Camsund, D., Heidorn, T. and Lindblad, P. (2014). "Design and analysis of LacI-repressed promoters and DNA-looping in a cyanobacterium." *J Biol Eng* **8**(1): 4.

Cao, Y. and Xian, M. (2011). "Production of phloroglucinol by *Escherichia coli* using a stationary-phase promoter." *Biotechnol Lett* **33**: 1853-1858.

Foster, J., Singh, A., Rothschild, L. and Sherman, L. (2007). "Growth-phase dependent differential gene expression in *Synechocystis* sp. strain PCC 6803 and regulation by a group 2 sigma factor." *Arch Microbiol* **187**: 265-279.

Gosse, J., Engel, B., Hui, J., Harwood, C. and Flickinger, M. (2010). "Progress Toward a Biomimetic Leaf: 4,000 h of Hydrogen Production by Coating-Stabilized Nongrowing Photosynthetic *Rhodopseudomonas palustris*." *Biotechnol Prog* **26**(4): 907-918.

Haddadin, F. and Harcum, S. (2005). "Transcriptome Profiles for High-Cell-Density Recombinant and Wild-Type *Escherichia coli*." *Biotechnol Bioeng* **90**(2): 127-153.

Huang, H., Camsund, D., Lindblad, P. and Heidorn, T. (2010). "Design and characterization of molecular tools for a Synthetic Biology approach towards developing cyanobacterial biotechnology." *Nucleic Acids Res* **38**(8): 2577-2593.

Jones, K., Kim, S. and Keasling, J. (2000). "Low-Copy Plasmids can Perform as Well as or Better than High-Copy Plasmids for Metabolic Engineering of Bacteria." *Metab Eng* **2**: 328-338.

Kahlon, S., Beeri, K., Ohkawa, H., Hihara, Y., Murik, O., Suzuki, I., Ogawa, T. and Kaplan, A. (2006). "A putative sensor kinase, Hik31, is involved in the response of *Synechocystis* sp. strain PCC 6803 to the presence of glucose." *Microbiology* **152**: 647-655.

Kaneko, T., Nakamura, Y., Sasamoto, S., Watanabe, A., Kohara, M., Matsumoto, M., Shimpo, S., Yamada, M. and Tabata, S. (2003). "Structural Analysis of Four Large Plasmids Harboring in a Unicellular Cyanobacterium, *Synechocystis* sp. PCC 6803." *DNA Res* **10**: 221-228.

Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M. and Tabata, S. (1996). "Sequence Analysis of the Genome of the Unicellular Cyanobacterium *Synechocystis* sp. Strain PCC6803. II. Sequence Determination of the Entire Genome and Assignment of Potential Protein-coding Regions." *DNA Res* **3**: 109-136.

Lee, C., Ow, D. and Oh, S. (2006). "Quantitative real-time polymerase chain reaction for determination of plasmid copy number in bacteria." *J Microbiol Methods* **65**(2): 258-267.

Markley, A. L., Begemann, M. B., Clarke, R. E., Gordon, G. C. and Pfleger, B. F. (2014). "Synthetic Biology Toolbox for Controlling Gene Expression in the Cyanobacterium Synechococcus sp. strain PCC 7002." *ACS Synth Biol*.

Marraccini, P., Bulteau, S., Cassier-Chauvat, C., Mermet-Bouvier, P. and Chauvat, F. (1993). "A conjugative plasmid vector for promoter analysis in several cyanobacteria of the genera *Synechococcus* and *Synechocystis*." *Plant Mol Biol* **23**: 905-909.

Melis, A. (2013). "Carbon partitioning in photosynthesis." *Curr Opin Chem Biol* **17**(3): 453-456.

Miksch, G., Bettenworth, F., Friehs, K., Flaschel, E., Saalbach, A., Twellmann, T. and Nattkemper, T. (2005). "Libraries of synthetic stationary-phase and stress promoters as a tool for fine-tuning of expression of recombinant proteins in *Escherichia coli*." *J Biotechnol* **120**: 25-37.

Mitschke, J., Georg, J., Scholz, I., Sharma, C. M., Dienst, D., Bantscheff, J., Voss, B., Steglich, C., Wilde, A., Vogel, J. and Hess, W. R. (2011). "An experimentally anchored map of transcriptional start sites in the model cyanobacterium Synechocystis sp. PCC6803." *Proc Natl Acad Sci U S A* **108**(5): 2124-2129.

Nakao, M., Okamoto, S., Kohara, M., Fujishiro, T., Fujisawa, T., Sato, S., Tabata, S., Kaneko, T. and Nakamura, Y. (2010). "CyanoBase: the cyanobacteria genome database update 2010." *Nucleic Acids Res* **38**: D379-D381.

Ng, W., Zentella, R., Wang, Y., Taylor, J. and Pakrasi, H. (2000). "phrA, the major photoreactivating factor in the cyanobacterium *Synechocystis* sp. strain PCC 6803 codes for a cyclobutane-pyrimidine-dimer-specific DNA photolyase." *Arch Microbiol* **173**: 412-417.

Shimada, T., Makinoshima, H., Ogawa, Y., Miki, T., Maeda, M. and Ishihama, A. (2004). "Classification and Strength Measurement of Stationary-Phase Promoters by Use of a Newly Developed Promoter Cloning Vector." *J Bacteriol* **186**(21): 7112-7122.

Singh, A., Elvitigala, T., Bhattacharyya-Pakrasi, M., Aurora, R., Ghosh, B. and Pakrasi, H. (2008). "Integration of Carbon and Nitrogen Metabolism with Energy Production is Crucial to Light Acclimation in the Cyanobacterium *Synechocystis*." *Plant Physiol* **148**: 467-478.

Summerfield, T., Nagarajan, S. and Sherman, L. (2011). "Gene expression under low-oxygen conditions in the cyanobacterium *Synechocystis* sp. PCC 6803 demonstrates Hik31-dependent and -independent responses." *Microbiology* **157**: 301-312.

Xu, W. and McFadden, B. (1997). "Sequence Analysis of Plasmid pCC5.2 from Cyanobacterium *Synechocystis* PCC 6803 That Replicates by a Rolling Circle Mechanism." *Plasmid* **37**: 95-104.

Xu, Y., Alvey, R., Byrne, P., Graham, J., Shen, G. and Bryant, D. (2011). "Expression of Genes in Cyanobacteria: Adaptation of Endogenous Plasmids as Platforms for High-Level Gene Expression in *Synechococcus* sp. PCC 7002." *Photosynth Res Protocols*. R. Carpentier, Springer Science+Business Media, LLC. **684:** 273-293.

Yang, X. and McFadden, B. (1993). "A Small Plasmid, pCA2.4, from the Cyanobacterium *Synechocystis* sp. Strain PCC 6803 Encodes a Rep Protein and Replicates by a Rolling Circle Mechanism." *J Bacteriol* **175**(13): 3981-3991.

Yang, X. and McFadden, B. (1994). "The complete DNA Sequence and Replication Analysis of the Plasmid PCB2.4 from the Cyanobacterium *Synechocystis* PCC 6803." *Plasmid* **31**: 131-137.

Zhang, F., Carothers, J. M. and Keasling, J. D. (2012). "Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids." *Nat Biotechnol* **30**(4): 354-359.

Zhou, Y., Chen, W. L., Wang, L. and Zhang, C. C. (2011). "Identification of the oriC region and its influence on heterocyst development in the filamentous cyanobacterium Anabaena sp. strain PCC 7120." *Microbiology* **157**(Pt 7): 1910-1919.

Table 4.1: Top genes ranked by Stationary Phase Promoter Score (SPPS).

ORF	Replicon	Annotation	Stationary Phase Promoter Score	Normalized Expression	Fold-Change (Autotrophic/ Mixotrophic)
slr9003	pCC5.2	Unknown	8.53	7.46	1.76 / 0.38
pSysA_116	pSysA	Unknown	7.37	3.82	4.19 / 2.90
slr9002	pCC5.2	Unknown	6.87	4.41	0.44 / 4.47
sll9006	pCC5.2	Unknown	6.41	4.24	1.02 / 3.31
pSysA_145	pSysA	Unknown	6.40	3.53	4.02 / 1.72
sll1982	Chromosome	putative transposase	6.32	4.85	2.28 / 0.66
slr9101	pCA2.4	replication protein A	6.31	6.55	0.28 / -0.77
ssr9005	pCC5.2	Unknown	6.03	3.46	-0.14 / 5.28
pSysA_39	pSysA	Unknown	5.97	3.44	2.89 / 2.17
sll5036	pSysM	Sulfide-quinone reductase	5.94	2.40	3.61 / 3.47
pSysA_27	pSysA	Unknown	5.92	3.32	2.96 / 2.24
ssl9001	pCC5.2	Unknown	5.90	3.58	0.05 / 4.59
sll8019	pSysG	Unknown	5.84	4.42	2.03 / 0.81
pSysA_25	pSysA	Unknown	5.83	3.17	2.99 / 2.33
slr0915	Chromosome	putative endonuclease	5.68	3.81	1.93 / 1.81
pCA24_1	pCA2.4	Unknown	5.61	3.88	-0.01 / 3.47
pSysA_24	pSysA	Unknown	5.60	3.11	2.81 / 2.18
ssr9004	pCC5.2	Unknown	5.44	3.07	-0.48 / 5.22
pSysA_34	pSysA	Unknown	5.44	2.52	3.10 / 2.72
pSysA_22	pSysA	Unknown	5.24	2.37	3.03 / 2.72

Table 4.2: Upstream sequences of high-SPPS genes.

ORF	Upstream sequence (-250 to -1 relative to start codon)
slr9003[a]	TCCAACAAAAAAAAGCTTTTCAGGAGGGAATTAAGATTGCTGCAGTAAAAAACGTA AGAAGTTTAGTTGACGCTAAAAAACTTACCTACAGACAATAACCCGGCCCAAAAAGC CAACAAAATACTTCAAAAATATTGTCTCTACTGTAGCTCTAAAAATTCCCAAAAGAA AAGCGGTCAACTCTTGAACCCGAGACCGCT
pSysA_116	ATATTTCTGGACGTGGAATGACTCTTGTTCAGGTTTATGACACTGTCAGCAATTGAGA TACTTTTGCTGATCGTTTCAGTCCCCTAACGGGGAAAAGAGGGTGTTGAACGAGCCA TGGGGTCACTATTGACCACCTCAGGCGCGTTTCAGTCCCCTGACGGGGAAAAGAGGG TGTTGAACCTTGGAAGAAGTTTCCCCTT
slr9002	ACGCACGATGACGTATGACCCTTTTAGCACGGTAGGGAGCGTGATAATCTTCTGCAA CACCTATATAGTATTGTTGCGATCGCGAGCGATGGCGTATGACCGGCAATAAGCTAC ACTGCGCCGATTCCAGCAAAGATAATCCCCTAAGCAACGCAATAATCTTCTGCGAAC CTTATATAAGGTTCTGCATATAACGCACG
sll9006	TTGGCTAGGGAATCCTTGGAAAATTCCCCTATCCCGGTAAGGAATCTTTCAAAGCCC AATACTTTAAGGAAGTGAACGGGGACGGTGAGAGTCTTCCGGCTAGCATCCTTGGCA CAAACCTTTCCATCTTCCCCGGCCAAATTTTGGAGCCTTTGGCTGTCCCTTGCCTGTA AAAATTCTGCACCGGTGGTGAAGTAATA
pSysA_145	ATATTTCTGGACGTGGAATGACTCTTGTTCAGGTTTATGACACTGTCAGCAATTGAGA TACTTTTGCTGATCGTTTCAGTCCCCTAACGGGGAAAAGAGGGTGTTGAACGAGCCA TGGGGTCACTATTGACCACCTCAGGCGCGTTTCAGTCCCCTGACGGGGAAAAGAGGG TGTTGAACCTTGGAAGAAGTTTCCCCTT
sll1982	TCCCTGGCAATTACGCTCAAAACGCAACTCTCGATTGTTCAAACAGAGTTGATAAAA CTGCTCATCGGAAAGGGATAGGCTGTCAAGTTTGACGGTTATGGCGGCGGGCATTGG CTTTACCCCAAAGGCATTGATGTGGCATTGGCTCCCATTCTAAGTGATGTTTACGGTG ACAGAATACCTATCGCTCTTCTATCATT
slr9101[b]	AAATATGGCATTTCATCTTTTCAGGTTTCCCCAAGGTTTCAACTTTCCTCCTTATATTT ATTACTGAGGGATAAGTCGCGGATGACAAAATTTGCTGAAACCCTTACCAGATAAGG CATAGAAGCCTATTGACAAAGTAGAAACCCTCTAGCTAAGCTTTGAGTGTCACTTCA AAAACTCAATATCTAGAGGGCTCCAAG
ssr9005[a]	TAAGGATGAAGTGCAGGGCATTATTGACCGCTACAGGGAAGACTTACTGGCAGGAA GACAGCTCCAAGATGTTCCCAGCTCCTACGAGGTCAAAACGGCGATCGCCATTCTGA CGGAGGCACTCAGCCTTAAAGCCAATGCCGGTGGAGCCATCAAAGCAAAAATCAGA GAAGCCCTAGCTATCTTGGAAAGGAACTGAA
pSysA_39	GGATTAGTAGAAGGAAAGATTGCCGACCGTTCAGGACAATATGCTGGGAGCGATGG TTTCAGTCCCGATCGCCGGGATTAGTAGAAGGAAAGCTGTTGGCGAAACAGGAAAA ATCTACGTAACAGAAGAGTTTCAGTCCCGATCGCCGGGATTAGTAGAAGGAAAGAA GTATTAAAGGAGTTGGGCGGTGGCTAGTAACA
sll5036	ATGCCTTCTTTTGCTAGGGAATCCTTTACTAATCCGATAGCTTCATCAAATCCTACCA GCAATTTTTTACTAAAGTAATACATTTTATATCTCTGATTTTTATTGAACTAGTCCTTG CCAAATCGAACAAGCACCAATATTATAAGAATATAACTACATAGTTGTATTCGTCAA TAGTTTTTGGGGGGAGGGAGTTTAAA
pSysA_27	CTTCAGCAGTTTCAGTCCCGATCGCCGGGATTAGTAGAAGGAAAGTAGACATTGTTG AAGAACTTGGACTGATGGAAGAAGTTTCAGTCCCGATCGCCGGGATTAGTAGAAGG AAAGAACGGGCTGGGTCTGGAACCGGAATTCGTGTCTATTGTTTCAGTCCCGATCGC CGGGATTAGTAGAAGGAAAGTCAAGTCTTC
ssl9001	TGCGTGTCAGCAACAGACCTATAGGTGTTGCAGTAGATTATCGCGTTCCCTACCGTGC TAGCTTTGCTCTGATAGGGTCAATGTGGTTTTTTGCCGGTCATACTTCATAACGCTTTT CTACGGGAGTACCCTAGGGTAGTCGCAGTAGATTATCGTGCTTCCCAGCGTGCTAAA ACTGGCTTGATAGGGGCAATGTAGGT

Table 4.2: Upstream sequences of high-SPPS genes (continued).

ORF	Upstream sequence (-250 to -1 relative to start codon)
sll8019	TTAAGAGAGGTAATTAACCTAACTTAACAAGAACATCGAGTTCTTAACGTACACCCCA GAAAAAGTTAAAGCCACCTGGCAAAGCGTGTTTCTCAGGCACGCCACAGGTAGCTAC ACAGACTAAAATCTTATGTTGTTAGTGTAGCATGCCAATTTGCCGGATAGCTCCTCCTG GGAAAAATTAGGAGAGTATCTAAGCA
pSysA_25	CTTCAGCAGTTTCAGTCCCGATCGCCGGGATTAGTAGAAGGAAAGTAGACATTGTTGA AGAACTTGGACTGATGGAAGAAGTTTCAGTCCCGATCGCCGGGATTAGTAGAAGGAA AGAACGGGCTGGGTCTGGAACCGGAATTCGTGTCTATTGTTTCAGTCCCGATCGCCGG GATTAGTAGAAGGAAAGTCAAGTCTTC
slr0915	GTTGTGACCATTGCAGTAAAGTACCGCCCGTATACTTCGAAAATCCCTAAAATTCTTA CTCTTCAGTGCATAGACTATGGGATGAATCTGCCCTAAAAATAAAGTTTGGCAAAAAT TCCCCCGATCAGTTATGATATTCGAAGCGACGCGGGATAGAGCAGTCTGGTAGCTCGT CGGGCTCAGGTCGCAAGATGTAAACC
pCA24_1[b]	TATGGTCATTCAACGCCCCCTAATTAGTCCCTAAACCCTGCCAAATATGGCATTTCATC TTTTCAGGTTTCCCCAAGGTTTCAACTTTCCTCCTTATATTTATTACTGAGGGATAAGTC GCGGATGACAAAATTTGCTGAAACCCTTACCAGATAAGGCATAGAAGCCTATTGACA AAGTAGAAACCCTCTAGCTAAGCT
pSysA_24[c]	GGATTAGTAGAAGGAAAGCCACCACCACCACCACGATTAATCAGCAATTATTAGTTTC AGTCCCGATCGCCGGGATTAGTAGAAGGAAAGCTGATAGATCGTAGCGGAATGCTAT GGGATGCCTTAGTTGTTTCAGTCCCGATCGCCGGGATTAGTAGAAGGAAAGTTCTGAG GTTCTTTCTAAAATTCTTCCCTATATT
ssr9004[a]	ACTAAGCGACATTATGGCCCCGCCAATATTGACCCCGATGGGCGATCGGCTATTTTTT CCCGGTGGTTTGAGCGGGATTCTATTTTGTATCACTCTGATACCGTATCCACTGAATCC TTATTAATTAATCAAGCCTAGGAACTGGATACCAAAAACAGGGAGTTATGATGGGAAT ATAATCCCGTTAACAGGCTAAACCC
pSysA_34	TCGCCGGGATTAGTAGAAGGAAAGACCACTCGACGATCAGATCTTCATCAATCAACCG AAGTTTCAGTCCCGATCGCCGGGATTAGTAGAAGGAAAGTCGGAGCCAGGCCCACCA AAAGCCGCTTGGTAGGATTGTTTCAGTCCCGATCGCCGGGATTAGTAGAAGGAAAGCC GACTTTTGGATCGATTAGAATCCGACG
pSysA_22[c]	CCTTAGTTGTTTCAGTCCCGATCGCCGGGATTAGTAGAAGGAAAGTTCTGAGGTTCTTT CTAAAATTCTTCCCTATATTTTGGTTTCAGTCCCGATCGCCGGGATTAGTAGAAGGAAA GCCACATTCATCGCTACAGACTTAGAAGGTAGTTACGAGTTTCAGTCCCGATCGCCGG GATTAGTAGAAGGAAAGCTAGCCT

[a] – These 3 ORFs appear to be in an operon, with slr9003 at the 5' end

[b] – The predicted ORFs slr9101 and PCA24_1 overlap, and the latter includes an additional 42 nucleotides at its 5' end.

[c] – These 2 ORFs appear to be in an operon, with pSysA_24 at the 5' end

Table 4.3: Number of genes up- and down-regulated at stationary phase by replicon at stationary phase vs. exponential phase in autotrophic and mixotrophic conditions.

Replicon	Total ORFs	Regulated Genes			
		Autotrophic		Mixotrophic	
		Up	Down	Up	Down
Chromosome	3,239	792	1278	658	1245
pSysA	134	105	2	123	2
pSysG	52	48	1	44	1
pSysM	141	104	9	95	13
pSysX	114	51	31	42	14
pCA2.4	4	1	2	3	1
pCB2.4	4	0	4	4	0
pCC5.2	6	3	2	5	0

Table 4.4: Effects of nutritional condition and growth phase on plasmid copy numbers per chromosome. For stationary phase, data are averaged across the early and later stationary phase time points. Primer sequences used to assess copy number via qPCR are also given.

Replicon	Copy Number per Chromosome			
	Autotrophic		Mixotrophic	
	Exponential	Stationary	Exponential	Stationary
pSysA	0.34 ± 0.00	0.33 ± 0.00	0.64 ± 0.01	0.60 ± 0.01
pSysG	0.64 ± 0.01	0.54 ± 0.01	0.72 ± 0.02	0.83 ± 0.01
pSysM	0.33 ± 0.00	0.31 ± 0.00	0.69 ± 0.01	0.49 ± 0.01
pSysX	0.65 ± 0.01	0.66 ± 0.01	1.24 ± 0.02	1.09 ± 0.01
pCA2.4	0.75 ± 0.01	5.41 ± 0.10	6.26 ± 0.06	7.39 ± 0.09
pCB2.4	0.40 ± 0.00	2.46 ± 0.02	3.74 ± 0.04	2.68 ± 0.02
pCC5.2	0.93 ± 0.01	3.72 ± 0.04	6.02 ± 0.05	7.33 ± 0.07
Primer Sequences				
Replicon	**Primer 1**		**Primer 2**	
Chromosome	CACCAGCACTCGCCTCACCG		CCTGGATGGTGGCACAGGCG	
pSysA	GAATCTGGCGGAGCACCTCGG		TGATGAAGCGGTGGTGGTGGC	
pSysG	GGTGGCTTCTGTGACCTTCTGCC		CCTACCAGGCGATCGTCGCC	
pSysM	GCTCGATCACAAGTCTGGCATTGGC		AGATAGCCGATTGACGGTACTGCGG	
pSysX	AGGAACATGAGGCTCGCATTCGC		TGCCTGGACCAATTCAACGTGCC	
pCA2.4	TGCCAGTGGCGGAGGTCTCT		CCAGCTCGCAATTCCTAACGGTCA	
pCB2.4	GGTGCGGTTAGCAACTTGTGCC		AATAGCACGCCTACTTCGTGACGG	
pCC5.2	CCAAGGGCACGGGAAAGACGG		TCTTCTCGCCGTTGCTCCACC	

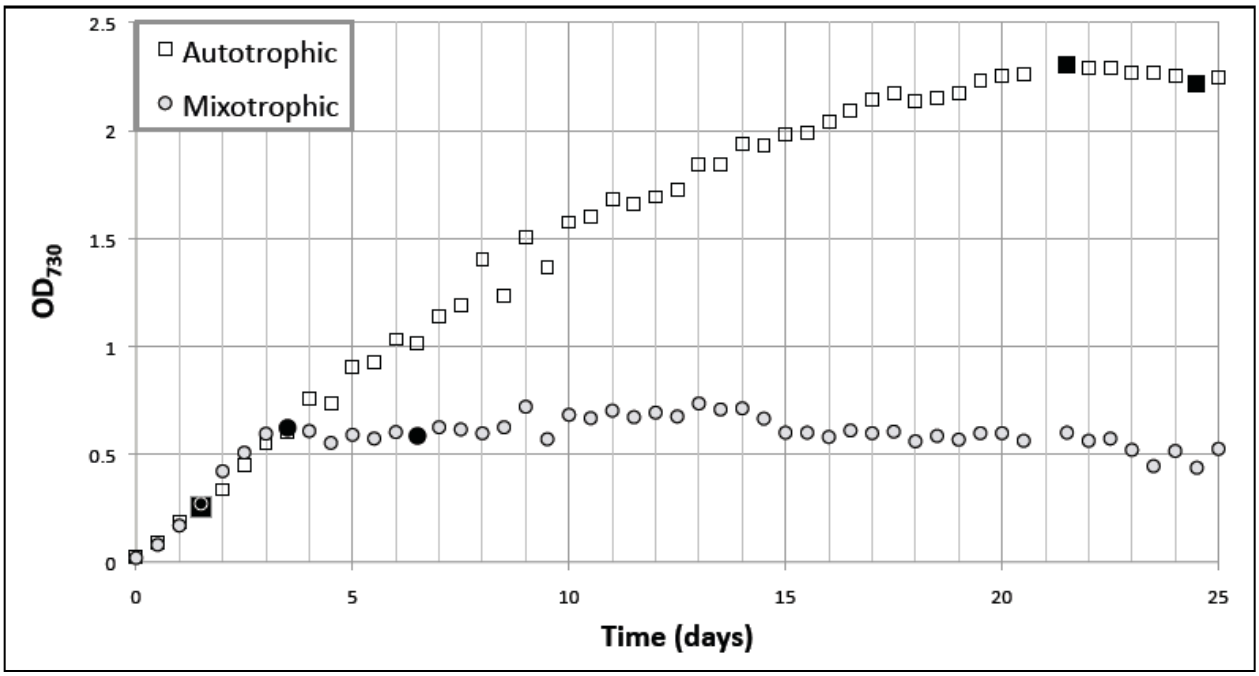


Figure 4.1: Growth curves of *Synechocystis* 6803 under autotrophic and mixotrophic conditions. Time points sampled for nucleic acid analysis are indicated with filled symbols. Duplicate cultures were grown in bubble column bioreactors bubbled with air + 5% CO_2 (autotrophic) or with air + 5 mM glucose (mixotrophic). Temperature was maintained at 30°C and light intensity at 100 μE m^{-2} s^{-1} from cool white fluorescent lamps.

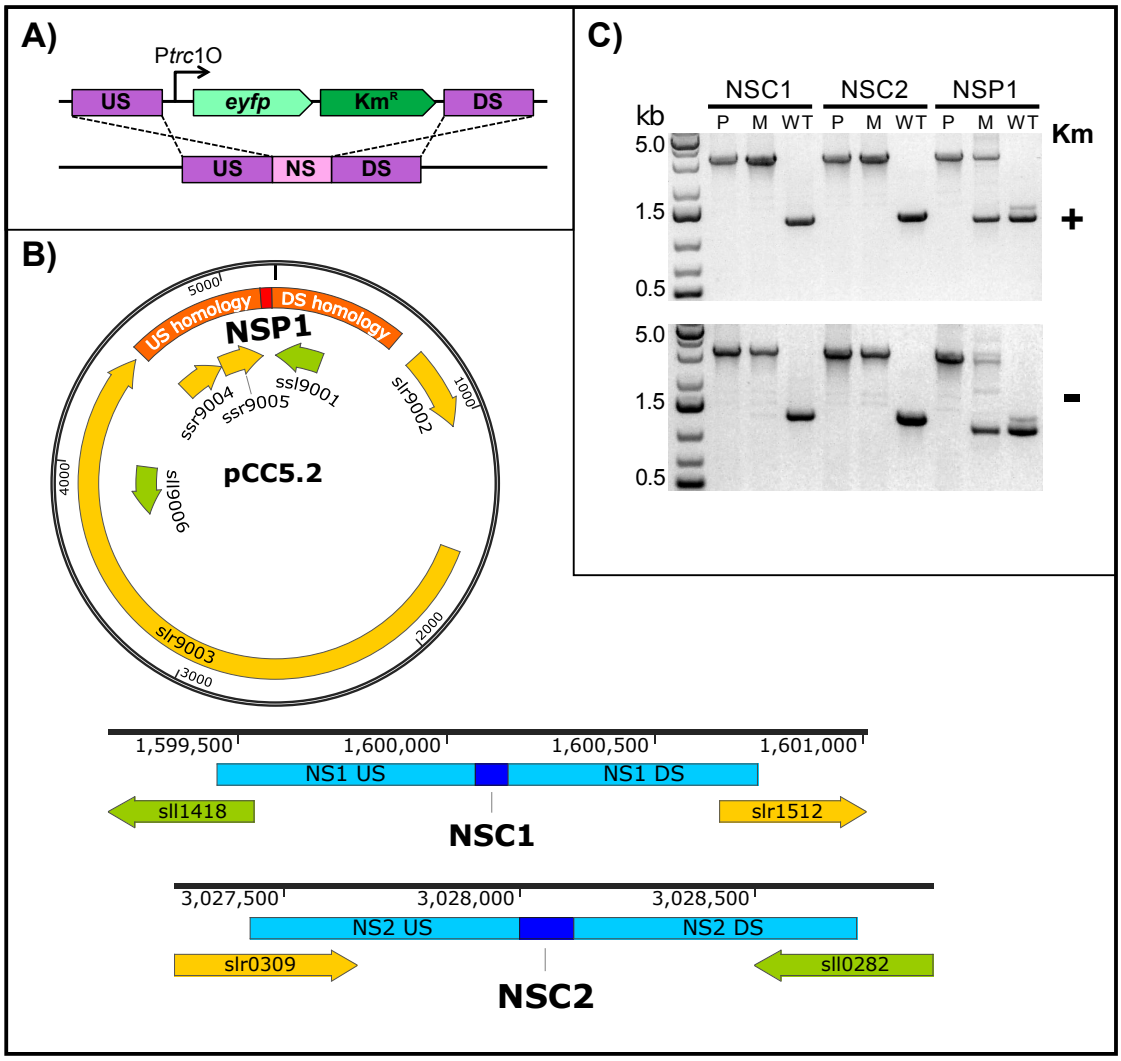


Figure 4.2: Selection of neutral sites on plasmid pCC5.2 and the chromosome of *Synechocystis* 6803. A) shows the P*trc*1O_*eyfp*_KmR cassette (Huang *et al.* 2010) and a general scheme for its insertion via double homologous recombination. Suicide vectors were constructed in the pUC118 backbone containing the 600 bp indicated on either side of an ~50-100 bp region to be replaced (the neutral site). B) shows the locations of the three neutral sites used in this study. Numbering on the chromosome corresponds to GenBank accession NC_000911. C) shows PCR confirmation of correct cassette insertion in all three mutants. Primers binding to the US and DS regions were used. The upper gel image is after growth in BG11 + 20 µg/mL of kanamycin, and the lower panel is after 1 month of growth with weekly subculturing in BG11 without antibiotic. "P", "M", and "WT" refer to use of suicide plasmid, mutant genomic DNA, or wild-type genomic DNA as PCR templates, respectively.

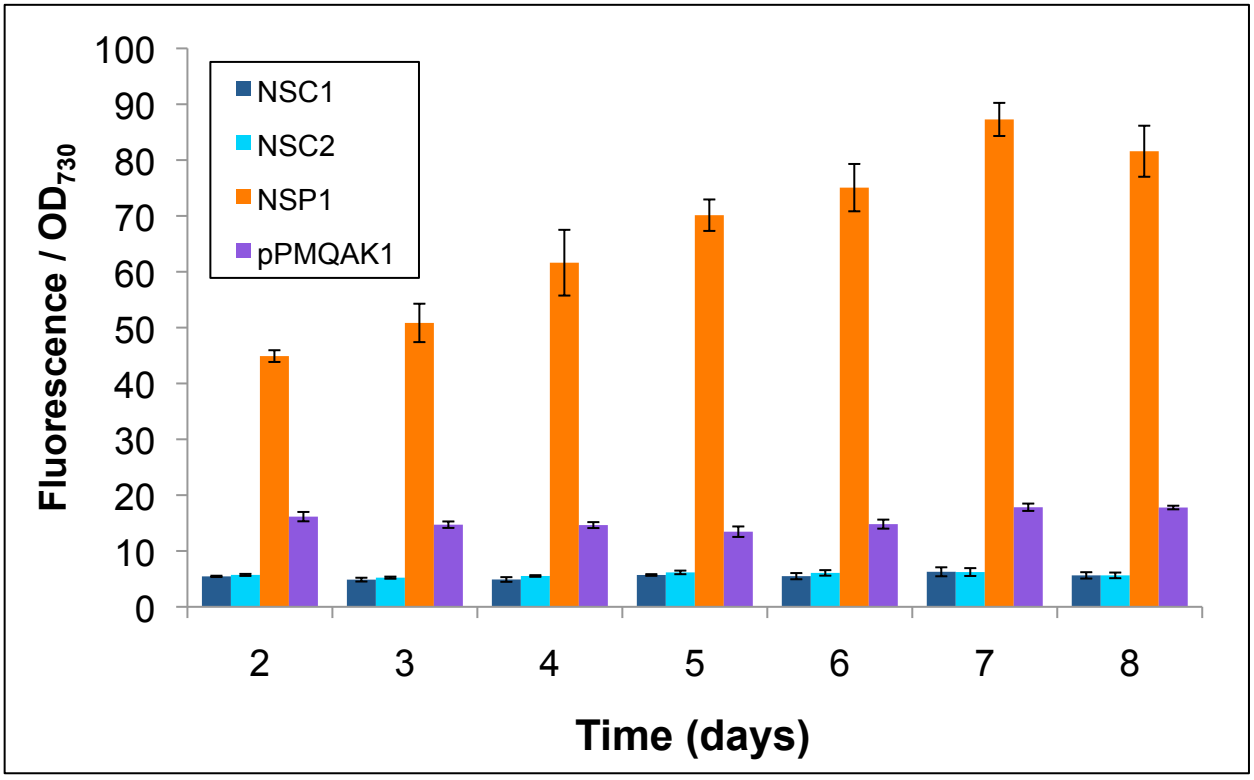


Figure 4.3: Different neutral sites influence expression level and timing. (A) Normalized fluorescence intensity over 8 days of growth for mutants expressing P_{trc}-EYFP in NSC1, NSC2, NSP1, or the medium copy-number plasmid pPMQAK1 (Huang *et al.* 2010). Error bars are ± SD for n=3 replicate cultures.

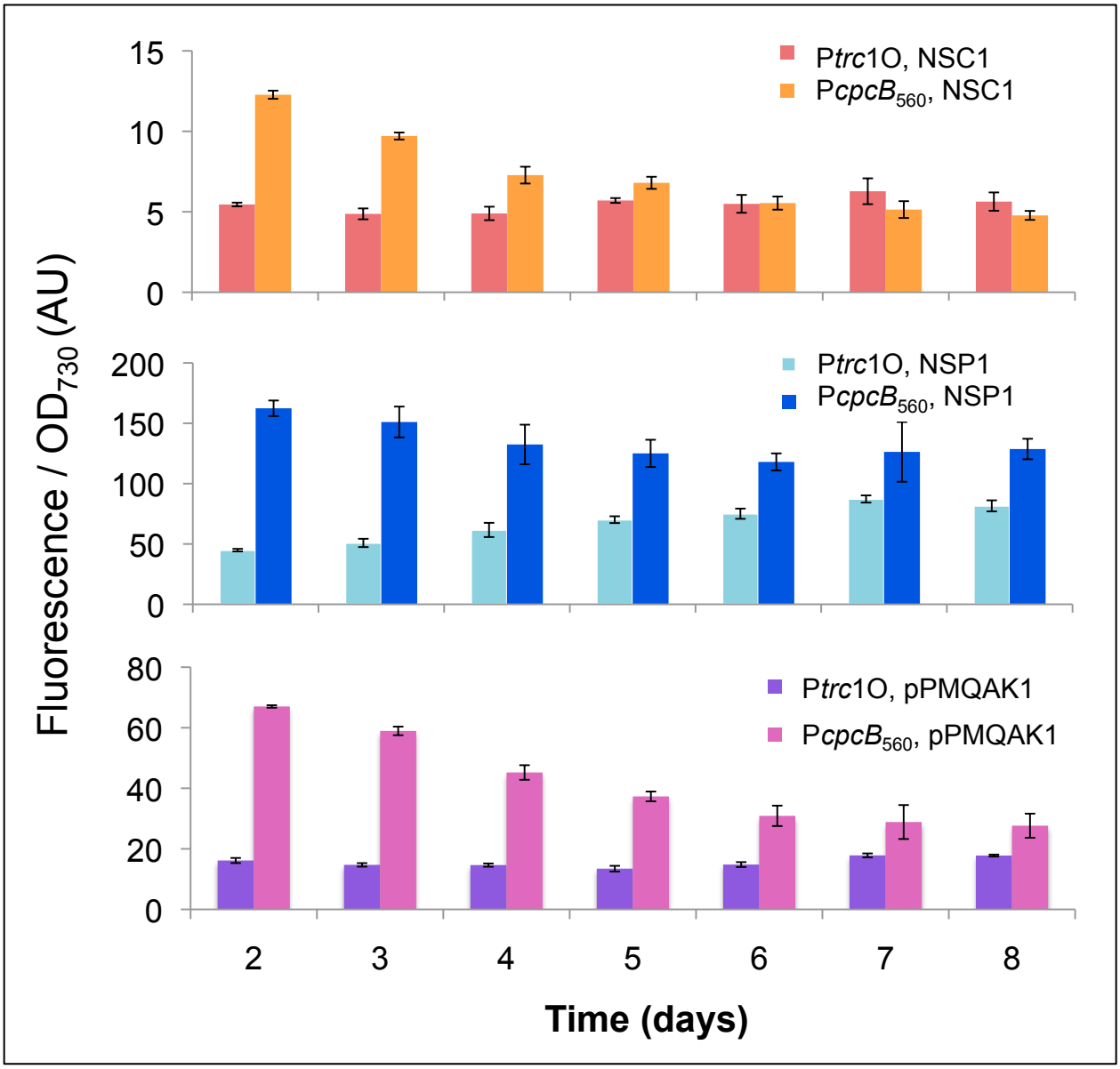


Figure 4.4: Composability of promoters with neutral sites. Fluorescence is shown from the EYFP cassette expressed on either NSC1 (upper chart), NSP1 (middle chart) or pPMQAK1 (lower chart). Error bars are ± SD for n=3 replicate cultures.

Chapter 5

Attempts to Overproduce Heptadecane in *Synechocystis* sp. PCC 6803

5.1 Introduction

5.1.1. Previous Attempts to Overproduce Cyanobacterial Alkanes

Since the initial discoveries of the genes responsible for alkane (Schirmer *et al.* 2010) and alkene (Mendez-Perez *et al.* 2011) production in cyanobacteria, a number of groups have attempted to overproduce these compounds in cyanobacteria and *E. coli*. However, success in these projects has been limited. The highest titer so far achieved in *E. coli* has been 7 μg mL^{-1} (Howard *et al.* 2013) and in cyanobacteria it has been 5.5 μg mL^{-1} OD_{730}^{-1}, from *Synechococcus* sp. PCC 7002 (Mendez-Perez *et al.* 2011). Synechocystis 6803 has been engineered to produce 1.2 μg mL^{-1} OD_{730}^{-1} of heptadecane and heptadecene, approximately 8 times more than the reference wild-type strain (Wang *et al.* 2013). In my own attempts to increase alkane production, low temperatures led to increased alkane production per cell (~2-fold, see chapter 2 of this dissertation).

A number of studies using metabolic engineering strategies, different environmental conditions, or alternative end products have sought to increase the output of either the ADO-type of PKS-type pathways from cyanobacteria (See chapter 2 of this work or Coates *et al.* (2014)). These studies are outlined in table 5.1. Wang *et al.* (2013) expressed two heterologous copies of the native *ado-far* cassette from alternate sites on the chromosome of *Synechocystis* 6803. This led to a more than 8-fold increase in heptadecane and heptadecene production to 1.3% of dry cell weight. They also eliminated *ddh* (*slr1556*) in an attempt to redirect carbon flux from pyruvate towards their end product, though their data do not make it clear that the deletion had any impact on their intended product yield. Also, while they added a strong *rbc* promoter to the 5' end of their over-expression cassette, more recent work has shown that several alternative transcripts exists within the wild-type *sll0208-sll0209* cassette encoding *ado* and *far,* and that *far* has a

separate promoter from *ado* (Klahn *et al.* 2014). It is also interesting to note that Wang *et al.* (2013) detected 3 different hydrocarbons in their wild-type strain: pentadecane, heptadecane, and heptadecene. In my own work and in other published work (Coates *et al.* 2014), only heptadecane has been detected in *Synechocystis* 6803, while the other two compounds have been detected in other cyanobacteria. Unsaturated hydrocarbons have only been detected in strains possessing the PKS-type pathways, which is absent in *Synechocystis* 6803. *Synechococcus* sp. PCC 7002 has also been engineered to overproduce the alkene products of its PKS-type pathway (see chapter 2 of this work). By replacing the promoter of the olefin synthase with the promoter of *psbA* from *Amaranthus hybridus*, they achieved a 2.5-fold increase in alkene production to 4.2 μg mL^{-1} OD_{730}^{-1} (Mendez-Perez *et al.* 2011) although this same group reported higher alkene production than this (5.5 μg mL^{-1} OD_{730}^{-1}) in their wild-type strain in a later publication (Mendez-Perez *et al.* 2014). This difference could be due to different growth conditions in the two experiments, as detailed growth conditions were not reported in the earlier study. Alkanes have also been produced using ADO from *Nostoc punctiforme* in *E. coli* (Howard *et al.* 2013). By replacing the aldehyde synthase activity of cyanobacterial *far* with *luxCED* from *Photorabdus luminescens* in a synthetic pathway, they were able to produce 7 μg mL^{-1} of *n*-alkanes and alkenes between 13 and 17 carbons in length. Attempts have also been made to improve alkane biosynthesis via protein engineering. One identified problem with the use of the ADO-type pathway in vitro is that the enzyme generates H_2O_2 as a byproduct, which in turn poisons the reaction. Fusing a catalase from *E. coli* to ADO from *Nostoc punctiforme*, led to an approximately 6-fold improvement in *in vitro* enzymatic activity (Andre *et al.* 2013).

As discussed in chapter 2 of this work, low temperature also leads to increased alkane and alkene production in both wild-type *Synechocystis* 6803 and *Synechocystis* 7002 (Mendez-

Perez *et al.* 2014). In either strain, these products are necessary for optimal growth at low temperature. In *Synechocystis* 6803, low temperature leads to enhanced production of heptadecane by about 2-fold (See chapter 2 of this work). In *Synechoccus* 7002, low temperature (30 C, vs. an optimal growth temperature of 38 C) leads to unsaturation of 1-nonadecene to 1,14-nonadecadiene, although the total alkene pool is unchanged. At more extreme temperature (22 C) at which the strain grows poorly, the 19:1 alkene is decreased but the 19:2 alkadiene remains abundant.

Attempts have also been made to repurpose elements of the ADO-type pathway for making other biofuel molecules. *Synechococcus elongatus* PCC 7942 has been engineered to overexpress *far* along with WS/DGAT from *Acinetobacter baylyii* under control of P*trc* (Kaiser *et al.* 2013). Although the latter overexpression alone did not enhance alkane yield from the strain, the combination of those genes led to accumulation of wax esters in lipid bodies inside the cell. However, following induction of this pathway by IPTG, the cultures became inviable, possibly due to membrane disruption by the wax esters, which accumulated in droplets visible in electron micrographs. *Synechocystis* 6803 has also been engineered to produce alternative end-products from its ADO-type pathway (Yao *et al.* 2014). By deleting *ado* and replacing it with alcohol-forming acyl-coA reductase from *Marinobacter aquaeolei* VT8 under control of P*petE*, fatty alcohols and also free fatty acids accumulated to as much as 16 mg gDW^{-1}. This is similar to the yield of alkane achieved by the same group in earlier work (1.3% DCW), as mentioned above (Wang *et al.* 2013). Building on the above work and using the engineering strategies outlined in the previous chapters, I have also attempted to overproduce heptadecane in *Synechocystis* 6803, although these experiments have met with little success. This chapter

documents the strategies I have used and attempts to give insight as to why I have so far been unsuccessful.

5.1.2. The Complete Pathway for Heptadecane Biosynthesis in *Synechocystis* sp. PCC 6803

Figure 5.1 shows the ADO-type pathway by which *Synechocystis* 6803 converts CO_2 to heptadecane and formate, with emphasis on energy requirements for this pathway. Although it is not shown in the figure here, *Synechocystis* 6803 includes a formate dehydrogenase complex that can oxidize formate to CO_2 and recover 1 NADPH, or ATP via cyclic electron flow (see chapter 2 of this work). Using this pathway requires 225 photons to synthesize one molecule of heptadecane from CO_2. Given the heat of combustion of heptadecane (Prosen *et al.* 1945), the quantum requirement of heptadecane biosynthesis, and the energy content of light theoretically extractable by photosynthesis (Walker 2009; Blankenship *et al.* 2011), the maximum possible efficiency of photosynthetic heptadecane production is given by equation 5.1.

$$\frac{2{,}713\ \text{kcal}}{\text{mol } C_{17}H_{36}} \times \frac{1\ C_{17}H_{36}}{225\ \text{photons}} \times \frac{1\ \text{mole PAR}}{100\ \text{kcal light}} \times 100\% = 12.0\%\ \text{efficiency} \qquad (5.1)$$

However, this does not account for the synthesis of enzymes, the growth of the organism, shading within cultures, or any other sources of inefficiency.

