DISSERTATION

DAY AND NIGHT FOR CYANOBACTERIA: SYSTEMS AND SYNTHETIC BIOLOGY APPROACHES TO UNDERSTANDING AND ENGINEERING *SYNECHOCYSTIS* SP. PCC 6803 UNDER DAY/NIGHT LIGHT CYCLES

Submitted by

Allison Jean Zimont Werner

Graduate Degree Program in Cell and Molecular Biology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2018

Doctoral Committee:

Advisor: Christie A.M. Peebles

Kenneth Reardon
Ashok Prasad
Adam Heuberger

ABSTRACT

DAY AND NIGHT FOR CYANOBACTERIA: SYSTEMS AND SYNTHETIC BIOLOGY APPROACHES TO UNDERSTANDING AND ENGINEERING *SYNECHOCYSTIS* SP. PCC 6803 UNDER DAY/NIGHT LIGHT CYCLES

Photosynthetic organisms—including plants, algae, and cyanobacteria—harness sunlight as an energy source to grow, utilizing atmospheric carbon dioxide in the process. This ability can be harnessed for the sustainable production of food, fuels, and chemicals, reducing demand for petrol-based products and overall greenhouse gas emissions. Photosynthetic success rests on the efficient and timely capture of sunlight. Natural day/night cycles subject these organisms to changing energy availability, presenting a fundamental question: How do phototrophs regulate metabolism to thrive under daily and dramatic changes in energy supply? This question has significant impact on the productivity of plants, algae, and cyanobacteria. Cyanobacteria have been extensively engineered for the production of biofuels, polymers, and valuable pigments under continuous-light (CL) laboratory conditions. However, industrial production requires outdoor cultivation under diurnal light/dark (LD) cycles, where yield improvements in engineered strains observed in CL are lost in LD cycles. The success of industrially-productive cyanobacteria biotechnology is limited by the lack of appropriate strain engineering tools and gap in knowledge of photosynthetic metabolism under daily day/night light cycles. The aim of this thesis is therefore to improve the feasibility of cyanobacteria biotechnology in industrially-relevant conditions by integrating aspects of diurnal LD cycles into genetic tools and by expanding the current knowledge of dynamic photosynthetic metabolism. The first part of this

thesis presents novel genetic engineering tools which enable light-entrained gene expression under diurnal LD cycles. The tools developed here enable engineering of temporally controlled chemical production under diurnal LD cycles, which we hypothesize will improve yield in outdoor cultivation environments. The second part of this thesis presents time-course characterization of growth and metabolite abundance under realistic diurnal LD cycles. Previous work was limited to on/off patterns of low light and restricted to detecting few metabolites. To expand the realism of light profiles and metabolite scope, a photobioreactor was engineered to supply sinusoidal patterns and intensity of light (sinLD cycles), and a multi-platform mass spectrometry workflow was developed to enable semi-comprehensive metabolite detection. Cyanobacteria growth under realistic diurnal sinLD cycles is presented for the first time, to our knowledge. We observe a short lag phase at the onset of day, followed by cell mass increase during the early day, cell division during afternoon and evening, and slight mass loss overnight. Further, comprehensive metabolite abundance every 30-120 minutes across a 24-hour diurnal sinLD cycle is presented. Insoluble C6 carbohydrates displayed sharp oscillations at the day/night transition; insoluble C5 carbohydrates and glucosamine display these in addition to abundance ‘re-sets’ at the night/day transition. Free amino acids and nucleic acids increase immediately upon transition to light during the lag phase, followed by gradual incorporation into protein during the mass accumulation phase. Metabolites involved in central metabolism did not oscillate to the same extent as other pathways. Accumulation of phosphoenolpyruvate but not pyruvate during the light phase suggests a potential bottleneck. Integration of the metabolomics data into genome-scale metabolic models to perform dynamic flux balance analysis could improve the method by which engineering targets are identified for production in outdoor conditions. Together, this thesis demonstrates the need for revision of the current approach to

cyanobacteria strain engineering. More broadly, this work highlights the dynamic nature of photosynthetic metabolism and motivates future investigations into metabolic regulation and metabolic flux under realistic day/night cycles.

ACKNOWLEDGEMENTS

Thank you to my advisor Professor Christie Peebles for teaching, mentoring, and trusting me. I am grateful to have been mentored by an advisor who bursts with ideas and optimism. Professor Peebles has created an environment where students are afforded the freedom to drive their science. This environment has allowed me to grow in the ability to design experiments and think critically. I feel fortunate to have been given the opportunity to pursue my doctorate degree in the environment Professor Peebles has created.