5.1.3. A Note of Caution About This Chapter

This chapter documents experiments that are unpublished elsewhere and are, for the most part, negative results. I view the primary audience of this chapter to be other researchers interested in successfully overproducing heptadecane or another product in *Synechocystis* 6803 or some other microbial chassis. Towards that end, I wish to make all the data I have generated available to such researchers in this openly accessible forum in the hope that my failures might enable their success. An important caveat to this chapter is that many of these experiments have

not been replicated and their results should be interpreted with caution. If you are such a researcher, I look forward to reading of your success in the pages of a prestigious journal some day!

5.2. Materials and Methods

5.2.1. OptForce

We used the metabolic model *i*Syn731 of *Synechocystis* 6803 that we created (see chapter 3 of this work) to predict strategies for optimizing alkane production using the OptForce algorithm (Ranganathan *et al.* 2010). While the details of this protocol are beyond the scope of this dissertation, the basic procedure is to use data from Flux Balance Analysis (FBA) to predict the minimal set of metabolic interventions required to increase production of a given metabolite. First, FVA (Flux Variability Analysis) is run to determine the range of each metabolic flux associated with maximal biomass production. Second, FBA is re-run with the added constraint of over-producing the metabolite of interest to a target level. Third, the flux ranges in the first and second scenarios are compared. Non-overlapping fluxes require some intervention to achieve the overproduction target. Finally, genetic interventions are prioritized by defining the minimal set of interventions required to achieve a given overproduction target. The final output of the algorithm is a group of so-called FORCE sets of genes/reactions that must be over-expressed as well as the yield of metabolite of interest achievable by doing this.

5.2.2. Alkane Extraction and Analysis

Approximately 2 mL of culture (OD730 ~0.5, 10^9 cells/mL, or 20 μg chlorophyll/mL) was pelleted by centrifugation and combined with 1 mL of ethyl acetate and 0.5 mL of 0.1 mm glass beads in a screw-top microcentrifuge tube. Cells were lysed in a bead beater for 3 cycles of

1 minute, with 5 minutes rest between cycles, or in some cases in an MP Bio FastPrep at its maximum speed setting. Glass beads and debris were pelleted by centrifugation for 10 minutes at 16,000 x *g* and then the upper ethyl acetate layer was removed for analysis. Chlorophyll *a* was determined on a DW-2000 spectrophotometer according to the formula [chl *a*] (μg/mL) = $16.29(A_{665}) - 8.24(A_{652})$ (Lichtenthaler 1987). Usually, the ethyl acetate extract was diluted 50-fold in methanol for chlorophyll determination as the exctinction coefficients determined by Lichtenthaler (1987) were in methanol. This method gave good agreement for chlorophyll concentration with extraction by pure methanol. However, the lower miscibility of ethyl acetate with water made it a preferred solvent for GC-MS analysis. We also attempted to use other solvents such as THF and hexane, but these solvents either created difficulties by rapid evaporation, were incompatible with plastic labware, or took a much longer time to extract alkanes from wet biomass. Alkanes were determined on an Agilent 6890 GC-MS fitted with a 12 meter DB5-MS column as previously (Schirmer *et al.* 2010) and quantified by comparison with an n-heptadecane standard (Sigma-Aldrich, St. Louis, MO). Extracts were also compared with other standards including n-pentadecane and heptadecene to confirm the identity of the product.

5.2.3 Growth Conditions

Growth conditions for the experiments outlined below are summarized in table 5.1. The details of each experiment are given in the results section.

5.2.4. Mutant Construction and Genetic Parts

A list of mutant strains constructed for these studies is given in table 5.2, and a complete list of genetic parts used for their construction is given in Table 5.3. The plasmids were assembled using Circular Polymerase Extension Cloning (CPEC) (Quan *et al.* 2011) The detailed protocol is given in chapter 4 of this work. Briefly, parts were amplified from their sources by

PCR using primers having 5' overlaps of between 15 and 30 base pairs, to achieve a part-to-part annealing temperature of around 55 C. In most cases, these primers were automatically designed using SnapGene (GSL Biotech). After necessary purification, parts were assembled in a second thermocycling reaction without additional primers. Phusion DNA polymerase (Thermo Scientific) was used for all PCR and CPEC reactions. For all mutants made, ~10^9 cells resuspended in 200 μL of fresh BG11 of either wild-type *Synechocystis* 6803 or the alkane-free mutant described in chapter 2 of this work was transformed with ~1 ug of plasmid DNA, then maintained overnight in dim light at 30 C. The next day, the cells were plated on selective media containing 2-5 μg/mL of gentamicin and colonies appeared after ~5-9 days. Correct insertions of overproduction cassettes were confirmed via colony PCR and each strain was re-streaked several times before analysis of alkanes.

5.3. Results and Discussion

5.3.1. Environmental Conditions Affecting Alkane Production

We assessed the effects of several different environmental conditions on heptadecane production in wild-type *Synechocystis* 6803. These conditions included growth temperature, light intensity, and macronutrient deprivation.

Figure 5.2 shows the effects of growth temperature on alkane production. We grew cultures for 2 days at 30, 25, or 20 C and then diluted those cultures to OD = 0.05. For the culture pre-incubated at 20 C, we transferred daughter cultures to both 20 C and 30 C. After 3 and 8 days, we took samples and measured the heptadecane content. As shown in chapter 2 of this work, low temperatures led to increased alkane production. Here, we also show that the effect on alkane content is temporary. After 3 days of growth at 30 C, a culture pre-incubated at

20 C had the same alkane content as a culture grown continuously at 30 C. After 3 or 8 days of growth at a low temperature of 20 or 25 C, *Synechocystis* 6803 produced approximately twice as much heptadecane as at 30 C.

Another possible role that we had considered for heptadecane was as a storage molecule or alternative reductant sink. Because other storage molecules such as glycogen and PHB are known to accumulate at stationary phase in *Synechocystis* 6803 (Panda *et al.* 2007), we reasoned that heptadecane might also accumulate. However, we did not find that this was the case. During a 20-day culture, the heptadecane content increased in nearly exact proportion to optical density (figure 5.3). Similarly to stationary phase, nutrient deprivation has often been used to accumulate secondary metabolite storage compounds. Towards this end, we examined the effect of macronutrient depletion on alkane content of *Synechocystis* 6803. While nutrient deprivation did lead to decreased growth, alkanes continued to accumulate at a rate similar to the still-growing cultures in complete BG11 media. In the microarray data presented in chapter 4, *ado* and *far* transcripts did not decrease at stationary phase under autotrophic conditions, but *far* transcripts did decrease slightly under mixotrophic conditions.

In chapter 2 of this work, we showed that an alkane-free mutant has increased cyclic electron flow in photosynthesis and grows poorly at low temperature, and that the wild-type strain also uses increased cyclic electron flow at low temperature. Increased cyclic electron flow has also been associated with high-light conditions in *Synechococcus* sp. PCC 7002 (Marathe *et al.* 2012). Thus, we were interested to see whether high light conditions might lead to increased alkane content and how high light might interact with low temperature. We grew wild-type cultures in BG11 media at a range of light intensities between 25 and 300 $\mu E\ m^{-2}\ s^{-1}$ and at either 20 or 30 C. While we again observed increased alkane content at low temperature, it does not

appear that high light led to over-accumulation of alkane. More heptadecane was produced per cell and per chlorophyll, but only at light intensities that caused chlorosis and/or poor growth.

5.3.2. Mutant Strains for Alkane Production

Several converging lines of evidence led us to attempt to overexpress a 5-gene cluster for alkane production from *Cyanothece* sp. PCC 7425 to increase alkane production in *Synechcocystis* 6803. First, this cluster, which contains *ado* and *far* in addition to *accA*, a short-chain dehydrogenase and a GTP cyclohydrolase is found to exist in many of the cyanobacterial strains that contain the ADO-type pathway (Klahn *et al.* 2014). We have identified the 4th and 5th genes of this pathway as *fabG* and *folE* via BLAST searches in preparation of a model of *Cyanothece* 7425 (see appendix chapter 2 of this work). In addition, AccA catalyzes the first committed step in fatty acid biosynthesis and over-expression of this gene has been shown to lead to accumulation of free fatty acids in *Synechocystis* 6803 (Liu *et al.* 2011). Finally, the OptForce algorithm identified both *accA* and *fabG* as potential targets for overproduction of alkane in *Synechocystis* 6803. With the reactions encoded by these 2 genes overexpressed, OptForce predicted that *Synechocystis* could potentially produce heptadecane from CO_2 at as much as 87% of its maximum theoretical yield. Based on this evidence, we created 4 strains (T2303-T2306) in the Δ*ado*Δ*far* background as outlined in Table 5.2. In each case, the strain contained P*psbA2* consisting of 250 bp immediately upstream of the start codon for *psbA2* from *Synechocystis* 6803 along with either *ado* and *far*; *ado, far,* and *accA; ado, far, accA,* and *fabG*; or *ado, far, accA, fabG*, and *folE* from *Cyanothece* sp. PCC 7425. In each case, the fragment started with the start codon of *ado* and ended with the stop codon of the final included gene. While *ado* and *far* were sufficient for alkane production in the mutant background, the production was only around 20% of that observed in the wild-type strain (figure 5.6). Addition of

accA and *fabG* both contributed to increased alkane production, but *folE* had a negative effect. However, none of these strains made as much heptadecane as the wild-type strain. One reason for this low production might be that, as discussed in chapter 4, P*psbA2* has often been used as a strong promoter, but its activity is not actually very high. According to my analysis of the microarray data presented in chapter 4 of this work, the *psbA2* transcript appears to be less abundant than the transcripts for *far* and *ado* in the wild-type strain based on microarray spot intensities.

In an attempt to improve output of the alkanes pathway, we constructed another set of mutants expressing the same alkane cluster as in T2306 and also in the knockout mutant background, but under control of several different promoters. Figure 5.7 shows the results of these experiments. Mutants with P*trc*1O and P$cpcB_{250}$ (T2310 and T2311) still gave lower alkane content that the wild-type, but similar output to T2306 with P*psbA2*. Mutants with P*pSysA_116* and P*slr9003* gave very low alkane production that was near the lower end of the detection range in these experiments. As discussed in chapter 4, these do not appear to be useful promoters for heterologous protein expression despite their apparent high activity in our microarray data. More recent data on transcription start sites in *Synechocystis* 6803 (Mitschke *et al.* 2011) as well as promoter engineering experiments (Markley *et al.* 2014; Zhou *et al.* 2014) also seem to indicate that this shorter *cpcB* 5' fragment does not actually contain the promoter for this gene, which resides farther upstream from the start codon. In light of recent data showing that additional copies of the native *ado* and *far* led to increased alkane production in *Synechocystis* 6803, we attempted to use the super-strong P$cpcB_{560}$ discussed in chapter 4 of this work to drive enhanced alkane production. In the wild-type background, we used this promoter to drive expression of alkane biosynthetic genes from the NSP1 locus (see table 5.2): *ado* and *far* from *Cyanothece*

7425, with and without the addition of *accA* and *fabG* from this same strain. These strains grew on plates under autotrophic conditions, but required glucose (5 mM) for growth in liquid BG-11. While this was a promising sign that they might be redirecting metabolic flux towards heptadecane, none of these strains produced detectably more alkane than the wild-type strain (Figure 5.8). It is possible that the metabolic burden of overexpressing multiple enzymes under such a strong promoter from the NSP1 locus was simply too high, limiting synthesis of other necessary proteins. Another possibility that should be tested is that the strains are exporting alkanes to the media under these conditions. So far, I have only tested for alkanes in the cell pellet of cultures of these strains.

One potential reason for the inability of all of these strains to produce more alkane is that the operon structure we believed to exist when designing them actually does not exist. More recent studies of transcription start sites in a number of different strains suggest that the *far* genes that reside downstream of *ado* in so many cyanobacterial strains have their own separate promoters (Klahn *et al.* 2014), and so it is not necessarily the case that placement of a strong promoter upstream of the cluster will lead to overexpression of the entire cluster. Further experiments will be needed to both understand the structure of gene clusters coding for alkane biosynthesis and for engineering balanced overexpression of all of the necessary genes. As tools for synthetic biology in cyanobacteria develop, including *in silico* models and parts for control of transcription and translation, this should become increasingly possible.

5.4. Conclusions

Although I have so far been unsuccessful in enhancing alkane production in *Synechocystis* 6803, this thesis represents significant progress towards that eventual goal.

Working with labmates and collaborators, I have developed tools for the *in silico* analysis of this promising strain (see chapter 3 and appendix chapter 2), Have advanced understanding of cyanobacterial carbon and light metabolism (see chapter 2 and appendix chapters 1, 3, and 4) and have created tools for the engineering of *Synechocystis* 6803 (see chapter 4). I have applied these tools in attempts to engineer enhanced heptadecane production and have generated some promising initial data that suggests that with further development, strains can be engineered to produce heptadecane in larger quantities, or any other biofuel molecule of interest.

5.5. References

Andre, C., Kim, S. W., Yu, X. H. and Shanklin, J. (2013). "Fusing catalase to an alkane-producing enzyme maintains enzymatic activity by converting the inhibitory byproduct H2O2 to the cosubstrate O2." *Proc Natl Acad Sci U S A* **110**(8): 3191-3196.

Blankenship, R. E., Tiede, D. M., Barber, J., Brudvig, G. W., Fleming, G., Ghirardi, M., Gunner, M. R., Junge, W., Kramer, D. M., Melis, A., Moore, T. A., Moser, C. C., Nocera, D. G., Nozik, A. J., Ort, D. R., Parson, W. W., Prince, R. C. and Sayre, R. T. (2011). "Comparing photosynthetic and photovoltaic efficiencies and recognizing the potential for improvement." *Science* **332**(6031): 805-809.

Coates, R. C., Podell, S., Korobeynikov, A., Lapidus, A., Pevzner, P., Sherman, D. H., Allen, E. E., Gerwick, L. and Gerwick, W. H. (2014). "Characterization of cyanobacterial hydrocarbon composition and distribution of biosynthetic pathways." *PLoS One* **9**(1): e85140.

Howard, T. P., Middelhaufe, S., Moore, K., Edner, C., Kolak, D. M., Taylor, G. N., Parker, D. A., Lee, R., Smirnoff, N., Aves, S. J. and Love, J. (2013). "Synthesis of customized petroleum-replica fuel molecules by targeted modification of free fatty acid pools in Escherichia coli." *Proc Natl Acad Sci U S A* **110**(19): 7636-7641.

Kaiser, B. K., Carleton, M., Hickman, J. W., Miller, C., Lawson, D., Budde, M., Warrener, P., Paredes, A., Mullapudi, S., Navarro, P., Cross, F. and Roberts, J. M. (2013). "Fatty aldehydes in cyanobacteria are a metabolically flexible precursor for a diversity of biofuel products." *PLoS One* **8**(3): e58307.

Klahn, S., Baumgartner, D., Pfreundt, U., Voigt, K., Schon, V., Steglich, C. and Hess, W. R. (2014). "Alkane Biosynthesis Genes in Cyanobacteria and Their Transcriptional Organization." *Front Bioeng Biotechnol* **2**: 24.

Lichtenthaler, H. K. (1987). "Chlorophylls and carotenoids: Pigments of photosynthetic biomembranes " *Methods in Enzymology* **148**: 350-382.

Liu, X., Sheng, J. and Curtiss, R., 3rd (2011). "Fatty acid production in genetically modified cyanobacteria." *Proc Natl Acad Sci U S A* **108**(17): 6899-6904.

Marathe, A. and Kallas, T. (2012). *Cyclic Electron Transfer Pathways in Syncechococcus sp. PCC 7002 Cyanobacteria During Photosynthesis At High Light Intensity*. Master of Science - Biology, The University of Wisconsin Oshkosh.

Markley, A. L., Begemann, M. B., Clarke, R. E., Gordon, G. C. and Pfleger, B. F. (2014). "Synthetic Biology Toolbox for Controlling Gene Expression in the Cyanobacterium Synechococcus sp. strain PCC 7002." *ACS Synth Biol*.

Mendez-Perez, D., Begemann, M. B. and Pfleger, B. F. (2011). "Modular synthase-encoding gene involved in alpha-olefin biosynthesis in Synechococcus sp. strain PCC 7002." *Appl Environ Microbiol* **77**(12): 4264-4267.

Mendez-Perez, D., Herman, N. A. and Pfleger, B. F. (2014). "A desaturase gene involved in the formation of 1,14-nonadecadiene in Synechococcus sp. strain PCC 7002." *Appl Environ Microbiol* **80**(19): 6073-6079.

Mitschke, J., Georg, J., Scholz, I., Sharma, C. M., Dienst, D., Bantscheff, J., Voss, B., Steglich, C., Wilde, A., Vogel, J. and Hess, W. R. (2011). "An experimentally anchored map of

transcriptional start sites in the model cyanobacterium Synechocystis sp. PCC6803." *Proc Natl Acad Sci U S A* **108**(5): 2124-2129.
Panda, B. and Mallick, N. (2007). "Enhanced poly-beta-hydroxybutyrate accumulation in a unicellular cyanobacterium, Synechocystis sp. PCC 6803." *Lett Appl Microbiol* **44**(2): 194-198.
Prosen, E. and Rossini, F. (1945). "Heats of combustion and formation of the paraffin hydrocarbons at 25° C." *Journal of Research of the National Bureau of Standards* **13**(263-267).
Quan, J. and Tian, J. (2011). "Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries." *Nat Protoc* **6**(2): 242-251.
Ranganathan, S., Suthers, P. F. and Maranas, C. D. (2010). "OptForce: An Optimization Procedure for Identifying All Genetic Manipulations Leading to Targeted Overproductions." *Plos Computational Biology* **6**(4).
Schirmer, A., Rude, M. A., Li, X., Popova, E. and del Cardayre, S. B. (2010). "Microbial biosynthesis of alkanes." *Science* **329**(5991): 559-562.
Walker, D. (2009). "Biofuels, facts, fantasy, and feasibility." *Journal of Applied Phycology* **21**: 509-517.
Wang, W., Liu, X. and Lu, X. (2013). "Engineering cyanobacteria to improve photosynthetic production of alka(e)nes." *Biotechnol Biofuels* **6**(1): 69.
Yao, L., Qi, F., Tan, X. and Lu, X. (2014). "Improved production of fatty alcohols in cyanobacteria by metabolic engineering." *Biotechnol Biofuels* **7**: 94.
Zhou, J., Zhang, H., Meng, H., Zhu, Y., Bao, G., Zhang, Y., Li, Y. and Ma, Y. (2014). "Discovery of a super-strong promoter enables efficient production of heterologous proteins in cyanobacteria." *Sci Rep* **4**: 4500.

Table 5.1: Previous attempts to overproduce cyanobacterial alkanes.

Host	Yield	Fold-change vs. wild-type	Reference	Notes
Metabolic Engineering				
Synechocystis 6803	1.2 μg mL^{-1} OD_{730}^{-1}	8.3	Wang *et al.* 2013	Overexpressed 2 extra copies of *far* and *ado* with P*rbc* and knocked out *ddh*.
E. coli	7 μg mL^{-1}	n/a	Howard *et al.* 2013	A synthetic pathway composed of *luxCED* from *Photorabdus luminescens* and *ado* from *Nostoc punctiforme* yielded mixed alkanes and alkenes between 13 and 17 carbons, as well as fatty alcohols.
Synechococcus 7002	4.2 μg mL^{-1} OD_{730}^{-1}	2.5	Mendez-Perez *et al.* 2011	Replaced WT promoter of *ols* with *PsbA* promoter. Yield given is for the sum of 19:1 and 19:2 alkenes.
E. coli	120 mol alkanes mol $protein^{-1}$	6	Andre *et al.* 2013	*In vitro* activity only. Fused catalase from *E. coli* to ADO from *Prochlorococcus marinus*.
Environmental Conditions				
Synechococcus 7002	5.5 μg mL^{-1} OD_{730}^{-1}	No change	Mendez-Perez *et al.* 2014	WT strain, growth at 22 C (vs. 38 C) increases the abundance of 19:2 alkadiene ~5-fold while decreasing the total alkene pool.
Synechocystis 6803	27 μg mL^{-1} OD_{730}^{-1}	n/a	This work (Chapter 2)	WT strain grown at 20 C under 250 μE m^{-2} s^{-1} of light in bubble column. In these conditions, the strain was highly chlorotic.
Alternative Products				
Synechococcus 7942	Lipid bodies – not quantified	n/a	Kaiser *et al.* 2013	Produced wax esters from fatty aldehyde using WS/DGAT from *Acinetobacter baylyii* and overexpression of *far* from P*trc*. Alkane yield was not enhanced by the latter overexpression. Wax esters accumulated in lipid bodies and led to cell death.
Synechocystis 6803	16 mg gDW^{-1}	n/a	Yao *et al.* 2014	By deleting *ado* and replacing it with alcohol-forming acyl-coA reductase from *Marinobacter aquaeolei* VT8 under P*petE*, fatty alcohols and also free fatty acids accumulated.

Table 5.2: Conditions and strains tested for alkane overproduction.

Environmental Conditions		
Experiment	**Result**	
Low temp (20, 25C)	Increased allocation by ~2-fold, with very little decrease in growth.	
Stationary phase	No change with growth phase.	
N/P/S starvation	Alkanes continued to accumulate after growth ceased at a similar rate to cultures grown in complete media.	
High light	Alkanes per cell were slightly higher at very high light intensities, but only at intensities that induced chlorosis/death.	
Genetic Interventions		
Strain	**Genotype**	**Results**
T2303	Δsll0208_9::P*psbA2_alkAB*$_{7425}$	~1/5th of WT alkane production
T2304	Δsll0208_9::P*psbA2_alkABC*$_{7425}$	~1/4th of WT alkane production
T2305	Δsll0208_9::P*psbA2_alkABCD*$_{7425}$	~1/2 of WT alkane production
T2306	Δsll0208_9::P*psbA2_alkABCDE*$_{7425}$	~1/4th of WT alkane production
T2310	Δsll0208_9::P*trc*1O_*alkABCDE*$_{7425}$	~1/4th of WT alkane production
T2311	Δsll0208_9::P*cpcB*$_{250}$_*alkABCDE*$_{7425}$	~1/6th of WT alkane production
T2312	Δsll0208_9::P*pSysA_116_alkABCDE*$_{7425}$	< 1/10th of WT alkane production
T2313	Δsll0208_9::P*slr9003_alkABC*$_{7425}$	< 1/10th of WT alkane production
T2408	NSP1::PcpcB$_{560}$_*alkAB*$_{7425}$	Strain grows poorly under autotrophic conditions, no increased alkane content in cells.
T2410	NSP1::PcpcB$_{560}$_*alkABCD*$_{7425}$	Strain grows poorly under autotrophic conditions, no increased alkane content in cells.

Table 5.3: Genetic parts used in these studies.

Part	**Source**	**Description**
P*cpcB*$_{560}$	*Synechocystis* sp. PCC 6803	Strong promoter (see chapter 4 of this work)
P*cpcB*$_{250}$	*Synechocystis* sp. PCC 6803	Putative strong promoter (see chapter 4 of this work). This fragment probably does not actually contain P*cpcB*, but only the downstream portions of the 5' UTR for this gene.
P*trc*1O	pPMQAK1	Strong promoter (see chapter 4 of this work)
P*pSysA116*	*Synechocystis* sp. PCC 6803	Putative strong promoter (See chapter 4 of this work)
P*slr9003*	*Synechocystis* sp. PCC 6803	Putative strong promoter (See chapter 4 of this work)
alkABCDE$_{7425}$ cluster	*Cyanothece* sp. PCC 7425	This 5-gene cluster contains *ado*, *far*, *accA*, *fabG*, and *folE*.
sll0208	*Synechocystis* sp. PCC 6803	This locus contains *ado*.
sll0209	*Synechocystis* sp. PCC 6803	This locus contains *far*.

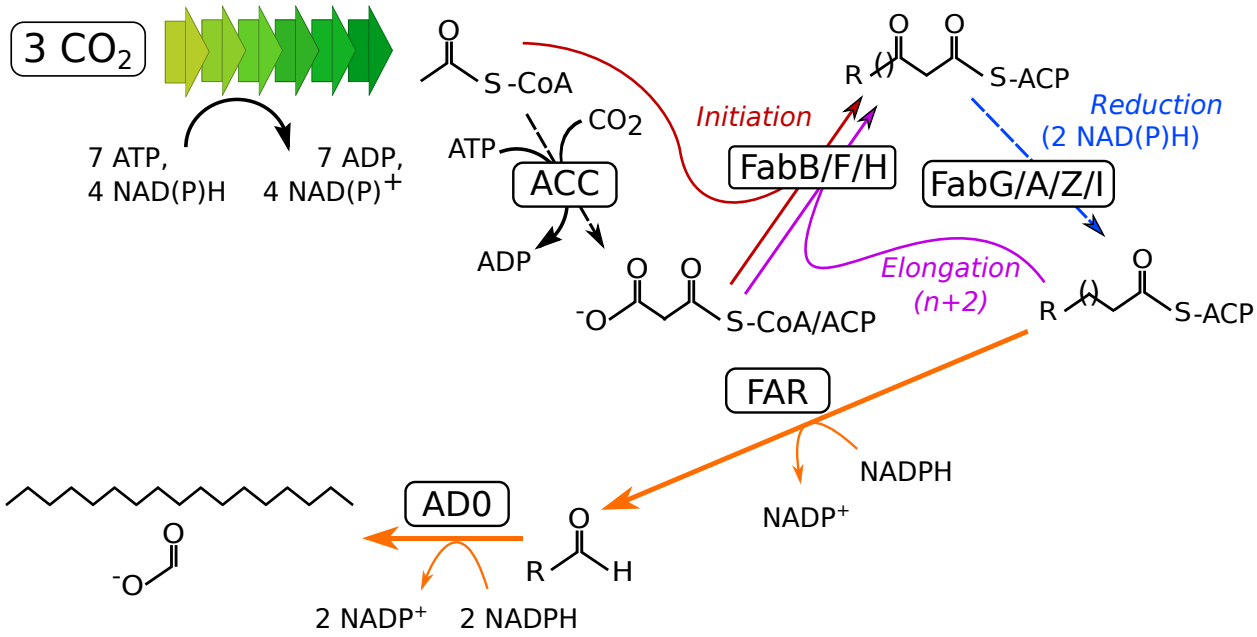


Figure 5.1: The pathway for production of n-heptadecane from CO2 in *Synechocystis* sp. PCC 6803. Synthesis of 1 molecule each of heptadecane and formate from CO_2 requires 57 NADPH and 72 ATP. These cofactors require a minimum of 225 photons for their photosynthesis, assuming that the formate generated is re-oxidized to CO_2 to recover NADPH.

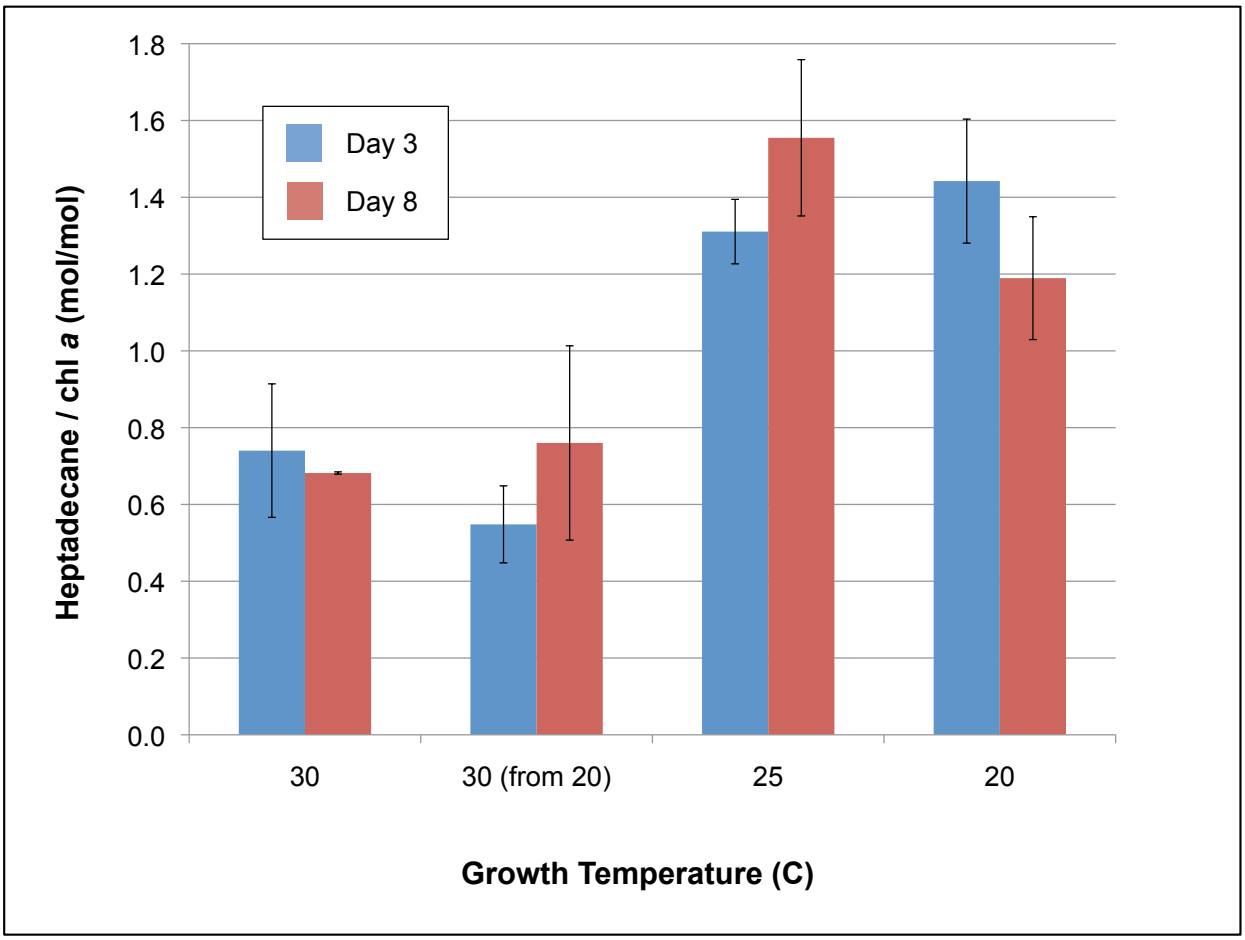


Figure 5.2: The effect of growth temperature on heptadecane content in *Synechocystis* 6803. Cultures were pre-incubated for 2 days at 20, 25, or 30 C, then diluted in fresh BG-11 media to OD = 0.05 and grown at that same temperature. Cultures pre-incubated at 20 C were also transferred to 30 C. All cultures were grown in shake flasks under 35 μE m^{-2} s^{-1} of white LED light. After 3 and 8 days, samples were taken, extracted with ethyl acetate, and heptadecane content was measured via GC-MS. Error bars are ± SD for n = 3 separate cultures.

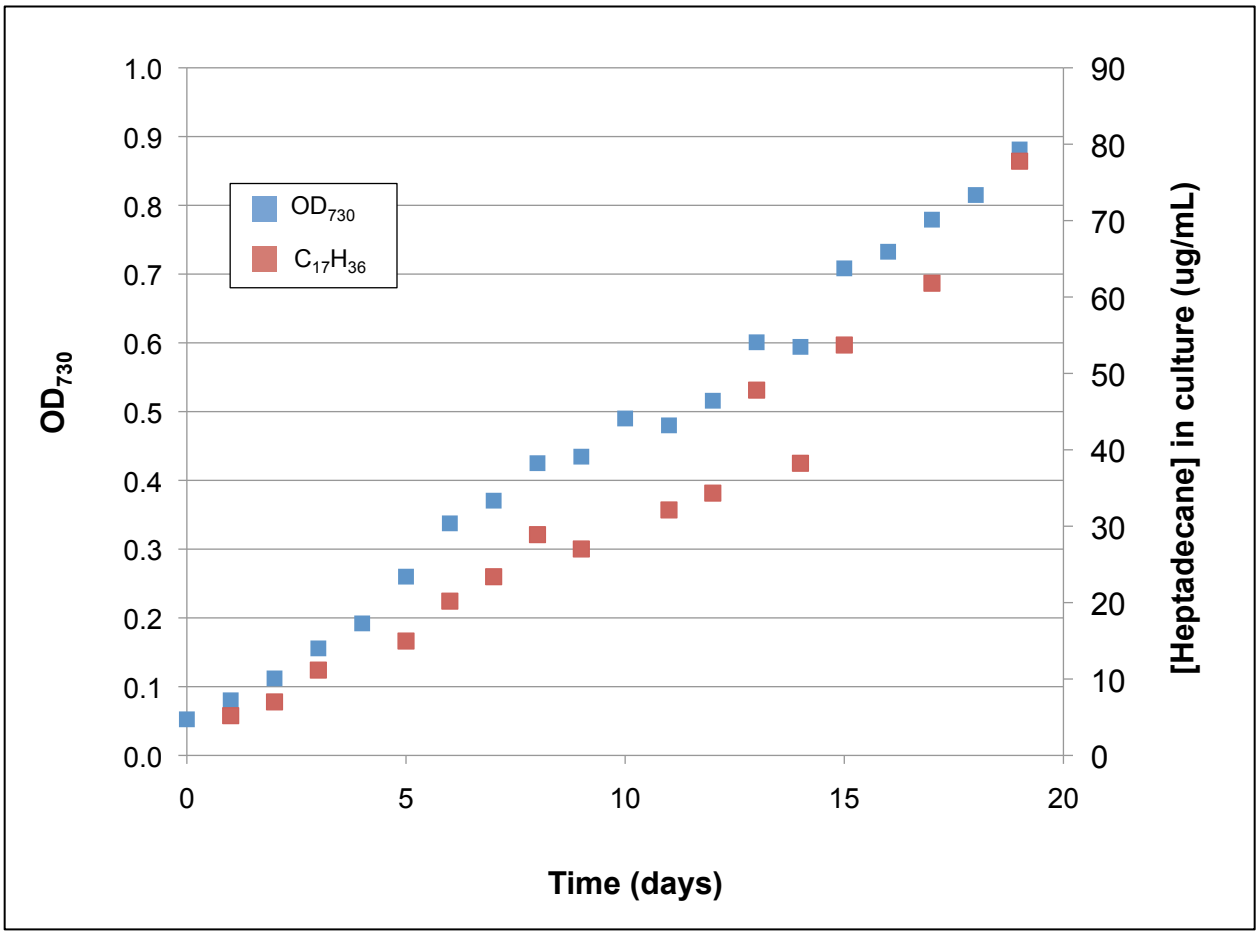


Figure 5.3: The effect of growth phase on heptadecane content in *Synechocystis* 6803. Cultures were grown in BG-11 media at 30 C in shake flasks, under ~50 µE m^{-2} s^{-1} of white fluorescent light. At specified intervals, samples were taken containing ~2 billion cells (estimated according to OD_{730}) extracted with ethyl acetate, and heptadecane content was measured via GC-MS. Data shown are the average of 2 separate cultures, although some data are missing and so some points represent only a single measurement.

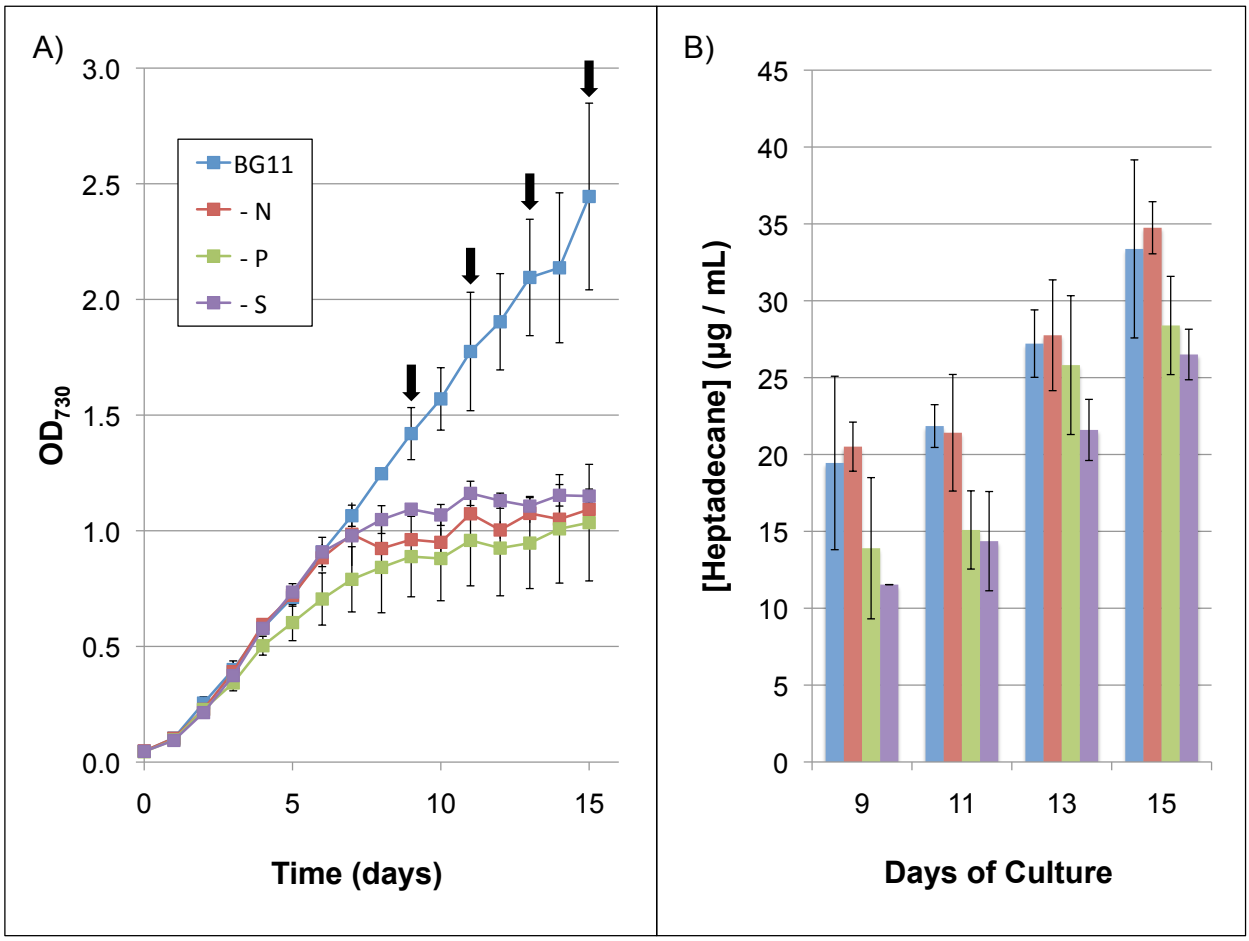


Figure 5.4: The effect of macronutrient deprivation on growth and alkane production in *Synechocystis* 6803. A single culture was grown for 5 days, then subcultured into depletion media. For N and S, the depletion media had 0.1X concentration of that nutrient, while the – P media was phosphate-free. OD730 was monitored daily and after 9, 11, 13, and 15 days, samples were taken extracted with ethyl acetate, and heptadecane content was measured via GC-MS. Error bars are ± SD for n = 3 separate cultures.

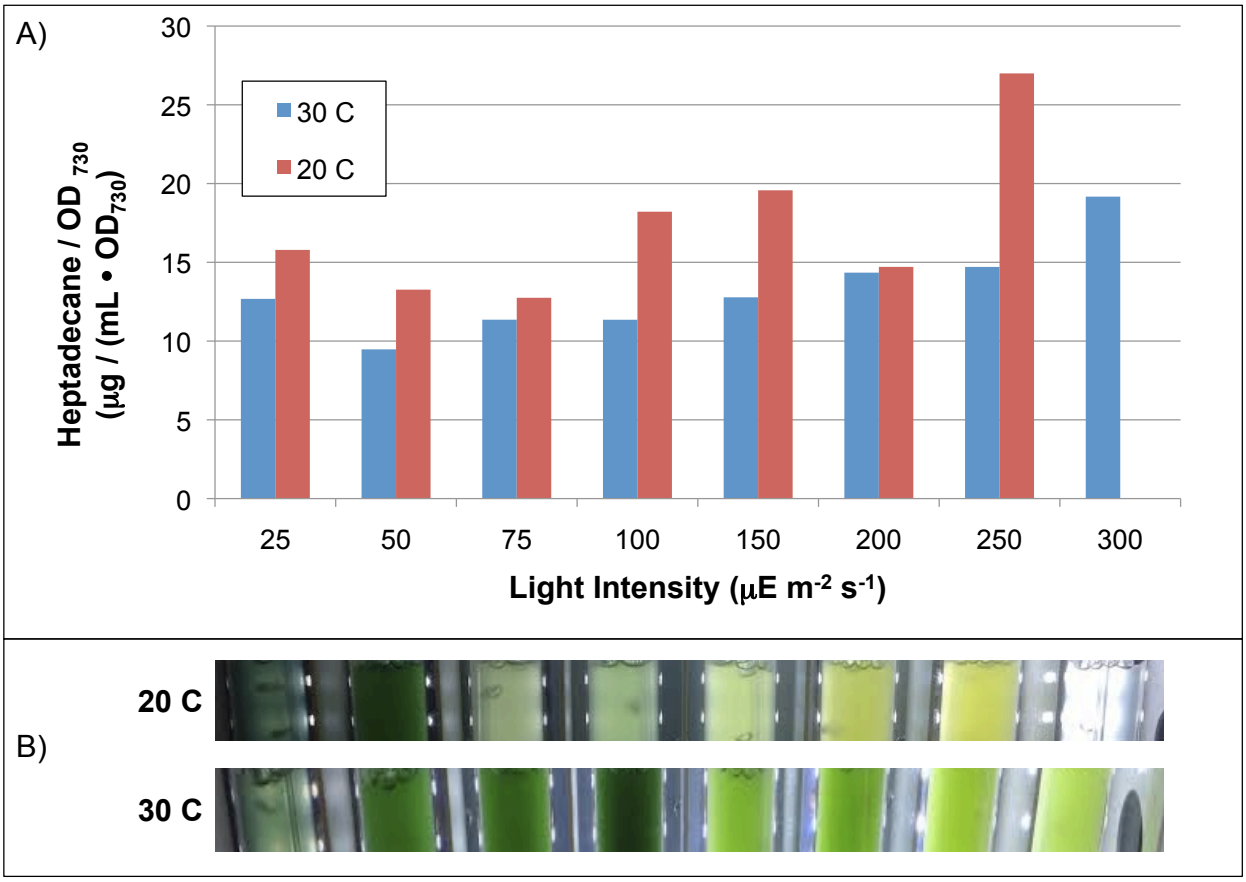


Figure 5.5: The effect of light intensity on alkane content in *Synechocystis* 6803. Cultures were grown at either 20 or 30 C and bubbled with air in the MC-1000 multicultivator (PSI, Czech Republic). Light intensity was varied from 25 to 300 µE m^{-2} s^{-1}. After several days of growth, samples were taken extracted with ethyl acetate, and heptadecane content was measured via GC-MS.

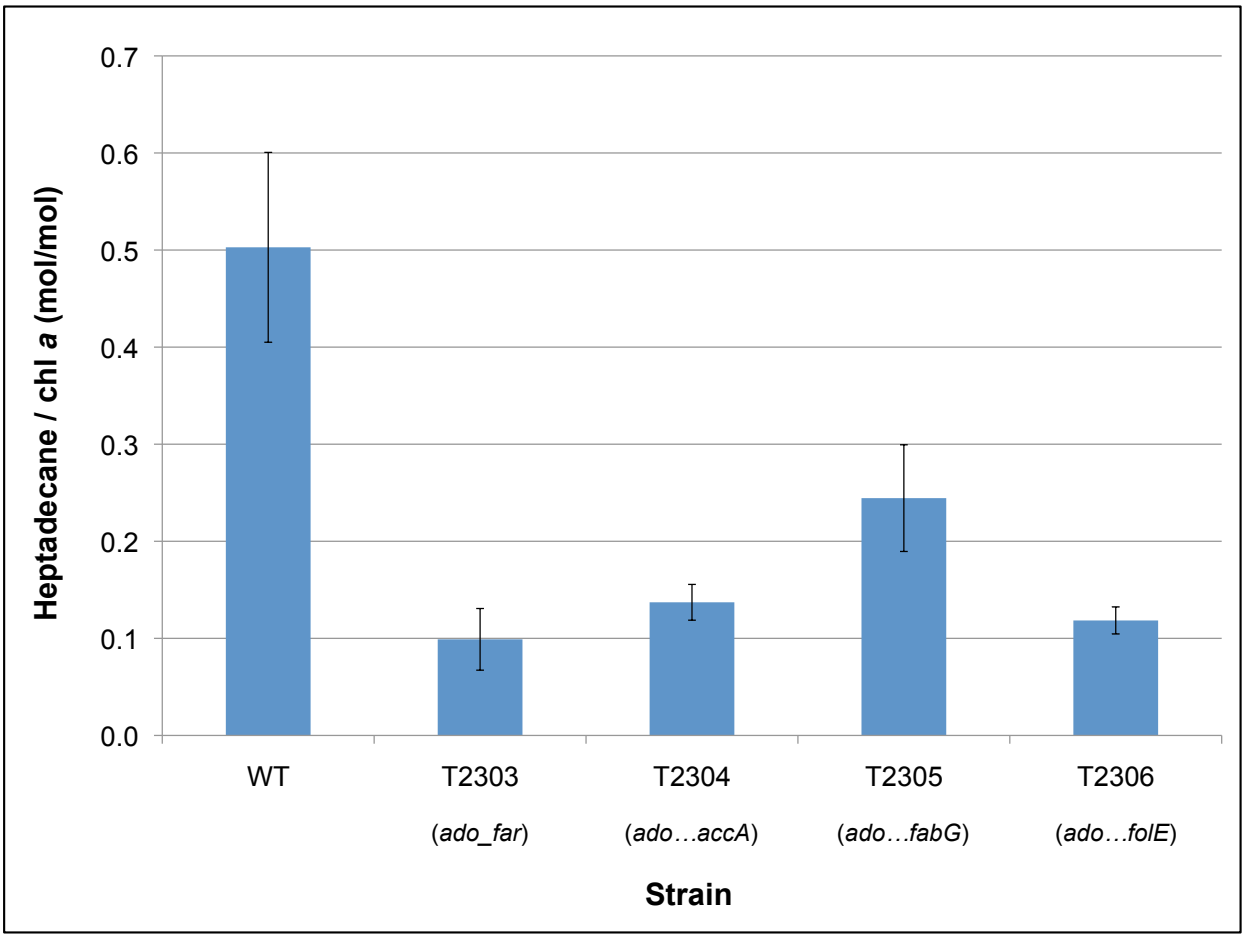


Figure 5.6: Production of alkanes in *Synechocystis* 6803 using the alkane biosynthesis cluster from *Cyanothece* 7425. Each strain was constructed in a Δ*ado*Δ*far* background and contained P*psbA2* driving either *ado* and *far* (T2303); *ado, far,* and *accA* (T2304)*; ado, far, accA,* and *fabG* (T2305); or *ado, far, accA, fabG*, and *folE* (T2306) from *Cyanothece* sp. PCC 7425. In each case, the fragment started with the start codon of *ado* and ended with the stop codon of the final included gene. Cultures were grown at 30 C in shake flasks with BG-11 media under ~50 μE m^{-2} s^{-1} of white fluorescent light. After several days of growth, samples were taken extracted with ethyl acetate, and heptadecane content was measured via GC-MS. Error bars represent ± SD for n = 3 separate cultures.

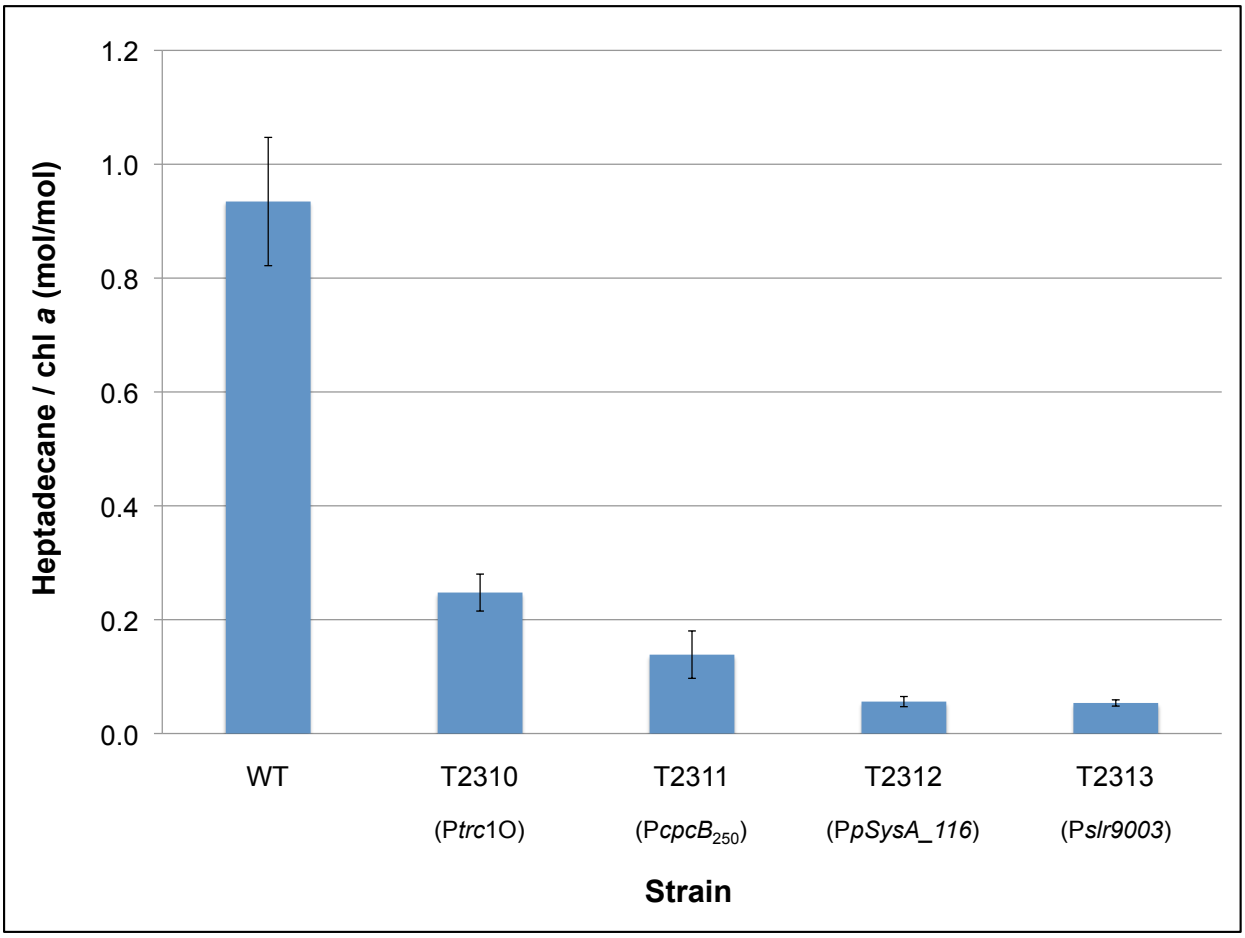


Figure 5.7: Production of alkanes using different promoters on alkane biosynthesis in *Synechocystis* 6803. Strain T2303 was modified to contain either P*trc1O*, P$cpcB_{250}$, P*pSysA_116*, or P*slr9003*. Cultures were grown at 30 C in shake flasks with BG-11 media under ~50 µE m^{-2} s^{-1} of white fluorescent light. After several days of growth, samples were taken extracted with ethyl acetate, and heptadecane content was measured via GC-MS. Error bars represent ± SD for n = 3 separate cultures.

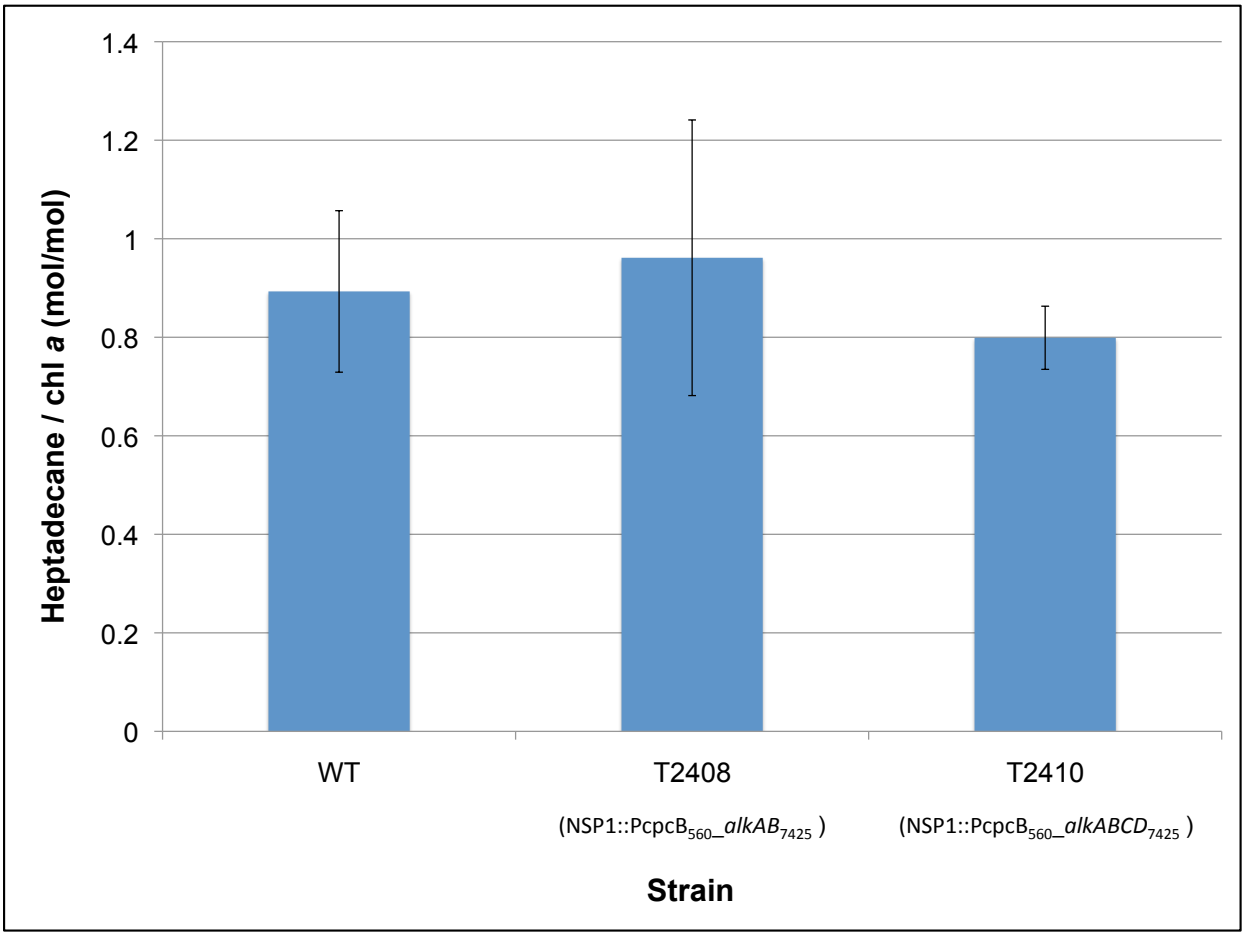


Figure 5.8: Production of alkanes in *Synechocystis* 6803 using the alkane biosynthesis cluster from *Cyanothece* 7425 and a neutral site on pCC5.2. Each strain was constructed in the wild-type background and contained P*cpcB*$_{560}$ driving *ado* and *far* from *Cyanothece* 7425, with (T2410) and without (T2408) the addition of *accA* and *fabG* from this same strain. Cultures were grown at 30 C in shake flasks with BG-11 media supplemented with 5 mM glucose under ~50 µE m^{-2} s^{-1} of white fluorescent light. After several days of growth, samples were taken extracted with ethyl acetate, and heptadecane content was measured via GC-MS. Error bars represent ± SD for n = 3 separate cultures.

Chapter 6

Future Directions

6.1. About this Chapter

Although the topics addressed in this brief chapter are also addressed elsewhere in this dissertation, they are collected here for the convenience of readers interested in what I believe is the future of these research topics. This chapter is roughly organized according to chapters of the dissertation.

6.2. Cyanobacterial Engineering for Alkane Overproduction

As detailed in chapter 5, our attempts so far to produce large amounts of heptadecane in *Synechocystis* sp. PCC 6803 have not been successful. However, recent progress on several fronts holds promise for these efforts. Two areas of particular interest are the regulation and mechanism of action of the enzymes responsible for alkane biosynthesis, AAR and ADO. Although AAR and ADO have been shown to be both necessary and sufficient for alkane production in a heterologous host (Schirmer *et al* 2010), it is not known what other genes might be involved in the process. It was recently shown that many cyanobacteria contain a genomic cluster associated with fatty acid biosynthesis and the ADO-type alkane biosynthesis pathway (Klahn *et al* 2014). While these genes often appear together in genomes, they appear not to be cotranscribed. Thus, refactoring and the creation of synthetic operons and other control circuits may be necessary or helpful to enhance alkane production. Further discussion of the transcription of genes associated with alkane biosynthesis can be found in chapter 5. In addition, the unusual deformylase mechanism of alkane biosynthesis used by the ADO-type pathway is still not fully understood. Very recently, it was discovered that ADO and AAR from *Nostoc punctiforme* appear to form a complex when expressed in *E. coli* (Warui *et al* 2014). Such a complex might

help to solve problems associated with the aldehyde intermediate that is both very insoluble and likely a very toxic compound to cells. It is possible that there are additional interaction partners and/or chaperone proteins that facilitate the integration of the final alkane product into the membrane. While we believe based on its chemistry that the alkane product would have to reside in the membrane, we have not shown this unequivocally and we do not know whether the hydrophobicity of this product would limit its overproduction. While hydrocarbons have an important role in helping cyanobacteria adapt to their environment at their natural concentration, the effects of higher concentrations remain unknown.

6.3. Synthetic Biology of Cyanobacteria

While chapter 1 of this dissertation is devoted to discussing this topic in much greater detail, I think this is an area that holds great promise and so deserves brief mention here. Great progress has been made in imagining and designing genetic circuits that confer all manner of functions in engineered biological systems. However, the bulk of this progress has been limited to *E. coli* and *S. cerevisiae*. Even basic tools like controllable promoters and plasmids with a range of copy numbers have only recently become available for use in cyanobacteria. As compared with heterotrophic hosts, cyanobacteria have a unique advantage in converting the pollutant carbon dioxide into useful and potentially profitable organic molecules. Tools that allow fine-tuning and autonomous control of transcription and translation in response to environmental cues will be especially important for harnessing the power of these promising bugs. For example, expression of energy-intensive pathways for biofuel synthesis should only occur when adequate energy from sunlight is available. For anaerobic production processes,

oxygenic photosynthesis must be separated, either in time or space, from the anaerobic portion of the process. Such tools are certainly within reach and could contribute greatly to the progress of this field. Control circuits that respond to the unique challenges of the cyanobacterial lifestyle such as diurnal rhythms will be especially important.

6.4. Cyclic Electron Flow as a Flexible Nutritional Strategy

Chapter 2 discusses the interaction between alkanes, cyclic electron flow, and chill tolerance in cyanobacteria. However, in the process of developing this chapter and conducting modeling studies, I came to believe that the role of cyclic electron flow is perhaps much broader than I had previously thought. CEF has often been regarded as a way to balance the ATP:NADPH ratio provided by linear electron flow (1.28:1) with the demands of carbon fixation (1.5:1) (Allen 2002). However, while carbon fixation is a major reductant sink, it is far from being the only reductant sink in a cell. In the modeling studies discussed in chapter 2 of this dissertation, CO_2 fixation by the Calvin Cycle required only about 59% of the energy produced in the form of ATP and NADPH by photosynthesis. This process accounted for 75% of NADPH produced and only 45% of ATP. While clearly the Calvin Cycle is a major consumer of cellular resources, it is far from being the only consumer, and so it is critical to consider the rest of metabolism, as well. During the development of figure 2.6, we experimented with different nitrogen sources and found that growth on ammonium instead of nitrate led to a nearly 2-fold higher optimum for the rate of cyclic electron flow. Presumably, this is because less NADPH was required to reduce nitrate to ammonium before incorporation into glutamate and other downstream metabolites. Thus, the overall ATP:NADPH ratio of metabolism was higher and more CEF was required for balanced growth. This view of cyclic electron flow as a valve that

regulates the balance between ATP and NADPH production is a significant departure from previous thinking, but we believe that it will lead to new insights into the function of this still controversial energy-generating pathway. Such an extension of this work is also appealing from my own perspective as a natural extension of the model, the development of which is described in chapter 3 of this dissertation. Flux balance modeling has often been associated with the field of metabolic engineering, but I believe that this technique can also have application in more basic sciences. For example, in this case we can use flux balance modeling to investigate the interaction between environmental conditions and modes of energy generation to meet the cell's metabolic needs.

6.5. References

Allen, J. (2002). "Photosynthesis of ATP-electrons, proton pumps, rotors, and poise." *Cell* **110**(3): 273-276.

Klahn, S., Baumgartner, D., Pfreundt, U., Voigt, K., Schon, V., Steglich, C. and Hess, W. R. (2014). "Alkane Biosynthesis Genes in Cyanobacteria and Their Transcriptional Organization." *Front Bioeng Biotechnol* **2**: 24.

Schirmer, A., Rude, M. A., Li, X., Popova, E. and del Cardayre, S. B. (2010). "Microbial biosynthesis of alkanes." *Science* **329**(5991): 559-562.

Warui, D. M., Pandelia, M. E., Rajakovich, L. J., Krebs, C., Bollinger, J. M. and Booker, S. J. (2014). "Efficient Delivery of Long-Chain Fatty Aldehydes from the Nostoc punctiforme Acyl-Acyl Carrier Protein Reductase to its Cognate Aldehyde Deformylating Oxygenase." *Biochemistry* 10.1021/bi500847u.

Appendix Chapter 1

^{13}C-MFA Delineates the Photomixotrophic Metabolism of *Synechocystis* sp. PCC 6803 Under Light- and Carbon-Sufficient Conditions

A1.1. Abstract of the Chapter

The topology of cyanobacterial central carbon metabolism remains under debate. For over 50 years, the lack of α-ketoglutarate dehydrogenase has led to the belief that cyanobacteria have an incomplete TCA cycle. Recent *in vitro* enzymatic experiments and *in silico* models suggest that this cycle may in fact be closed. In this study, we employed ^{13}C isotopomers to delineate central pathways in cyanobacterium *Synechocystis* sp. PCC 6803. By tracing the incorporation of supplemented glutamate into the downstream metabolites in the TCA cycle, we observed a direct *in vivo* transformation of α-ketoglutarate to succinate. In addition, isotopic tracing of glyoxylate didn't show a functional glyoxylate shunt and glyoxylate was used for glycine synthesis. The photomixotrophic central carbon metabolism was then profiled with ^{13}C-MFA under light and carbon sufficient conditions. We observed that 1) the *in vivo* flux through the TCA cycle reactions (α-ketoglutarate → succinate) was minimal (<2%); 2) the relative flux of CO_2 fixation was six times higher than that of glucose utilization; 3) the relative flux through the oxidative pentose phosphate pathway was low (<2 %). Our ^{13}C-MFA results improve the understanding of the versatile metabolism in cyanobacteria and will shed a light on their application for biosynthesis of various valuable chemical compounds.

A1.2. Introduction

Synechocystis sp. PCC 6803 is a naturally transformable cyanobacterium (Grigorieva *et al.* 1982) and a model organism for studying photosynthesis (Berry *et al.* 2002). *Synechocystis* 6803 and other *cyanobacterial species* are promising phototrophic cell factories for synthesis of valuable chemicals and biofuels (Deng *et al.* 1999; Chisti 2008; Atsumi *et al.* 2009; Lindberg *et*

al. 2010; Lan *et al.* 2011). To explore cyanobacterial metabolism for biotechnology applications, genomics and transcriptomics approaches have been used to study *Synechocystis* 6803 (Yoshikawa *et al.* 2013). Complementing these approaches, fluxomics tools (flux balance analysis, FBA, and ^{13}C-metabolic flux analysis, ^{13}C-MFA) are also powerful in deciphering genome functions and unraveling cell phenotype in phototrophs under autotrophic, mixotrophic, and heterotrophic metabolisms (Yang *et al.* 2002; Yang *et al.* 2002; Shastri *et al.* 2005; Knoop *et al.* 2010; McKinlay *et al.* 2010; McKinlay *et al.* 2011; Yoshikawa *et al.* 2011; Young *et al.* 2011; Nogales *et al.* 2012; Saha *et al.* 2012; Knoop *et al.* 2013). These multi-omics studies have improved our understanding and application of cyanobacterial cell factories (Kohlstedt *et al.* 2010).

Nevertheless, cyanobacterial metabolism is still not completely resolved. Due to the lack of α-ketoglutarate dehydrogenase, cyanobacteria were thought to have an incomplete tricarboxylic acid (TCA) cycle (Smith *et al.* 1967; Pearce *et al.* 1969). This assumption has been employed in most cyanobacterial metabolic models so far (Yang *et al.* 2002; Shastri *et al.* 2005; Young *et al.* 2011). Recently, a pair of enzymes from *Synechococcus* sp. PCC 7002, α-ketoglutarate decarboxylase and succinic semialdehyde dehydrogenase, were found to transform α-ketoglutarate into succinate *in vitro* (Zhang *et al.* 2011). These two enzymes have homologues throughout the cyanobacterial phylum. Contemporaneously, Nogales *et al.* (2012) identified an overlapping GABA (γ-aminobutyric acid) shunt *in silico* that could also complete the TCA cycle via GABA and succinic semialdehyde. Such a pathway in cyanobacteria would help explain previous observations that α-ketoglutarate added to cultures of a *Synechocystis* double mutant strain (knockout succinate dehydrogenase and fumarate reductase) led to accumulation of

succinate (Cooley *et al.* 2000). However, an *in vivo* flux from α-ketoglutarate to succinate has not been measured (Knoop *et al.* 2010).

Another open question has been whether the glyoxylate shunt was active or even existed. Glyoxylate shunt activity in some cyanobacterial species were reported (Pearce *et al.* 1967; Eley 1988) and thus were included in the metabolic models of *Synechocystis* 6803 (Yang *et al.* 2002; Shastri *et al.* 2005; Fu 2009; Montagud *et al.* 2010). But homologues encoding isocitrate lyase and malate synthase have still not been found. Moreover, the oxidative pentose phosphate pathway (OPP pathway) in *Synechocystis* 6803 has been considered inactive under light conditions (Yang *et al.* 2002), but this pathway was recently proved to be highly active in photoautotrophic metabolism (Young *et al.* 2011). The previous application of ^{13}C-MFA to photomixotrophic metabolism in *Synechocystis* 6803 was operated in a CO_2-limited culture lacking carbonate and atmospheric CO_2 (Yang *et al.* 2002). Moreover, the application of ^{13}C-MFA requires the attainment of a metabolic steady state. In a photobioreactor, cyanobacteria are continuously moving between "light" (near surface of bioreactor) and "dark" zones (depending on mixing, cell density, and reactor size/geometry). In the light zone, cells fix CO_2 and accumulate glycogen, while they may use this storage component (or glucose in the medium) for "heterotrophic" metabolism in the dark zone. To minimize such heterogeneous growth, this study has performed ^{13}C-MFA experiments using small culture volume (<50mL) and low biomass density (OD_{730}<0.5) to ensure cell metabolism under light and carbon sufficient conditions. This approach may provide a better understanding of photomixotrophic metabolism in *Synechocystis* 6803.

A1.3. Materials and Methods

A1.3.1. Photomixotrophic Culture

Synechocystis 6803 cultures were grown in a modified BG-11 medium depleted of ^{12}C. Ferric ammonium citrate was replaced with ferric ammonium sulfate (Katoh *et al.* 2001). ^{13}C was supplied as ~2 g/L $NaH^{13}CO_3$ and 5 g/L glucose (U-$^{13}C_6$ or 1-$^{13}C_1$). The purity of ^{13}C-substrates was >98% (Cambridge Isotope Laboratories, Tewksbury, MA, USA.). Inocula from an unlabeled *Synechocystis* 6803 autotrophic culture (OD_{730} = ~0.9) was added into 30 mL ^{13}C-labeled medium in 100 mL serum bottles, which were then sealed with rubber septa to prevent atmospheric CO_2 intrusion.

All cultures were started with only a 0.5 % inoculation volume to minimize the inoculation effect. Cell growth was under continuous illumination (~50 μmol photons m^{-2} s^{-1}) on a shaker at 150 rpm at 30 °C. Cell density was monitored by a UV-Vis spectrophotometer (GENESYS, Thermo Scientific) at 730 nm. The conversion ratio between OD_{730} and dry biomass weight was 1 unit OD_{730} = 0.45 g dry cell weight L^{-1}. Total Organic Carbon Analyzer (inorganic carbon measurement mode) with non-dispersive infrared detector (Shimadzu Corporation, Japan) was used to determine sodium bicarbonate concentration in the culture supernatant. Enzyme kits (R-Biopharm, Darmstadt, Germany) were used to measure the glucose concentrations in the culture.

A1.3.2. Isotopic Dilution Experiments

Isotopic dilution experiments were employed to identify the presence of certain pathways *in vivo*. To investigate the structure of the TCA cycle, we used glutamate (instead of α-ketoglutarate) as the tracer since cyanobacteria exhibited very low capability to uptake α-ketoglutarate (Vázquez-Bermúdez *et al.* 2000). To examine the presence of the glyoxylate shunt,

we used glyoxylate as the tracer. Specifically, unlabeled glutamate (10 mM) or unlabeled glyoxylate (15 mM) was added into ^{13}C-labeled cultures (grown on $NaH^{13}CO_3$ and U-$^{13}C_6$ glucose) during the exponential growth phase (OD_{730} = ~0.4). After 30 minutes of incubation with a respective tracer, samples from two biological replicates were harvested and free metabolites were extracted. To identify whether *Synechocystis* 6803 used glutamate or glyoxylate for biomass synthesis, *Synechocystis* 6803 cultures were grown with a respective unlabeled tracer (10 mM glutamate or 15 mM glyoxylate), $NaH^{13}CO_3$ and U-$^{13}C_6$ glucose for 48 hours (OD_{730} reached ~0.4). Samples from two biological replicates were then collected to analyze the ^{12}C incorporation into proteinogenic amino acids.

A1.3.3. Metabolite Extraction and GC-MS Analysis

Isotopomer analysis of free metabolites and proteinogenic amino acids is based on previous reports (Meadows *et al.* 2008; Tang *et al.* 2009; You *et al.* 2012). GC-MS analysis had three technical replicates per sample.

Intracellular free metabolites were used to qualitatively characterize functional pathways. Supporting information Figs. A1.S1A, B, and C illustrate the molecular structure of TMS-derivatized amino acids, succinate, and α-ketoglutarate used for analysis. The fragment $[M-15]^+$, minus a methyl group from the TMS group, includes the labeling information of the entire molecule. The $[M-15]^+$ fragment, together with $[M-43]^+$ or $[M-117]^+$(minus the α-carboxyl group from a metabolite), was used for GC-MS analysis.

Proteinogenic amino acids were used to determine the function and quantify the metabolic fluxes. The mass fragments of ten key amino acids provided sufficient constraints for flux calculations (Pingitore *et al.* 2007; Tang *et al.* 2007; Tang *et al.* 2009). The fragments ($[M-57]^+$, $[M-159]^+$ or $[M-85]^+$, and f302) were used for flux analysis (Antoniewicz *et al.*

2007)_ENREF_1. In addition, because of overlap peaks and product degradations, several amino acids (proline, arginine, cysteine, and tryptophan) were not analyzed. The isotopic labeling data are shown as mass fractions, i.e., M0, M1, M2, etc., representing fragments containing unlabeled, singly labeled, and doubly labeled metabolites, etc.

A1.2.4. ^{13}C-Metabolic Flux Analysis

^{13}C-MFA was used to quantify *in vivo* fluxes through the central metabolic network in *Synechocystis* 6803. Photomixotrophic cultures were grown on 1-$^{13}C_1$ glucose and $NaH^{13}CO_3$. Biomass was collected during the exponential growth phase for proteinogenic amino acids analysis. The metabolic network of *Synechocystis* 6803 was reconstructed based on tracer experiments and previous reports (Pelroy *et al.* 1972; Smith 1983; Robert Tabita 2004; Bauwe *et al.* 2010; Young *et al.* 2011) that included glycolysis, the Calvin Cycle, complete TCA Cycle, glyoxylate shunt, and photorespiration pathways (Table A1.S1). In our ^{13}C-MFA, relative metabolic fluxes through the central metabolism were profiled with the assumption that the Calvin cycle flux from Ru5P to RuBP was 100. The minimization of a quadratic function that calculated the difference between predicted and measured isotopomer patterns solved the relative metabolic fluxes (Table A1.S2). The biomass composition (Table A1.S1) was based on a previous report (Saha *et al.* 2012). Reaction reversibility was characterized by the exchange coefficient exch, defined as $v^{exch}\beta = \frac{\text{exch}}{1-\text{exch}}$, where v^{exch} was the exchange flux and β was the exchange constant (Dauner *et al.* 2001). In this study, β was equal to 500 and exch ranged from 0 to 1. The forward flux ($v^{forward}$) and backward flux ($v^{backward}$) in the model were transformed from the v^{exch} and the net flux, v^{net}, using the following formulation (Wiechert *et al.* 1997):

$$\begin{pmatrix} \mathbf{v}^{\mathbf{forward}} \\ \mathbf{v}^{\mathbf{backward}} \end{pmatrix} = \begin{pmatrix} \mathbf{v}^{\mathbf{exch}} - \mathbf{min}(-\mathbf{v}^{\mathbf{net}}, \mathbf{0}) \\ \mathbf{v}^{\mathbf{exch}} - \mathbf{min}(\mathbf{v}^{\mathbf{net}}, \mathbf{0}) \end{pmatrix}.$$

The optimization for ^{13}C-MFA was performed as follows:

$$\min(\mathbf{M_{exp}} - \mathbf{M_{sim}}(\mathbf{v}))^{T}(\mathbf{M_{exp}} - \mathbf{M_{sim}}(\mathbf{v})) \quad \text{(A1.1)}$$

s.t. $$\mathbf{v} \in (\text{lb}, \text{ub}) \quad \text{(A1.2)}$$

$$\mathbf{S} \cdot \mathbf{v} = \mathbf{0} \quad \text{(A1.3)}$$

$$\mathbf{A_i} \cdot \mathbf{X_i} = \mathbf{B_i} \cdot \mathbf{Y_i} \ (\text{i}=1, 2, 3, 4, 5) \quad \text{(A1.4)}$$

Equation A1.1 is the quadratic error function that was optimized and M_{exp} is the vector of experimentally measured labeling patterns of amino acids. M_{sim} is the counterpart of the simulated data as a function of fluxes. *v* is the flux vector that is to be determined. Equation A1.2 gives the boundary conditions of the flux variables. Equation A1.3 represents the metabolite balances. Equation A1.4 represents the elementary metabolite unit (EMU) balance, where X_i and Y_i represent the unknown and known EMU variables of size i, respectively, and A_i and B_i are matrices of linear functions of the fluxes (Antoniewicz *et al.* 2007; Leighty *et al.* 2012).

The MATLAB optimization solver 'fmincon' was employed to minimize the quadratic error function. To avoid local minima, 100 initial guesses were randomly generated, and the solution set that minimized the objective function was used as the best fit. The Monte Carlo method was used to calculate 95% confidence intervals (Zhao *et al.* 2003). The measured isotopomer data was perturbed 1000 times with normally distributed noise within measurement error, and the optimization solver was restarted with the optimal solution. The determination of confidence intervals of the fluxes (95%) was based on 1000 simulations, and confidence intervals were used to calculate standard deviations.

A1.4. Results

A1.4.1. Photomixotrophic Biomass Growth and Metabolic Pseudo-Steady State

Figure A1.1 shows the growth curves in serum bottles and shake flasks. Cell doubling times were similar in both containers. The similarity of growth indicates that O_2 accumulation in the serum bottle headspace had minimal effect on photomixotrophic growth. During the early growth phase, the specific growth rate in the early exponential phase was 0.079 h^{-1}. After cultivation in serum bottles for 75 hours, cell growth slowed down and the culture pH rose from 8 to 10.

To determine a pseudo-steady state metabolic period for ^{13}C-MFA, biomass samples from serum bottles were collected at different time points to analyze amino acid labeling (Table A1.S2). The labeling patterns in biomass protein were relatively stable (standard deviation were below 0.01) between samples taken within the first 48-hour cultures (Table A1.S2). Thereby, our ^{13}C-MFA was based on ^{13}C-biomass samples taken at early growth phase (OD_{730} = ~0.4, Table A1.S2).

A1.4.2. ^{13}C-based Pathway Investigation

Based on isotopic dilution of downstream metabolites after incubating cells with unlabeled precursors, *in vivo* enzyme functions were investigated in tracer experiments. Prior to ^{12}C-glutamate pulse treatment, α-ketoglutarate (Figure A1.2A), succinate (Figure A1.2B), and malate (Figure A1.2C) were nearly fully labeled (M5 for α-ketoglutarate; M4 for succinate and malate) in ^{13}C labeled cultures. After 30-minute of incubation with unlabeled glutamate, ^{12}C carbon from glutamate was incorporated into the downstream metabolites of α-ketoglutarate. ^{12}C abundance increased, to over 65% in succinate, 90% in α-ketoglutarate, and 30% in malate. Mass spectra of these metabolites before and after glutamate addition are shown in Supporting

information Fig. A1.S1. After incubation with unlabeled glutamate for 40 hours, all amino acids, except glutamate, from biomass protein remained fully ^{13}C labeled (Figure A1.3 and A1.S2).

Labeled cultures pulsed with unlabeled glyoxylate showed a shift from fully ^{13}C to ^{12}C in free glycine and glyoxylate (M2 to M0, Figure A1.4A). However, no significant shift was observed in succinate and α-ketoglutarate, both of which are downstream metabolites of malate. After the labeled culture was incubated with unlabeled glyoxylate for 40 hours, ^{12}C was only incorporated into proteinogenic glycine (Figure A1.4B), while other proteinogenic amino acids remained highly labeled.

A1.4.3. Flux Analysis Results

^{13}C-MFA results are sensitive to model network construction, the labeling patterns of substrates, and the completeness of isotopomer data. In this study, the isotopic dilution experiments were employed to reconstruct a ^{13}C-MFA network with a complete TCA cycle and an active glyoxylate shunt in *Synechocystis* 6803. Singly labeled glucose and fully labeled bicarbonate were used to generate unique isotopomer data in amino acids. Via advanced EMU simulations and isotopomer information from different MS fragments, ^{13}C-MFA profiled the photomixotrophic metabolism under light/carbon sufficient conditions. Relative flux distributions, exchange coefficients for reversible reactions, and 95% confidence intervals are shown in Figure A1.5 and Supporting information Table A1.S1. The simulated fluxes fit the isotopomer data well ($r^2 > 0.99$, Supporting information Fig. A1.S3).

^{13}C-MFA indicated that *Synechocystis* 6803 had a high CO_2 fixation flux through the Calvin cycle (~100) than the glucose uptake flux (~18) in the early photomixotrophic growth phase. Consistent with this observation, less than 0.1 g L^{-1} glucose was consumed during early growth phase. In contrast, previous ^{13}C-MFA of photomixotrophic metabolism in *Synechocystis*

6803 under CO_2 limiting conditions found the glucose uptake flux to be ~50 (Yang *et al.* 2002). In another study, when cell culture was dense (OD_{730} up to 20), *Synechocystis* 6803 utilized significantly more glucose than CO_2 for its growth (Varman *et al.* 2013). Thereby, CO_2 and light conditions can significantly affect the photomixotrophic metabolism in *Synechocystis* 6803.

Under photomixotrophic conditions with sufficient light and carbon sources, the flux from α-ketoglutarate to succinate was not significant (< 2% of total CO_2 uptake). Most of the flux from α-ketoglutarate went to glutamate biosynthesis. The glyoxylate shunt did not show a measureable flux (<0.1% of total CO_2 uptake). Additionally, the OPP pathway showed a measurable flux (1.9% of total CO_2 uptake), which played a minor role in C5 carbon synthesis and NADPH production. The flux through photorespiration, however, was limited to 0.1% of total CO_2 uptake. Although the confidence intervals (Table A1.S1) of these anaplerotic reactions (PEP + CO_2 →OAA; MAL → CO_2 + PYR) were larger than those of other fluxes, malic enzyme showed significant flux and was the main route for supplying pyruvate.

A1.5. Discussion

A1.5.1. TCA Cycle Metabolism

Cyanobacteria are prokaryotic bacteria responsible for the conversion of the early atmosphere into our current oxygen-rich atmosphere (Riding 2006). Primitive anaerobic prokaryotes developed two separate TCA pathways: the reductive branch (oxaloacetate to succinate) and the oxidative branch (citrate to *α-ketoglutarate*) (Figure A1.2D.i). Some anaerobic bacteria, such as *Clostridium acetobutylicum*, use a bifurcated TCA cycle that terminates at succinate (Figure A1.2D.ii). As atmospheric oxygen levels rose, the two branches linked to complete the TCA cycle. For example, TCA cycle in facultative anaerobes (e.g., *E. coli*) can be

complete if oxygen is present (Figure A1.2D.iii). A phototrophic bacterium, *Chlorobaculum tepidum,* employs a reverse TCA cycle (Feng *et al.* 2010) seen in Figure A1.2D.iv.

In our study, the labeling patterns of free metabolites indicated that the pathway for converting α-ketoglutarate to succinate can be complete under glutamate addition conditions. Significant amounts of unlabeled α-ketoglutarate, succinate, and malate (Figure A1.2A, B and C) were observed after unlabeled glutamate was added into ^{13}C labeled cultures. Since α-ketoglutarate dehydrogenase activity has never been shown to exist in cyanobacteria, we presume that this conversion was accomplished by a newly discovered pathway through succinic semialdehyde (Knoop *et al.* 2010; Zhang *et al.* 2011; Nogales *et al.* 2012). On the other hand, key proteinogenic amino acids, e.g., aspartate (derived from oxaloacetate), alanine (derived from pyruvate), and serine (derived from 3-phosphoglycerate), had very little ^{12}C incorporation (< 5%) from glutamate after two-day incubation with unlabeled glutamate (Figure A1.3). These results qualitatively indicated that the flux from α-ketoglutarate towards the complete TCA cycle was very small compared to other fluxes (e.g., fluxes through glycolysis and Calvin Cycle). The low conversion from α-ketoglutarate to its TCA cycle downstream metabolites was also observed in a *Synechococcus elongatus* PCC 7942 mutant (with an engineered α-ketoglutarate permease), in which α-ketoglutarate was mainly converted into glutamate and glutamine instead of TCA cycle downstream metabolites (Vázquez-Bermúdez *et al.* 2000).

Although our ^{13}C-study cannot distinguish whether the conversion of α-ketoglutarate to succinate was via α-ketoglutarate decarboxylase or the GABA shunt (Nogales *et al.* 2012), this reaction may be notable in cyanobacterial metabolism only under certain conditions (e.g., with the presence of large amount of glutamate or α-ketoglutarate). The poor activity of this reaction may also explain why the previous tracer studies did not observe the conversion of α-

ketoglutarate to succinate in which they used an assay to detect α-ketoglutarate dehydrogenase activity (Pearce *et al.* 1969), as opposed to the decarboxylase activity that has been more recently observed to convert α-ketoglutarate to succinate (Zhang *et al.* 2011).

Although many cyanobacterial species appear to have a complete TCA cycle pathway, it may not be adapted to carry a large flux. A recent FBA model indicates that a complete cyanobacterial TCA cycle via AKG dehydrogenase may reduce biomass growth due to the unnecessary metabolic burden for the synthesis of multi-protein enzymes (Knoop *et al.* 2013). For organisms that obtain sufficient reducing equivalents from light reactions, the use of a complete TCA cycle to oxidize carbon is unnecessary. Thereby, the complete TCA pathways in *Synechocystis* 6803 may serve only to regenerate intermediates or fine-tune the metabolic balance under certain photomixotrophic conditions (such as the presence of extracellular glutamate).

A1.5.2. The Glyoxylate Shunt

This study also examined the presence of glyoxylate shunt and determined its function in *Synechocystis* 6803. Previous metabolic models predicted that *Synechocystis* 6803 contains a bacterial-like glyoxylate shunt (Yang *et al.* 2002). However, *Synechocystis* 6803 and nearly all other sequenced cyanobacteria, lack homologues of known genes that encode isocitrate lyase and malate synthase. Some ^{13}C-MFA (Young *et al.* 2011) and FBA (Shastri *et al.* 2005; Yoshikawa *et al.* 2011) studies have also suggested that the glyoxylate shunt in *Synechocystis* 6803 was incomplete under photoautotrophic and photomixotrophic conditions. In our tracer experiments with the addition of unlabeled glyoxylate during the exponential phase, we observed the uptake of glyoxylate and its conversion to glycine (Figure A1.4A). However, in the proteinogenic amino acids of ^{13}C-cultures grown with ^{12}C glyoxylate (Figure A1.4B), we did not see significant ^{12}C

accumulation in proteinogenic amino acids downstream of malate (i.e., the end-product of the glyoxylate shunt), including alanine and aspartate (Figure A1.4B). Statistically, ^{13}C-MFA showed that the *in vivo* flux through the presumed glyoxylate shunt was essentially zero (Figure A1.5). This observation of the glyoxylate shunt is supported by a recent enzymatic test using crude extracts of *Synechocystis* 6803 cells, in which no isocitrate lyase activity was detected (Knoop *et al.* 2013).

A1.5.3. Malic Enzyme Activity

Under continuous light illumination, the malic enzyme is important for optimal *Synechocystis* 6803 growth. Its gene expression (*slr0721*) is high under photomixotrophic conditions compared to photoautotrophic conditions (Yoshikawa *et al.* 2013). Moreover, ^{13}C-MFA revealed significant malic enzyme flux in *Synechocystis* 6803 under photoautotrophic (Young *et al.* 2011) and CO_2 limited photomixotrophic cultures (Yang *et al.* 2002). Previous reports indicated that the malic enzyme reaction (Malate→Pyruvate +CO_2 +NADPH) is instrumental in a carbon concentration mechanism akin to that in C4 plants, and this enzyme may indirectly transport NADPH between different cell locations. In this study, high malic enzyme activity was observed when the bicarbonate and reduced carbon source were sufficient. In fact, deletion of malic enzyme gene significantly reduces *Synechocystis* 6803 growth under both autotrophic and glucose-based mixotrophic conditions, while the growth can be recovered by providing pyruvate (Bricker *et al.* 2004). Thereby, high flux through malic enzyme (~31) is likely to serve as a key route for pyruvate synthesis when pyruvate kinase is inhibited by ATP a negative allosteric inhibitor under photosynthetic conditions (Bricker *et al.* 2004).