I am grateful to my committee members Ken Reardon, Ashok Prasad, and Adam Heuberger for exceptional guidance. You have each provided me with unique critique, recommendation, and perspective which supported the scientific excellence of this dissertation, as well as my personal success. Thank you for the attention and energy you have given me and my work over the past four years.

Corey Broeckling, Director of the CSU Proteomics and Metabolomics Facility (PMF), deserves special acknowledgement for mentoring me in the art of mass spectrometry. I am grateful to have been mentored by someone so knowledgeable and altruistic. PMF was invaluable to the success of this dissertation. The staff—Linxing, Lisa, Sarah, and Emily—ran the mass spectrometry instruments for my metabolomics work and were helpful as I sought to learn the techniques.

Thank you to Peebles' lab students and alumni, Jake Sebesta, Shea Moore-Farell, Ian Cheah, and Steve Albers, for reading manuscripts, critiquing presentations, providing advice, and generally being great lab-mates. Thank you to the undergraduate researchers who I have been

fortunate to work beside: Aidan Ceney, Makayla Hohn, Anthony Roulier, Olivia Smith, Katelyn Oliver, and Dylan Miller.

Fort Collins has been a most gracious host for my graduate studies. I encountered cheerfully helpful individuals all across the school and city which enabled my experience to be accelerated and enjoyable. Thank you to Tammy, Angela, Julie, and Dee for being kindest and attentive custodial team I will likely ever meet; your companionship meant a great deal to me. Thank you to Tim Gonzales for never growing weary with my requests for your endless supply of fix-it power; your support was critical to the success of the photobioreactor engineering project. Thank you to Charlene Spencer and Denise Morgan for the excellent administrative work you provided. Thank you to my program directors Howard Liber and Carol Wilusz; you went above and beyond to assist me in times of need. Thank you to the baristas of Mugs Coffee Lounge for the routine kindness and excellent Americanos you provided on a daily basis.

I was fortunate to find a group of peers who sustained me through my studies. Laylaa Ramos, Talia Bickett, Aliena Debelak, Hailey Sedam, Hannah Berry and Katie Teegarten have been a source of friendship, support, and encouragement through all the milestones of a PhD: first-year coursework, second-year prelim exams, third-year blues, PhD defenses, and all the times in-between. Allie Holschbach has been a source of valuable advice and a most enjoyable cycling partner. Ashley and Adam Heck, Mallory Shields and Rachel West have been the greatest neighbors and ski buddies one could ask for; I will be sad to leave the halfway house neighborhood. I am grateful to, and owe a great deal of my sanity to, these and the many other wonderful individuals with whom I have become friends during my time at CSU.

Thank you to my parents, Ben and Kathy Zimont, for being the first to teach me the joy of exploring the natural world and continuing to support that joy throughout my education.

Thank you to my sister, Meg Zimont, for providing comedic relief and reminding me there is more to life than science. Thank you to my grandparents Grams and Gramps, to my parents-in-law Maureen and Tom, and to my Werner siblings Casey, Haley, and Kaitlyn Werner and Frank McCoy for being a source of love and support.

Finally, thank you to my husband Jordan, for being a constant source of good spirits, balanced perspective, and maintaining the humble belief that I would succeed in this endeavor.

DEDICATION

To my family.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 – INTRODUCTION

SUMMARY

The purpose of this chapter is to introduce the reader to the themes and concepts presented in this thesis: cyanobacteria biotechnology, diurnal day/dark cycles, and systems biology. A broad introduction to biotechnology and cyanobacteria is first presented to orient the novice. A detailed overview of synthetic biology and metabolic engineering tools and applications to cyanobacteria are then presented. The overarching theme of this work—diurnal light/dark cycles—is next introduced, with a thorough review of previous work done studying cyanobacteria under these environmental conditions. Special attention to circadian rhythms and systems biology investigations is provided. Next, metabolomics is introduced and mass spectrometry platforms and data analysis strategies utilized are defined. Finally, the contributions of this thesis are summarized for each chapter.

CYANOBACTERIA BIOREFINERIES

General history of biotechnology

The use of microorganisms to improve human life predates modern science. Common products today—including cheese, bread and beer—were invented by necessity using biotechnology thousands of years ago. Today, microorganisms are used to produce biodegradable plastics, life-saving drugs, and environmentally sustainable energy sources. Biotechnology, most simply, is the use of biology to improve human health and/or environment [1] and has profoundly influenced human society for thousands of years.