A1.5.4. The Oxidative Pentose Phosphate Pathway

The OPP pathway is an important NADPH synthesis route in heterotrophic organisms. Since photosynthetic light reactions produce significant amounts of NADPH, the OPP pathway becomes futile in photoautotrophic metabolism. A mutant (Δ*zwf*) of *Synechococcus elongatus* PCC 7942, that lacks the OPP pathway enzymes, exhibited a similar growth rate to the wild-type strain under autotrophic conditions (Scanlan *et al.* 1995). Moreover, glucose in phototrophic cultures has been shown to either increase or have a small effect on key OPP pathway enzyme transcriptions (Yang *et al.* 2002; Kahlon *et al.* 2006; Yoshikawa *et al.* 2013). Taken together, these data indicate that the OPP pathway is dispensable under light conditions. However, the OPP pathway mutant described above exhibits decreased viability under dark incubations (Scanlan *et al.* 1995). FBA models also predicted that the OPP pathway was active only under light-limited conditions (Shastri *et al.* 2005). Our experiments measured a low flux (~1.5) through the OPP pathway under early photomixotrophic growth conditions. However, biomass samples collected from the late growth phase (Table A1.S2) showed a higher unlabeled proteinogenic histidine (M0 fraction, Table A1.S2), indicating that more glucose was directed to OPP for ribose-5-phosphate synthesis (precursor of histidine). These results suggest a flexibility of the OPP pathway in balancing NADPH under different growth conditions.

A1.5.5. Limitations of Our ^{13}C-MFA Techniques for Cyanobacterial Study

^{13}C-MFA accuracy is highly dependent on metabolic model construction. Nevertheless, incomplete annotations, errors, or inconsistencies are prevalent in cyanobacterial genome databases, rendering it difficult to generate a comprehensive metabolic network for ^{13}C-MFA. Tracer experiments were used here to examine the structure of a cyanobacterial metabolic network. Since the key intermediate tracers (e.g., ^{13}C-glutamate and ^{13}C-glyoxylate) are

prohibitively expensive, an inverse tracer labeling approach was employed to save experimental costs. A ^{13}C-culture background was first built with commonly used ^{13}C-substrates (bicarbonate and glucose). Unlabeled glutamate or glyoxylate were then added into the ^{13}C-cultures as tracers. Their incoropration into downstream metabolites was used to determine pathway functions. Since the addition of intermediate tracers may change cell metabolism, the results from isotopic dilution experiments were only used to build a more accurate ^{13}C-MFA model.

In addition, the application of ^{13}C-MFA requires the attainment of a steady state with minimal labeling changes in the central metabolism. Therefore, metabolisms under photoautotrophic or circadian conditions cannot be analyzed using steady-state ^{13}C-MFA. Although an advanced isotopic non-stationary MFA (Young *et al.* 2011) has been developed to capture the transient states of metabolic networks, it is difficult to precisely measure the labeling patterns of low-abundant and unstable free metabolites. It is also difficult to resolve the unexpected labeling kinetics of free metabolites caused by metabolic channeling (Young *et al.* 2011). Furthermore, the degradation-regeneration of certain cellular polymers (e.g., cyanophycin) may exchange carbons between free metabolites and macromolecules, interfering with non-stationary metabolite labeling (Huege *et al.* 2011). Therefore, a steady-state ^{13}C-MFA is more suitable to probe cyanobacteria metabolisms. *Synechocystis* 6803's metabolism is diverse and affected by both light and carbon conditions, so the observed results pertain solely to photomixotrophic conditions when light and carbon sources are sufficient.

A1.6. Conclusion

This study used ^{13}C-metabolism analysis to delineate the photomixotrophic metabolism of *Synechocystis* 6803. ^{13}C-analysis confirmed the *in vivo* conversion of α-ketoglutarate to

succinate when an additional source was supplied to increase the α-ketoglutarate pool size (e.g., glutamate) while this flux under photomixotrophic conditions is negligible compared to all other fluxes in the model. Glyoxylate was discovered as a potential source for glycine synthesis, while the activity of the glyoxylate shunt was not observed. Under photomixotrophic conditions, malic enzyme, rather than pyruvate kinase, is fundamental route for pyruvate synthesis. Oxidative pentose phosphate pathway flux is low when light and inorganic carbons are sufficient. These findings complement information of previous multiple-omics studies which have shown the ^{13}C-tools to greatly advance the understanding of cellular metabolisms (Tang *et al.* 2012). This study also suggests that the photomixotrophic metabolism of cyanobacterial cell factories can efficiently incorporate both sugar and CO_2 for biosynthesis, resulting higher growth rate and biomass density.

A1.7. Supporting Information

The following supporting information is available online with the originally published version of this article (You *et al.* 2014) at DOI: 10.1002/biot.201300477.

Figure A1.S1: Mass spectra of several free metabolites from *Synechocystis* 6803.

Figure A1.S2: Long-term tracer experiments and proteinogenic amino acid analysis.

Figure A1.S3: ^{13}C-MFA model fitting.

Table A1.S1: Complete list of estimated fluxes in central metabolism and exchange coeffecients.

Table A1.S2: Comparison of experimental and simulated labeling profiles of proteinogenic amino acids from mixotrophic culture grown on $NaH^{13}CO_3$ and 1-^{13}C glucose.

Item A1.S1: Model Assumption Description.

Item A1.S2: Calculation of glucose consumption rates.

File A1.S1: Biomass Formation Equation.

File A1.S2: Carbon transition in reactions included in the ^{13}C-MFA.

A1.8. References

Antoniewicz, M. R., Kelleher, J. K. and Stephanopoulos, G. (2007). "Accurate Assessment of Amino Acid Mass Isotopomer Distributions for Metabolic Flux Analysis." *Analytical Chemistry* **79**(19): 7554-7559.

Antoniewicz, M. R., Kelleher, J. K. and Stephanopoulos, G. (2007). "Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions." *Metabolic Engineering* **9**(1): 68-86.

Atsumi, S., Higashide, W. and Liao, J. C. (2009). "Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde." *Nat Biotech* **27**(12): 1177-1180.

Bauwe, H., Hagemann, M. and Fernie, A. R. (2010). "Photorespiration: players, partners and origin." *Trends in plant science* **15**(6): 330-336.

Berry, S., Schneider, D., Vermaas, W. F. and Rögner, M. (2002). "Electron transport routes in whole cells of *Synechocystis* sp. strain PCC 6803: the role of the cytochrome bd-type oxidase." *Biochemistry* **41**(10): 3422-3429.

Bricker, T. M., Zhang, S., Laborde, S. M., Mayer, P. R., Frankel, L. K. and Moroney, J. V. (2004). "The Malic Enzyme Is Required for Optimal Photoautotrophic Growth of *Synechocystis* sp. Strain PCC 6803 under Continuous Light but Not under a Diurnal Light Regimen." *Journal of Bacteriology* **186**(23): 8144-8148.

Chisti, Y. (2008). "Biodiesel from microalgae beats bioethanol." *Trends in Biotechnology* **26**(3): 126-131.

Cooley, J. W., Howitt, C. A. and Vermaas, W. F. J. (2000). "Succinate:Quinol Oxidoreductases in the Cyanobacterium *Synechocystis* sp. Strain PCC 6803: Presence and Function in Metabolism and Electron Transport." *Journal of Bacteriology* **182**(3): 714-722.

Dauner, M., Bailey, J. E. and Sauer, U. (2001). "Metabolic flux analysis with a comprehensive isotopomer model in *Bacillus subtilis*." *Biotechnology and Bioengineering* **76**(2): 144-156.

Deng, M.-D. and Coleman, J. R. (1999). "Ethanol Synthesis by Genetic Engineering in Cyanobacteria." *Applied and Environmental Microbiology* **65**(2): 523-528.

Eley, J. H. (1988). "Glyoxylate cycle enzyme activities in the cyanobacterim *Anacystis nidulans*." *Journal of Phycology* **24**(4): 586-588.

Feng, X., Tang, K.-H., Blankenship, R. E. and Tang, Y. J. (2010). "Metabolic flux analysis of the mixotrophic metabolisms in the green sulfur bacterium *Chlorobaculum tepidum*." *Journal of Biological Chemistry* **285**(45): 39544-39550.

Fu, P. (2009). "Genome-scale modeling of *Synechocystis* sp. PCC 6803 and prediction of pathway insertion." *Journal of Chemical Technology & Biotechnology* **84**(4): 473-483.

Grigorieva, G. and Shestakov, S. (1982). "Transformation in the cyanobacterium *Synechocystis* sp. 6803." *Fems Microbiology Letters* **13**(4): 367-370.

Huege, J., Goetze, J., Schwarz, D., Bauwe, H., Hagemann, M. and Kopka, J. (2011). "Modulation of the Major Paths of Carbon in Photorespiratory Mutants of *Synechocystis*." *PLoS One* **6**(1): e16278.

Kahlon, S., Beeri, K., Ohkawa, H., Hihara, Y., Murik, O., Suzuki, I., Ogawa, T. and Kaplan, A. (2006). "A putative sensor kinase, Hik31, is involved in the response of *Synechocystis* sp. strain PCC 6803 to the presence of glucose." *Microbiology* **152**(3): 647-655.

Katoh, H., Hagino, N. and Ogawa, T. (2001). "Iron-Binding Activity of FutA1 Subunit of an ABC-type Iron Transporter in the Cyanobacterium *Synechocystis* sp. Strain PCC 6803." *Plant and Cell Physiology* **42**(8): 823-827.

Knoop, H., Gründel, M., Zilliges, Y., Lehmann, R., Hoffmann, S., Lockau, W. and Steuer, R. (2013). "Flux Balance Analysis of Cyanobacterial Metabolism: The Metabolic Network of *Synechocystis* sp. PCC 6803." *PLoS Comput Biol* **9**(6): e1003081.

Knoop, H., Zilliges, Y., Lockau, W. and Steuer, R. (2010). "The Metabolic Network of *Synechocystis* sp. PCC 6803: Systemic Properties of Autotrophic Growth." *Plant Physiology* **154**(1): 410-422.

Kohlstedt, M., Becker, J. and Wittmann, C. (2010). "Metabolic fluxes and beyond—systems biology understanding and engineering of microbial metabolism." *Applied Microbiology and Biotechnology* **88**(5): 1065-1075.

Lan, E. I. and Liao, J. C. (2011). "Metabolic engineering of cyanobacteria for 1-butanol production from carbon dioxide." *Metabolic Engineering* **13**(4): 353-363.

Leighty, R. W. and Antoniewicz, M. R. (2012). "Parallel labeling experiments with [U-^{13}C]glucose validate *E. coli* metabolic network model for ^{13}C metabolic flux analysis." *Metabolic Engineering* **14**(5): 533-541.

Lindberg, P., Park, S. and Melis, A. (2010). "Engineering a platform for photosynthetic isoprene production in cyanobacteria, using Synechocystis as the model organism." *Metabolic Engineering* **12**(1): 70-79.

McKinlay, J. B. and Harwood, C. S. (2010). "Carbon dioxide fixation as a central redox cofactor recycling mechanism in bacteria." *Proceedings of the National Academy of Sciences* **107**(11669-11675).

McKinlay, J. B. and Harwood, C. S. (2011). "Calvin Cycle Flux, Pathway Constraints, and Substrate Oxidation State Together Determine the H_2 Biofuel Yield in Photoheterotrophic Bacteria." *MBio* **2**(2).

Meadows, A. L., Kong, B., Berdichevsky, M., Roy, S., Rosiva, R., Blanch, H. W. and Clark, D. S. (2008). "Metabolic and Morphological Differences between Rapidly Proliferating Cancerous and Normal Breast Epithelial Cells." *Biotechnology Progress* **24**(2): 334-341.

Montagud, A., Navarro, E., Fernandez de Cordoba, P., Urchueguia, J. and Patil, K. (2010). "Reconstruction and analysis of genome-scale metabolic model of a photosynthetic bacterium." *BMC Systems Biology* **4**(1): 156.

Nogales, J., Gudmundsson, S., Knight, E. M., Palsson, B. O. and Thiele, I. (2012). "Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis." *Proceedings of the National Academy of Sciences* **109**(7): 2678-2683.

Pearce, J. and Carr, N. (1967). "The metabolism of acetate by the blue-green algae, Anabaena variabilis and Anacystis nidulans." *J Gen Microbiol* **49**: 301 - 313.

Pearce, J., Leach, C. K. and Carr, N. G. (1969). "The Incomplete Tricarboxylic Acid Cycle in the Blue-green Alga *Anabaena Variabilis*." *Journal of General Microbiology* **55**(3): 371-378.

Pelroy, R. A. and Bassham, J. A. (1972). "Photosynthetic and dark carbon metabolism in unicellular blue-green algae." *Archiv für Mikrobiologie* **86**(1): 25-38.

Pingitore, F., Tang, Y. J., Kruppa, G. H. and Keasling, J. D. (2007). "Analysis of amino acid isotopomers using FT-ICR MS." *Analytical Chemistry* **79**(6): 2483-2490.

Riding, R. (2006). "Cyanobacterial calcification, carbon dioxide concentrating mechanisms, and Proterozoic–Cambrian changes in atmospheric composition." *Geobiology* **4**(4): 299-316.

Robert Tabita, F. (2004). "The Biochemistry and Molecular Regulation of Carbon Dioxide Metabolism in Cyanobacteria." *The Molecular Biology of Cyanobacteria*. D. Bryant, Springer Netherlands. **1:** 437-467.

Saha, R., Verseput, A. T., Berla, B. M., Mueller, T. J., Pakrasi, H. B. and Maranas, C. D. (2012). "Reconstruction and Comparison of the Metabolic Potential of Cyanobacteria *Cyanothece* sp. ATCC 51142 and *Synechocystis* sp. PCC 6803." *PLoS One* **7**(10): e48285.

Scanlan, D. J., Sundaram, S., Newman, J., Mann, N. H. and Carr, N. G. (1995). "Characterization of a *zwf* mutant of *Synechococcus* sp. strain PCC 7942." *Journal of Bacteriology* **177**(9): 2550-2553.

Shastri, A. A. and Morgan, J. A. (2005). "Flux Balance Analysis of Photoautotrophic Metabolism." *Biotechnology Progress* **21**(6): 1617-1626.

Smith, A. J. (1983). "Modes of cyanobacterial carbon metabolism." *Annales de l'Institut Pasteur / Microbiologie* **134**(1, Supplement B): 93-113.

Smith, A. J., London, J. and Stanier, R. Y. (1967). "Biochemical Basis of Obligate Autotrophy in Blue-Green Algae and Thiobacilli." *Journal of Bacteriology* **94**(4): 972-983.

Tang, J. K.-H., You, L., Blankenship, R. E. and Tang, Y. J. (2012). "Recent advances in mapping environmental microbial metabolisms through ^{13}C isotopic fingerprints." *Journal of The Royal Society Interface* **9**: 2767-2780.

Tang, Y. J., Hwang, J. S., Wemmer, D. and Keasling, J. D. (2007). "The *Shewanella oneidensis* MR-1 fluxome under various oxygen conditions." *Applied and Environmental Microbiology* **73**(3): 718-729.

Tang, Y. J., Martin, H. G., Dehal, P. S., Deutschbauer, A., Llora, X., Meadows, A., Arkin, A. and Keasling, J. D. (2009). "Metabolic flux analysis of *Shewanella* spp. reveals evolutionary robustness in central carbon metabolism." *Biotechnol Bioeng.* **102**(4): 1161-1169.

Tang, Y. J., Shui, W. Q., Myers, S., Feng, X., Bertozzi, C. and Keasling, J. D. (2009). "Central metabolism in *Mycobacterium smegmatis* during the transition from O_2-rich to O_2-poor conditions as studied by isotopomer-assisted metabolite analysis." *Biotechnology Letters* **31**(8): 1233-1240.

Varman, A. M., Xiao, Y., Pakrasi, H. B. and Tang, Y. J. (2013). "Metabolic Engineering of *Synechocystis* sp. Strain PCC 6803 for Isobutanol Production." *Applied and Environmental Microbiology* **79**(3): 908-914.

Vázquez-Bermúdez, M. F., Herrero, A. and Flores, E. (2000). "Uptake of 2-Oxoglutarate in *Synechococcus* Strains Transformed with the *Escherichia coli kgtP* Gene." *Journal of Bacteriology* **182**(1): 211-215.

Wiechert, W., Siefke, C., de Graaf, A. A. and Marx, A. (1997). "Bidirectional reaction steps in metabolic networks: II. Flux estimation and statistical analysis." *Biotechnology and Bioengineering* **55**(1): 118-135.

Yang, C., Hua, Q. and Shimizu, K. (2002). "Metabolic flux analysis in *Synechocystis* using isotope distribution from ^{13}C-labeled glucose." *Metab Eng* **4**: 202 - 216.

Yang, C., Hua, Q. and Shimizu, K. (2002). "Quantitative analysis of intracellular metabolic fluxes using GC-MS and two-dimensional NMR spectroscopy." *Journal of Bioscience and Bioengineering* **93**(1): 78-87.

Yang, C. Y., Hua, Q. H. and Shimizu, K. S. (2002). "Integration of the information from gene expression and metabolic fluxes for the analysis of the regulatory mechanisms in *Synechocystis*." *Applied Microbiology and Biotechnology* **58**(6): 813-822.

Yoshikawa, K., Hirasawa, T., Ogawa, K., Hidaka, Y., Nakajima, T., Furusawa, C. and Shimizu, H. (2013). "Integrated transcriptomic and metabolomic analysis of the central metabolism of *Synechocystis* sp. PCC 6803 under different trophic conditions." *Biotechnology Journal* **8**(5): 571-580.

Yoshikawa, K., Kojima, Y., Nakajima, T., Furusawa, C., Hirasawa, T. and Shimizu, H. (2011). "Reconstruction and verification of a genome-scale metabolic model for *Synechocystis* sp. PCC6803." *Applied Microbiology and Biotechnology* **92**(2): 347-358.

You, L., Berla, B., He, L., Pakrasi, H. B. and Tang, Y. J. (2014). "13C-MFA delineates the photomixotrophic metabolism of Synechocystis sp. PCC 6803 under light- and carbon-sufficient conditions." *Biotechnol J* **9**(5): 684-692.

You, L., Page, L., Feng, X., Berla, B., Pakrasi, H. B. and Tang, Y. J. (2012). "Metabolic Pathway Confirmation and Discovery Through ^{13}C-labeling of Proteinogenic Amino Acids." *J Vis Exp*(59): e3583.

Young, J. D., Shastri, A. A., Stephanopoulos, G. and Morgan, J. A. (2011). "Mapping photoautotrophic metabolism with isotopically nonstationary ^{13}C flux analysis." *Metabolic Engineering* **13**(6): 656-665.

Zhang, S. and Bryant, D. A. (2011). "The Tricarboxylic Acid Cycle in Cyanobacteria." *Science* **334**(6062): 1551-1553.

Zhao, J. and Shimizu, K. (2003). "Metabolic flux analysis of *Escherichia coli* K12 grown on ^{13}C-labeled acetate and glucose using GC-MS and powerful flux calculation method." *Journal of Biotechnology* **101**: 101-117.

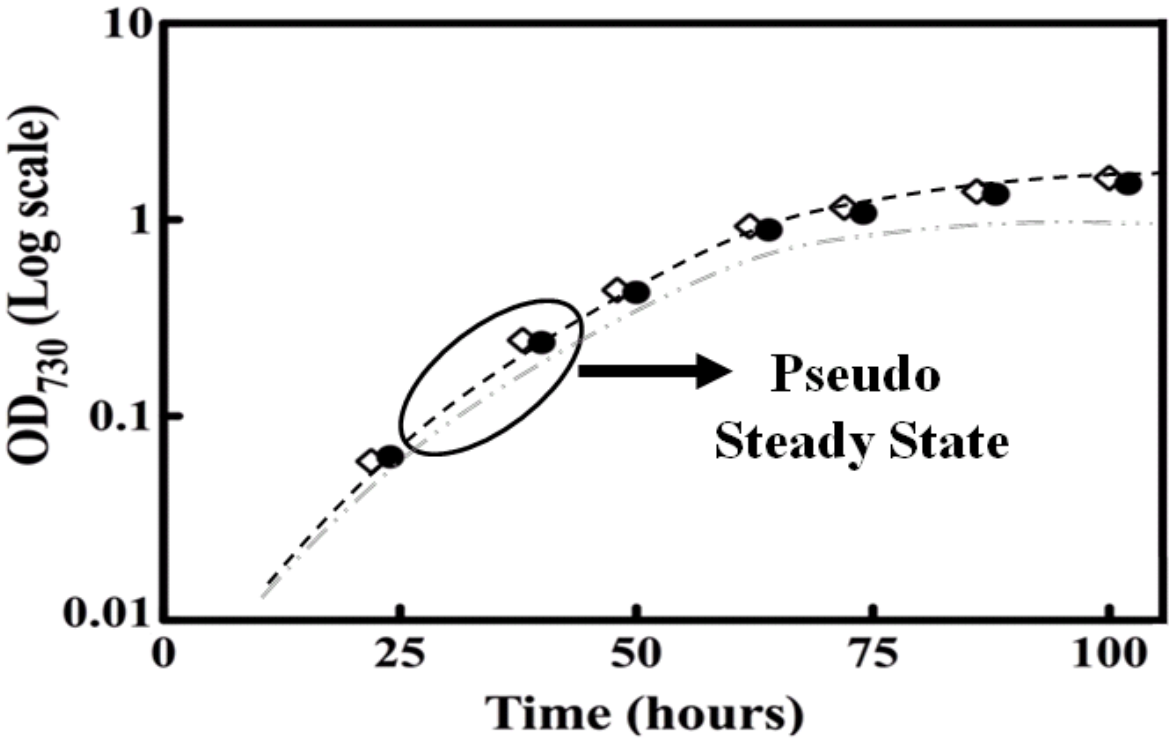


Figure A1.1: Representative growth curve of *Synechocystis* 6803 under photomixotrophic conditions. The cultures were grown in shake flasks (open circles) or serum bottles (closed circles) in modified liquid BG-11 medium supplemented with 5 g/L glucose and ~2 g/L sodium bicarbonate. The circle highlights the metabolic pseudo-steady state period of cultivation. The dot-dash line represents the growth curve under photoautotrophic conditions (in shake flasks). Each symbol represents the mean value of biological triplicate cultures.

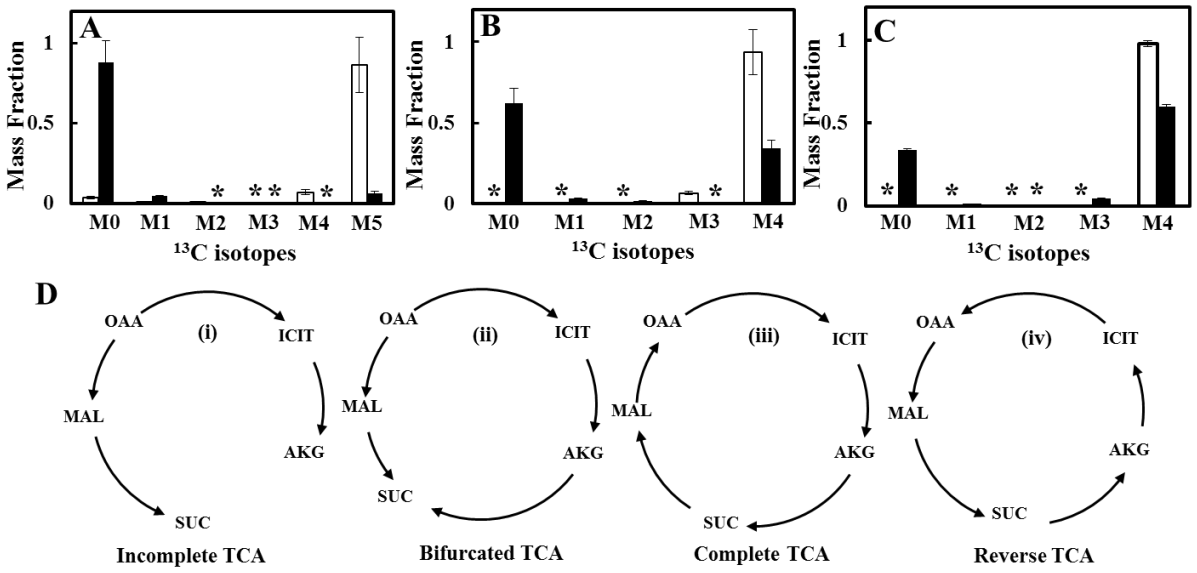


Figure A1.2: **Tracing the *Synechocystis* 6803 TCA pathway by isotopic dilution experiments.** A, B, and C show the isotopomer distributions of the $[M-15]^+$ fragment in α-ketoglutarate (AKG), succinate (SUC), and malate (MAL). Data are from biological duplicates and technical triplicates. *Synechocystis* 6803 was grown in the labeled medium with U-^{13}C glucose and $NaH^{13}CO_3$. Unlabeled glutamate was added during OD_{730} = ~0.4. Biomass samples were harvested after 30-min incubation from cultures with (■) or without (□) unlabeled glutamate. D shows different scenarios of microbial TCA pathways. Error bars indicate standard deviations of averages from two biological and three technical replicates. Stars in the figures indicate 0 values. Other abbreviations used are ICT for isocitrate and OAA for oxaloacetate.

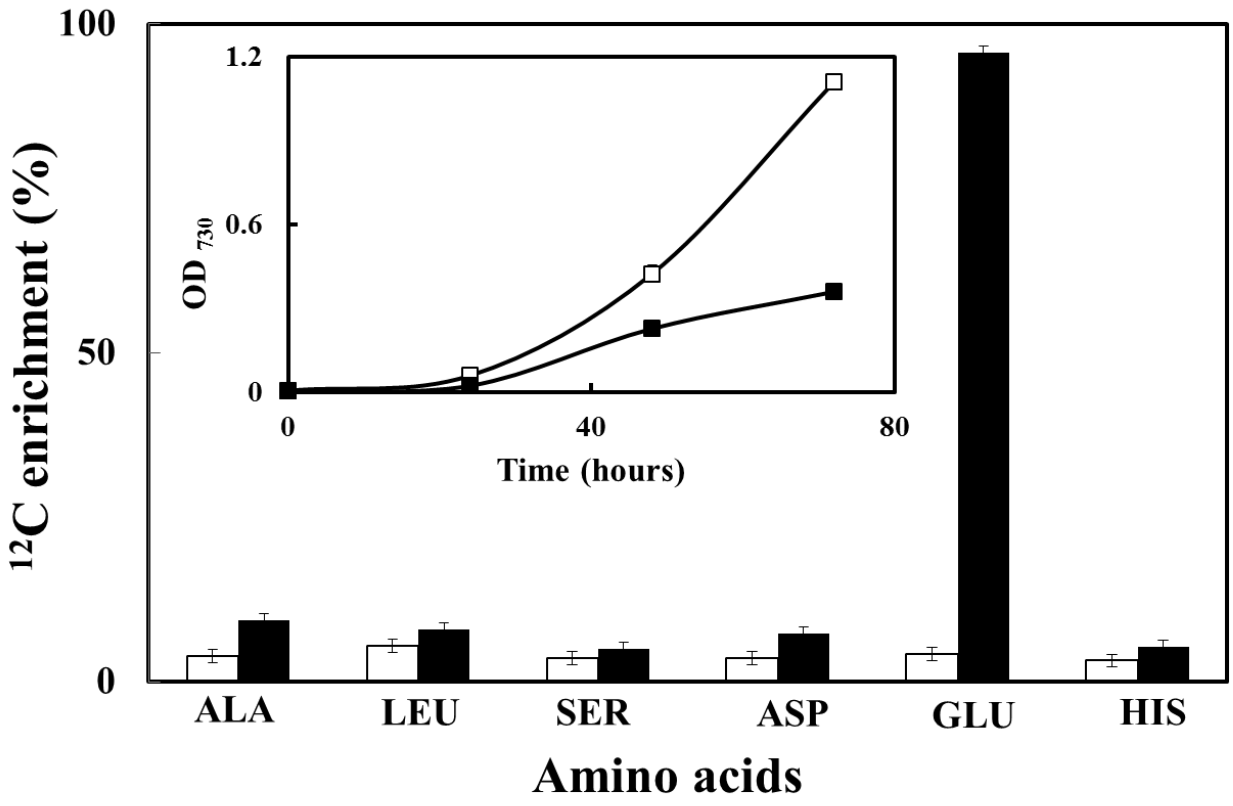


Figure A1.3: Long-term tracer experiment results with unlabeled glutamate. Biomass samples (OD_{730} = ~0.4, data from biomass duplicates and technical triplicates) were harvested after a 48-hour incubation in fully labeled photomixotrophic cultures with (■) or without (□) unlabeled glutamate. The ^{12}C-enrichment is calculated by $\sum_{i=0}^{n} \frac{i}{n} M_{(n-i)}$ (n is the total carbon number of an amino acid; *M* the relative molar concentration of mass isotopomer n-i). The error bars in the figure represent the standard deviation among samples (n=2). The inset figure shows growth of cultures with (■) or without (□) unlabeled glutamate (i.e., glutamate showed inhibition to cyanobacterial growth).

Abbreviations: ALA, alanine; ASP, aspartate; GLU, glutamate; HIS, histidine; LEU, leucine; SER, serine.

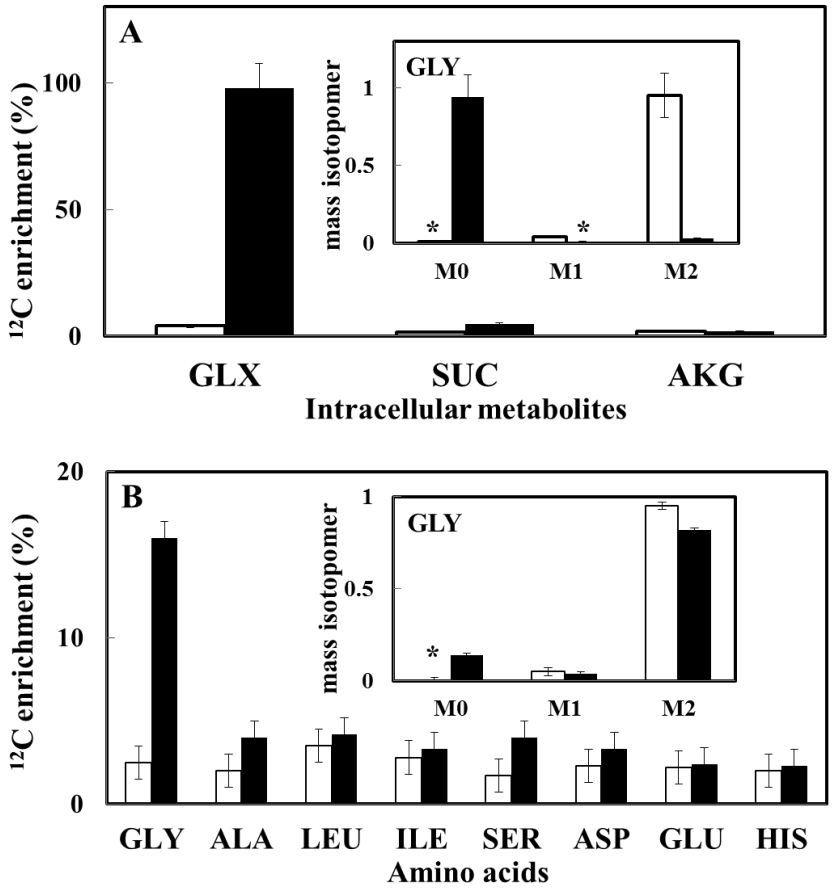


Figure A1.4: Tracing *Synechocystis* 6803 glyoxylate shunt by tracer experiments. A shows the ^{12}C-ratio of free metabolites. Biomass samples were harvested after 30-min incubation in fully labeled cultures (OD_{730} = ~0.4) with (■) or without (□) unlabeled glyoxylate. The inset shows the isotopomer distribution of free glycine. B shows the ^{12}C-enrichment of proteinogenic amino acids (n=2). Biomass samples were harvested after a 48-hour incubation in fully labeled cultures with (■) or without (□) unlabeled glyoxylate. The inset shows the isotopomer distribution of proteinogenic glycine. The error bars represent the standard deviations of averages from two biological and three technical replicates. Stars in the figure indicate that the MS peaks cannot be detected by GC-MS.

Abbreviations: AKG, α-ketoglutarate; ALA, alanine; ASP, aspartate; GLU, glutamate; GLX, glyoxylate; GLY, glycine; HIS, histidine; ICT, isocitrate; ILE, isoleucine; LEU, leucine; MAL, malate; OAA, oxaloacetate; SER, serine; SUC, succinate.

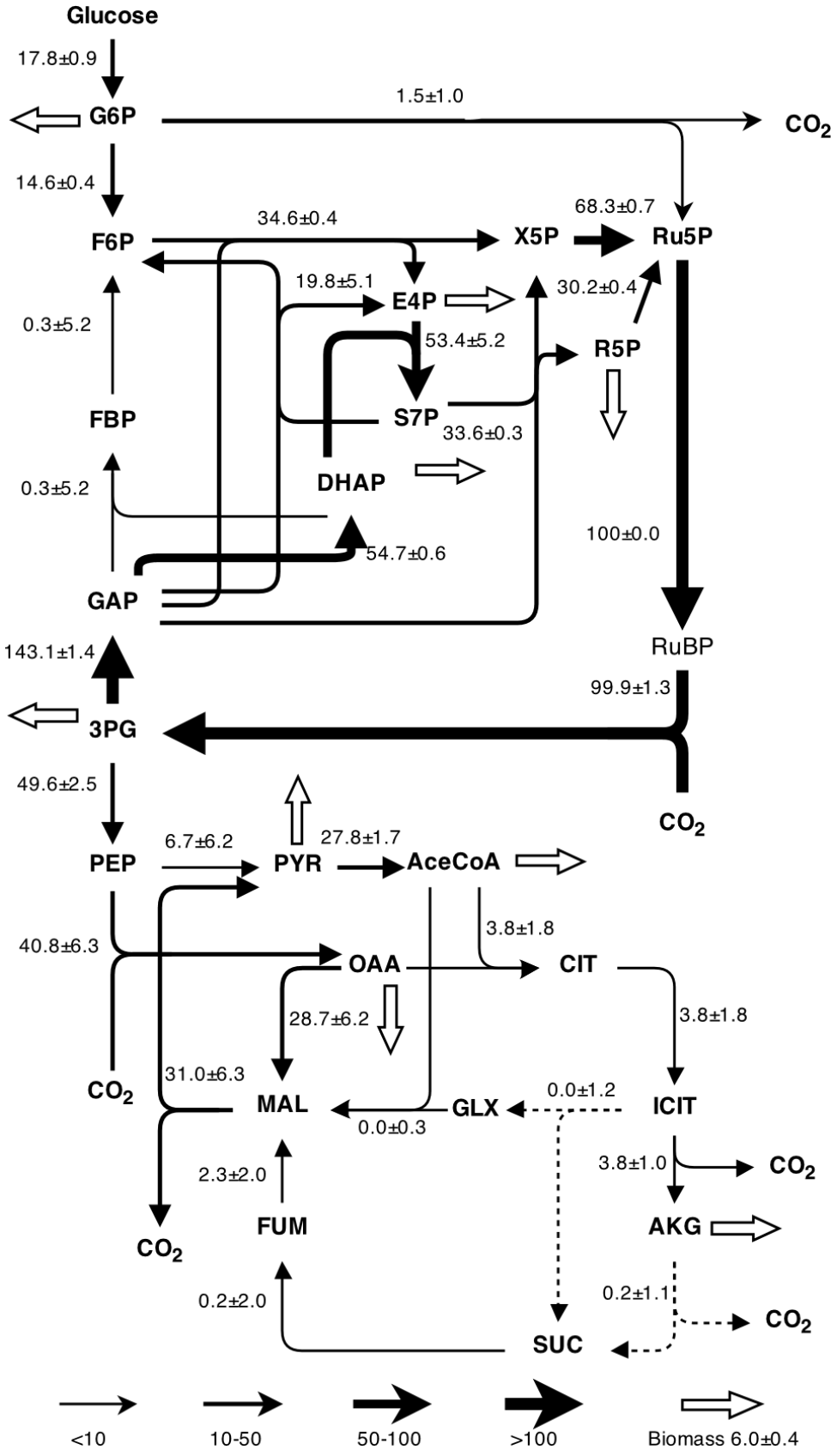

Glucose
17.8±0.9
G6P
1.5±1.0
CO2
14.6±0.4
F6P
34.6±0.4
X5P
68.3±0.7
Ru5P
19.8±5.1
E4P
30.2±0.4
0.3±5.2
53.4±5.2
R5P
FBP
S7P
33.6±0.3
DHAP
0.3±5.2
100±0.0
54.7±0.6
GAP
RuBP
143.1±1.4
99.9±1.3
3PG
49.6±2.5
CO2
6.7±6.2
27.8±1.7
PEP
PYR
AceCoA
3.8±1.8
40.8±6.3
OAA
CIT
28.7±6.2
3.8±1.8
31.0±6.3
CO2
MAL
GLX
0.0±1.2
ICIT
0.0±0.3
3.8±1.0
CO2
2.3±2.0
FUM
AKG
CO2
0.2±2.0
0.2±1.1
CO2
SUC
<10
10-50
50-100
>100
Biomass 6.0±0.4

Figure A1.5: (Previousorry.

Page). Flux distribution in the central metabolism of *Synechocystis* 6803 under photomixotrophic conditions. All the estimated relative flux rates are shown beside the pathways, which are normalized to the Calvin cycle flux (note: Ru5P → RuBP flux was assumed to be 100). The standard deviation of each flux was shown in the figure, which was calculated based on the 95% confidence intervals (Table A1.S1). The grey arrows represent the fluxes to biomass. The estimated glucose consumption rate was 0.24 mmol g^{-1} dry cell weight $hour^{-1}$ (Supporting information S2).

Abbreviations: 3PG, 3-phosphoglycerate; AKG, α-ketoglutarate; ALA, alanine; ASP, aspartate; CIT, citrate; DHAP, dihydroxyacetone phosphate; E4P, erythrose 4-phosphate; F6P, Fructose 6-phosphate; FUM, fumarate; G6P, glucose 6-phosphate; GAP, glyceraldehyde 3-phosphate; GLC, glycolate; GLU, glutamate; GLX, glyoxylate; GLY, glycine; HIS, histidine; ICT, isocitrate; ILE, isoleucine; LEU, leucine; MAL, malate; MTHF, 5,10-Methylenetetrahydrofolate; OAA, oxaloacetate; PEP, phosphoenolpyruvate; PYR, pyruvate; R5P, Ribose 5-phosphate; Ru5P, ribulose-5-phosphate; RuBP, ribulose-1,5-diphosphate; S7P, sedoheptulose-7- phosphate; SER, serine; SUC, succinate; THF, tetrahydrofolate; X5P, xylulose-5-phosphate

Appendix Chapter 2

Rapid Construction of Metabolic Models for a Family of Cyanobacteria Using a Multiple Source Annotation Workflow

A2.1. Abstract of the Chapter

A2.1.1. Background

Cyanobacteria are photoautotrophic prokaryotes that exhibit robust growth under diverse environmental conditions with minimal nutritional requirements. They can use solar energy to convert CO_2 and other reduced carbon sources into biofuels and chemical products. The genus *Cyanothece* includes unicellular nitrogen-fixing cyanobacteria that have been shown to offer high levels of hydrogen production and nitrogen fixation. The reconstruction of quality genome-scale metabolic models for organisms with limited annotation resources remains a challenging task.

A2.1.2. Results

Here we reconstruct and subsequently analyze and compare the metabolism of five *Cyanothece* strains, namely *Cyanothece* sp. PCC 7424, 7425, 7822, 8801 and 8802, as the genome-scale metabolic reconstructions *i*Cyc792, *i*Cyn731, *i*Cyj826, *i*Cyp752, and *i*Cyh755 respectively. We compare these phylogenetically related *Cyanothece* strains to assess their bio-production potential. A systematic workflow is introduced for integrating and prioritizing annotation information from the Universal Protein Resource (Uniprot), NCBI Protein Clusters, and the Rapid Annotations using Subsystems Technology (RAST) method. The genome-scale metabolic models include fully traced photosynthesis reactions and respiratory chains, as well as balanced reactions and GPR associations. Metabolic differences between the organisms are highlighted such as the non-fermentative pathway for alcohol production found in only *Cyanothece* 7424, 8801, and 8802.

A2.1.3. Conclusions

Our development workflow provides a path for constructing models using information from curated models of related organisms and reviewed gene annotations. This effort lays the foundation for the expedient construction of curated metabolic models for organisms that, while not being the target of comprehensive research, have a sequenced genome and are related to an organism with a curated metabolic model. Organism specific models, such as the five presented in this paper, can be used to identify optimal genetic manipulations for targeted metabolite overproduction as well as to investigate the biology of diverse organisms.

A2.2. Introduction

Genome-scale models (GSMs) are the collection of gene to protein to reaction associations (GPRs), charge and elementally balanced reactions, and constraints on molecular functions found within a cell (Price *et al.* 2004; Reed *et al.* 2006; Feist *et al.* 2009; Thiele *et al.* 2010). The constraints placed on molecular function define the possible phenotypes of an organism under specific conditions (Price *et al.* 2004). There are a number of applications for GSMs beyond the prediction of wildtype phenotypes in varying environments. These include the identification of optimal gene and medium modifications, non-native routes for metabolite production, and lethal gene deletions (Carneiro *et al.* 2006; Suthers *et al.* 2009; Ranganathan *et al.* 2010; Ranganathan *et al.* 2010; Zomorrodi *et al.* 2012). A genome-scale model of *Cyanothece* ATCC 51142, *i*Cyt773, was recently published (Saha *et al.* 2012). It contains 4 compartments, with 811 metabolites and 946 charge and elementally balanced reactions. *i*Cyt773 is an improvement upon the previously published *i*Cce806 model (Vu *et al.* 2012), and contains 43

genes and 266 reactions unique from *i*Cce806 (Saha *et al.* 2012). Further comparison of the two models can be found in the work by Saha *et al.* (2012). *i*Cyt773 also models the diurnal rhythm of *Cyanothece* metabolism. Since *Cyanothece* ATCC 51142 is closely related to all five *Cyanothece* species discussed in this paper (Bandyopadhyay *et al.* 2011), it was used in the development of the reconstructions for five organisms, *Cyanothece* PCC 7424, 7425, 7822, 8801, and 8802, as *i*Cyc792, *i*Cyn731, *i*Cyj826, *i*Cyp752, and *i*Cyh755 respectively (all five models are included in Additional Files 1 and 2). All models were named using their associated KEGG organism code. The objective of this study is to expediently generate models for a collection of members of a genus, using as a foundation an existing high-quality metabolic model for a representative member of the genus, while integrating information from a range of available sources.