Our ancestors made great societal advances using the unique capabilities of microorganisms in the earliest forms of biotechnology (even though they did not understand that microorganisms existed). Neolithic dairy farmers preserved dairy products by fermenting sour milk into cheese at least 7,500 years ago [2]. As early as 2000 BC, vinegar was made by fermenting sugar-rich foods such as barley and fruit [3]. Vinegar was used by nearly every early civilization as a food preserving agent, disinfectant, beverage, and condiment [3]. Bread was leavened with yeast and provided a compact and enjoyable way to store and transport cereals. The earliest archeological evidence of fermented beverages date back to 1200 BC in ancient China, but may likely be even older [4]. Fermented beverages, commonly made from barley, fruit, or rice, contain ethanol which acts as a pain-reducer and disinfectant and were likely used as a safer alternative to contaminated water sources.

The age of Classical Biotechnology, spanning the mid-1800s to mid-1900s, was a period of remarkable breakthrough in understanding the underlying principles of biology and genetics. Louis Pasteur discovered that microorganisms were responsible for fermentation in 1857. Gregor John Mendel proposed that internal and invisible information was responsible for observable traits, marking the ideological conception of genes presented as Mendel's "Laws of Inheritance" in 1865. A few years later, Fredrich Miescher isolated "nuclein" from white blood cells. Nuclear DNA organizational structure was later observed and the term "chromosome" was coined by Heinrich Wilhelm Gottfried Von Waldeyer-Hartz in 1888. This rapid series of discoveries is truly remarkable! In 1953 the structure of DNA was solved by James Watson and Francis Crick using Rosalind Franklin's X-ray crystallography. With the organization and structure of DNA in hand, tremendous advancement in genetics, molecular biology, and biotechnology were achieved.

Today, Modern Biotechnology addresses the challenges facing society using advanced tools and ever-expanding knowledge. Pest-resistant, drought-resistant, and higher-producing crops have revolutionized agriculture. Antibiotics, vaccines, anti-cancer drugs, gene therapies, and improved diagnostics have revolutionized human health care. Biosynthesis of chemicals such as dyes, plastics, polymers, and fuels promise to reduce petroleum use and decrease overall greenhouse gas emissions.

A common approach to mass production of these important products is to engineer biology to produce the chemical of interest (*e.g.* a therapeutic drug, a polymer to make plastic bottles, a biofuel replacement for jet fuel, etc.). This approach is advantageous when it produces less waste and requires less energy input as compared with traditional chemical synthesis. Take the anti-cancer drug Paclitaxel (Taxol™) as an example. Taxol was discovered and isolated from the bark of Pacific and/or European Yew trees. Taxol has a complex structure consisting of interlocked rings and beautiful, winding stereochemistry. These properties make it difficult and costly to synthesize chemically (*i.e.* 'free-floating' in a tube or reactor). Isolation from the tree alone is impractical and unsustainable because the global demand (~1040 kg/year in 2014) far exceeds the Yew tree supply. Thus, production using biotechnology is a promising solution. After much research and development, the majority of the drug is now produced using plant cell cultures: plant cells are cultured in specific conditions which allow them to function as microscopic plant cell factories to produce Taxol [5]. Research groups and private companies are working to further increase the amount of Taxol produced using metabolic engineering techniques. This technique is widely applied for the production of other pharmaceuticals.

There are many biological hosts (the organism being engineered, also called a chassis) which can be used for chemical production. Commercially, *E. coli* and yeast are the most

common hosts due largely to the well-established toolset and vast knowledge base available. *E. coli* alone produces ~30% of all therapeutic proteins, including FDA approved drugs treating Crohn's disease, multiple sclerosis (MS), and macular degeneration [6]. Yeast is widely used for biodegradable polymers, platform chemicals, and ethanol [7–9]. However, these hosts have drawbacks. Both *E. coli* and yeast are heterotrophic, which means they require a sugar substrate to grow. The addition of sugar substrate can be a significant material cost and represents one key drawback of using these microorganisms as hosts. Photosynthetic microorganisms, on the other hand, do not require the addition of a sugar substrate, because they use atmospheric carbon dioxide and sunlight to grow. Cyanobacteria are one type of photosynthetic microorganism which can be used for a wide range of biotechnology applications.

Cyanobacteria

Cyanobacteria are ancient, productive, and diverse microorganisms that profoundly influence global biodiversity and underpin Earth's carbon cycle. They have gained attention in the past several decades as a biotechnology chassis primarily because they are photosynthetic: cyanobacteria convert sunlight and carbon dioxide into chemicals, and thus do not require the addition of a sugar substrate. A brief introduction to the Cyanobacteria phylum follows.

Fossilized cyanobacteria date back 3.5 billion years, presenting them as one of the earliest organisms known to exist. Geobiologists postulate that oxygen production from cyanobacteria, in combination with other geological events, led to the transformation of the Earth's atmosphere into the oxygen-rich atmosphere we know today [10].

Cyanobacteria are the progenitors of the chloroplast—the power house organelle of plants and photosynthetic organisms. Likely, a non-photosynthetic eukaryote engulfed a

cyanobacterium; then, that cell was then engulfed by another eukaryote resulting in a photosynthetic eukaryotic cell, like that which comprises plants [11,12]. Plants are of great interest to society, but they are much more complex than bacteria, and therefore more difficult to study. Because of the direct link between cyanobacteria and plants, cyanobacteria serve as great model organisms for understanding plant photosynthesis.

The key difference between cyanobacteria and algae is that cyanobacteria are single-celled bacteria while algae are eukaryotic and can be large multicellular organisms (*i.e.* giant kelp and brown algae). It can be confusing because cyanobacteria have internal thylakoid membranes which deviate from the classical definition of a bacterium. Conversely, there are many cases of unicellular algae! Diatoms are another distinct type of algae primarily due to their strong silica cell wall. Together, these photosynthetic microorganisms—cyanobacteria, algae, and diatoms—are commonly referred to as phytoplankton.

Phytoplankton are the foundation of the ocean ecosystem connecting biological life with Earth's primary energy source: the Sun. Together, phytoplankton contribute *half* of the Earth's primary productivity despite only comprising 1% of photosynthetic biomass [13]! Cyanobacteria inhabit a shockingly diverse range of ecosystems including hot deserts, polar deserts, Antarctic lakes, fresh water lakes, salt water oceans, savannahs, tropical rainforests, and virtually any rock surface [14–16]. The diversity of ecosystems in which cyanobacteria thrive is a testament to both their resiliency and metabolic diversity!

Oxygenic photosynthesis

Plants and cyanobacteria carry out oxygenic photosynthesis, and thus support the oxygen-requiring life on Earth. Oxygenic photosynthesis is defined as "a non-cyclic photosynthetic electron chain where the initial electron donor is water and, as a consequence, molecular oxygen

is liberated as a byproduct" [17]. Understanding the molecular mechanisms of photosynthesis is critical to understanding and engineering cyanobacteria.

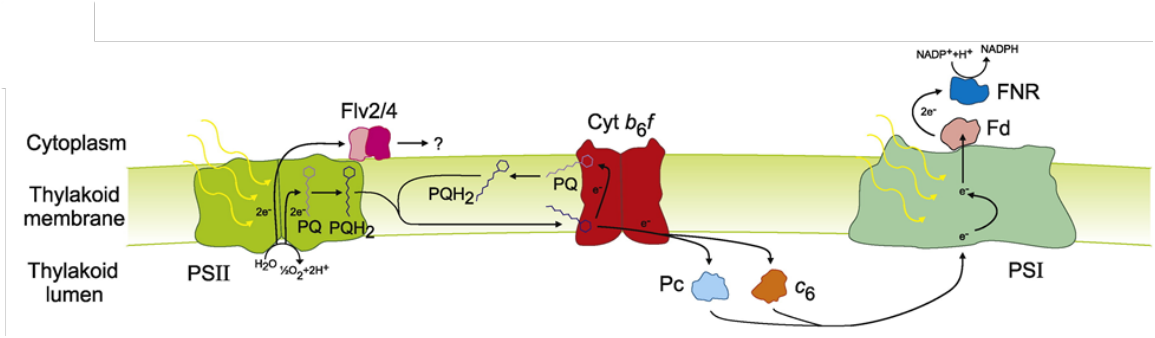

Figure 1-1. Linear photosynthetic electron transport. Image modified from [18]. ATP synthase not pictured. PSII: Photosystem II; PQ: plastoquinone; PQH_2: plastoquinol; Flv2/4: flavodiiron proteins; Cyt b_6f: cytochrome b_6f; Pc: plastocyanin; Fd: ferrodoxin; FNR: Ferrodoxin-NADP(+) reductase; PSI: Photosystem I.