The genus *Cyanothece* belongs to the phylum of Cyanobacteria. Cyanobacteria have a number of properties that make them ideal candidates for bio-production. Photosynthetic cyanobacteria bypass the need for sugar carbon substrates while having higher solar energy conversion efficiencies (i.e., 3-9%) than C3 (2.4%) and C4 plants (3.7%) (Dismukes *et al.* 2008). *Cyanothece* generate not only hydrogen (Tamagnini *et al.* 2002; Min *et al.* 2010; Bandyopadhyay *et al.* 2011; Melnicki *et al.* 2012) but also fix atmospheric nitrogen by temporally segregating the photosynthesis and nitrogenase activities (Welsh *et al.* 2008; Stockel *et al.* 2011). In addition, *Cyanothece* possess the potential to grow in air and can be easily fixed to solid matrices (Hall *et al.* 1995). All five species discussed in this paper are capable of fixing nitrogen and producing hydrogen, while *Cyanothece* sp. PCC 7425 is the only species that is not capable of accomplishing this task in an aerobic environment (Bandyopadhyay *et al.* 2011). PCC

7425 also varies in a number of physical characteristics, enough so that it has been suggested that it should be reclassified to another genus pending further review (Porta *et al.* 2000).

Cyanothece PCC 7424, 7425, 7822, 8801, and 8802, were all sequenced following the promising discoveries made concerning the metabolic capabilities of *Cyanothece* ATCC 51142 (Bandyopadhyay *et al.* 2011). These five species exhibit unique metabolic characteristics that motivated the development of five separate reconstructions. Fragments of a butanol producing pathway have been postulated to exist in all strains through an inspection of the *Cyanothece* genomes (Wu *et al.* 2010). Metabolic capabilities such as the alkane biosynthetic pathway and alternative pathways for breaking down arginine across species (Bandyopadhyay *et al.* 2011) have been hypothesized to exist as well. Given differences in metabolism, developed genetic systems (Min *et al.* 2010), and variations in growth characteristics, phenotype, and rhythms of nitrogen fixation and respiration (Bandyopadhyay *et al.* 2012), it is important to globally assess the metabolic repertoire of each strain separately.

There exist numerous databases devoted to gene annotations for a wide variety of organisms (Kanehisa *et al.* 2000; Gillespie *et al.* 2011; Kanehisa *et al.* 2012; The Uniprot Consortium 2012). However, the number of gene annotations is skewed towards a handful of extensively studied organisms. *Escherichia coli* K-12, the strain modeled in the *i*AF1260 metabolic reconstruction (Feist *et al.* 2007), has approximately 16 times the number of reviewed annotations (4,326) in the Universal Protein Resource (Uniprot) compared to *Cyanothece* PCC 7424 (271) (The Uniprot Consortium 2012). For most (microbial) organisms Uniprot contains only a small subset of required gene annotations (i.e., 200-300). Faced with this paucity of organism-specific gene annotation information, most metabolic reconstructions rely on a single database (i.e., typically KEGG) from which to pull gene annotations (Dal'Molin *et al.* 2011;

Balagurunathan *et al.* 2012; Licona-Cassani *et al.* 2012; The Uniprot Consortium 2012). This may introduce errors in the reconstruction as absent functionalities could be included in the model due to permissive homology cutoffs or errors in the original annotation source. In addition, specific and non-specific references to the same metabolite (e.g. D-Glucose vs. α-D-Glucose) and generic or unbalanced reactions (Dal'Molin *et al.* 2011) may also affect the consistency of the reconstruction. Integrating and contrasting information from multiple databases can remedy many of these shortcomings.

A systematic workflow is put forth that addresses the aforementioned challenges. It allows for the parallel reconstruction of genome-scale models for organisms that have a sequenced genome and are closely related to a species with a curated genome-scale model. Using this workflow, reconstructions were developed for all five *Cyanothece* species using *i*Cyt773 and reviewed annotations from Uniprot (The Uniprot Consortium 2012), NCBI Protein Clusters (Klimke *et al.* 2009), and the Rapid Annotations using Subsystems Technology (RAST) method (Aziz *et al.* 2008). These annotations were used to retrieve charge and elementally balanced reactions from both the *i*Cyt773 model and the SEED database (Overbeek R *et al.* 2005) for the construction of draft models. No reconciliation between the *i*Cyt773 and SEED reactions or metabolites was required as *i*Cyt773 was initially constructed using SEED notation when possible. The five models are all capable of producing biomass using the *i*Cyt773 biomass equations under diverse nutrient conditions. All five models are free of thermodynamically infeasible cycles, and the fractions of reactions mapped to specific genes (i.e., GPRs) are within the range of manually curated reconstructions. The use of multiple annotation sources helps to mitigate errors that may arise from a single source. Unlike automated draft models (i.e., Model SEED (Henry *et al.* 2010)), organism-specific metabolites such as pigments are included in the

biomass equation and light reactions are fully traced. This workflow is also more adept at excluding metabolites present in related species but absent in the reconstructed organism. For example, menaquinone and ubiquinone are known to not exist within *Cyanothece* (Collins *et al.* 1981), but are often pulled into draft models generated by automated software.

A2.3. Results and Discussion

A2.3.1. Model Comparisons

The five models were developed by combining reactions from the curated metabolic model, *i*Cyt773, with reactions taken from the SEED database whose presence in that organism were confirmed by reviewed annotations. The statistics for the five developed models are shown in Table A2.1 (See Additional Files 1 and 2 for model files). The model development workflow identified reactions that are in some cases unique to each reconstruction. However, closely related *Cyanothece* 8801 and 8802 have no unique reactions though they do contain a set of 30 reactions that are not found in any other reconstruction. All five models contain four compartments; cytosol, periplasm, thylakoid lumen, and carboxysome. The number of genes present in each reconstruction is similar to the number of open reading frames (ORFs) associated with the *i*Cyt773 and *i*Syn731 models. Figure A2.1 shows that the percent of non-exchange reactions without associated genes falls within ranges comparable to those of numerous manually curated models (Feist *et al.* 2007; Saha *et al.* 2012; Vu *et al.* 2012; Knoop *et al.* 2013). Biomass yields were also calculated for each of the five models using the same photoautotrophic conditions used to calculate the biomass yield for *i*Cyt773 (Saha *et al.* 2012). All five models had an identical yield of 0.026 mole biomass / mole carbon fixed.

Figure A2.2 shows the number of reactions shared between *i*Cyt773 and each one of the models. A total of 922 reactions from *i*Cyt773 are shared with at least one of the five models while 169 reactions have been added to all five models during the SEED reaction retrieval step of the workflow. The removal of these 169 reactions only affects biomass production in *i*Cyn731. It does not grow when the reactions are removed since one of the reactions is essential as it is the only Fe(II) oxidoreductase present within iCyn731. The other four models contain another Fe(II) oxidoreductase.. The number of reactions shared between each of the five models and *i*Cyt773 (Figure A2.2A) generally matches the phylogenetic relationships between the organisms (Bandyopadhyay *et al.* 2011). *Cyanothece* 7425, which is the farthest removed of the five species from *Cyanothece* 51142, also has the fewest identified homologs with *Cyanothece* 51142. The two most closely related pairs, *Cyanothece* 7424/7822, and 8801/8802, have the highest reaction similarities (see Figure A2.2B) while the farthest removed species, *Cyanothece* 7425, has the lowest similarity. This divergence lends support to the possibility of reclassification (Porta *et al.* 2000).

A2.3.2. Model Validation Using Published Findings

The effect of a gene knockout on an organism's phenotype is frequently used in assessing GSM quality (Knoop *et al.* 2010; Saha *et al.* 2012). However, unlike the CyanoMutants database for *Synechocystis* PCC 6803 (Nakamura *et al.* 1999; Nakao *et al.* 2010), none of the five species have a detailed repository of known mutants. The Δ*nifK* mutant for *Cyanothece* 7822 was shown to not be able to grow without the presence of combined nitrogen (nitrate) (Min *et al.* 2010). This finding implies the critical involvement of *nifK* in the fixation of nitrogen. In *i*Cyj826 this gene is involved in the GPR of the nitrogen fixation reaction present within the model. Given that the GPR describes *nifK* as one of three critical subunits of the enzyme, its deletion results in the

inability for that reaction to carry flux. Upon its removal from *i*Cyj826, the model is unable to grow without the addition of nitrate or ammonium, showing that the model reacts to the knockout in the same manner as the organism does *in vivo*.

Despite the many similarities between the 5 species, significant differences also exist (Bandyopadhyay *et al.* 2011). Genes that code for isocitrate lyase and malate synthase (glyoxylate shunt) are present only in *Cyanothece* 7424 and 7822 as reflected in the models. 2-oxoglutarate decarboxylase and succinic semialdehyde dehydrogenase, found in many cyanobacteria, complete the TCA cycle despite the absence of 2-oxoglutarate dehydrogenase (Zhang *et al.* 2011). Both of the enzymes in the alternate pathway are present within *i*Cyt773, and were transferred to all five models. The associated genes are also bidirectional best hits with the two genes in *Synechococcus* PCC 7002 that are associated with the aforementioned enzymes (Zhang *et al.* 2011). *i*Cyn731, *i*Cyp752, and *i*Cyh755 all contain an alkane biosynthetic pathway similar to what is present within *i*Cyt773. While *i*Cyt773 contains the pathway that enables the production of pentadecane from Hexadecenoyl-ACP, Schirmer *et al*. have measured heptadecane but not pentadecane production from *Cyanothece* 7425 (Schirmer *et al.* 2010). *i*Cyn731 contains only heptadecane production, while *i*Cyp752, and *i*Cyh755 contain pathways for both pentadecane and heptadecane (neither specific literature evidence neither in support nor in conflict with this was found). The two enzymes required, Hexadecenoyl-ACP reductase and Hexadecenal decarbonylase (EC 1.2.1.80 and 4.1.99.5 respectively per *i*Cyt773), have no corresponding annotations or orthologous genes in *Cyanothece* 7424 or 7822 (Schirmer *et al.* 2010).

Polyhydroxyalkanoates (PHAs) are a complex family of polyesters that can be synthesized by a wide variety of bacteria (Steinbuchel *et al.* 1995). *Cyanothece* 7424, 7425 and

7822 all contain the enzymes keto-thiolase and acetoacetyl-CoA reductase, which are necessary for the synthesis of polyhydroxyalkanoic acids (Steinbuchel *et al.* 1995; Rehm *et al.* 1999; Philip *et al.* 2007). There are RAST and unreviewed Uniprot annotations that identify genes within each of these three organisms associated with a PHA synthase. The non-fermentative pathway for higher alcohols exist in the 7424, 8801, and 8802 strains (Bandyopadhyay *et al.* 2011). The same pathway has been seen in *E. coli* (Atsumi *et al.* 2008; Clomburg *et al.* 2010) after the addition of the *kivD* gene from *Lactococcus lactis* (de la Plaza *et al.* 2004) and the adh2 gene from *Saccharomyces cerevisiae* (Russell *et al.* 1983). The pathway uses the 2-keto acid intermediates of amino acid biosynthesis and diverts them towards the synthesis of alcohols (Atsumi *et al.* 2008). The *kivD* gene encodes a 2-keto acid decarboxylase that acts on a wide range of substrates and enables the conversion of the 2-keto acids into aldehydes. The workflow identified genes in *Cyanothece* 7424, 8801 and 8802 which are bidirectional best hits with the *kivD* gene from *Lactococcus lactis*, and also annotated as being associated with the same EC number as *kivD*. An alcohol dehydrogenase, such as adh2, then converts these aldehydes into alcohols. The *adhA* gene (slr1192) in *Synechocystis* 6803 has been found to have wide substrate specificity that includes the aldehydes associated with butanol and propanol (Vidal *et al.* 2009). All five species contained a gene that was a bidirectional best hit with slr1192. While both the forward and reverse BLAST searches for *Cyanothece* 7425 had e-values in the order of 10^{-28} and percent identities of 30%, the searches, both forward and reverse, for the other four organisms had e-values ranging between 10^{-138} and 10^{-153} with percent identities ranging from 58 to 61%. The presence of orthologs to both a 2-keto acid decarboxylase and alcohol dehydrogenase with wide ranges of specificity in *Cyanothece* 7424, 8801, and 8802 provides annotation evidence for

the hypothesized presence of non-fermentative higher alcohol pathways (Bandyopadhyay *et al.* 2011).

Significant variations in nitrogen metabolism between the five species has been documented (Bandyopadhyay *et al.* 2011). Arginine decarboxylase is present in all 5 reconstructions, but differences arise in the subsequent agmatine catabolism. *Cyanothece* 51142 does not contain the associated genes for the conversion of agmatine to putrescine, and this is reflected in the *i*Cyt773 model (Bandyopadhyay *et al.* 2011; Saha *et al.* 2012) as these reactions are absent. Both *i*Cyc792 and *i*Cyj826 contain agmatinase and urease. The proposed pathway for agmatine breakdown into putrescine in *Cyanothece* 7425, 8801, and 8802 is through N-carbamoylputrescine. The two reactions required for this degradation can be found in all 3 associated models. Finally, as predicted by Bandyopadhyay *et al*. (2011), *i*Cyc792, *i*Cyj826, *i*Cyp752, and *i*Cyh755 contain the reactions required to break putrescine down into spermidine and spermine.

A2.3.3. Validation of Proposed Reconstruction Workflow

Additional reactions retrieved using reviewed annotations have provided a number of insights into the five species that would not have been either found or confirmed if reactions were only pulled from *i*Cyt773. The diverging nitrogen metabolism reactions were retrieved using SEED, as agmatine is the preferred polyamine for *Cyanothece* 51142 (Bandyopadhyay *et al.* 2011). An alternative butanol pathway is present in varying stages of completion in the five models. While butanol can be produced from a 2-keto acid as previously discussed, it can also be produced through the coenzyme A dependent pathway (Ezeji *et al.* 2007; Papoutsakis 2008). The coenzyme A dependent pathway was found to exist within a *Clostridium* species (Sillers *et al.* 2008; Yu *et al.* 2011). Figure A2.3 shows the comparative level of completion of the

fermentative butanol pathway within each of the 5 species. *Cyanothece* 7425 is the only organism to contain the complete pathway. The alcohol dehydrogenase exists within the models given the identification of homologs to the *Synechocystis adhA* gene (Vidal *et al.* 2009). The 7424/7822 and 8801/8802 pairs have the same enzymes. Figure A2.3 also shows e-values for the BLAST searches between the genes and the genes in the fermentative butanol pathway of *Clostridium acetobutylicum* ATCC 824. Given the lower e-value for Butanoyl-CoA dehydrogenase, it is the gene most likely to not be present or functional within *Cyanothece* 7425. The enzymes present in the five pathways mirror the phylogenetic relationship trends of the five species in a manner comparable to what was initially seen in the reaction similarities from Figure A2.2, as well as with the glyoxylate shunt and nitrogen metabolism.

The proposed workflow also served to complete unfinished pathways from *i*Cyt773. All five models are capable of converting galactose-1-phosphate to fructose-6-phosphate as in *i*Cyt773. Three of the models, *i*Cyn731, *i*Cyj826, and *i*Cyh755, also include the reaction that converts galactose into galactose-1-phosphate, enabling them to process galactose in the glycolysis pathway. Tetrahydrobiopterin (BH4) is a pteridine compound that acts as a cofactor for nitric oxide synthases and aromatic amino acid hydrolases in higher animals (Thony *et al.* 2000). Pteridine glycosides have been found in cyanobacteria, although their function is still unknown (Choi *et al.* 2001), and the first isolated pteridine glycosyltransferase from *Synechococcus elongatus* PCC 7942 acted on BH4 (Chung *et al.* 2000). Even though *i*Cyt773 does not contain the complete BH4 pathway as described by Thony *et al.* (2000), our workflow completed the pathway in all five species, identifying a gene that is a bidirectional best hit with the gene in *Synechococcus elongatus* PCC 7942. The reaction was not included in the models, as it does not exist within the SEED reaction database. All enzymes that were retrieved from

annotations but were not included in the model because of a lack of associated reaction in the subset of the SEED database used for model development are listed in Additional File 3.

Reactions not transferred from *i*Cyt773 offer insight into divergences between the metabolism of the new organism and the reference model. Two of the reactions that were not transferred from *i*Cyt773 to the models for *Cyanothece* 7424 and 7822 are responsible for the conversion of hexa- or octadecenoyl-ACP to *n*-hepta- or pentadecane. As previously mentioned it is accepted that the alkane biosynthetic pathway does not exist in these organisms (Schirmer *et al.* 2010). Another compound that is generally not found in the 5 species is xanthine, a purine base involved in the breakdown of purine ribonucleotides such as inosine-5'-phosphate and xanthosine-5'-phosphate, into uric acid. *i*Cyt773 can produce xanthine from either hypoxanthine or xanthosine, *i*Cyc792 only contains the reactions for production from xanthosine and cannot break xanthine down into uric acid. *i*Cyn731 only contains the reactions for production from hypoxanthine, but can convert xanthine into uric acid. The other three species do not contain any reactions involving xanthine and thus cannot process purine ribonucleotides through this pathway. Six reactions involved in transporting metabolites between the cytoplasm and periplasm or extracellular space were not transferred, such as molybdate transport via the ABC system. Given the likelihood that such reactions still exist within the other *Cyanothece* strains, it is possible that the associated GPR in *i*Cyt773 should be reevaluated for these reactions.

A2.3.4. Comparisons with Other Model Development Methods

Current model development methods can be generally characterized as manual, semi-automated, or automated. The workflow presented in this paper is best classified as semi-automated. This workflow allows for more expedited model development while avoiding some of the sources of error plaguing automated model generation and allowing for a wide range of

customization. This workflow can be adapted for use with any models, annotation sources, and additional reaction sets given annotation availability and user preferences.

Many draft models are nowadays generated through the identification and comparison of homologs with the GPRs of curated models (Sun *et al.* 2009; Pinchuk *et al.* 2010; Sun *et al.* 2010; Hamilton *et al.* 2012). Hamilton *et al.* (2012) identified the possibility for bidirectional BLAST searches to identify false positive ortholog pairs. The E-value cutoff for the searches performed for the test was 10^{-5}. Here we use a more conservative cutoff of 10^{-30} to safeguard against such instances. When the cutoff was relaxed from 10^{-30} to 10^{-5} for the bidirectional BLAST between *Cyanothece* 51142 and the five species there were between 280 and 403 additional best-hit pairs for each of the organisms. The number of these pairs that involved genes present in *i*Cyt773 varied between 15 for *Cyanothece* 7424 and 8801, and 26 for *Cyanothece* 7425. The reliance of manually constructed models on reviewing every annotation and manually curating the model greatly increases the time spent on development. This workflow helps to mitigate the need for manual review of each annotation by only using annotations that are reviewed or are derived from reviewed sources. Manual curation can then be reserved for certain key steps. Some of these models only include additional reactions beyond those retrieved from the curated models if the reactions are required for biomass production (Sun *et al.* 2009; Sun *et al.* 2010; Hamilton *et al.* 2012). This restricts the inclusion of reactions unique to either that organism or a subset of organisms that the reference models do not belong to. This introduces the risk of not including secondary metabolism pathways, which could be of great interest. The workflow presented here aims to overcome this through the use of outside annotations to retrieve SEED reactions.

There are a number of approaches for the automated development of metabolic reconstructions (Henry *et al.* 2010; Liao *et al.* 2011; Reyes *et al.* 2012; Vitkin *et al.* 2012) affording significant gains in development time, however, at the expense of some omissions and erroneous additions. The *Cyanothece* models created using the MIRAGE method contain generalized lipids along with a non-specific acceptor metabolite (Vitkin *et al.* 2012). Both the KBase and MIRAGE models constructed for *Cyanothece* 7424 contain menaquinone and ubiquinone, compounds shown to not exist within that organism (Collins *et al.* 1981). Conversely, there are a number of metabolites present in the biomass composition of the five reconstructed models that do not exist within either in the KBase or MIRAGE models (i.e., 22 specific lipid metabolites, 4 pigments and cyanophycin). The model produced through KBase also does not contain the pigment β-carotene. Many of these models do not have specified compartments apart from cytoplasm and extracellular space (Henry *et al.* 2010; Reyes *et al.* 2012; Vitkin *et al.* 2012). Automated model development can exclude unique metabolic pathways if they are absent from the training set of models. Specifically, both the MIRAGE and KBase models generally lack light reactions.

Other methods that combine manual and automated steps provide their own distinct approach to model reconstruction. The RAVEN toolbox (Agren *et al.* 2013) allows for the curation of a reconstruction from models of related species using homologs identified through BLAST bidirectional best hits, and additional unique functions supplied through annotations taken from KEGG Orthology (Kanehisa *et al.* 2000). This method was employed for the construction of the *Penicillium chrysogenum* model *i*AL1006 (Agren *et al.* 2013). Our workflow can currently pull from up to three sources, with the ability to quickly expand the number of sources sampled, resulting in more identified EC numbers with higher confidence.

A2.4. Conclusions

In this paper we presented a workflow that was used to rapidly develop curated models for five *Cyanothece* strains using the previously published *i*Cyt773 model and reviewed annotations from numerous sources. The comparisons between these five models line up with the established phylogenetic relationships between the species. Specific reactions that were both kept from being taken from *i*Cyt773 or added from the SEED database demonstrate the efficacy of this workflow and provide insights into the metabolism of the five species, as well as suggesting areas for further curation in the *i*Cyt773 model. This workflow can easily be adapted to work with other metabolic models, annotation sources, and reaction databases. All five models (*i*Cyc792, *i*Cyn731, *i*Cyj826, *i*Cyp752, and *i*Cyh755) are included in the supplementary material.

A2.5. Methods

A2.5.1. Draft Model Development

Draft models for the five organisms were developed using the workflow shown in Figure A2.4, which uses a combination of reviewed gene annotations and identified homologs between the new organism and *Cyanothece* 51142. Reactions that were determined to exist in both *Cyanothece* 51142 and the organism being modeled were transferred from *i*Cyt773 to the draft model. This reaction sharing was established through a comparison of homologs between the two genomes. These homologs were determined using a bidirectional BLAST search between the genomes of *Cyanothece* 51142 and the organism, using an E-value cut off of 10^{-30} and the requirement of mutual best hits. The Boolean logic given by each GPR in *i*Cyt773 was evaluated using these bidirectional hits. If the organism contained the homologs required to satisfy the

logic and encode the protein, the reaction was transferred to the draft model. This only requires one isozyme to be present within the organism (i.e. if the associated genes for a reaction are listed as "gene A OR gene B OR gene C", only one of the three genes must have a homolog), yet requires that all genes that code for an essential protein complex have a homolog. These identified reactions were added to the draft model with the GPRs modified to reflect the homologs present in the organism. Both the homology searches and identification of reactions to be included in the model were automated steps.

Reviewed annotations retrieved from Uniprot (The Uniprot Consortium 2012), NCBI Protein Clusters (Klimke *et al.* 2009), and RAST (Aziz *et al.* 2008), are used to support the inclusion of additional reactions into the draft models. An automated process was used to retrieve annotations that reference specific enzyme commission (EC) numbers, along with the EC numbers associated with the reactions retrieved using bidirectional BLAST. Only specific EC numbers were used to avoid the inclusion of unnecessary reactions. For some genes the annotations are inconsistent. These discrepancies are resolved through a manual multi-step procedure shown in Figure A2.4. First the EC numbers are checked to confirm that they have not been transferred to a new number. An example of this transfer of EC numbers can be seen with the annotations for the *Cyanothece* 7424 gene, PCC7424_1895. Both Uniprot and NCBI Clusters assigned the EC 2.5.1.75 to the gene, whereas the RAST method assigned the EC number 2.5.1.8. Despite the apparent mismatch, EC 2.5.1.8 had previously been transferred to 2.5.1.75, resolving any conflict between the annotations. If the enzymes are uniquely classified, a search of literature, specifically the InterPro database (Hunter *et al.* 2012), is then performed to validate their existence (or non-existence) in the organism. The *Cyanothece* 7424 gene, PCC7424_2477 has an associated annotation of 1.1.1.29 from *i*Cyt773, whereas RAST assigns both 1.1.1.26 and

1.1.1.81 to the gene. InterPro states that the 1.1.1.26 enzyme belongs to a protein family that is found in hyperthermophilic archaea, thus ruling out its existence in *Cyanothece* 7424. After using the InterPro information to rule out a possible associated enzyme, the annotation is resolved through order of confidence (described below), and 1.1.1.29 is attributed to the gene. Next, any enzymes that are associated with generic metabolites, or metabolites known to not be found within the organism, are removed. Such filtering can be seen with the *Cyanothece* 7425 gene Cyan7425_1569. Both the model and RAST annotation suggest that succinate dehydrogenase (1.3.99.1) is associated with this gene. However NCBI Protein Clusters suggests enzyme 1.3.5.1, which is a succinate dehydrogenase specific to ubiquinone. As ubiquinone is not present within *Cyanothece* (Collins *et al.* 1981), this conflict is resolved. The list of all reactions removed from each model for containing generic metabolites is included in Additional File 4. If discrepancies still exist, annotation resolutions are made based on a confidence order of *i*Cyt773, Uniprot, NCBI, and RAST. The order of confidence is derived from the likelihood that a source has been manually reviewed and is applicable to the individual gene in question. *i*Cyt773 GPR relationships were curated specifically for a *Cyanothece* model. Uniprot reviewed annotations are manually annotated individually (The Uniprot Consortium 2012), while the protein cluster annotations used in this study are curated as a group of related genes (Klimke *et al.* 2009), and RAST annotations are developed using the manually curated FIGfams (Aziz *et al.* 2008; Meyer *et al.* 2009). Lower confidence is placed in these annotations, as it is possible that the automated RAST program could improperly assign annotations in some cases. If all of the enzymes proposed by the other annotation sources are contained within the list of enzymes found to relate to the gene through inspection of *i*Cyt773, the annotation is not listed as conflicting and the enzymes from the model are used. There were on average between 40 and 50 genes with

conflicting annotations. Between 55 and 70% of conflicts required order of confidence to resolve. Using multiple sources allows for the identification of probable errors in the databases. These annotations can also reveal errors in other databases not used in the model development. One such example is gene PCC7424_2817 in the *Cyanothece* 7424 genome. All sources used in this paper, along with KEGG Orthology (Kanehisa *et al.* 2012), indicate that the enzyme associated with this gene is 2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylic-acid synthase (EC 2.2.1.9). Both the KEGG and REFSEQ (Pruitt *et al.* 2012) annotations list the same enzyme name, but list the EC number as 4.1.1.71 (associated with 2-oxoglutarate decarboxylase).

Subsequently, this resolved list of EC numbers is referenced against the *i*Cyt773 model. Reactions with a matching EC number are retained, and the remaining EC numbers are used to retrieve reactions from the SEED database (Overbeek R *et al.* 2005). Reactions are only taken from the subset used by the SEED service for GapFilling (Satish Kumar *et al.* 2007), as these reactions are confirmed to be charge and elementally balanced. Those EC numbers that did not have an associated reaction within this set of SEED reactions and were therefore not included within the models are compiled in Additional File 3. All duplicate reactions retrieved from *i*Cyt773 are removed while the remaining reactions necessary for photosynthesis are included. These reactions are known to exist within the organisms, as they can grow autotrophically. Any oxidative phosphorylation reactions or diffusion transport reactions that had not previously been added to the model are appended given their obvious essentiality. This set of reactions constitutes the draft model. All steps in draft model development are automated except for the EC annotation reconciliation. The time required to complete this step is reduced as more models are developed, and results can be applied to related organisms.

A2.5.2. Biomass and Removal of Thermodynamically Infeasible Cycles

The four biomass descriptions developed for the *i*Cyt773 model were used in the 5 models (Saha *et al.* 2012). Initially, all draft models were not capable of producing biomass. A subset of reactions from iCyt773 needed to be included in the draft models to allow for the generation of biomass. A mixed integer linear program was used to determine the minimal set of additional reactions required for the production of biomass. All alternative solutions within two reactions of the global minimum were found, and every reaction was examined for evidence suggesting its existence within the organism. Given the necessity of their inclusion for biomass production even reactions with no identified evidence were included in the models. In situations with several alternate solutions, the solution that contained the most reactions with evidence for their inclusion was chosen. Necessary reactions, which could not have previously been included in the models as they did not have associated enzymes or genes, were added at this point. Between 3 and 8 reactions with a GPR in *i*Cyt773 that did not have direct literature or annotation evidence were included in order to produce biomass. A substantial number of these reactions did not have both a gene and enzyme associated in *i*Cyt773, which would lower their chance to be included during the initial stages of draft model development (See Additional File 5 for a full list of reactions included in this step). While the initial reaction set was generated for the production of 1% of the maximum biomass when all *i*Cyt773 reactions were included, the inclusion of two reactions expected to be present in all models, the exchange reaction for oxygen and the diffusive transport of carbon dioxide between the periplasm and cytoplasm, allowed for biomass production exceeding 90% of the maximum. The 7425 model requires an additional two reactions to produce maximum biomass, but the other four models are capable of such production with the addition of the carbon dioxide transport and oxygen exchange reactions. This

process was performed for both autotrophic and heterotrophic growth conditions. For autotrophic growth, 16 reactions were added to *i*Cyc792*,* 24 to *i*Cyn731, and 18 to *i*Cyj826, *i*Cyp752, and *i*Cyh755. The same approach was used for heterotrophic growth, where only *i*Cyn731 required the inclusion of one reaction to grow under heterotrophic conditions.

The models were further modified to avoid the presence of thermodynamically infeasible cycles. Flux variability analysis was performed to identify unbounded reaction fluxes. Given the absence of thermodynamically infeasible cycles within *i*Cyt773, added reactions from SEED were solely responsible for the creation of any cycles. The number of SEED reactions present in cycles varied between 39 in *i*Cyh755 and 51 in both *i*Cyn731 and *i*Cyj826. Three steps were taken to modify the SEED reactions involved in the cycles. First the Gibbs free energy values provided by SEED were examined. Any reactions where the entire free energy value range, factoring in error, was more than 4 kcal/mol removed from zero was restricted to the directionality specified by Gibbs free energy. Any SEED reactions whose fluxes still hit the bounds were restricted to the direction opposite of the cycle. The annotations of the few SEED reactions that were still involved in cycles were inspected. All of these reactions were supported solely by RAST annotations. Given this lower confidence due to the single-source annotation, the reactions, (between 4 and 10 for each model) were removed. Additional File 6 lists all SEED reactions found in cycles, along with any reaction modifications made to eliminate the cycles.

A2.5.3. GPR Development

GPR relationships were primarily derived from either the previous bidirectional blast analysis of *i*Cyt773 reactions or the analysis of retrieved annotations. Bidirectional best hits were previously used to evaluate the presence of each reaction in the new organism. If a reaction is

added to the model, the GPR for every isozyme or complete subunit that is present is translated to the list of genes for the new organism.

The GPR relationships for reactions retrieved from SEED were developed by applying the Autograph method (Notebaart *et al.* 2006). All genes that were linked to an enzyme through an annotation were used for the GPR for each reaction associated with that enzyme. If there are RAST annotations for each of these genes with the correct EC annotation, then they are used for the comparison. For all five species there were no ECs for which this was not the case. Genes that shared the same annotation designation were determined to be isozymes while those with different names were seen to be subunits of a protein. There is a small subset of reactions in the models that were taken from *i*Cyt773 because of either proof of their existence (e.g. photosynthetic reactions) or their requirement for biomass production. Many of these GPR relationships are missing a small number of bidirectional best hits. For these genes the BLAST cutoff was reduced to 10^{-10}. These few additional best hits aided in the resolution of many of the remaining reactions, leaving between 6 and 13 of the reactions without a transferred GPR.

A2.5.4. Model Simulations and Analysis

Flux balance analysis (Jeffrey D Orth *et al.* 2010) was used in both the model development and model validation phases to determine flux distribution under varying conditions.

Maximize $v_{biomass}$

Subject to

$$\sum_{j=1}^{M} S_{ij} v_j = 0, \ \forall i \in 1,...,N \qquad \text{(A2.1)}$$

$$v_{j,min} \leq v_j \leq v_{j,max}, \ \forall j \in 1,...,M \qquad \text{(A2.2)}$$

Where S_{ij} is the stoichiometric coefficient for metabolite i in reaction j, $v_{j,min}$ and $v_{j,max}$ denote the minimum and maximum flux values for reaction j, while v_j represents the flux value of reaction j. N and M denote the total number of metabolites and reactions respectively.

A mixed integer linear program was used in the determination of a minimal set of reactions necessary for biomass production using the following formulation.

$$Minimize \sum_{j}^{M} y_j$$

Subject to

$$\sum_{j=1}^{M} S_{ij} v_j = 0, \forall i \in 1,...,N \quad \text{(A2.3)}$$

$$v_{j,min}\, y_j \leq v_j \leq v_{j,max} y_j, \forall j \in 1,...,M \quad \text{(A2.4)}$$

$$v_{Biomass} \geq v_{Biomass}^{min} \quad \text{(A2.5)}$$

All reactions were assigned a binary variable y_j, which when equal to zero eliminates flux through reaction j. The value of y for all reactions present in the draft model was fixed at one. Biomass production was fixed at greater than 1% of the maximum value when all *i*Cyt773 reactions were included, and the number of included reactions was minimized.

Flux variability analysis was used for identification of reactions present within cycles, and used the following formulation.

$$\begin{bmatrix} Max / Min\ v_j \\ \text{Subject to} \\ \sum_{j=1}^{M} S_{ij} v_j = 0, \forall i \in 1,...,N \\ v_{j,min} \leq v_j \leq v_{j,max}, \forall j \in 1,...,M \end{bmatrix} \quad \text{(A2.6)}$$

$$\forall j \in 1,...,M \quad \text{(A2.7)}$$

No constraints were placed on the biomass growth so as to identify all possible cycles within the model. This analysis was performed iteratively after each series of modifications was made to the reactions present within the cycles.

The reaction similarity between any two models is calculated using the following formula,

$$Similarity = \sqrt{\frac{A^2}{(A+B)(A+C)}} \quad \text{(A2.8)}$$

A denotes the total number of shared reactions between the two organisms, whereas B and C represent the number of unique reactions in each model.

CPLEX solver (version 12.3 IBM ILOG) was used in the GAMS (version 23.3.3, GAMS Development Corporation) environment for solving the optimization models. All computations were carried out on Intel Xeon X5675 Six-Core 3.06 GHz processors that are a part of the lionxf cluster, which was built and operated by the Research Computing and Cyberinfrastructure Group of The Pennsylvania State University. All codes used in model development were written using the Python programming language.

A2.6. Supplemental Material

The following supporting information is available online with the originally published version of this article (Mueller *et al.* 2013) at DOI:10.1186/1752-0509-7-142.

Additional File 1 (.xlsx): All five genome-scale reconstructions (*i*Cyc792, *i*Cyn731, *i*Cyj826, *i*Cyp752, and *i*Cyh755), along with GPR and metabolite information.

Additional File 2 (.zip): A zipped file containing all five genome-scale reconstructions (*i*Cyc792, *i*Cyn731, *i*Cyj826, *i*Cyp752, and *i*Cyh755), along with GPR and metabolite information in SBML format.

Additional File 3 (.xlsx): List of retrieved EC numbers not associated with any reactions in the SEED subset of reactions used for draft model development.

Additional File 4 (.xlsx): List of reactions removed from the five models for containing generic metabolites.

Additional File 5 (.xlsx): Set of reactions added to each model to insure biomass production, along with associated support for inclusion.

Additional File 6 (.xlsx): All reaction modifications made to eliminate thermodynamically infeasible cycles.

A2.7. References

Agren, R., Liu, L. M., Shoaie, S., Vongsangnak, W., Nookaew, I. and Nielsen, J. (2013). "The RAVEN Toolbox and Its Use for Generating a Genome-scale Metabolic Model for Penicillium chrysogenum." *PLOS Computational Biology* **9**(3).

Atsumi, S., Hanai, T. and Liao, J. C. (2008). "Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels." *Nature* **451**(7174): 86-U13.

Aziz, R., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R., Formsma, K., Gerdes, S., Glass, E., Kubal, M., Meyer, F., Olsen, G., Olson, R., Osterman, A., Overbeek, R., McNeil, L., Paarmann, D., Paczian, T., Parrello, B., Pusch, G., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A. and Zagnitko, O. (2008). "The RAST Server: Rapid Annotations using Subsystems Technology." *BMC Genomics*.

Balagurunathan, B., Jonnalagadda, S., Tan, L. and Srinivasan, R. (2012). "Reconstruction and analysis of a genome-scale metabolic model for Scheffersomyces stipitis." *Microb Cell Fact* **11**: 27.

Bandyopadhyay, A., Elvitigala, T., Liberton, M. and Pakrasi, H. B. (2012). "Variations in the rhythms of respiration and nitrogen fixation in members of the unicellular diazotrophic cyanobacterial genus Cyanothece." *Plant Physiol*.

Bandyopadhyay, A., Elvitigala, T., Welsh, E., Stöckel, J., Liberton, M., Min, H., Sherman, L. A. and Pakrasi, H. B. (2011). "Novel Metabolic Attributes of the Genus Cyanothece, Comprising a Group of Unicellular Nitrogen-Fixing Cyanobacteria." *mBio* **2**(5).

Carneiro, S., Rocha, I. and Ferreira, E. (2006). "Application of a genome-scale metabolic model to the inference of nutritional requirements and metabolic bottlenecks during recombinant protein production in Escherichia coli." *Microbial Cell Factories* **5**.

Choi, Y. K., Hwang, Y. K. and Park, Y. S. (2001). "Molecular cloning and disruption of a novel gene encoding UDP-glucose: tetrahydrobiopterin alpha-glucosyltransferase in the cyanobacterium Synechococcus sp. PCC 7942." *FEBS Lett* **502**(3): 73-78.

Chung, H. J., Kim, Y. A., Kim, Y. J., Choi, Y. K., Hwang, Y. K. and Park, Y. S. (2000). "Purification and characterization of UDP-glucose:tetrahydrobiopterin glucosyltransferase from Synechococcus sp. PCC 7942." *Biochim Biophys Acta* **1524**(2-3): 183-188.

Clomburg, J. M. and Gonzalez, R. (2010). "Biofuel production in Escherichia coli: the role of metabolic engineering and synthetic biology." *Appl Microbiol Biotechnol* **86**(2): 419-434.

Collins, M. D. and Jones, D. (1981). "Distribution of Isoprenoid Quinone Structural Types in Bacteria and Their Taxonomic Implications." *Microbiological Reviews* **45**(2): 316-354.

Dal'Molin, C. G., Quek, L. E., Palfreyman, R. W. and Nielsen, L. K. (2011). "AlgaGEM--a genome-scale metabolic reconstruction of algae based on the Chlamydomonas reinhardtii genome." *BMC Genomics* **12 Suppl 4**: S5.

de la Plaza, M., Fernandez de Palencia, P., Pelaez, C. and Requena, T. (2004). "Biochemical and molecular characterization of alpha-ketoisovalerate decarboxylase, an enzyme involved in the formation of aldehydes from amino acids by Lactococcus lactis." *FEMS Microbiol Lett* **238**(2): 367-374.

Dismukes, G. C., Carrieri, D., Bennette, N., Ananyev, G. M. and Posewitz, M. C. (2008). "Aquatic phototrophs: efficient alternatives to land-based crops for biofuels." *Curr Opin Biotechnol* **19**(3): 235-240.

Ezeji, T. C., Qureshi, N. and Blaschek, H. P. (2007). "Bioproduction of butanol from biomass: from genes to bioreactors." *Curr Opin Biotechnol* **18**(3): 220-227.

Feist, A. M., Henry, C., Reed, J. L., Krummenacker, M., Joyce, A., Karp, P., Broadbelt, L., Hatzimanikatis, V. and Palsson, B. O. (2007). "A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information." *Molecular Systems Biology* **3**(121).

Feist, A. M., Herrgård, M. J., Thiele, I., Reed, J. L. and BO, P. (2009). "Reconstruction of Biochemical Networks in Microbial Organisms." *Nat. Rev Microbiol* **7**(2): 129-143.

Gillespie, J. J., Wattam, A. R., Cammer, S. A., Gabbard, J. L., Shukla, M. P., Dalay, O., Driscoll, T., Hix, D., Mane, S. P., Mao, C., Nordberg, E. K., Scott, M., Schulman, J. R., Snyder, E. E., Sullivan, D. E., Wang, C., Warren, A., Williams, K. P., Xue, T., Yoo, H. S., Zhang, C., Zhang, Y., Will, R., Kenyon, R. W. and Sobral, B. W. (2011). "PATRIC: the comprehensive bacterial bioinformatics resource with a focus on human pathogenic species." *Infect Immun* **79**(11): 4286-4298.

Hall, D., Markov, S., Watanabe, Y. and Rao, K. (1995). "The potential applications of cyanobacterial photosynthesis for clean technologies." *Photosynthesis Research* **46**: 159-167.

Hamilton, J. J. and Reed, J. L. (2012). "Identification of Functional Differences in Metabolic Networks Using Comparative Genomics and Constraint-Based Models." *PLoS One* **7**(4).

Henry, C. S., DeJongh, M., Best, A. A., Frybarger, P. M., Linsay, B. and Stevens, R. L. (2010). "High-throughput generation, optimization and analysis of genome-scale metabolic models." *Nat Biotechnol* **28**(9): 977-982.

Hunter, S., Jones, P., Mitchell, A., Apweiler, R., Attwood, T. K., Bateman, A., Bernard, T., Binns, D., Bork, P., Burge, S., de Castro, E., Coggill, P., Corbett, M., Das, U., Daugherty, L., Duquenne, L., Finn, R. D., Fraser, M., Gough, J., Haft, D., Hulo, N., Kahn, D., Kelly, E., Letunic, I., Lonsdale, D., Lopez, R., Madera, M., Maslen, J., McAnulla, C., McDowall, J., McMenamin, C., Mi, H., Mutowo-Muellenet, P., Mulder, N., Natale, D., Orengo, C., Pesseat, S., Punta, M., Quinn, A. F., Rivoire, C., Sangrador-Vegas, A., Selengut, J. D., Sigrist, C. J., Scheremetjew, M., Tate, J., Thimmajanarthanan, M., Thomas, P. D., Wu, C. H., Yeats, C. and Yong, S. Y. (2012). "InterPro in 2011: new developments in the family and domain prediction database." *Nucleic Acids Res* **40**(Database issue): D306-312.

Jeffrey D Orth, Ines Thiele and Palsson, B. O. (2010). "What is flux balance analysis?" *Nature Biotechnology* **28**(3).

Kanehisa, M. and Goto, S. (2000). "KEGG: Kyoto Encyclopedia of Genes and Genomes." *Nucleic Acids Research* **28**: 27-30.

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. and Tanabe, M. (2012). "KEGG for integration and interpretation of large-scale molecular datasets." *Nucleic Acids Research*.

Klimke, W., Agarwala, R., Badretdin, A., Chetvernin, S., Ciufo, S., Fedorov, B., Kiryutin, B., O'Neill, K., Resch, W., Resenchuk, S., Schafer, S., Tolstoy, I. and Tatusova, T. (2009). "The National Center for Biotechnology Information's Protein Clusters Database." *Nucleic Acids Research*.

Knoop, H., Grundel, M., Zilliges, Y., Lehmann, R., Hoffmann, S., Lockau, W. and Steuer, R. (2013). "Flux balance analysis of cyanobacterial metabolism: the metabolic network of Synechocystis sp. PCC 6803." *PLoS Comput Biol* **9**(6): e1003081.

Knoop, H., Zilliges, Y., Lockau, W. and Steuer, R. (2010). "The metabolic network of Synechocystis sp. PCC 6803: systemic properties of autotrophic growth." *Plant Physiol* **154**(1): 410-422.

Liao, Y. C., Chen, J. C., Tsai, M. H., Tang, Y. H., Chen, F. C. and Hsiung, C. A. (2011). "MrBac: a web server for draft metabolic network reconstructions for bacteria." *Bioeng Bugs* **2**(5): 284-287.

Licona-Cassani, C., Marcellin, E., Quek, L., Jacob, S. and Nielsen, L. (2012). "Reconstruction of the Saccharopolyspora erythraea genome-scale model and its use for enhancing erythromycin production." *Antonie van Leeuwenhoek* **102**(3): 493-502.

Melnicki, M. R., Pinchuk, G. E., Hill, E. A., Kucek, L. A., Fredrickson, J. K., Konopka, A. and Beliaev, A. S. (2012). "Sustained H(2) production driven by photosynthetic water splitting in a unicellular cyanobacterium." *mBio* **3**(4): e00197-00112.

Meyer, F., Overbeek, R. and Rodriguez, A. (2009). "FIGfams: yet another set of protein families." *Nucleic Acids Res* **37**(20): 6643-6654.

Min, H. and Sherman, L. A. (2010). "Genetic transformation and mutagenesis via single-stranded DNA in the unicellular, diazotrophic cyanobacteria of the genus Cyanothece." *Appl Environ Microbiol* **76**(22): 7641-7645.

Min, H. and Sherman, L. A. (2010). "Hydrogen production by the unicellular, diazotrophic cyanobacterium Cyanothece sp. strain ATCC 51142 under conditions of continuous light." *Appl Environ Microbiol* **76**(13): 4293-4301.

Mueller, T. J., Berla, B. M., Pakrasi, H. B. and Maranas, C. D. (2013). "Rapid construction of metabolic models for a family of Cyanobacteria using a multiple source annotation workflow." *BMC Syst Biol* **7**: 142.

Nakamura, Y., Kaneko, T., Miyajima, N. and Tabata, S. (1999). "Extension of CyanoBase. CyanoMutants: repository of mutant information on Synechocystis sp. strain PCC6803." *Nucleic Acids Res* **27**(1): 66-68.

Nakao, M., Okamoto, S., Kohara, M., Fujishiro, T., Fujisawa, T., Sato, S., Tabata, S., Kaneko, T. and Nakamura, Y. (2010). "CyanoBase: the cyanobacteria genome database update 2010." *Nucleic Acids Res* **38**(Database issue): D379-381.

Notebaart, R. A., van Enckevort, F. H., Francke, C., Siezen, R. J. and Teusink, B. (2006). "Accelerating the reconstruction of genome-scale metabolic networks." *BMC Bioinformatics* **7**: 296.

Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, de Crécy-Lagard V, Diaz N, Disz T, Edwards R, Fonstein M, Frank ED, Gerdes S, Glass EM, Goesmann A, Hanson A, Iwata-Reuyl D, Jensen R, Jamshidi N, Krause L, Kubal M, Larsen N, Linke B, McHardy AC, Meyer F, Neuweger H, Olsen G, Olson R, Osterman A, Portnoy V, Pusch GD, Rodionov DA, Rückert C, Steiner J, Stevens R, Thiele I, Vassieva O, Ye Y, Zagnitko O and V., V. (2005). "The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes." *Nucleic Acids Res* **33**(17): 5691-5702.

Papoutsakis, E. T. (2008). "Engineering solventogenic clostridia." *Curr Opin Biotechnol* **19**(5): 420-429.

Philip, S., Keshavarz, T. and Roy, I. (2007). "Polyhydroxyalkanoates: biodegradable polymers with a range of applications." *Journal of Chemical Technology and Biotechnology* **82**(3): 233-247.

Pinchuk, G. E., Hill, E. A., Geydebrekht, O. V., De Ingeniis, J., Zhang, X., Osterman, A., Scott, J. H., Reed, S. B., Romine, M. F., Konopka, A. E., Beliaev, A. S., Fredrickson, J. K. and Reed, J. L. (2010). "Constraint-based model of Shewanella oneidensis MR-1 metabolism: a tool for data analysis and hypothesis generation." *PLoS Comput Biol* **6**(6): e1000822.

Porta, D., Rippka, R. and Hernandez-Marine, M. (2000). "Unusual ultrastructural features in three strains of Cyanothece (cyanobacteria)." *Arch Microbiol* **173**(2): 154-163.

Price, N. D., Reed, J. L. and Palsson, B. O. (2004). "Genome-scale models of microbial cells: evaluating the consequences of constraints." *Nat Rev Microbiol* **2**(11): 886-897.

Pruitt, K. D., Tatusova, T., Brown, G. R. and Maglott, D. R. (2012). "NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy." *Nucleic Acids Res* **40**(Database issue): D130-135.

Ranganathan, S. and Maranas, C. D. (2010). "Microbial 1-butanol production: Identification of non-native production routes and in silico engineering interventions." *Biotechnol J* **5**(7): 716-725.

Ranganathan, S., Suthers, P. and Maranas, C. D. (2010). "OptForce: An Optimization Procedure for Identifying All Genetic Manipulations Leading to Targeted Overproductions." *PLOS Computational Biology*.

Reed, J. L., Famili, I., Thiele, I. and Palsson, B. O. (2006). "Towards multidimensional genome annotation." *Nat Rev Genet* **7**(2): 130-141.

Rehm, B. H. and Steinbuchel, A. (1999). "Biochemical and genetic analysis of PHA synthases and other proteins required for PHA synthesis." *Int J Biol Macromol* **25**(1-3): 3-19.

Reyes, R., Gamermann, D., Montagud, A., Fuente, D., Triana, J., Urchueguia, J. F. and de Cordoba, P. F. (2012). "Automation on the generation of genome-scale metabolic models." *J Comput Biol* **19**(12): 1295-1306.

Russell, D. W., Smith, M., Williamson, V. M. and Young, E. T. (1983). "Nucleotide sequence of the yeast alcohol dehydrogenase II gene." *J Biol Chem* **258**(4): 2674-2682.

Saha, R., Verseput, A. T., Berla, B. M., Mueller, T. J., Pakrasi, H. B. and Maranas, C. D. (2012). "Reconstruction and Comparison of the Metabolic Potential of Cyanobacteria Cyanothece sp. ATCC 51142 and Synechocystis sp. PCC 6803." *PLoS One* **7**(10): e48285.

Satish Kumar, V., Dasika, M. S. and Maranas, C. D. (2007). "Optimization based automated curation of metabolic reconstructions." *BMC Bioinformatics* **8**: 212.

Schirmer, A., Rude, M. A., Li, X., Popova, E. and del Cardayre, S. B. (2010). "Microbial biosynthesis of alkanes." *Science* **329**(5991): 559-562.

Sillers, R., Chow, A., Tracy, B. and Papoutsakis, E. T. (2008). "Metabolic engineering of the non-sporulating, non-solventogenic Clostridium acetobutylicum strain M5 to produce butanol without acetone demonstrate the robustness of the acid-formation pathways and the importance of the electron balance." *Metab Eng* **10**(6): 321-332.

Steinbuchel, A. and Valentin, H. E. (1995). "Diversity of Bacterial Polyhydroxyalkanoic Acids." *Fems Microbiology Letters* **128**(3): 219-228.

Stockel, J., Jacobs, J. M., Elvitigala, T. R., Liberton, M., Welsh, E. A., Polpitiya, A. D., Gritsenko, M. A., Nicora, C. D., Koppenaal, D. W., Smith, R. D. and Pakrasi, H. B. (2011). "Diurnal rhythms result in significant changes in the cellular protein complement in the cyanobacterium Cyanothece 51142." *PLoS One* **6**(2): e16680.

Sun, J., Haveman, S. A., Bui, O., Fahland, T. R. and Lovley, D. R. (2010). "Constraint-based modeling analysis of the metabolism of two Pelobacter species." *BMC Syst Biol* **4**: 174.
Sun, J., Sayyar, B., Butler, J. E., Pharkya, P., Fahland, T. R., Famili, I., Schilling, C. H., Lovley, D. R. and Mahadevan, R. (2009). "Genome-scale constraint-based modeling of Geobacter metallireducens." *BMC Syst Biol* **3**: 15.
Suthers, P. F., Zomorrodi, A. and Maranas, C. D. (2009). "Genome-scale gene/reaction essentiality and synthetic lethality analysis." *Mol Syst Biol* **5**: 301.
Tamagnini, P., Axelsson, R., Lindberg, P., Oxelfelt, F., Wunschiers, R. and Lindblad, P. (2002). "Hydrogenases and hydrogen metabolism of cyanobacteria." *Microbiol Mol Biol Rev* **66**(1): 1-20, table of contents.
The Uniprot Consortium (2012). "Reorganizing the protein space at the Universal Protein Resource (UniProt)." *Nucleic Acids Research*.
Thiele, I. and Palsson, B. O. (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." *Nat Protoc* **5**(1): 93-121.
Thony, B., Auerbach, G. and Blau, N. (2000). "Tetrahydrobiopterin biosynthesis, regeneration and functions." *Biochem J* **347 Pt 1**: 1-16.
Vidal, R., Lopez-Maury, L., Guerrero, M. G. and Florencio, F. J. (2009). "Characterization of an Alcohol Dehydrogenase from the Cyanobacterium Synechocystis sp Strain PCC 6803 That Responds to Environmental Stress Conditions via the Hik34-Rre1 Two-Component System." *Journal of Bacteriology* **191**(13): 4383-4391.
Vitkin, E. and Shlomi, T. (2012). "MIRAGE: a functional genomics-based approach for metabolic network model reconstruction and its application to cyanobacteria networks." *Genome Biol* **13**(11): R111.
Vu, T., Stolyar, S., Pinchuk, G., Hill, E., Kucek, L. A., Brown, R., Lipton, M., Osterman, A., Fredrickson, J., Konopka, A., Beliaev, A. and JL, R. (2012). "Genomescale modeling of light-driven reductant partitioning and carbon fluxes in diazotrophic unicellular cyanobacterium Cyanothece sp. ATCC 51142." *PLOS Computational Biology* **8**(4).
Welsh, E. A., Liberton, M., Stockel, J., Loh, T., Elvitigala, T., Wang, C., Wollam, A., Fulton, R. S., Clifton, S. W., Jacobs, J. M., Aurora, R., Ghosh, B. K., Sherman, L. A., Smith, R. D., Wilson, R. K. and Pakrasi, H. B. (2008). "The genome of Cyanothece 51142, a unicellular diazotrophic cyanobacterium important in the marine nitrogen cycle." *Proc Natl Acad Sci U S A* **105**(39): 15094-15099.
Wu, B., Zhang, B., Feng, X., Rubens, J. R., Huang, R., Hicks, L. M., Pakrasi, H. B. and Tang, Y. J. (2010). "Alternative isoleucine synthesis pathway in cyanobacterial species." *Microbiology* **156**(Pt 2): 596-602.
Yu, M., Zhang, Y., Tang, I. C. and Yang, S. T. (2011). "Metabolic engineering of Clostridium tyrobutyricum for n-butanol production." *Metab Eng* **13**(4): 373-382.
Zhang, S. Y. and Bryant, D. A. (2011). "The Tricarboxylic Acid Cycle in Cyanobacteria." *Science* **334**(6062): 1551-1553.
Zomorrodi, A. R., Suthers, P. F., Ranganathan, S. and Maranas, C. D. (2012). "Mathematical optimization applications in metabolic networks." *Metab Eng* **14**(6): 672-686.

Table A2.1: Statistics for the five developed models. Genes, reactions, and metabolites for each of the five models are listed, along with reactions that are unique to that reconstruction.

	Strain - Reconstruction				
	7424 - *i*Cyc792	**7425 - *i*Cyn731**	**7822 - *i*Cyj826**	**8801 - *i*Cyp752**	**8802 - *i*Cyh755**
Reactions	1242	1306	1258	1172	1161
Metabolites	1107	1160	1110	994	973
Genes	792	731	826	752	755
Unique Reactions	41	149	40	0	0

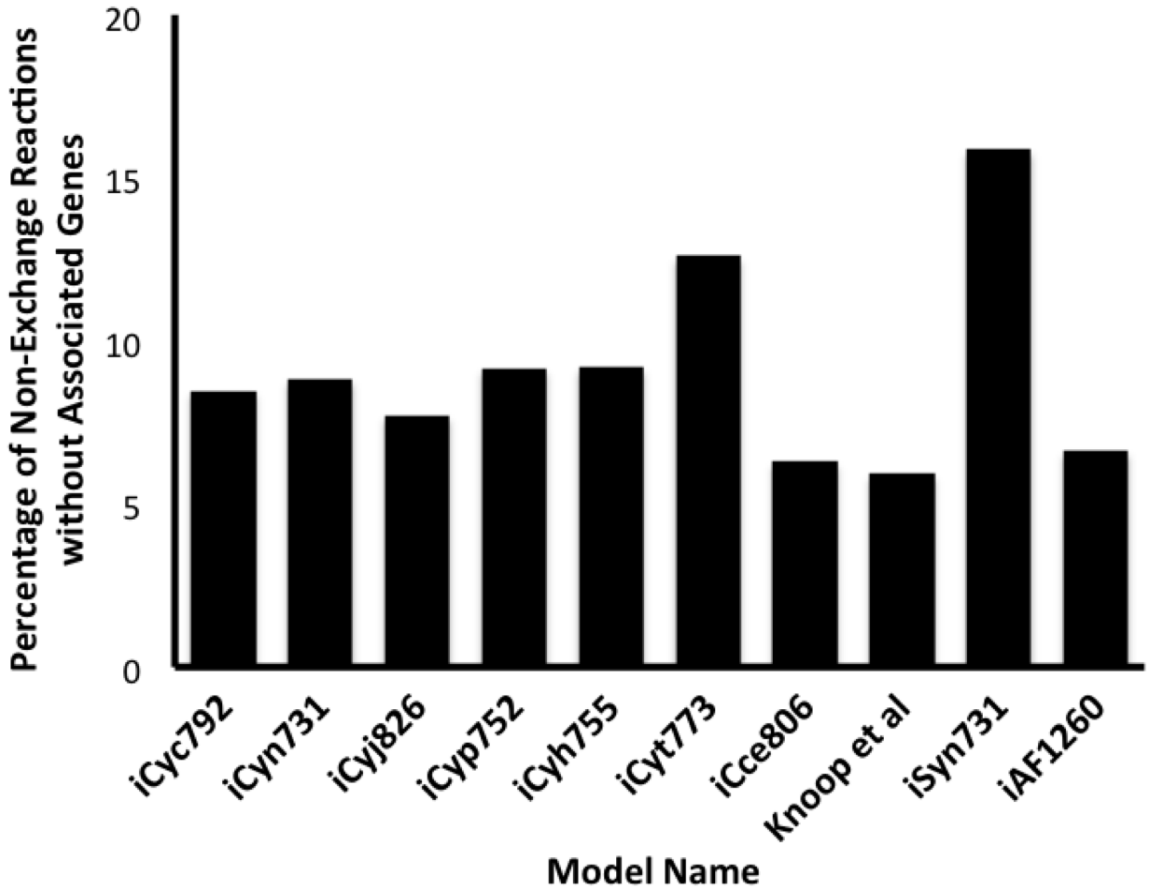
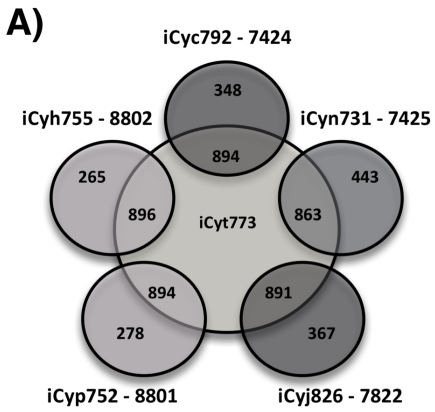


Figure A2.1: Comparisons of five species models with previously curated models. Comparison of the number of non-exchange reactions without associated genes between the five models and five curated models, *i*Cyt773, *i*Syn731 (Saha *et al*. 2012), *i*Cce806 (Vu *et al*. 2012), *i*AF1260 (Feist *et al.* 2007), and the *Synechocystis* PCC 6803 model developed by Knoop *et al*. (2013). The model-organism correlations are *i*Cyt773 and *i*Cce806: *Cyanothece* ATCC 51142, *i*Syn731 and Knoop et al: *Synechocystis* PCC 6803, and *i*AF1260: *Escherichia coli* K-12 MG1655.



B)

	iCyc792 - 7424	iCyn731 - 7425	iCyj826 - 7822	iCyp752 - 8801	iCyh755 - 8802
iCyc792 - 7424	1.000	.871	.938	.902	.894
iCyn731 - 7425		1.000	.875	.863	.860
iCyj826 - 7822			1.000	.912	.906
iCyp752 - 8801				1.000	.991
iCyh755 - 8802					1.000

Figure A2.2: Comparison of reaction similarity to phylogenetic relationships. (A) Venn Diagram comparing the number of reactions each model shares with the *i*Cyt773 model (B) Similarity matrix for the five models. See Methods for description of the similarity calculation done to compare reactions between two models. Both model names and organism numbers are included.

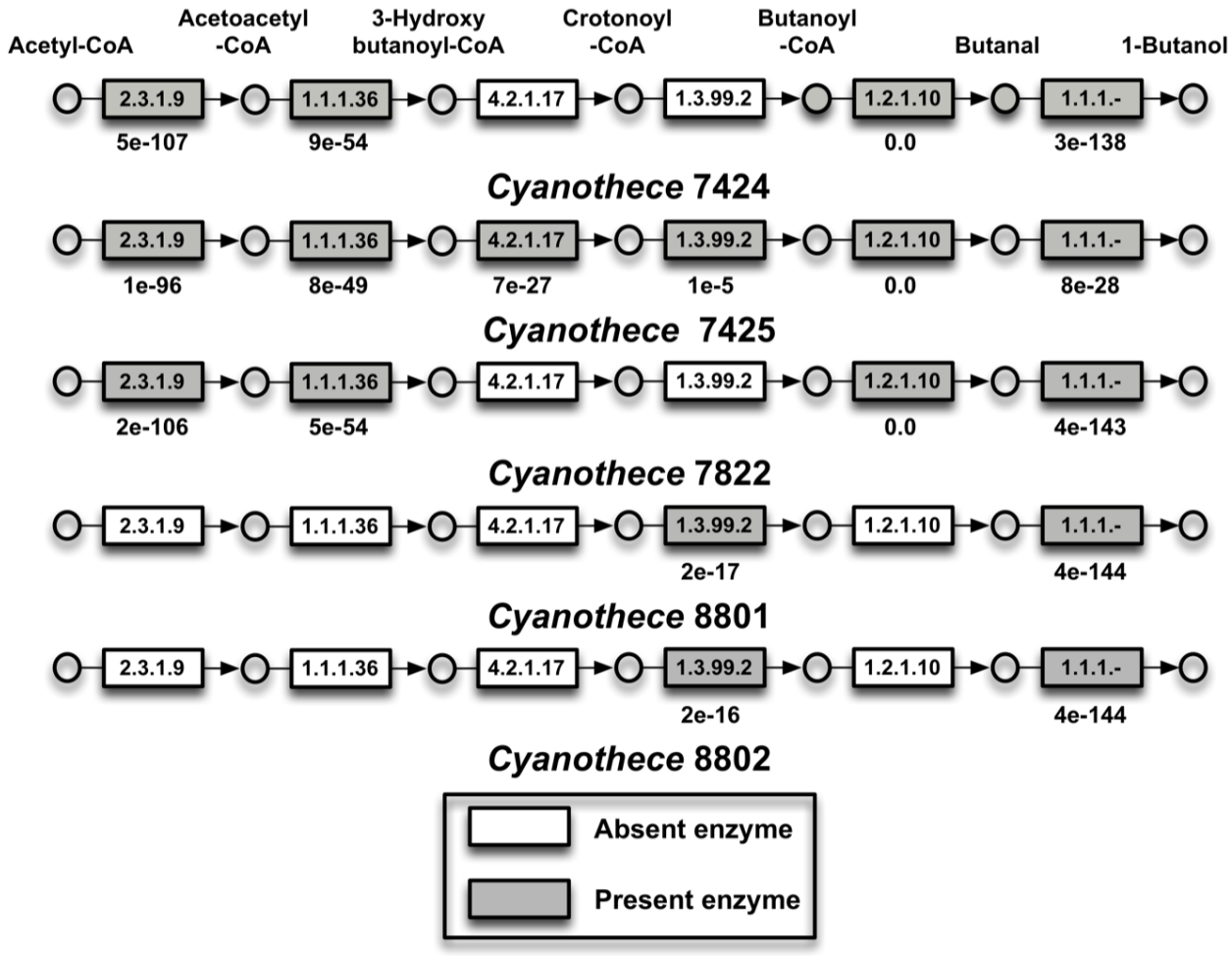


Figure A2.3: Comparison of fermentative butanol pathway enzymes present in each of the 5 species. The enzymes highlighted are present in the organism's reconstruction along with the associated reaction. Listed e-values are for BLAST searches between the genes of the five species and the associated gene in *Clostridium acetobutylicum* and the *adhA* gene in *Synechocystis* 6803. EC-gene relationships: 2.3.1.9: CA_C2873, 1.1.1.36: CA_C2708, 4.2.1.17: CA_C2712, 1.3.99.2: CA_C2711, 1.2.1.10: CA_P0035, 1.1.1.-: slr1192.

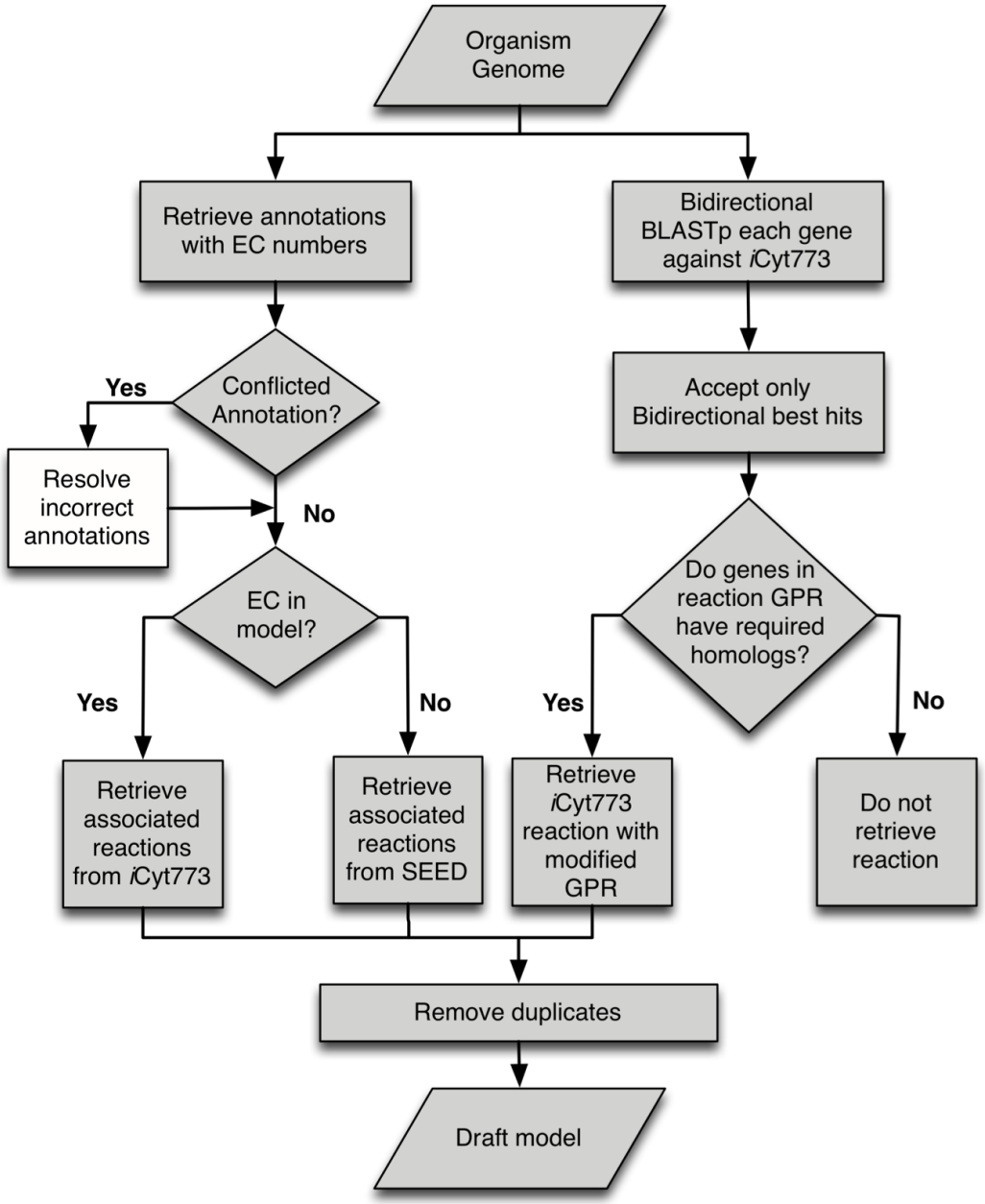


Figure A2.4: Workflow for development of draft models. These models are developed from a sequenced genome and curated genome-scale model of related organism. The right hand side outlines the steps required to evaluate the reactions in *i*Cyt773 for their presence in the other organisms. The steps to retrieve gene annotations and resolve any conflicts are shown on the left hand side. The steps in gray were automated, whereas the manually performed step, the resolution of conflicting annotations, is shown in white.

Appendix Chapter 3

Metabolic Pathway Confirmation and Discovery through ^{13}C-Labeling of Proteinogenic Amino Acids

A3.1. Introduction

Microbes have complex metabolic pathways that can be investigated using biochemistry and functional genomics methods. One important technique to examine cell central metabolism and discover new enzymes is ^{13}C-assisted metabolism analysis (Zamboni *et al.* 2009). This technique is based on isotopic labeling, whereby microbes are fed with ^{13}C-labeled substrates. By tracing the atom transition paths between metabolites in the biochemical network, we can determine functional pathways and estimate the carbon flux through these pathways.

As a complementary method to transcriptomics and proteomics, approaches for isotopomer-assisted metabolic pathway analysis contain three major steps, as shown in figure A3.1 (Tang *et al.* 2009). **First**, we grow cells with ^{13}C labeled substrates. In this step, the composition of the medium and the selection of labeled substrates are two key factors. To avoid measurement noise from non-labeled carbon in nutrient supplements, a minimal medium with a specific carbon source is required. Further, the choice of a labeled substrate is based on how effectively it will elucidate the pathway being analyzed. Because novel enzymes often involve different reaction stereochemistry or new intermediate products, in general, singly labeled carbon substrates are more informative for detection of novel pathways than uniformly labeled ones for detection of novel pathways. By using a substrate with a single labeled carbon, we can easily trace the fate of the labeled carbon from reactant to products, while with multiple labeled carbons substrates may confound the carbon tracing (Tang *et al.* 2007; Tang *et al.* 2009). **Second**, we analyze amino acid labeling patterns using GC-MS. Amino acids are abundant in protein and thus can be obtained from biomass hydrolysis. Amino acids can be derivatized by N-(tert-butyldimethylsilyl)-N- methyltrifluoroacetamide (TBDMS) before GC separation. TBDMS-derivatized amino acids fragment in MS and result in characteristic arrays of fragments. Based

on the mass to charge (m/z) ratio of fragmented and unfragmented amino acids, we can deduce the possible labeling patterns of the central metabolites that are precursors of the amino acids. **Third**, we trace ^{13}C transitions in the proposed pathways and, based on the final labeled data, confirm whether these pathways are active (Tang *et al.* 2009). Measurement of amino acids can provide isotopic labeling information about crucial precursor metabolites in central metabolism (Figure A3.2). These key metabolic nodes reflect the operation of carbon fluxes in associated central pathways.

^{13}C-assisted metabolism analysis via proteinogenic amino acids is widely used for functional characterization of poorly-characterized microbial metabolisms (Zamboni *et al.* 2009). For example, we have used this technique to investigate photosynthetic microbes including cyanobacteria for mixotrophic or heterotrophic carbon utilization (Feng *et al.* 2010). In this paper, we will use *Cyanothece* sp. ATCC 51142 as the model strain to demonstrate the use of labeled carbon substrates for discovering new enzymatic functions. On the other hand, ^{13}C-assisted metabolism analysis can be significantly improved by including measurements of intracellular metabolites besides amino acids. Direct measurement of intracellular metabolites allows the investigation of complex metabolisms with coverage of more pathways. If a rich medium is used for cell culture, measurement of intracellular metabolites, instead of amino acids, also reduces the interference in labeling data that arises from exogenous non-labeled amino acids. In this paper, we also briefly introduce mass spectrometry techniques to measure the labeling pattern of intermediate metabolites in central metabolic pathways.

A3.2. Protocol

A3.2.1. Cell Culture

1.i) Grow cells in minimal medium with trace elements, salts, vitamins, and specifically labeled carbon substrates that are best for pathway investigation. Use either shaking flasks or bioreactors for cell culture.

Note: Organic nutrients, such as yeast extract, may interfere with the measurement of amino acid labeling and thus cannot be present in the culture medium.

1.ii) Monitor cell growth by the optical density of the culture at an optimal wavelength (e.g., OD_{730} for *Cyanothece* 51142) with a UV/Vis spectrophotometer.

1.iii) Cells can first be grown in a non-labeled medium. The middle-log growth phase cells are preferred to be used for inoculation (3% (v/v) by volume inoculation ratio) of the labeled medium. The labeled culture should be sub-cultured (3% (v/v) by volume inoculation ratio) in the same labeled medium to further dilute non-labeled carbon from the initial inoculum.

A3.2.2. Amino Acid Extraction

2.i) Harvest sub-cultured cells (10mL) in the middle-log growth phase by centrifugation (10 min, 8000×g).

2.ii) Resuspend the pellet in 1.5mL of 6M HCl and transfer it to a clear glass, screw-top GC vial. Cap the vials and place them in a 100^{o}C oven for 24 hours to hydrolyze the biomass proteins into amino acids. Hydrolysis of biomass pellets can yield 16 of the 20 common amino acids. Cysteine and tryptophan are degraded, and glutamine and asparagine are converted to glutamate and aspartate, respectively (Dauner *et al.* 2000).

2.iii) Centrifuge the amino acid solution at 20,000×g for 5 min using 2 ml Eppendorf tubes, and transfer the supernatants to new GC vials. This step removes solid particles in the hydrolysis solution.

2.iv) Remove the GC vial lids and dry the samples completely under a stream of air using a Thermo Scientific Reacti-Vap evaporator (note: a freeze dryer can also be used to dry samples). This step can be done overnight.

A3.2.3. Amino Acid Derivatization and GC-MS Conditions

Analysis of amino acids or charged/highly polar metabolites via gas chromatography requires that these metabolites be derivatized, so that the amino acids are volatile and can be separated by gas chromatography (Tang *et al.* 2009).

3.i) Dissolve the dried samples with 150 µL of tetrahydrofuran (THF) and 150 µL of TBDMS reagent.

3.ii) Incubate all samples in an oven or a water bath between 65 and 80^{o}C for 1 hour. Vortex occasionally to make sure the metabolites in the vial are dissolved.

3.iii) Centrifuge the samples at 20,000×g for 10 min, and then transfer the supernatant to new GC vials. The supernatant should be a clear and yellowish solution. Due to saturation of the detectors, GC-MS measurement accuracy can be affected by the high concentration of injected TBDMS derivatized amino acids (these samples often show dark brown color), therefore, we should dilute these samples using THF before GC-MS measurement (Wittmann 2007).

3.iv) Analyze the samples by GC-MS (use a 1:5 or 1:10 split ratio, injection volume = 1 µL, carrier gas helium = 1.2 mL/min). Use the following GC temperature program: Hold at 150^{o}C for 2 minutes, increase at 3^{o}C per min to 280^{o}C, increase at 20^{o}C per min to 300^{o}C, and then hold for

5 minutes. Solvent delay can be set as ~5 min (for a 30 meter GC column). The range of the mass to charge ratio (m/z) in MS can be set between 60 and 500.

A3.2.4. GC-MS Data Analysis

Note: TBDMS-derivatized amino acid measurements can be affected by isotope discrimination in GC separation, since light isotopes move slightly faster than heavy isotopes in a GC column.

4.i) To reduce potential measurement errors, average the mass spectrum of the whole denoted amino acid peak range (Wittmann 2007).

4.ii) The GC and MS spectra of TBDMS derivatized metabolites have been reported before.(Antoniewicz *et al.* 2007) The GC retention time and the unique m/z peaks for each amino acid are illustrated in Figure A3.3. In general, TBDMS-derivatized amino acids are clearly cracked by MS into two charged fragments: fragment $(M-57)^+$, containing the entire amino acid, and fragment $(M-159)^+$, which lacks the α carboxyl group of the amino acid. For leucine and isoleucine, the (M-57)+ peak was overlapped by other mass peaks. We suggest using fragment (M-15)+ to analyze the entire amino acid labeling. Also, the $(f302)^+$ group is often detected in most amino acids, and it contains only the first (α-carboxyl group) and second carbons in an amino acid backbone. However, because this MS peak often has high noise-to-signal ratios, $(f302)^+$ is not recommended for quantitatively analyzing the metabolic fluxes (Antoniewicz *et al.* 2007).

4.iii) Derivatization of amino acids or central metabolites introduces significant amounts of naturally-labeled isotopes, including ^{13}C (1.13%), ^{18}O (0.20%), ^{29}Si (4.70%), and ^{30}Si (3.09%). The measurement noise from natural isotopes in the raw mass isotopomer spectrum can be corrected by using published software (Dauner *et al.* 2000; Wahl *et al.* 2004). The final isotopic

labeling data are reported as mass fractions, i.e., M_0, M_1, M_2, M_3 and M_4 (representing fragments containing zero to four ^{13}C labeled carbons).

A3.2.5. Pathway Analysis Using Labeled Amino Acid Data

An increasing number of genome sequences for non-model microbial species are being published each year. However, functional characterization of these species has lagged far behind the pace of genomic sequencing. ^{13}C-labeling experiments can play important roles in the confirmation and discovery of metabolic pathways in these non-model organisms.

In this protocol, measurement of amino acids can provide isotopic labeling information about eight crucial precursor metabolites: 2-oxo-glutarate, 3-P-glycerate, acetyl-CoA, erythrose-4-P, oxaloacetate, phosphoenolpyruvate, pyruvate, and ribose-5-P. Although the m/z ratio gives just the overall amount of labeling of MS ions, we can partially assess the isotopomer distributions of amino acids by examining the m/z ratios of both unfragmented $(M-57)^+$ and fragmented amino acids $(M-159)^+$ or $(f302)^+$. Furthermore, we can perform several cell cultures with a chemically identical medium but substrates that have different labeling patterns (1^{st} position labeled, 2^{nd} position labeled, etc.). The labeling information about amino acids from these experiments can be integrated to decode the actual carbon transition routes through the central metabolic pathways.

For pathway analysis, the choice of a labeled substrate is important. In general, singly labeled carbon substrates are easier to use in tracing central pathways. Also, such singly labeled substrates are more informative to elucidate unique molecule structures in metabolites than uniformly labeled substrates (Tang *et al.* 2007). For example, the (*Re*)-type citrate synthase has different reaction stereochemistry from normal citrate synthase, and thus causes citrate to have different chemical structures. On the other hand, substrates are different in their suitability to

detect their associated pathways. Glucose is best for detecting the split ratio between the glycolysis and pentose phosphate pathways, while pyruvate or acetate are best for analyzing the TCA cycle and some amino acid pathways. Therefore, it is always good to use different substrate and different tracer experiments to investigate the overall picture of cell metabolism.

By investigating only a few key amino acids produced from well-designed ^{13}C tracer experiments, we may reveal several unique pathways or enzyme activities in the central metabolism without performing sophisticated ^{13}C-metabolic flux analysis. However, the outcome of the labeling experiments should be further confirmed using other biochemistry methods:

5.i) Entner–Doudoroff pathway: [1-^{13}C] glucose can be used as the carbon source. If the pathway is active, serine labeling will be significantly lower than labeling in alanine (Tang *et al.* 2009).

5.ii) Branched TCA cycle: [1-^{13}C] pyruvate can be used as the carbon source. If the TCA cycle is broken, aspartate can be labeled by two carbons, while glutamate is labeled with only one carbon (Feng *et al.* 2009; Tang *et al.* 2010).

5.iii) CO_2 fixation (i.e., Calvin cycle activity in mixotrophic metabolism): Non-labeled CO_2 with labeled carbon substrates can be used as the carbon sources. If the Calvin cycle is functional, serine and histidine labeling will be significantly diluted, compared to other amino acids (Feng *et al.* 2010). Such a method can quantify microbial CO_2 fixation when organic carbon sources are present in the medium.

5.iv) Oxidative pentose phosphate pathway: [1-^{13}C] glucose can be used as the carbon source. If the pathway is active, non-labeled alanine will be > 50% (Feng *et al.* 2009).

5.v) Anaplerotic pathway (e.g., pyruate → oxaloacetate): $^{13}CO_2$ can be used as the carbon source. If the pathway is active, aspartate labeling will be enriched (Feng *et al.* 2010).

5.vi) (*Re*)-citrate synthase: [1-^{13}C] pyruvate can be used as the carbon source. If the enzyme is active, glutamate is labeled in β-carboxyl group (Tang *et al.* 2007; Tang *et al.* 2009).

5.vii) Citramalate pathway: [1-^{13}C] pyruvate, [2-^{13}C] glycerol, or [1-^{13}C] acetate can be used as the carbon source. If the pathway is active, leucine and isoleucine labeling amounts are identical (Tang *et al.* 2007).

5.viii) Serine-isocitrate lyase cycle: [1-^{13}C] pyruvate or [1-^{13}C] lactate can be used as the carbon source. If the pathway is active, the third position carbon in serine will be labeled (Tang *et al.* 2007).

5.ix) Utilization of nutrients (e.g., exogenous amino acids) by organisms: a culture medium with fully labeled carbon substrates and non-labeled amino acids can be used. If the cells selectively transport and utilize these supplemented non-labeled nutrients, we will see significant labeling dilution of these amino acids in the protein. This method can be used to investigate which nutrient supplements the cell prefers (Zhuang *et al.* 2011).

A3.3. Representative Results

Recent bioenergy studies have revived interests in using novel phototrophic microorganisms for bioenergy production and CO_2 capture. In the past years, quite a few ^{13}C-assisted metabolism analyses, including ^{13}C-Metabolic Flux Analyses (^{13}C-MFA), have been applied to investigate central metabolisms in phototrophic bacteria, because biochemical knowledge of the central metabolic pathways is not well-founded in these non-model organisms (Erb *et al.* 2007; Tang *et al.* 2009; Feng *et al.* 2010; McKinlay *et al.* 2010; Tang *et al.* 2010; McKinlay *et al.* 2011). Here, we present an example of the discovery of an alternate isoleucine pathway in *Cyanothece* 51142 (Wu *et al.* 2010). *Cyanothece* 51142 does not contain the enzyme

(EC 4.3.1.19, threonine ammonia-lyase), which catalyzes conversion of threonine to 2-ketobutyrate in the typical isoleucine synthesis pathway. To resolve the isoleucine pathway, we grew *Cyanothece* 51142 (20mL) in ASP2 medium (Reddy *et al.* 1993) with 54 mM glycerol (2-^{13}C, >98%). *Cyanothece* 51142 utilizes 2nd position labeled glycerol as the main carbon source and we observe that threonine and alanine (whose precursor is pyruvate) have one labeled carbon, while isoleucine was labeled with three carbons. Therefore, synthesis in *Cyanothece* 51142 cannot be derived from the threonine route employed by most organisms (Figure A3.4). On the other hand, leucine and isoleucine have identical labeling patterns based on fragment (M-15)$^+$ and fragment (M-159)$^+$. For example, the isotopomer data from [M-15]$^+$ (containing unfragmented amino acids) showed identical labeling for leucine (M_0=0.01, M_1=0.03, M_2=0.21, M_3=0.69) and isoleucine (M_0=0.01, M_1=0.03, M_2=0.24, M_3=0.67). Thus leucine and isoleucine must be synthesized from the same precursors (i.e., pyruvate and acetyl-CoA). This observation is consistent with the labeled carbon transition in the citramalate pathway for isoleucine synthesis. To further confirm this pathway, we searched the Joint Genome Institute database and found the presence of a citramalate synthase *CimA* (cce_0248) in *Cyanothece* 51142*,* and the expression of this gene was also revealed by RT-PCR.

A3.4. Discussion

We demonstrate that ^{13}C-isotope labeling is a useful technique for determining pathways in microorganisms under defined growth conditions. The experimental protocol consists of feeding the cell with a labeled substrate and measuring the resulting isotopic labeling patterns in the synthesized amino acids. The labeling information can be integrated with genomic

information to identify novel pathways, and it can also be used to decipher absolute carbon fluxes via metabolic flux analysis (Zamboni *et al.* 2009). Therefore, this technique can be used in analyzing microorganisms related to biofuel, ecological and medical applications.

This technique has several limitations. ***First***, it is suited only to analysis of carbon metabolism using organic carbon substrates, as it cannot directly resolve metabolism in autotrophic metabolisms when CO_2 is used as the sole carbon source. Bacterial culture using CO_2 as the only carbon source labels all amino acids to the same extent as the input $^{12}CO_2/^{13}CO_2$ mixture (Shastri *et al.* 2007). This makes pathway analysis impossible, as ^{13}C-assisted metabolism analysis has to be inferred from a rearrangement of ^{13}C concentrations in metabolites by different metabolic pathways. ***Second***, this paper presents solely qualitative results discriminating between "active" and "non-active" pathways. Precise quantification of metabolism (^{13}C-MFA) requires a sophisticated modeling approach to decipher metabolic fluxes from isotopomer data. ***Third***, ^{13}C-metabolism analysis is limited by technical challenges in measuring low abundance and unstable intracellular metabolites such as phosphate metabolites. The scope of metabolism analysis can be significantly extended and more pathways can be covered by measuring free metabolites besides amino acids. Broader study of measured metabolites requires both highly-efficient metabolite extraction methods and highly-sensitive analytical platforms. LC-MS, FT-ICR MS, and CE-MS have been used for identifying the labeling patterns of free metabolites in the central metabolism, and provide more insight into these active pathways (Tang *et al.* 2009). ***Fourth***, ^{13}C-assisted pathway analysis is best done in minimal medium, because addition of non-labeled nutrient supplements leads to falsely lower labeling concentrations and complicates quantitative ^{13}C-MFA. Also, if metabolism analysis is

based on proteinogenic amino acids, then cells may utilize exogenous amino acids extensively for protein synthesis and dilute the labeling of proteinogenic amino acids (Tang *et al.* 2009).

A3.5. Supplementary Material

The following supporting information is available online with the originally published version of this article at DOI:10.3791/3583

Video protocol: An 8-minute video describing this protocol is available with the online version of this article.

A3.5. References

Antoniewicz, M. R., Kelleher, J. K. and Stephanopoulos, G. (2007). "Accurate assessment of amino acid mass isotopomer distributions for metabolic flux analysis." *Anal. Chem.* **79**(19): 7554-7559.

Dauner, M. and Sauer, U. (2000). "GC-MS analysis of amino acids rapidly provides rich information for isotopomer balancing." *Biotechnology Progress* **16**: 642-649.

Erb, T. J., Berg, I. A., Brecht, V., Müller, M., Fuchs, G. and Alber, B. E. (2007). "Synthesis of C5-dicarboxylic acids from C2-units involving crotonyl-CoA carboxylase/reductase: The ethylmalonyl-CoA pathway." *PNAS* **104**(25): 10631-10636

Feng, X., Banerjee, A., Berla, B., Page, L., Wu, B., Pakrasi, H. B. and Tang, Y. J. (2010). "Mixotrophic and photoheterotrophic metabolisms in Cyanothece sp. ATCC 51142 under continuous light." *Microbiology* **156**(8): 2566 - 2574.

Feng, X., Mouttaki, H., Lin, L., Huang, R., Wu, B., Hemme, C. L., He, Z., Zhang, B., Hicks, L. M., Xu, J., Zhou, J. and Tang, Y. J. (2009). "Characterization of the Central Metabolic Pathways in Thermoanaerobacter sp. X514 via Isotopomer-Assisted Metabolite Analysis." *Appl. Environ. Microbiol.* **75**(15): 5001-5008.