Thylakoid membranes house the electron transport chain (ETC) in cyanobacteria (Figure 1-1). Phycobilisomes are protein-chromophore complexes situated atop the thylakoid membranes with the purpose of capturing and funneling light energy (photons). Absorbed photon energy is transferred to Photosystem II (PSII) reaction centers as excited electrons. At the reaction center, two electrons split H_2O into 2 H^+ and ½ O_2 in the thylakoid lumen. Here begins a series of redox chemistries and electron transfers which comprise the electron transport chain.

Two electrons reduce plastiquinone (PQ) to plastiquinol (PQH_2). PQH_2 diffuses through the membrane to the cytochrome *b_6f* complex (cyt *b_6f*) where it transfers energy, is deprotonated and oxidized back to PQ, cyt *b_6f* transfers the remaining electron to plastocyanin (PC), PC transfers one electron to Photosystem I (PSI) which in turn transfers the electron to ferredoxin (Fd) and then ferredoxin NADP-reductase (FNR). FNR uses this energy to catalyze the formation of NADPH from $NADP^+$. The proton gradient across the thylakoid membrane is used to power ATP synthase, which catalyzes the formation of ATP from ADP. The energy ratios

(ATP/ADP and NADPH/$NADP^+$) and redox status (PQ vs. PQH_2) are being discovered as critically important to metabolic regulation in cyanobacteria.

Conversion of light energy to chemical energy via the ETC has been called the "light reactions" and CO_2 assimilation via the Calvin Benson Bassham (CBB) cycle has been called the "dark reactions". This terminology can be misleading because the so-called "dark reactions" can occur during periods of light. The "dark reaction" nomenclature can be rationalized as far as CO_2 fixation does not *directly* require light, but does indeed require the chemical energy (ATP and NADPH) produced by the light-harvesting ETC.

Gene expression overview

Gene expression is regulated at many points along the central dogma of molecular biology process (Figure 1-2). The regulation points are pertinent to biological, genetic, and metabolic engineers because they control how engineering strategies fare. Here, a brief description of each regulatory element is provided. In the following section, genetic tools developed for cyanobacteria are discussed.

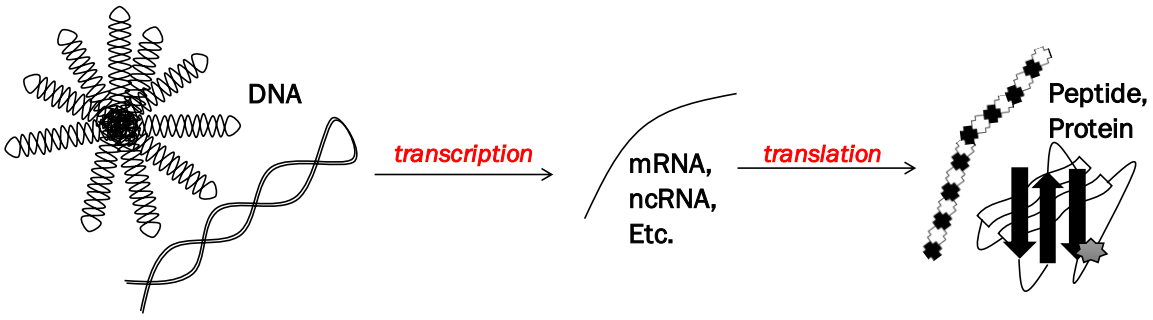

Figure 1-2. Gene expression overview. DNA is transcribed into RNA and RNA is translated into peptides which are folded into functional proteins. Image created by A. Werner.

Chromosome topology regulates the rate of transcription by controlling promoter accessibility. Eukaryotic nuclear DNA is structured on histones and extensive research has demonstrated that post-translational modifications (PTMs) to histones wildly affect gene

expression. It has long been accepted that bacteria do not organize their DNA into protein-assisted chromatin-like structures. However, the advent of "ultrastructure resolution" high-voltage cryo-electron tomography (cryoET) challenges this notion. In 2016, Murata and colleagues applied cryoET to study DNA dynamics in *Synechococcus elongatus* PCC 7942. They observed chromosome compaction resembling chromatin organization and the presence of paired polyphosphate bodies (PPBs) physically associated with compacted chromosomes [19]. Understanding the extent to which, and how, chromosome topology controls gene expression in cyanobacteria—and all bacteria—is an area that needs to be explored further. The role of chromosome topology on gene expression in cyanobacteria is discussed in the context of circadian rhythms in the following section.