Feng, X., Tang, K.-H., Blankenship, R. E. and Tang, Y. J. (2010). "Metabolic flux analysis of the mixotrophic metabolisms in the green sulfur bacterium *Chlorobaculum tepidum*." *J Biol Chem* **285**(45): 35104-35112.

McKinlay, J. B. and Harwood, C. S. (2010). "Carbon dioxide fixation as a central redox cofactor recycling mechanism in bacteria." *Proc Natl Acad Sci U S A* **107**(11669-11675).

McKinlay, J. B. and Harwood, C. S. (2011). "Calvin cycle flux, pathway constraints, and substrate oxidation state together determine the H_2 biofuel yield in photoheterotrophic bacteria." *MBio* **2**(e00323-00310).

Reddy, K. J., Haskell, J. B., Sherman, D. M. and Sherman, L. A. (1993). "Unicellular, aerobic nitrogen-fixing cyanobacteria of the genus Cyanothece." *J Bacteriol.* **175**(5): 1284-1292.

Shastri, A. A. and Morgan, J. A. (2007). "A transient isotopic labeling methodology for 13C metabolic flux analysis of photoautotrophic microorganisms." *Phytochemistry* **68**(16-18): 2302-2312.

Tang, K.-H., Feng, X., Tang, Y. J. and Blankenship, R. E. (2009). "Carbohydrate metabolism and carbon fixation in Roseobacter denitrificans OCh114." *PLoS One* **4**(10): e7233.

Tang, K.-H., Feng, X., Zhuang, W.-Q., Alvarez-Cohen, L., Blankenship, R. E. and Tang, Y. J. (2010). "Carbon flow of Heliobacterium modesticaldum is more related to Firmicutes than to the green sulfur bacteria." *J Biol Chem* **285**: 35104-35112.

Tang, Y. J., Chakraborty, R., Martin, H. G., Chu, J., Hazen, T. C. and Keasling, J. D. (2007). "Flux analysis of central metabolic pathways in *Geobacter metallireducens* during reduction of soluble Fe(III)-NTA." *Appl. Environ. Microbiol.* **73**(12): 3859-3864.

Tang, Y. J., Martin, H. G., Deutschbauer, A., Feng, X., Huang, R., Llora, X., Arkin, A. and Keasling, J. D. (2009). "Invariability of central metabolic flux distribution in *Shewanella oneidensis* MR-1 under environmental or genetic perturbations " *Biotechnol Prog.* **25**(5): 1254-1259

Tang, Y. J., Martin, H. G., Myers, S., Rodriguez, S., Baidoo, E. K. and Keasling, J. D. (2009). "Advances in analysis of microbial metabolic fluxes via ^{13}C isotopic labeling." *Mass Spectrom Rev* **28**(2): 362-375.

Tang, Y. J., Meadows, A. L., Kirby, J. and Keasling, J. D. (2007). "Anaerobic central metabolic pathways in *Shewanella oneidensis* MR-1 reinterpreted in the light of isotopic metabolite labeling." *Journal of Bacteriology* **189**(3): 894-901.

Tang, Y. J., Pingitore, F., Mukhopadhyay, A., Phan, R., Hazen, T. C. and Keasling, J. D. (2007). "Pathway confirmation and flux analysis of central metabolic pathways in *Desulfovibrio vulgaris* Hildenborough using GC-MS and FT-ICR mass spectrometry." *Journal of Bacteriology* **189**(3): 940-949.

Tang, Y. J., Yi, S., Zhuang, W., Zinder, S. H., Keasling, J. D. and Alvarez-Cohen, L. (2009). "Investigation of carbon metabolism in "*Dehalococcoides ethenogenes*" strain 195 via isotopic and transcriptomic analysisa." *J. Bacteriol.* **191**(16): 5224-5231.

Wahl, S. A., Dauner, M. and Wiechert, W. (2004). "New tools for mass isotopomer data evaluation in ^{13}C flux analysis: mass isotope correction, data consistency checking, and precursor relationships." *Biotechnology and Bioengineering* **85**(3): 259-268.

Wittmann, C. (2007). "Fluxome analysis using GC-MS." *Microbial Cell Factories* **6**: 6.

Wu, B., Zhang, B., Feng, X., Rubens, J. R., Huang, R., Hicks, L. M., Pakrasi, H. B. and Tang, Y. J. (2010). "Alternative isoleucine synthesis pathway in cyanobacterial species." *Microbiology* **156**: 596-602.

Zamboni, N., Fendt, S. M., Ruhl, M. and Sauer, U. (2009). "C-13-based metabolic flux analysis." *Nature Protocols* **4**(6): 878-892.

Zamboni, N. and Sauer, U. (2009). "Novel biological insights through metabolomics and 13C-flux analysis." *Curr Opin Microbiol.* **12**: 553-558.

Zhuang, W. Q., San, Y., Feng, X., Zinder, S. H., Tang, Y. J. and Alvarez-Cohen, L. (2011). "Selective utilization of exogenous amino acids by Dehalococcoides ethenogenes strain 195 and the enhancement resulted to dechloronation activity. ." *Appl Environ Microbiol*(Accepted).

Table A3.1: Specific reagents and equipment

Name of the reagent	Company	Catalogue number	Comments
TBDMS	Sigma-Aldrich	19915	-
THF	Sigma-Aldrich	34865	-
Labeled carbon substrate	Cambridge Isotope Laboratories	Depend on the experimental requirement	Website: http://www.isotope.com
Gas chromatograph	Agilent Technologies	Hewlett-Packard, model 7890A	-
GC Columns	J&W Scientific, Folsom, CA	DB5 (30m)	-
Mass spectrometer	Agilent Technologies	5975C	-
Reacti-Vap Evaporator	Thermo Scientific	TS-18825	For drying amino acid samples

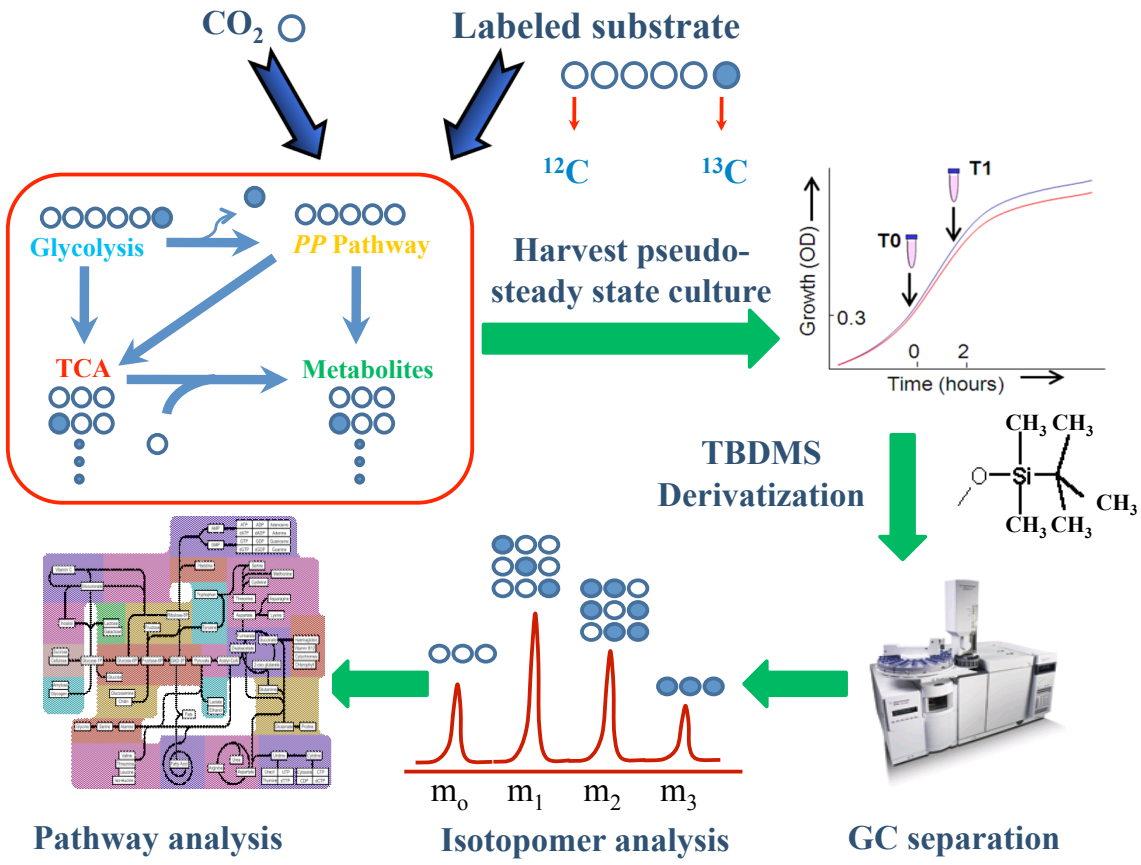


Figure A3.1. Steps for ^{13}C-assisted pathway analysis.

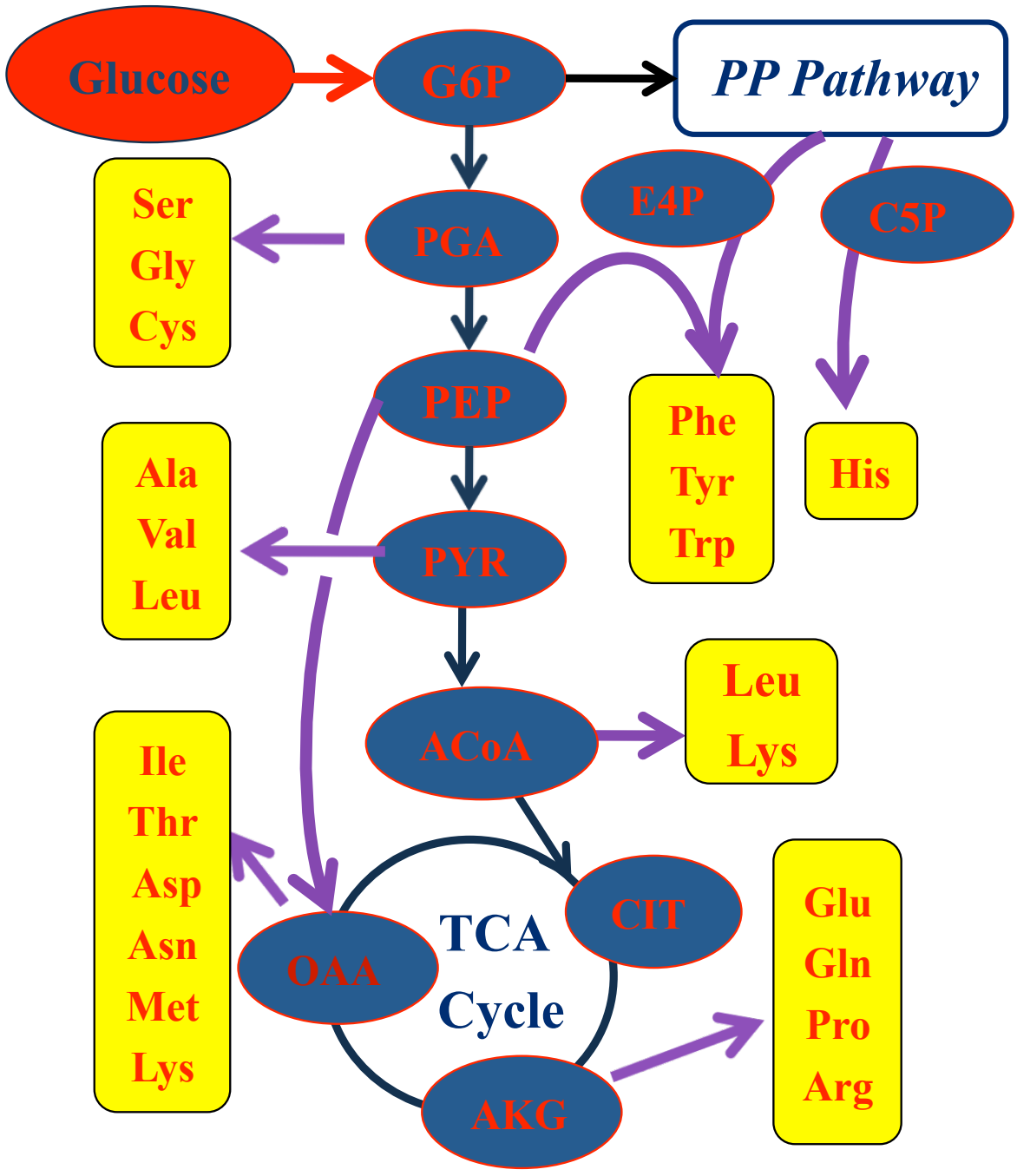


Figure A3.2. Key amino acids used for acquiring the labeling pattern of their central metabolic precursors. (ACoA, acetyl-CoA; AKG, α-Ketoglutarate; C5P, ribose 5-phosphate; CIT, citrate; E4P, erythrose 4-phosphate; G6P, glucose 6-phosphate; OAA, oxaloacetate; PEP, phosphoenolpyruvate; PGA, 3-phosphoglycerate; PYR, pyruvate.).

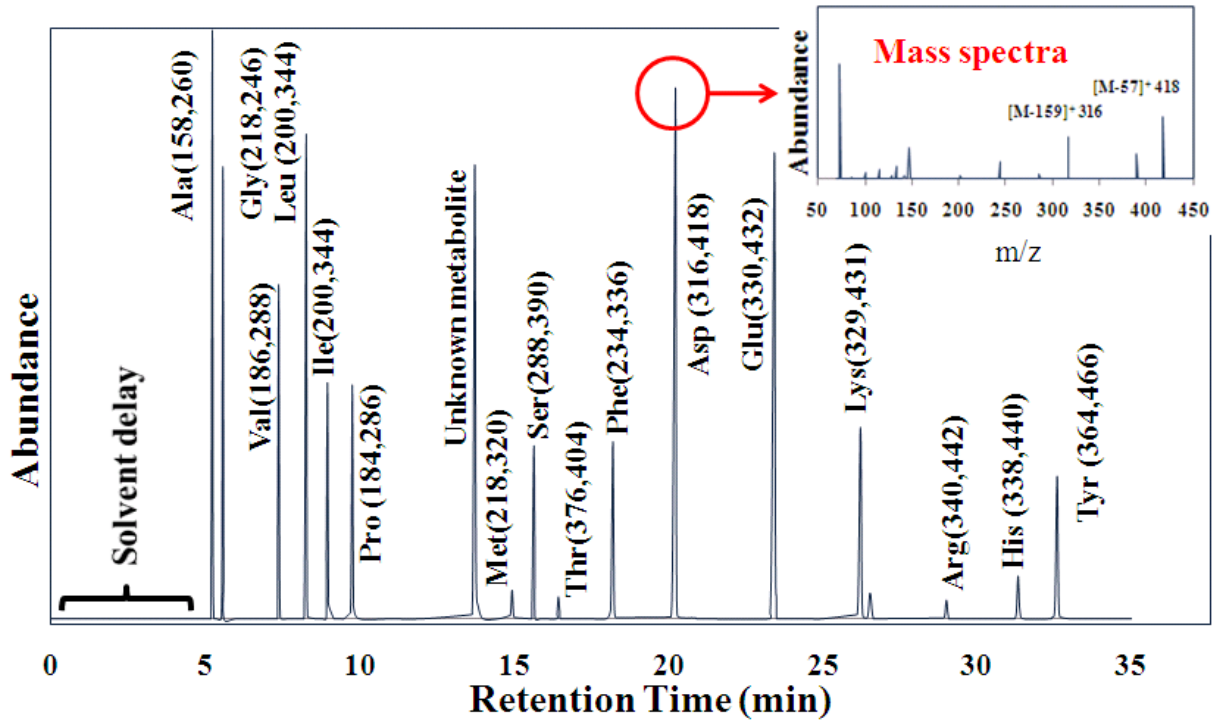


Figure A3.3. Gas chromatography peaks for 17 amino acids (arginine fragmentation cannot be deciphered).

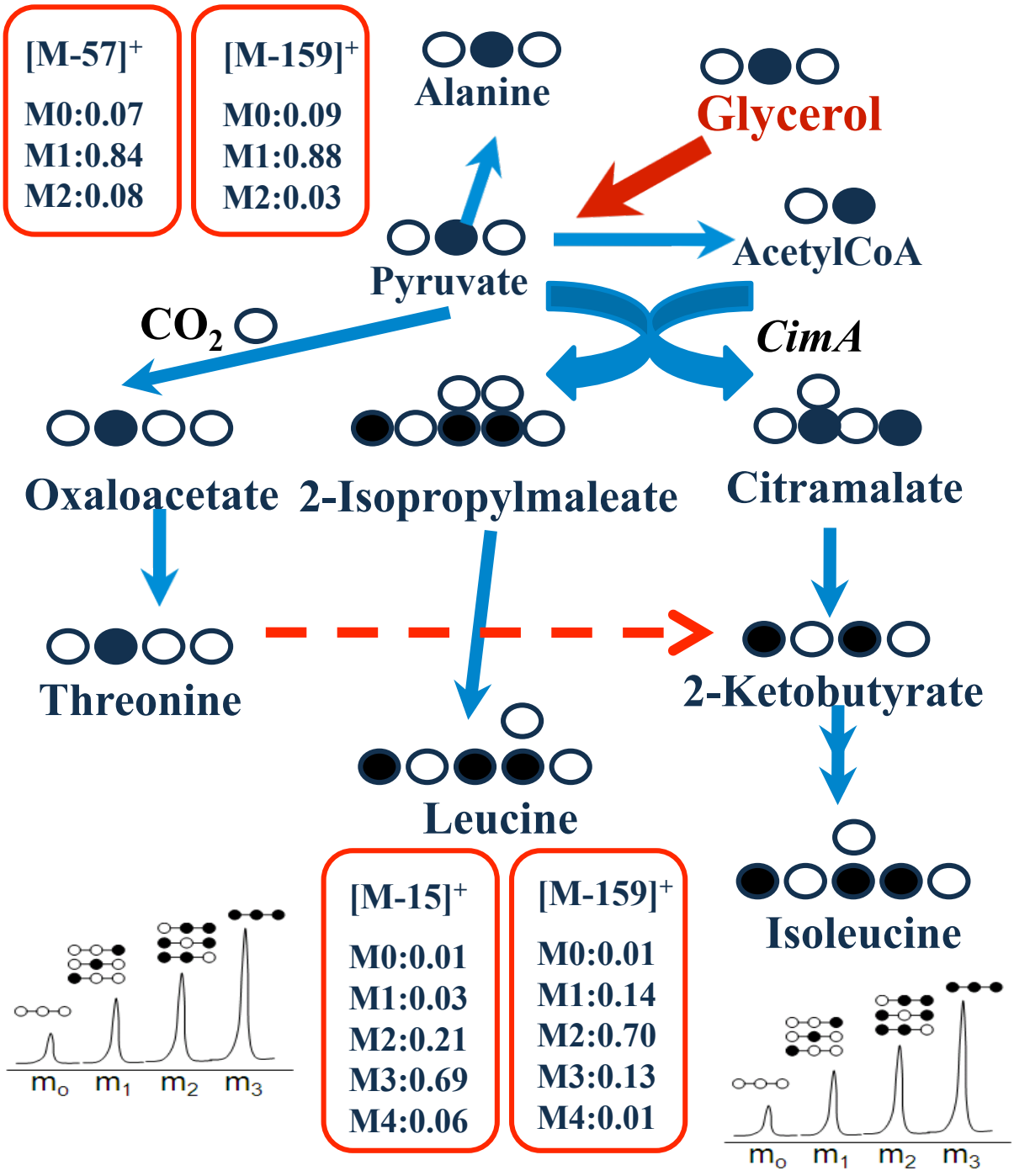


Figure A3.4. Labeling transitions in isoleucine pathways in *Cyanothece* 51142 (Wu *et al.* 2010).

Appendix Chapter 4:

Mixotrophic and Photoheterotrophic Metabolism in *Cyanothece* sp. ATCC 51142 Under Continuous Light

A4.1. Chapter Summary

The unicellular diazotrophic cyanobacterium, *Cyanothece* sp. ATCC 51142 (*Cyanothece* 51142) is able to grow aerobically under nitrogen-fixing conditions with alternating light-dark cycles or continuous illumination. This study investigated the effects of carbon and nitrogen sources on *Cyanothece* 51142 metabolism via ^{13}C-assisted metabolite analysis and biochemical measurements. Under continuous light (50 μmol photons/m^2/s) and nitrogen-fixing conditions, we found that glycerol addition promoted aerobic biomass growth (by twofold) and nitrogenase-dependent hydrogen production [up to 25 μmol H_2 (mg chlorophyll)$^{-1}$ h^{-1}], but strongly reduced phototrophic CO_2 utilization. Under nitrogen-sufficient conditions, *Cyanothece* 51142 was able to metabolize glycerol photoheterotrophically, and the activity of light dependent reactions (e.g. oxygen evolution) was not significantly reduced. In contrast, *Synechocystis* sp. PCC 6803 showed apparent mixotrophic metabolism under similar growth conditions. Isotopomer analysis also detected that *Cyanothece* 51142 was able to fix CO_2 via anaplerotic pathways, and to take up glucose and pyruvate for mixotrophic biomass synthesis.

A4.2. Introduction

Rising concerns about global warming due to the greenhouse effect have renewed research focused on biological capture of CO_2. Cyanobacteria have versatile metabolic capabilities, which allow them to grow under autotrophic, heterotrophic, and mixotrophic conditions (Bottomley & Baalen, 1978; Eiler, 2006; Yang *et al.*, 2002). More importantly, some cyanobacteria can capture solar energy to fix nitrogen and generate H_2, thereby serving as a source of biofertilizer and biofuel, while simultaneously consuming atmospheric CO_2 (Bernat *et*

al., 2009; Dutta *et al.*, 2005; Fay, 1992; Madamwar *et al.*, 2000; Tamagnini *et al.*, 2007; Tuli *et al.*, 1996). *Cyanothece,* sp. ATCC 51142 (*Cyanothece* 51142), a unicellular diazotrophic cyanobacterium, is able to grow aerobically under nitrogen-fixing conditions and has been recognized as contributing to the marine nitrogen cycle (Zehr et al., 2001). The recent sequencing of the *Cyanothece* 51142 genome and its transcriptional analysis have uncovered the diurnally oscillatory metabolism of the bacterium in alternating light-dark cycles (photosynthesis during the day and nitrogen fixation at night) (Stöckel *et al.*, 2008; Toepel *et al.*, 2008; Welsh *et al.*, 2008). In general, cyanobacteria use spatial or temporal separation of oxygen-sensitive nitrogen-fixation and oxygen-evolving photosynthesis as a strategy for diazotrophic growth (Benemann & Weare, 1974; Fay, 1992). Interestingly, *Cyanothece* 51142 demonstrates simultaneous N_2 fixation and O_2 evolution under continuous-light conditions, though it appears to be unicellular (Colon-Lopez *et al.*, 1997; Huang & Chow., 1986). For example, a recent study on transcriptional and translational regulation of continuously-illuminated *Cyanothece* has revealed strong synthesis capability for nitrogenase and circadian expression of 10% of its genes (Toepel *et al.*, 2008). Furthermore, *Cyanothece* strains usually utilize exogenous carbon substrates for mixotrophic growth under light conditions and for heterotrophic growth under dark conditions (Reddy *et al.,* 1993). Carbon substrates are key factors controlling the efficiency of cyanobacterial aerobic growth and hydrogen production (Berman-Frank *et al.*, 2003; Reddy *et al.*, 1993; Tamagnini *et al.*, 2007). Genome analysis studies have revealed that *Cyanothece* 51142 has a unique gene cluster on its linear chromosome that contains key genes involved in glucose and pyruvate metabolism (Welsh *et al.*, 2008). However, the ability of this strain to metabolize glucose or pyruvate remains unknown.

To quantitatively examine the effect of carbon and nitrogen sources on *Cyanothece* central metabolism, this study investigated the effects of three carbon sources (glucose, glycerol, and pyruvate as representatives of sugar, lipid derivatives, and organic acids from central metabolic pathways, respectively) on *Cyanothece* 51142 growth and metabolism. Two nitrogen sources other than N_2, ammonia and nitrate, were also examined. Precise readouts on metabolic state and activity were based on ^{13}C-assisted metabolite analysis integrated with biochemical assays and the gene expression patterns obtained by reverse transcription PCR (RT-PCR) (Fong *et al.*, 2006; Pingitore *et al.*, 2007; Tang *et al.*, 2007c; Tang *et al.*, 2009; Wu *et al.*, 2010). Our work demonstrates that ^{13}C-assisted metabolite analysis can be used as a high throughput tool to study cyanobacterial metabolisms. Superior to the traditional ^{14}C method (Bottomley & Baalen, 1978), the non-radioactive ^{13}C method can provide rich information about which carbons within a metabolite are labeled, and thus enable an in-depth understanding of carbon utilization and metabolic regulation in *Cyanothece* 51142.

A4.3. Materials and Methods

A4.3.1. Bacterial Strains and Growth Conditions

Cyanothece 51142 was first grown in 150 mL Erlenmeyer flasks fed with ASP2 medium (Reddy *et al.*, 1993) without nitrate. Ambient CO_2 provided the sole carbon source. For experiments examining the effect of nitrogen sources, 18 mM $NaNO_3$ or 17 mM NH_4Cl was added into the medium. Cultures were grown aerobically under continuous light (50 μmol photons$\cdot m^{-1} \cdot s^{-1}$) on a shaker at 150 r. p. m. and 30°C. Cells at late-mid exponential phase were sub-cultured into different culture media with various nitrogen and carbon sources. Isotopically-

labeled carbon substrates (Cambridge Isotope Laboratories, Andover, MA) were used for mixotrophic growth, including 54 mM glycerol (2-^{13}C, >98%), 26 mM glucose (U-^{13}C, >98%) and 11 mM sodium pyruvate (3-^{13}C, >98%). For tracer experiments, a 3% inoculum from unlabeled stock culture was used to inoculate a 50 mL medium containing labeled carbon sources. At the mid-exponential phase of growth, a 3% inoculum from the first isotopic labeled culture was used to inoculate 50 mL sub-cultures with the same medium to remove the effect of unlabeled carbon introduced from the initial inoculum. Cell growth was monitored by a UV-Vis spectrometer (GENESYS, Thermo Scientific, USA) at 730 nm. To perform a comparative study, a glucose tolerant *Synechocystis* strain PCC 6803 (a model cyanobacterium for studying fundamental processes of photosynthetic metabolism) was also cultured in BG11 medium (pH=7.6) under the same growth conditions (continuous light and 30 ^{o}C, Stanier *et al.*, 1971). The BG11 medium was supplemented with 6 mM glucose (U-^{13}C, >98%) to support mixotrophic growth. *Synechocystis* PCC 6803 was also sub-cultured in the same labeled medium twice before sampling for ^{13}C-labeled metabolite analysis.

A4.3.2. Metabolite and Photosynthetic Activity Analysis

To analyse metabolites in *Cyanothece* 51142, biomass was harvested at the mid-exponential phase of growth (~90 h) by centrifugation at 7,000 rpm for 15 min at 10˚C. The concentrations of pyruvate, glucose and glycerol were analyzed with enzymatic assay kits (R-Biopharm). To measure hydrogen produced by *Cyanothece* 51142, 20 ml of culture solution was taken from the culture flask after three days and transferred into a 35.2 ml glass vial sealed with a rubber septum and kept under continuous light (50 μmol photons m^{-2} s^{-1}). A modified protocol was used to quantify hydrogen (Rey *et al.*, 2007). Briefly, hydrogen that accumulated in the headspace of the sealed culture vials (for 5 h) was withdrawn with a Hamilton gas-tight syringe

and quantified on an Agilent 6890N Gas Chromatograph with a molseive 5A 60/80 column [inner dimensions 6'×1/8" (1830x3.17 mm)] and Thermal Conductivity Detector. Injection, oven, and detector temperatures were 100°C, 50°C, and 100°C, respectively. Argon was the carrier gas (flow rate of 65 ml min^{-1}). All measurements included three biological replicates.

Photosynthesis activities were determined based on measurements of chlorophyll fluorescence and oxygen evolution. Chlorophyll fluorescence profiles of photosystem II (PSII) of *Cyanothece* 51142 under different nutrient conditions were detected by a FL100 fluorometer (Photon Systems Instruments, Brno, Czech Republic) as described previously (Roose & Pakrasi, 2004). All samples taken for measurement were diluted to OD_{730} ~0.2 using cell-free ASP2 medium. The samples were first adapted for 3 min in total darkness. During the measurement (performed at room temperature), the fluorometer emitted saturating light pulses to determine the fluorescence yield of the samples. The photosynthesis activity was derived by the maximum quantum yield (F_v/F_m) according to the formula $F_v/F_m = (F_m - F_0)/F_m$, where F_0 is initial fluorescence and F_m is maximum fluorescence at the beginning of measurement (Krause & Weis, 1991).

Oxygen evolution rates of *Cyanothece* 51142 grown under different nutrient conditions were measured with a Hansatech oxygen electrode. Assays were performed at 30 °C on whole cells in ASP2 media with a saturating light intensity of 8,250 μmol photons m^{-2} s^{-1} for 2 min at a 2.5 ml reaction volume. For each reaction, the chlorophyll concentration of each sample was diluted and ~6 μg mL^{-1}. The oxygen evolution rates [μmol O_2 (mg chlorophyll)$^{-1}$ h^{-1}] were then measured and normalized based on chlorophyll concentration.

A4.3.3. RNA Extraction and RT-PCR

The bacteria grown under different cultural conditions were harvested at mid-exponential phase according to the corresponding growth curves. The total RNA was extracted by using a PureLinkRNA Mini kit (Invitrogen), following the manufacturer's instruction. cDNA was synthesized from ~2 μg RNA by using a High-Capacity cDNA Reverse Transcription Kit (Invitrogen). The primers for RT-PCR were designed using Primer Premier 5 software (PREMIER Biosoft) and analyzed by OligoAnalyzer 3.0 software (Integrated DNA Technologies). The forward primer 5'-AGCGGTGGAGTATGTGGT-3' and reverse primer 5'-GGCTGGGTTTGATGAGATT-3' were employed to amplify a 16S rRNA gene as a control. The forward primer 5'-CCGACTACACTCCGAAAG-3' and reverse primer 5'-ACGTAACGCCCGTAATGC-3' were used to amplify the Rubisco (*rbcL*) gene and the forward primer 5'-TAATCACGAAACGGGAG-3' and reverse primer 5'-CACCACATCAGCGTATTG-3' to amplify the *prk* gene. The PCRs were conducted with the following cycle conditions: 2 min of activation of the polymerase at 94 °C, followed by 30 cycles consisting of 1 min at 94 °C, 30 s at 53 °C and 2 min at 72 °C; finally, a 10 min extension was performed at 72 °C. The final PCR product was observed directly on 2% agarose gels after electrophoresis.

A4.3.4. Isotopic Analysis

The preparation and isotopic analysis of proteogenic amino acids were performed as previously described (Tang *et al.*, 2007a,b). In brief, exponentially growing biomass from ~20 ml culture was collected by centrifugation (8,000×*g*, 10 min, 4˚C) and hydrolyzed in 6 M HCl at 100°C for 24 h. The amino acid mix was dried and derivatized in tetrahydrofuran (THF) and N-(tert-butyl dimethylsilyl)-N-methyl-trifluoroacetamide (Sigma-Aldrich) at 70°C for 1 h. A gas chromatograph (Hewlett-Packard model 7890A, Agilent Technologies) equipped with a DB5-

MS column (J&W Scientific) and a mass spectrometer (5975C, Agilent Technologies) were used for analyzing amino acid labeling profiles. The ion $[M-57]^+$ from unfragmented amino acid was detected and mass fractions of key amino acids were calculated (Wahl *et al.*, 2004). The substrate utilization ratios R (reflecting the degree of mixotrophic metabolism) for an amino acid X was calculated from the labeling patterns of proteogenic amino acids by the following equation:

$$\text{Amino acid X: } \frac{0.98 \times n \times V_{sub} + 0.01 \times V_{CO2}}{m \times V_{sub} + V_{CO2}} = \frac{(\sum_{i=1}^{C} i \times M_i)}{C} \Longrightarrow \quad R = \frac{V_{sub}}{V_{CO2}} \qquad \text{(A4.1)}$$

where the ratio R indicates the utilization of labeled carbon substrate over unlabeled CO_2 for producing an amino acid X (and its precursors). M_i is the GC-MS isotopomer fraction for amino acid X (i.e., M_0 is the unlabeled fraction, M_1 is the singly labeled fraction, M_2 is the doubly labeled fraction, M_3 is the triply labeled fraction, etc), C is the total number of carbon atoms in the amino acid molecule, V_{sub} is the carbon flux from ^{13}C labeled substrate, V_{CO2} is the carbon flux from CO_2, *0.98* is the purity of the labeled carbon substrate; *0.01* is the natural abundance of ^{13}C, m is the total number of carbons in the substrate molecule, and n is the total number of labeled carbons in the substrate molecule. R indicates the amount of labeled carbon that percolated through the central metabolic networks (Figure A4.1).

A4.4. Results

A4.4.1. Cell Growth with Different Carbon and Nitrogen Sources

Figure A4.2 and Supplementary Figure A4.S1 show the effect of carbon and nitrogen substrates on the growth of *Cyanothece* 51142 under continuous light. Biomass growth was significantly enhanced by the addition of glycerol to ASP2 medium. For example, glycerol addition doubled the specific growth rate from 0.28 to 0.63 day^{-1} under N_2-fixing conditions. These results are consistent with an earlier report on two *Cyanothece* strains (Reddy *et al.*, 1993). On the other hand, *Cyanothece* growth was not apparently enhanced by either glucose or pyruvate (Supplementary Figure A4.S1), and a high concentration of pyruvate (64 mM) inhibited *Cyanothece* growth. Compared with nitrogen fixing cultures, the presence of nitrate salts in the growth media increased *Cyanothece* autotrophic growth rates from 0.28 day^{-1} (N_2-fixation condition) to 0.37 day^{-1} (nitrate-sufficient condition). Similarly, the presence of glycerol enhanced growth rate by approximately two-fold (from 0.60 to 1.02 day^{-1}). As expected, high concentrations of ammonium salts (17 mM) fully inhibited growth (data not shown) because of their well-known deleterious effect on the photosystems of cyanobacteria (Drath *et al.*, 2008; Dai *et al.*, 2008).

A4.4.2. Isotopic Analysis of Amino Acids

^{13}C enrichment patterns in key metabolites were used to estimate the relative utilization of labeled carbon substrates (i.e., glucose, pyruvate, and glycerol) and CO_2 for metabolite synthesis under mixotrophic growth. Figure A4.1 shows the central metabolic pathways in *Cyanothece* 51142 (http://www.genome.jp/kegg/). The labeling of five amino acids was analyzed: histidine (precursors: ribose-5-phosphate and 5,10-methyl-THF), synthesized from the Calvin cycle and pentose phosphate pathway; serine (precursor: 3-phosphoglycerate, a product

from the Calvin cycle); alanine (precursor: pyruvate, originated from carbon substrate or CO_2 fixation); and aspartate and glutamate (precursors: oxaloacetate and 2-oxoglutarate, respectively, synthesized from the citric acid cycle). Under nitrate-sufficient conditions, glycerol could be used as the sole carbon source for synthesis of alanine, serine, and histidine (as indicated by R values approaching infinity). This indicates that the cell was undergoing completely heterotrophic metabolism. R values of some of the key amino acids in glucose and pyruvate cultures were positive and thus these two carbon sources were actually utilized for biomass synthesis (Table 1). However, their measured R values were between 0 and 0.3, which indicated that CO_2 was the main carbon source for metabolite synthesis. This result was consistent with the fact that glucose and pyruvate did not apparently improve the biomass growth. Compared with nitrogen-sufficient conditions, nitrogen-fixing conditions further limited glucose and glycerol utilization, as shown by the decreased labeling fractions of three key amino acids (i.e., alanine, serine, and histidine, Table A4.1).

A4.4.3. Nitrogenase-Dependent H_2 Production, Photosynthesis and Calvin Cycle Activity

Hydrogen production was under continuous light with different carbon substrates (N_2 as the sole nitrogen source) was measured in the exponential (day 4) and stationary (day 9) growth phases (Supplementary Figure A4.S2). In the exponential growth phase under nitrogen fixing conditions, hydrogen production rates were as follows: glycerol, 25±6 μmol H_2 (mg chlorophyll)$^{-1}$ h^{-1}; glucose, 13±9 μmol H_2 (mg chlorophyll)$^{-1}$ h^{-1}; pyruvate, 4±2 μmol H_2 (mg chlorophyll)$^{-1}$ h^{-1} ; and under photoautotrophic conditions; 5±1 μmol H_2 (mg chlorophyll)$^{-1}$ h^{-1}. Under all nitrate or ammonium chloride conditions, hydrogen production was not detected, regardless of the carbon substrate.

The measurement of photosynthetic parameters (Fig. A4.3) suggested that, compared with photoautotrophic conditions, addition of an exogenous carbon source (glycerol, glucose, or pyruvate) did not strongly suppress the maximal quantum yield of PSII (F_v/F_m) or the oxygen evolution rate. Nitrate-sufficient conditions enhanced the oxygen evolution rates by two- to threefold compared with nitrogen-fixing conditions, while the changes of quantum yields of PSII were much less significant (10-30%). Gene expression in the carbon fixation pathway was also determined (Fig. A4.4). RT-PCR results indicated that two key enzymes in Calvin Cycle ([Rubisco, (*rbcL*) and phosphoribulokinase (*prk*)] were functional under conditions of growth with glyverol or glucose. The above measurements confirmed that the light-dependent reactions were active under all culture conditions, even though carbon substrates reduced the relative contribution of CO_2 fixation to biomass synthesis.

A4.5. Discussion

A4.5.1. Carbon Substrate Utilization and Regulation

In continuous light, *Cyanothece* 51142 can efficiently utilize glycerol for aerobic growth. Based on the measurement of carbon substrates in the culture medium during the exponential growth phase, the uptake rates for glycerol were 0.22±0.05 g (g dry biomass)$^{-1}$ day^{-1} under nitrogen fixing and 0.35±0.06 g (g dry biomass)$^{-1}$ day^{-1} under nitrate-sufficient conditions. Glycerol promoted *Cyanothece* 51142 growth because it provided carbon and energy sources. Under nitrate-sufficient conditions, the high values of R for the serine, alanine and histidine labeling data indicated that 3-phosphoglycerate, pyruvate and ribose-5-phosphate nodes in the central metabolic pathways (Figure A4.1) originating completely from glycerol, while the

contribution of CO_2 photofixation to these metabolite nodes was negligible. As a comparison, the glucose-tolerant strain of *Synechocystis* sp. 6803 was cultured with fully labeled glucose under continuous light and nitrogen-sufficient conditions (Supplementary Figure A4.S3). The measured R values (Table 1) for serine (0.87), alanine (0.92) and histidine (1.73) indicated that *Synechocystis* 6803 had a typical mixotrophic growth. In general, cyanobacterial heterotrophic growth has been reported only under three conditions: complete darkness, dim light and pulses of light (Anderson & McIntosh, 1991; Van Baalen *et al.*, 1971). When the light is sufficient for photoautotrophy, *Cyanothece* photoheterotrophic growth is only achieved by addition of PSII inhibitors (Reddy *et al.*, 1993). This study shows that rapidly growing *Cyanothece* 51142 cells can shift their metabolic strategies from mixotrophic or autotrophic growth to photoheterotrophic growth, possibly because maximal utilization of an energy-rich carbon substrate (glycerol) can reduce energy costs related to CO_2 fixation (fixation of one CO_2 consumes two ATP and one NADPH) and building block synthesis, so that maximal biomass growth can be achieved.

On the other hand, glucose was not apparently consumed by *Cyanothece* 51142, as the consumed concentrations were below 1 mM in all experiments. In the [U-^{13}C] glucose experiments (Table 1), all five amino acids contained labeled carbons, which indicated that the labeled glucose had percolated through all the entire central metabolic pathways, thereby confirming the ability of *Cyanothece* 51142 to metabolize glucose. The R values of all key amino acids were below 0.05 for both nitrogen-fixation and nitrate-sufficient conditions, suggesting that a large fraction of the carbon in the biomass had originated from CO_2 fixation. In contrast, glucose is the most favorable carbon source for *Synechocystis* species (Yang *et al.*, 2002), and the R values (Table 1) from key amino acids were around ~0.4-1.7. While both *Synechocystis* 6803 and *Cyanothece* 51142 have completely annotated central pathways for

glucose metabolism, *Synechocystis* 6803 contains a glucose transporter (gene code Sll0771) that shares a sequence relationship with mammalian glucose transporters (Bottomley & Baalen, 1978; Flores & Schmetterer, 1986; Schmetterer, 1990). So far, the presence of a glucose transporter in *Cyanothece* 51142 has not been rigorously verified. From the genome database (DOE Joint Genome Institute, www.jgi.doe.gov/), a gene (cce_3842) has been identified as a glucose transport protein that shared weak (25%) amino acid identity with the Sll0771 protein of *Synechocystis* PCC 6803. Based on the glucose-dependent growth data, we conclude that the enzymes involved in glucose transport or utilization in *Cyanothece* 51142 may not be as efficient as those of *Synechocystis* PCC 6803.

Analysis of labeled pyruvate-grown *Cyanothece* cells showed that serine (precursor 3-phosphoglycerate) and histidine (precursor ribose-5-phosphate) were completely unlabeled (R=0). Such a labeling profile suggests that CO_2 was used as the sole carbon source for synthesis of metabolites in glycolysis and the pentose phosphate pathway (i.e., there was no gluconeogenesis activity). Pyruvate was used only to synthesize alanine (R=0.3~0.6) and metabolites in the tricarboxylic acid cycle: (pyruvate→oxaloacetate→Asp) (pyruvate→acetylCoA→citrate→2-oxoglutarate→Glutamate), as reflected by the labeled carbon present in glutamate and aspartic acid. Interestingly, the R values for alanine (0.60) and glutamate (1.25) were higher under nitrogen-fixing conditions than under nitrate-sufficient conditions, indicating that relatively more labeled pyruvate was used for glutamate synthesis under these conditions. The nitrogen fixation was via nitrogenase: $N_2 + 6\ H^+ + 6\ e^- \rightarrow 2\ NH_3$, and the nitrogenase-generated ammonium was assimilated into amino acids through the glutamine synthetase/glutamate synthase pathway (Postgate, 1998). Utilization of supplemented pyruvate for glutamate synthesis could facilitate the nitrogen fixation process.

The enzyme RuBisCO is known to be the rate-limiting factor in the Calvin Cycle for capturing CO_2 to synthesize three-carbon sugars (glycerate 3-phosphate) (Atsumi *et al.*, 2009). We examined RuBisCO (*rbcL*) and phosphoribulokinase (*prk*) gene expression to reveal the metabolic regulation in the Calvin cycle at transcription level. Under photoautotrophic, mixotrophic, and heterotrophic growth conditions, expression of the two genes was clearly observed. Although Calvin cycle genes were expressed, *Cyanothece* 51142 still grew heterotrophically in the presence of glycerol and nitrate, based on the isotopomer data (no apparent incorporation of CO_2 from the Calvin cycle). These inconsistencies indicate that ^{13}C-assisted metabolite analysis provides a direct readout of actual metabolic status, while gene expression results alone cannot be relied upon, as there are many points of possible post-transcriptional regulation.

Furthermore, *Cyanothece* 51142 can fix CO_2 via anaplerotic pathways (i.e., C4 carbon fixation) (Slack & Hatch, 1967). In the presence of glycerol and under nitrate-sufficient conditions (Table 1), R ratios for aspartate synthesis was 1.53, much smaller than the R ratios ($R=\infty$) for Ala, Ser, and His. This indicates that, even though phototrophic CO_2 fixation was significantly inhibited, CO_2 was utilized for the synthesis of C4 metabolites in the tricarboxylic acid cycle via anaplerotic pathways: (1) PEP+CO_2→oxaloacetate (catalyzed by phosphoenolpyruvate carboxylase or phosphoenolpyruvate carboxykinase) or (2) pyruvate+CO_2→malate (catalyzed by malic oxidoreductase). Such anaplerotic pathways synthesized key TCA cycle metabolites such as oxaloacetate and succinate (precursors for chlorophyll).

Meanwhile, CO_2 was generated by two reactions (i.e., pyruvate→acetylCoA+CO_2; isocitrate→2-oxoglutarate+CO_2), which are essential steps for glutamate synthesis. These

catabolic processes cause the loss of unlabeled carbon when the 2nd position labeled glycerol is used as the main carbon source. Therefore, the coefficients V_{CO2} (CO_2 utilization flux) and R (carbon utilization ratio) were both negative for glutamate synthesis (Equation 1) in glycerol supplemented cultures (under both nitrogen fixation and nitrate-sufficient conditions) (Table 1).

A4.5.2. Photosynthesis Activity

Photosynthesis activity was estimated by the F_v/F_m parameter (maximum quantum efficiency of photosystem II) (Pirintsos *et al.*, 2009). When glycerol or glucose was utilized, the maximum quantum yield F_v/F_m (i.e, efficiency of PSII) in *Cyanothece* 51142 was not significantly affected (changes were within ~30%, Figure A4.3A). Although chlorophyll fluorescence estimation is not an accurate method for determination of absolute PSII activity (Schreiber *et al.*, 1995; Ting & Owens, 1992), we have used it in our study as a tool only to confirm active photon capture in the light-harvesting antenna complexes of PSII under both heterotrophic and mixotrophic conditions.

Oxygen evolution was measured as one molecule of the pigment chlorophyll absorbs one photon and uses its energy to generate NADPH, ATP, and O_2 in the light-dependent reactions (Kaftan *et al.*, 1999). The oxygen evolution rates in *Cyanothece* 51142 rose by 2-3 fold under all nitrate-sufficient conditions compared to corresponding nitrogen fixation conditions (Fig. A4.3B). The significantly higher rates of oxygen evolution indicated that the photosynthetic process of water splitting was more active and provided more energy (ATP and NADPH) to support biomass growth.

Finally, precise determination of the photosynthesic activity of *Cyanothece* 51442 is difficult, as its metabolic behavior fluctuates under continuous light due to its circadian rhythm (Colon-Lopez *et al.*, 1997; Toepel *et al.*, 2008). The photoreaction activity data in Fig. A4.3

represent only qualitative (not quantitative) evidence to support the presence of active light-dependent reactions under all culture conditions.

A4.5.3. Nitrogen Utilization and Nitrogenase-Dependent Hydrogen Production

Under anaerobic conditions (using argon gas to flush the culture), hydrogen production rates of *Cyanothece* 51142 were as high as 100 μmol (mg chlorophyll)$^{-1}$ hr^{-1} (Data not shown). Under aerobic conditions, the hydrogen production enzyme (hydrogenase) is completely inactivated by oxygen (Tamagnini *et al.*, 2007). *Cyanothece* 51142 uses nitrogenase for both nitrogen fixation and hydrogen production. Nitrate, ammonium and some amino acids inhibit nitrogenase activity and thus fully prohibit aerobic hydrogen production by cyanobacteria (Rawson, 1985). Furthermore, NH_4^+ is a direct nitrogen source (nitrate is reduced to NH_4^+) that can be incorporated into biomass via glutamine/glutamate synthase (Muro-Pastor *et al.*, 2005). *Cyanothece* 51142, however, only grows at low concentrations of NH_4^+ (below 1 mM) because of an inhibition effect (Galmozzi *et al.*, 2007; Rawson, 1985). Nitrogen fixation is an energy demanding process: $N_2+8H^++\mathbf{8e^-}+\mathbf{16ATP}\rightarrow 2NH_3+H_2+16ADP+16P_i$. The addition of glycerol reduces CO_2 fixation via the Calvin Cycle, so more energy (ATP and NADH) can be directed to nitrogen fixation, thus promoting hydrogen production by 4-5 fold (Dutta *et al.*, 2005; Madamwar *et al.*, 2000). Glucose and pyruvate cannot significantly promote hydrogen production because their utilization is very low and their effect on the energy economy limited. Hydrogen production rates dropped for all mixotrophic cultures of *Cyanothece* 51142 after 9 days, suggesting that inhibitory metabolites that reduced nitrogenase activity accumulated during cultivation (Atsumi *et al.*, 2009; Nyström, 2004). Finally, the coexistence of oxygen-evolving photosynthesis and oxygen-sensitive nitrogen fixation (indicated by hydrogen evolution) is an attractive characteristic in some cyanobacteria (Benemann & Weare, 1974; Huang & Chow.,

1986). Unlike filamentous cyanobacterial species, in which nitrogen fixation and oxygenic photosynthesis are spatially segregated (Berman-Frank *et al.*, 2001), *Cyanothece* 51142 is able to maintain activities for N_2 fixation, respiration, and photosynthesis within the same cell under continuous light. This strain not only has a strong ability to scavenge intracellular oxygen and synthesize nitrogenase (Colon-Lopez *et al.*, 1997; Fay, 1992), but also develops a highly circadian mechanism for nitrogen fixation (Elvitigala *et al.*, 2009).

This study improves our understanding of *Cyanothece* 51142 physiology with different carbon and nitrogen sources as well as its potential application for hydrogen production. In general, exogenous carbon substrates may improve cellular growth but have strong negative effects on CO_2 fixation. Continuously illuminated *Cyanothece* 51142 shows simultaneous oxygen evolution and nitrogenase-dependent hydrogen production, while hydrogen production can be significantly enhanced by the addition of glycerol. A comparison of metabolic status under autotrophic, mixotrophic and heterotrophic growth conditions indicated that *Cyanothece* 51142 has an inherent metabolic strategy for maximal biomass production at low energy cost. Finally, this study has further confirmed that ^{13}C-assisted metabolite analysis is a high-throughput method which can provide new and precise information to understand a biological system.

A4.6. Supplementary Materials

The following supporting information is available online with the originally published version of this article (Feng *et al.* 2010) at DOI: 10.1099/mic.0.038232-0.

Supplementary Fig. S1: The growth of *Cyanothece* 51142 in the presence of different carbon and nitrogen substrates under continuous light.

Supplementary Fig. S2: Hydrogen production under mixotrophic conditions

Supplementary Fig. S3: Growth of *Synechocystis* 6803 in the presence of glucose under continuous light

A4.7. References

Anderson, S. L. & McIntosh, L. (1991). Light-activated heterotrophic growth of the cyanobacterium *Synechocystis* sp. strain PCC 6803: a blue-light-requiring process. *Journal of Bacteriology* 173, 2761-2767.

Atsumi, S., Higashide, W. & Liao, J. C. (2009). Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde. *Nature Biotechnology* 27, 1177-U1142.

Benemann, J. R. & Weare, N. M. (1974). Hydrogen Evolution by Nitrogen-Fixing *Anabaena cylindrica* Cultures. *Science* 184, 174-175.

Berman-Frank, I., Lundgren, P., Chen, Y.-B., Küpper, H., Kolber, Z., Bergman, B. & Falkowski, P. (2001). Segregation of Nitrogen Fixation and Oxygenic Photosynthesis in the Marine Cyanobacterium *Trichodesmium*. *Science* 294, 1534-1537.

Berman-Frank, I., Lundgren, P. & Falkowski, P. (2003). Nitrogen fixation and photosynthetic oxygen evolution in cyanobacteria. *Research in Microbiology* 154, 157-164.

Bernat, G., Waschewski, N. & Rogner, M. (2009). Towards efficient hydrogen production: the impact of antenna size and external factors on electron transport dynamics in *Synechocystis* PCC 6803. *Photosynth Res* 99, 205-216.

Bottomley, P. J. & Baalen, C. V. (1978). Characteristics of Heterotrophic Growth in the Blue-Green Alga *Nostoc* sp. Strain Mac. *Journal of General Microbiology* 107, 309-318.

Colon-Lopez, M. S., Sherman, D. M. & Sherman, L. A. (1997). Transcriptional and translational regulation of nitrogenase in light-dark- and continuous-light grown cultures of the unicellular cyanobacterium *Cyanothece* sp. strain ATCC 51142. *Journal of Bacteriology* 179, 4319-4327.

Dai, G., Deblois, C. P., Liu, S., Juneau, P. & Qiu, B. (2008). Differential sensitivity of five cyanobacterial strains to ammonium toxicity and its inhibitory mechanism on the photosynthesis of rice-field cyanobacterium Ge-Xian-Mi (Nostoc). *Aquatic toxicology* 89, 113-121.

Drath, M., Kloft, N., Batschauer, A., Marin, K., Novak, J. & Forchhammer, K. (2008). Ammonia triggers photodamage of photosystem II in the cyanobacterium *Synechocystis* sp strain PCC 6803. *Plant Physiology* 147, 206-215.

Dutta, D., De, D., Chaudhuri, S. & Bhattacharya, S. K. (2005). Hydrogen production by Cyanobacteria. *Microb Cell Fact* 4.

Eiler, A. (2006). Evidence for the Ubiquity of Mixotrophic Bacteria in the Upper Ocean: Implications and Consequences. *Applied and Environmental Microbiology* 72, 7431-7437.

Elvitigala, T., Stöckel, J., Ghosh, B. K. & Pakrasi, H. B. (2009). Effect of continuous light on diurnal rhythms in *Cyanothece* sp. ATCC 51142. *BMC Genomics* 10, 226.

Fay, P. (1992). Oxygen relations of nitrogen fixation in cyanobacteria. *Microbiological Reviews* 56, 340-373.

Feng X, Bandyopadhyay A, Berla B, Page L, Wu B, Pakrasi HB, Tang YJ (2010). Mixotrophic and photoheterotrophic metabolism in *Cyanothece* sp. ATCC 51142 under continuous light. *Microbiology* 156, 2566-2574.

Flores, E. & Schmetterer, G. (1986). Interaction of fructose with the glucose permease of the cyanobacterium *Synechocystis* sp. strain PCC 6803. *Journal of Bacteriology* 66, 693-696.

Fong, S. S., Nanchen, A., Palsson, B. O. & Sauer, U. (2006). Latent pathway activation and increased pathway capacity enable *Escherichia coli* adaptation to loss of key metabolic enzymes. *Journal of Biological Chemistry* 281, 8024-8033.

Galmozzi, C. V., Fernandez-Avila, M. J., Reyes, J. C., Florencio, F. J. & Muro-Pastor, M. I. (2007). The ammonium-inactivated cyanobacterial glutamine synthetase I is reactivated in vivo by a mechanism involving proteolytic removal of its inactivating factors. *Mol Microbiol* 65, 166-179.

Huang, T. C. & Chow., T. J. (1986). New type of N_2-fixing unicellular cyanobacterium (blue-green alga). *FEMS Microbiology Letter* 36, 109-110.

Kaftan, D., Meszaros, T., Whitmarsh, J. & Nedbal, L. (1999). Characterization of Photosystem II Activity and Heterogeneity during the Cell Cycle of the Green Alga *Scenedesmus quadricauda*. *Plant Physiology* 120, 433-441.

Krause, G. H. & Weis, E. (1991). Chlorophyll fluorescence and photosynthesis: the basics. *Ann Rev Plant Physiol Plant Mol Biol* 42, 313-349.

Madamwar, D., Garg, N. & Shah, V. (2000). Cyanobacterial hydrogen production. *World J Microbiol Biotechnol* 16, 757-767.

Muro-Pastor, M. I., Reyes, J. C. & Florencio, F. J. (2005). Ammonium assimilation in cyanobacteria. *Photosynthesis Research* 83, 135-150.

Nyström, T. (2004). Stationary phase physiology. *Annual Review of Microbiology* 58, 161-181.

Pingitore, F., Tang, Y. J., Kruppa, G. H. & Keasling, J. D. (2007). Analysis of amino acid isotopomers using FT-ICR MS. *Analytical Chemistry* 79, 2483-2490.

Pirintsos, S. A., Munzi, S., Loppi, S. & Kotzabasis, K. (2009). Do polyamines alter the sensitivity of lichens to nitrogen stress? *Ecotoxicology and Environmental Safety* 72, 1331-1336.

Postgate, J. (1998).Nitrogen Fixation, 3rd Edition: Cambridge University Press, Cambridge UK.

Rawson, D. M. (1985). The effects of exogenous amino acids on growth and nitrogenase activity in the cyanobacterium *Anabaena cylindrica* PCC 7122. *Journal of General Microbiology* 134, 2549-2544.

Reddy, K. J., Haskell, J. B., Sherman, D. M. & Sherman, L. A. (1993). Unicellular, aerobic nitrogen-fixing cyanobacteria of the genus *Cyanothece*. *J Bacteriol* 175, 1284-1292.

Rey, F. E., Heiniger, E. K. & Harwood, C. S. (2007). Redirection of Metabolism for Biological Hydrogen Production. *Applied and Environmental Microbiology* 73, 1665-1671.

Roose, J. L. & Pakrasi, H. B. (2004). Evidence that D1 processing is required for manganese binding and extrinsic protein assembly into photosystem II. *J Biol Chem* 279, 45417-45422.

Schmetterer, G. R. (1990). Sequence conservation among the glucose transporter from the cyanobacterium *Synechocystis* sp. PCC 6803 and mammalian glucose transporters. *Plant Molecular Biology* 14, 697-706.

Schreiber, U., Endo, T., Mi, H. & Asada, K. (1995). Quenching Analysis of Chlorophyll Fluorescence by the Saturation Pulse Method: Particular Aspects Relating to the Study of Eukaryotic Algae and Cyanobacteria. *Plant and Cell Physiology* 36, 873-882.

Slack, C. R. & Hatch, M. D. (1967). Comparative studies on the activity of carboxylases and other enzymes in relation to the new pathway of photosynthetic carbon dioxide fixation in tropical grasses. *Biochem J* 103, 660-665.

Stanier, R. Y., Kunisawa, R., Mandel, M. & Cohen-Bazire, G. (1971). Purification and properties of unicellular blue-green algae (order Chroococcales). *Bacteriol Rev* 35, 171-205.

Stöckel, J., Welsh, E. A., Liberton, M., Kunnvakkam, R., Aurora, R. & Pakrasi, H. B. (2008). Global transcriptomic analysis of *Cyanothece* 51142 reveals robust diurnal oscillation of central metabolic processes. *Proc Natl Acad Sci* 105, 6456-6461.

Tamagnini, P., Leitao, E., Oliveira, P., Ferreira, D., Pinto, F., Harris, D. J., Heidorn, T. & Lindblad, P. (2007). Cyanobacterial hydrogenases: diversity, regulation and applications. *Fems Microbiology Reviews* 31, 692-720.

Tang, Y. J., Hwang, J. S., Wemmer, D. & Keasling, J. D. (2007a). The *Shewanella oneidensis* MR-1 fluxome under various oxygen conditions. *Applied and Environmental Microbiology* 73, 718-729.

Tang, Y. J., Meadows, A. L., Kirby, J. & Keasling, J. D. (2007b). Anaerobic central metabolic pathways in *Shewanella oneidensis* MR-1 reinterpreted in the light of isotopic metabolite labeling. *Journal of Bacteriology* 189, 894-901.

Tang, Y. J., Pingitore, F., Mukhopadhyay, A., Phan, R., Hazen, T. C. & Keasling, J. D. (2007c). Pathway confirmation and flux analysis of central metabolic pathways in *Desulfovibrio vulgaris* Hildenborough using GC-MS and FT-ICR mass spectrometry. *Journal of Bacteriology* 189, 940-949.

Tang, Y. J., Martin, H. G., Myers, S., Rodriguez, S., Baidoo, E. E. K. & Keasling, J. D. (2009). Advances in metabolic network and flux analysis of microorganisms via 13C isotopic labeling. *Mass Spectrometry Reviews* 28, 362-375.

Ting, C. S. & Owens, T. G. (1992,). Limitations of the pulse-modulated technique for measuring the fluorescence characteristics of algae. *Plant Physiol* 100, 367-373.

Toepel, J., Welsh, E., Summerfield, T. C., Pakrasi, H. B. & Sherman, L. A. (2008). Differential transcriptional analysis of the cyanobacterium *Cyanothece* sp strain ATCC 51142 during light-dark and continuous-light growth. *J Bacteriol* 190, 3904-3913.

Tuli, R., Naithani, S. & Misra, H. S. (1996). Cyanobacterial photosynthesis and the problem of oxygen in nitrogen-fixation: A molecular genetic view. *Journal of Scientific & Industrial Research* 55, 638-657.

Van Baalen, C., Hoare, D. S. & Brandt, E. (1971). Heterotrophic growth of blue-green algae in dim light. *Journal of Bacteriology* 105, 685-689.

Wahl, S. A., Dauner, M. & Wiechert, W. (2004). New tools for mass isotopomer data evaluation in ^{13}C flux analysis: mass isotope correction, data consistency checking, and precursor relationships. *Biotechnology and Bioengineering* 85, 259-268.

Welsh, E. A., Liberton, M., Stöckel, J. & other authors (2008). The genome of *Cyanothece* 51142, a unicellular diazotrophic cyanobacterium important in the marine nitrogen cycle. *Proc Natl Acad Sci USA* 105, 15094-15099.

Wu, B., Zhang, B., Feng, X., Rubens, J. R., Huang, R., Hicks, L. M., Pakrasi, H. B. & Tang, Y. J. (2010). An Alternate Isoleucine Biosynthesis Pathway involves Citramalate Synthase in Cyanothece sp. ATCC 51142. *Microbiology* 156, 596-602.

Yang, C., Hua, Q. & Shimizu, K. (2002). Metabolic flux analysis in *Synechocystis* using isotope distribution from ^{13}C-labeled glucose. *Metab Eng* 4, 202-216.

Table A4.1. Isotopic analysis of the labeling profiles of amino acids in *Cyanothece* 51142 and *Synechocystis* 6803 under different growth conditions (the standard error for GC-MS measurement is below 0.02, technical replicates, n=2).

Amino Acids	$[M-57]^+$	N_2						$NaNO_3$						*PCC* 6803 (NO_3-medium)	
		Glucose	R^1	Pyruvate	*R*	Glycerol	*R*	Glucose	*R*	Pyruvate	*R*	Glycerol	*R*	Glucose	***R***
Ala	M0	0.67		0.41		0.19		0.61		0.51		0.07		0.04	
	M1	0.19	***0.032***	0.55	***0.597***	0.71	***4.2***	0.19	***0.042***	0.48	***0.327***	0.85	+∞	0.05	***0.92***
	M2	0.11		0.03		0.10		0.17		0.01		0.07		0.28	
Ser	M0	0.65		0.98		0.20		0.58		0.97		0.08		0.04	
	M1	0.22	***0.033***	0.02	***0***	0.72	***3.7***	0.22	***0.046***	0.03	***0***	0.81	+∞	0.06	***0.87***
	M2	0.10		0		0.09		0.16		0		0.10		0.28	
Asp	M0	0.58		0.54		0.10		0.59		0.94		0.07		0.04	
	M1	0.24	***0.030***	0.43	***0.195***	0.64	***2.2***	0.20	***0.032***	0.06	***0.005***	0.78	***1.53***	0.05	***0.44***
	M2	0.11		0.04		0.25		0.17		0		0.15		0.19	
	M3	0.06		0		0.01		0.03		0		0		0.47	
Glu^2	M0	0.43		0.15		0.02		0.38		0.47		0.01		0.02	
	M1	0.26	***0.041***	0.44	***1.25***	0.14	***- 1.78***	0.22	***0.051***	0.49	***0.170***	0.15	***- 2.11***	0.02	***0.76***
	M2	0.21		0.37		0.62		0.27		0.04		0.74		0.04	
	M3	0.07		0.04		0.21		0.09		0		0.10		0.07	
	M4	0.02		0		0.01		0.03		0		0		0.53	
His	M0	0.44		0.91		0.05		0.33		0.92		0.01		0.01	
	M1	0.28		0.08		0.28		0.24		0.08		0.21		0.01	
	M2	0.17	***0.032***	0	***0***	0.50	***2.83***	0.22	***0.049***	0	***0***	0.55	+∞	0.02	***1.73***
	M3	0.07		0		0.16		0.12		0		0.20		0.03	
	M4	0.03		0		0		0.07		0		0.03		0.06	
	M5	0		0		0		0.01		0		0		0.22	

[1] - Bold values were the carbon substrate (glycerol, pyruvate, or glucose) utilization ratios (substrate/CO_2 fixation) for amino acid synthesis calculated according to Equation (1).

[2] - The glutamate synthesis pathway involved the loss of two carbons from pyruvate to ketoglutarate. Such a microbial process changed the labeling enrichment, and the negative value indicated the net loss of unlabeled CO_2.

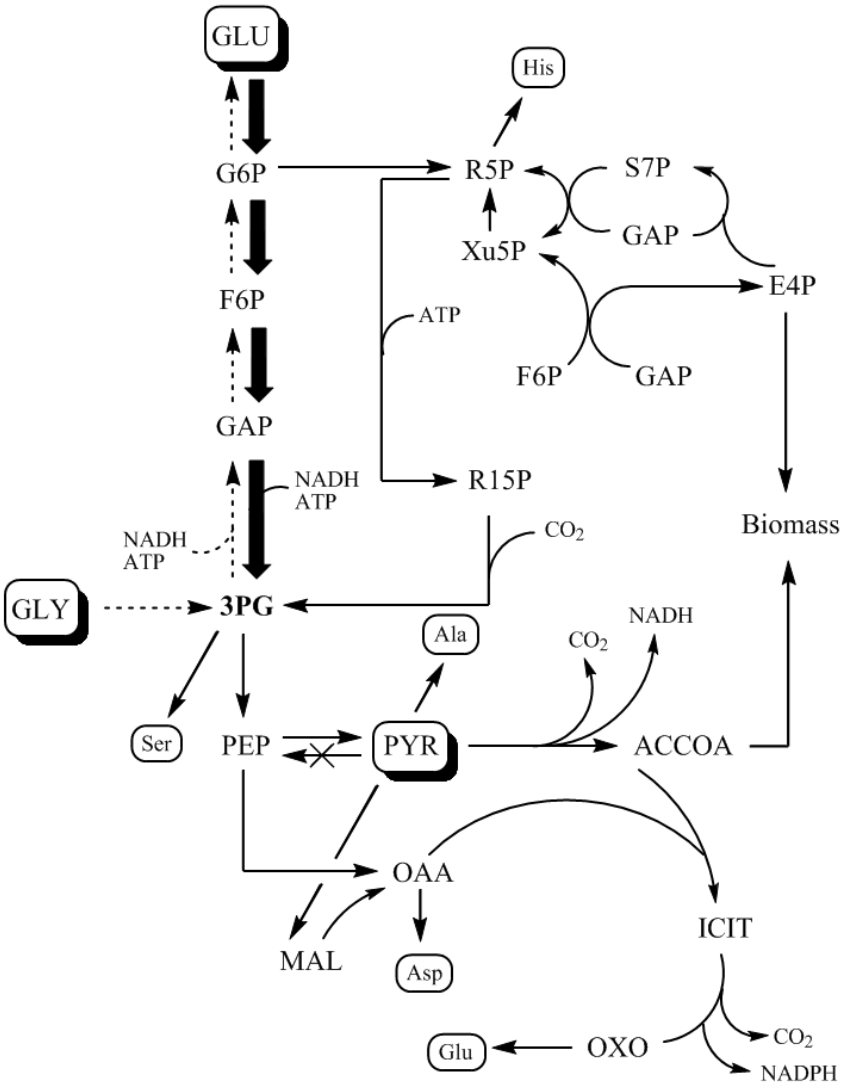


Figure A4.1. Central metabolic pathways of *Cyanothece* 51142 with glucose, glycerol, and pyruvate as carbon substrates. The dashed line shows the metabolic pathway with glycerol as carbon substrate; the bold line indicates glucose; the solid line shows the common pathway for all carbon conditions. Abbreviations: ACCOA, acetyl-coenzyme A; Ala, alanine; E4P, erythrose-4-phosphate; F6P, fructose-6-phosphate; G6P, glucose-6-phosphate; GAP, glyceraldehyde 3-phosphate; 3PG, 3-phosphoglycerate; GLY, glycerol; GLU, glucose; His, histidine; ICIT, citrate/isocitrate; MAL, malate; OAA, oxaloacetate; OXO, 2-oxoglutarate; PEP, phosphoenolpyruvate; PYR, pyruvate; R5P, ribose-5-phosphate (or ribulose-5-phosphate); R15P, ribulose-1,5-bisphosphate; S7P, sedoheptulose-7-phosphate; Ser, serine; Xu5P: xylulose-5-phosphate.

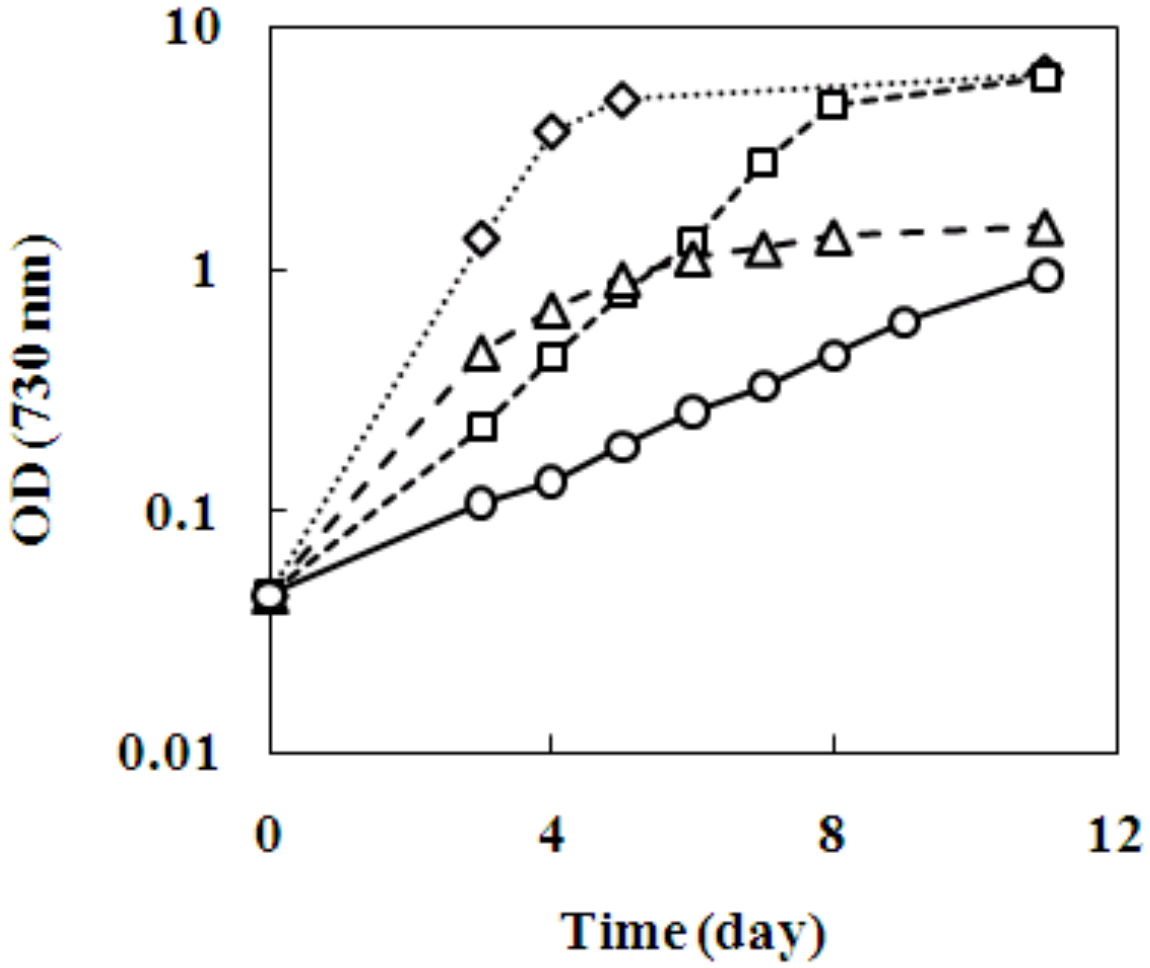


Figure A4.2. *Cyanothece* 51142 growth curves under different nitrogen and carbon sources (biological replicates, n=3). Diamond: Gly+Nitrate; Square: Gly+N_2; Triangle: CO_2+ Nitrate; Circle: CO_2+N_2. The glycerol-growing samples were taken at day four when the remained glycerol in the culture medium was sufficient for biomass growth (>30mM).

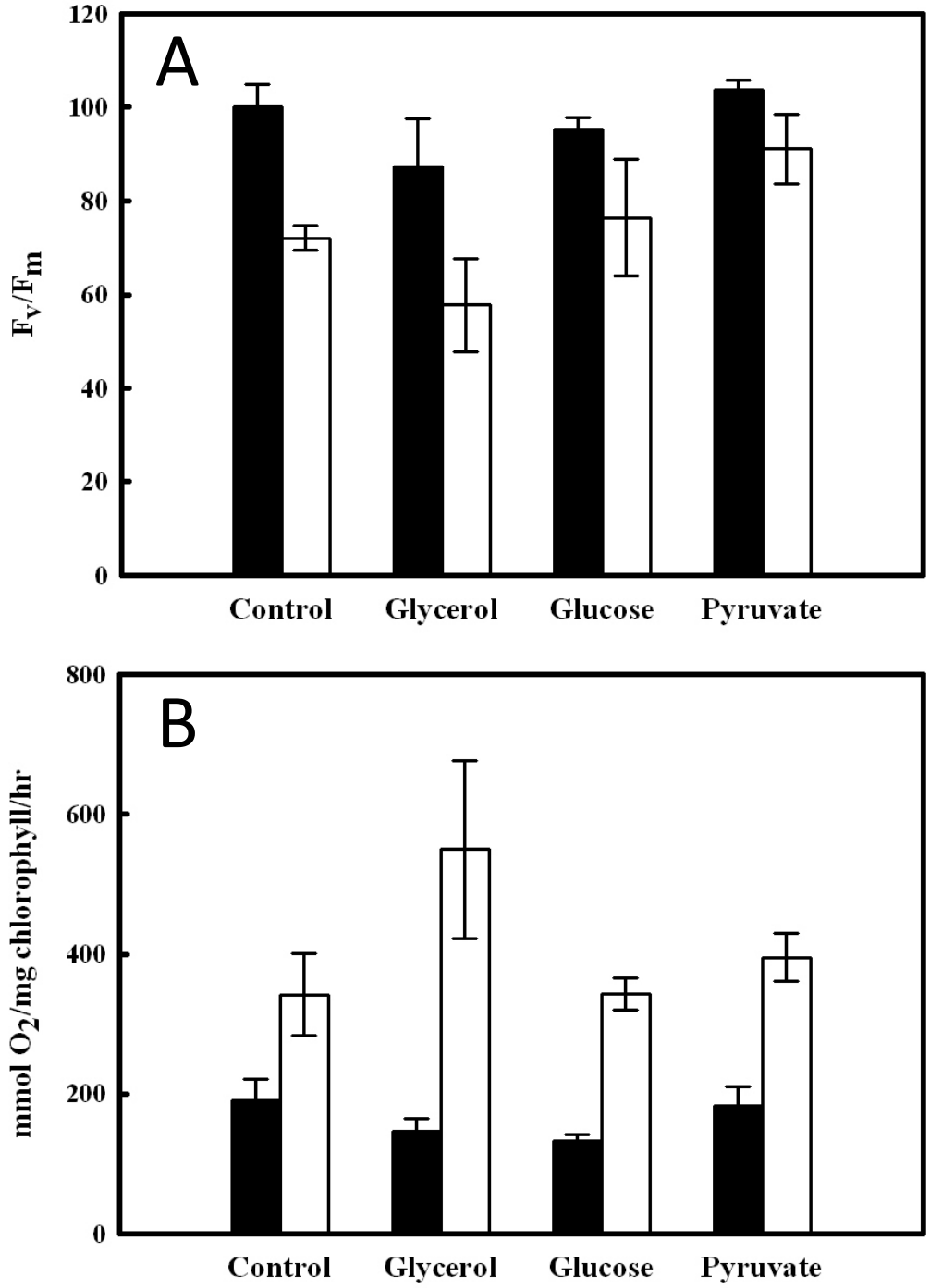


Figure A4.3. Maximum quantum yields (Figure 2A) of PSII and oxygen evolution rates (Figure 2B) in *Cyanothece* 51142 under different growth conditions. All samples were taken at the exponential growth phase based on the growth curve. Black column, N_2 as nitrogen source; white column, $NaNO_3$ as nitrogen source. Data are for 3 biological replicates.

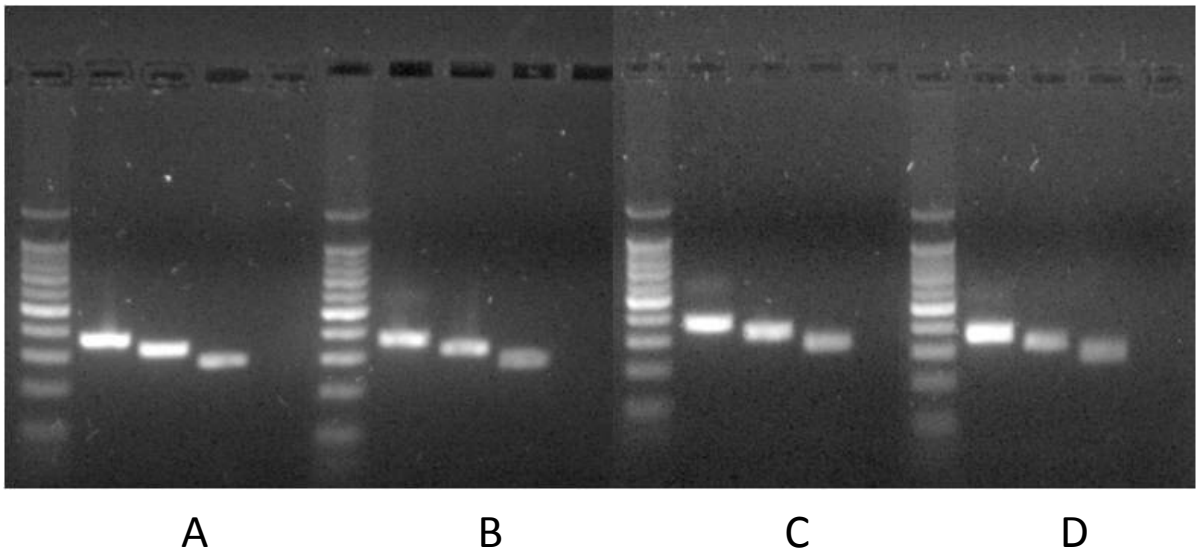


Figure A4.4. Reverse transcription PCR (RT-PCR) study for ribulose-1,5-bisphosphate carboxylase oxygenase (*rbcL*) and phosphoribulokinase (*prk*) under different mixotrophic growth conditions. (A) CO_2+N_2; (B) CO_2+$NaNO_3$; (C) glycerol+$NaNO_3$; (D) glucose+$NaNO_3$. The 16S rRNA gene was used as the internal reference; the no template control (NTC) was added under each mixotrophic growth conditions.

Curriculum Vitae Bertram Michael Berla

PhD (December, 2014) in chemical engineering with expertise in cyanobacterial metabolism

Research Highlights

- Created a system for high protein expression in cyanobacteria using endogenous plasmids.
- Identified the physiological role for a new class of cyanobacterial diesel-like secondary metabolites in cold-stress adaptation. These alkanes also modulate the rate of cyclic electron flow in photosynthesis.
- Resolved a 45-year conflict in the literature about cyanobacterial central metabolism using stable isotope labeling. Our experiments provided the first direct, *in vivo* evidence that a complete TCA cycle exists in cyanobacteria.
- Created a high-quality, comprehensive metabolic model of *Synechocystis* 6803 that predicted engineering strategies for biofuel production as well as the consequences of increased photosynthetic cyclic electron flow in a mutant.

Education:

- PhD: Energy, Environmental, and Chemical Engineering, Washington University in St. Louis. Dissertation topic: "Metabolic Engineering of Cyanobacteria for Photosynthetic Production of Drop-In Liquid Fuels"
- BS *Cum Laude* (2005) in Biology Honors and Chemistry, University of Illinois, Urbana-Champaign.

Research Experience:

- PhD Thesis project*: ""Metabolic Engineering of Cyanobacteria for Photosynthetic Production of Drop-In Liquid Fuels", May.2008 – Present*

    - Developed genetic tools for a model cyanobacterium.
    - Demonstrated that diesel-like molecules produced by cyanobacteria are necessary for cold tolerance and modulate cyclic electron flow.
    - Engineered cyanobacteria to produce high levels of heterologous proteins.
    - Standardized a GC-MS assay for n-alkane production from cyanobacteria with >10X higher sample throughput than the standard method.
    - Collaborated with computational biologists to build a comprehensive, high quality, genome-scale metabolic model of a photosynthetic bacterium.
    - Mentored a group of undergraduate researchers in the IGEM (international Genetically Engineered Machine) competition.

- Research Technician*, Donald Danforth Plant Science Center, labs of Daniel Schachtman and Leslie Hicks, Jan.2006 – Jul.2008*

    - Developed a real-time assay of transpiration in maize leaves.
    - Investigated the control of drought response by small molecules in maize xylem sap, identifying signals that led to drought tolerance and resistance.
    - Mastered a variety of techniques including gene expression assays, ion chromatography, mass spectrometry, porometry, proteomics.
    - Mentored undergraduate summer interns.
    - Analyzed the *Brassica juncea* root proteome during exposure to cadmium and optimized workflows for enhanced proteome coverage using iTRAQ.

- Undergraduate Thesis*, Univ. of Illinois Urbana-Champaign Aug.2001 – May.2005*

    - Designed and implemented a model system to study how rising CO_2 affects C_3/C_4 plants.
    - Performed surveys of insect populations in the corn-soybean ecosystem.

Skills:

Molecular Biology and Cloning
Transcriptomics/Proteomics
Analytical Chemistry
Protein Biochemistry
Metabolic Engineering
Basic computer programming
Laboratory Safety Officer
Lab Webmaster

Publications:

- **Berla BM**, Saha R, Maranas CD, Pakrasi HB (2015) "Cyanobacterial Alkanes Promote Growth in Cold Stress and Modulate Cyclic Photophosphorylation" *In preparation.*
- Ng A, **Berla BM**, Pakrasi HB (2015) "Neutral sites on endogenous plasmids in *Synechocystis* sp. PCC 6803 enable increased protein expression and are composable with strong promoters" *In preparation*.
- You L, **Berla BM**, He L, Pakrasi HB Tang YJ (2014) "13C-MFA delineates the photomixotrophic metabolism of Synechocystis sp. PCC 6803 under light- and carbon-sufficient conditions" *Biotechnol J* doi: 10.1002/biot.201300477
- **Berla BM**, Saha R, Immethun CM, Moon TS, Maranas CD, Pakrasi HB (2013) "Synthetic Biology of Cyanobacteria: Unique Challenges and Opportunities" *Frontiers in Microbiology* **4**:246 doi: 10.3389/fmicb.2013.00246
- Mueller T, **Berla BM**, Pakrasi HB, Maranas CD (2013) "Rapid Construction of Metabolic Models for a Family of Cyanobacteria Using a Multiple Source Annotation Workflow" *BMC Syst Biol* doi: 10.1186/1752-0509-7-142
- **Berla BM**, Pakrasi HB (2012) "Up-regulation of plasmid-encoded genes during stationary phase in *Synechocystis* sp. PCC 6803, a cyanobacterium", *Appl Environ Microbiol* doi: 10.1128/AEM.01174-12

- Saha R, Verseput AT, **Berla BM**, Mueller TJ, Pakrasi HB, Maranas CD (2012) "Reconstruction and Comparison of the Metabolic Potential of Cyanobacteria *Cyanothece* sp. ATCC 51142 and *Synechocystis* sp. PCC 6803" *PloS ONE* 7(10):e48285
- You L, **Berla BM**, Feng X, Page L, Pakrasi HB, Tang YJ (2012) "Metabolic pathway discovery and quantification via 13C-isotopomer analysis of proteinogenic amino acids" *J Vis Exp* doi: 10.3791/3583
- Feng X, Bandyopadhyay A, **Berla BM**, Page L, Wu B, Pakrasi HB, Tang YJ (2010) "Mixotrophic and photoheterotrophic metabolism in *Cyanothece* sp. PCC 51142 under continuous light" *Microbiology* 156(8):2566-2574
- Alvarez S, **Berla BM**, Sheffield J, Cahoon RE, Jez JM, Hicks LM. (2009) "Comprehensive analysis of the *Brassica juncea* root proteome in response to cadmium exposure by complementary proteomic approaches" *Proteomics* 9(9):2419-31
- Ernst L, Goodger JQ, Alvarez S, Marsh EL, **Berla B**, Lockhart E, Jung J, Li P, Bohnert HJ, Schachtman DP (2010) "Sulphate as a xylem-borne chemical signal precedes the expression of ABA biosynthetic genes in maize roots" *J Exp Bot* 61(12):3395-405.

Recent Presentations:

- **Berla BM**, Saha R, Pakrasi HB. October 26, 2014. Annual Midwest/Southeast Photosynthesis Meeting, Turkey Run, IN "Cyanobacterial Alkanes Enable Growth and Modulate Cyclic Electron Flow Under Cold-Stress."
- **Berla BM,** Yu J, Ng A**,** Pakrasi HB. July 21, 2014. Society for Industrial Microbiology and Biotechnology Annual Meeting, St. Louis, MO "Cyanobacterial Alkanes Enable Low-Temperature Growth" Poster.
- **Berla BM**, Pakrasi HB. April 4, 2014. Graduate Student Research Award Lecture, Washington University Dept. of Energy, Environmental, and Chemical Engineering, St. Louis, MO "Upregulation of plasmid genes during stationary phase in the cyanobacterium Synechocystis sp. strain PCC6803 and potential use of these plasmids for biofuel engineering."
- Yu J, **Berla BM**, Landry BP, Mueller T, Saha R, Sherman LA, Maranas CD, Pakrasi HB. February 10, 2014. Genomic Science Contractors-Grantees Meeting XII, Bethesda, MD "Use of Systems Biology Approaches to Develop Advanced Biofuel-Synthesizing Cyanobacterial Strains" Poster.
- Saha R, Verseput AT, **Berla BM**, Mueller TJ, Pakrasi HB, Maranas CD. August 7 2013. 11th Workshop on Cyanobacteria, St. Louis, MO "Comparative Genome-Scale Modeling of the Metabolic Potential of Cyanobacteria *Cyanothece* sp. ATCC 51142 and *Synechocystis* sp. PCC 6803" Poster.
- Kottapalli J, Luskin J, Shih R, Sossenheimer P, **Berla BM**, Bhattacharyya M, Duan N, Liberton M, Yu J, Zhao LX, Zhang F, Pakrasi HB. August 7 2013. 11th Workshop on Cyanobacteria, St. Louis, MO "Converting *E. coli* into a Nitrogen Bio-Fertilizer Using a Cyanobacterial *nif* Cluster: an iGEM project" Poster.
- You L, **Berla BM**, He L, Pakrasi HB, Tang YJ. August 7 2013. 11th Workshop on Cyanobacteria, St. Louis, MO "*In Vivo* Quantification of Flux Through A Cyanobacterial TCA Cycle" Poster.