Promoters regulate the rate of transcription initiation by influencing RNA polymerase binding, open complex formation, and RNA polymerase clearance. *Trans*-acting transcription factors or RNA polymerase (RNAP) sigma factors bind to specific *cis*-acting promoter elements upstream of the transcription start site (TSS) to recruit RNAP and initiate transcription. TSSs are used to identify promoter sequences. "Promoter bashing" experiments have identified two conserved regions at -10 and -35 base pairs upstream of the TSS. The -10 element is highly conserved across domains and consists of six nucleotides—the so-called "TATA box"—which interact directly with RNAP sigma factor to recruit the holozyme and initiate transcription.

Transcription initiation rates are also heavily affected by transcription factors (TFs) which bind to **enhancer and silencer** DNA elements. Enhancer and silencer sequences can be thousands of base pairs upstream of ORFs or immediately adjacent depending on the specific gene. Activators and repressors are types of transcription factors (TFs) that modulate transcription rates in seemingly endless ways that are still being explored. Munch and colleagues

complied an online tool to predict TF binding sites (TFBSs) in prokaryotes [20], but the conservation of other prokaryotic TFBSs to cyanobacteria has yet to be fully characterized.

Ribosome binding sites (RBS) regulate the rate of mRNA translation via accessibility and/or recruitment of the ribosome. A set of cyanobacterial RBS have been systematically compared using the fluorescent reporters EYFP and mTAgBFP [21] and computational tools exist to predict RBS strength [22]. Recently, a study comparing multiple RBSs for the expression of three proteins report that RBS performance changes dramatically depending on the downstream ORF, suggesting the importance of mRNA folding dynamics on the translation initiation rate [23]. Despite inclusion of mRNA folding thermodynamics into computational tools, secondary structure of mRNA transcripts still influence the translational initiation efficiency and could result in the poor predictive power in *Synechocystis* sp. PCC 6803 [24]. Together, these observations suggest further study into translation initiation at the RBS—and alternate translation initiation methods—are necessary to improve protections in cyanobacteria and establish a modular RBS tool-set.

Codon usage within the ORF regulates peptide expression via selective discrimination of degenerate bases in a yet-undefined way. Because the genetic code is degenerate there are multiple genetic codes which will produce the same protein. In 1980, Richard Grantham first hypothesized that degenerate codon usage, or 'coding strategy', is consistent within a species [25]. This is now widely accepted and well documented in a range of species. However, the underlying reason for varying coding strategies is still unclear. Hypotheses for an organism's codon usage include: (1) evolutionary metabolic discrimination, (2) optimization of mRNA secondary structure, (3) codon and anti-codon harmonization, (4) replication/transcription rate control via nucleotide availability, (5) translation rate control via tRNA availability. Engineering

heterologous protein expression requires selecting the proper codon usage for the host. Many commercial entities offer codon optimization services (https://www.atum.bio/services/genegps, http://bioinformatics.ua.pt/eugene/, and https://www.idtdna.com/CodonOpt to name a few).

Finally, **post-translational modifications (PTMs)** of amino acids regulate protein activity. PTMs can affect protein folding, conformation, and/or activity and play key roles both in modulating protein function and in signal transduction networks. In cyanobacteria, PTMs play critical roles in regulating events such as circadian rhythms, global carbon partitioning, photosynthesis, protein splicing, and many others (see [26] for review). However, the PTM landscape in cyanobacteria is still being explored and is largely uncharacterized.

Synechocystis sp. PCC 6803 as a model cyanobacterium

The cyanobacterium *Synechocystis* species (sp.) Pasteur Culture Collection (PCC) Number 6803 (*Synechocystis* sp. PCC 6803, hereafter referred to as *S.* 6803) has become a model fresh-water cyanobacterium for scientific study and engineering applications in-part because it was the first cyanobacterium to have its genome fully-sequenced and publicly-available [27]. *S.* 6803 has a ~3.57 megabase pairs (Mb) circular chromosome and seven native plasmids totaling ~0.38 Mb. For comparison, the human nuclear genome is 3,235 Mb and the maize corn genome is ~2,500 Mb (haploid, [28]). *S.* 6803 cells are approximately 1.5 micrometers in diameter and house subcellular compartments such as thylakoid membranes, protein carboxysome micro-compartments, polyphosphate bodies, lipid bodies, polyhydroxyalkanoate granules, and other cytoplasmic inclusions, depending on the growth environment (Figure 1-3).

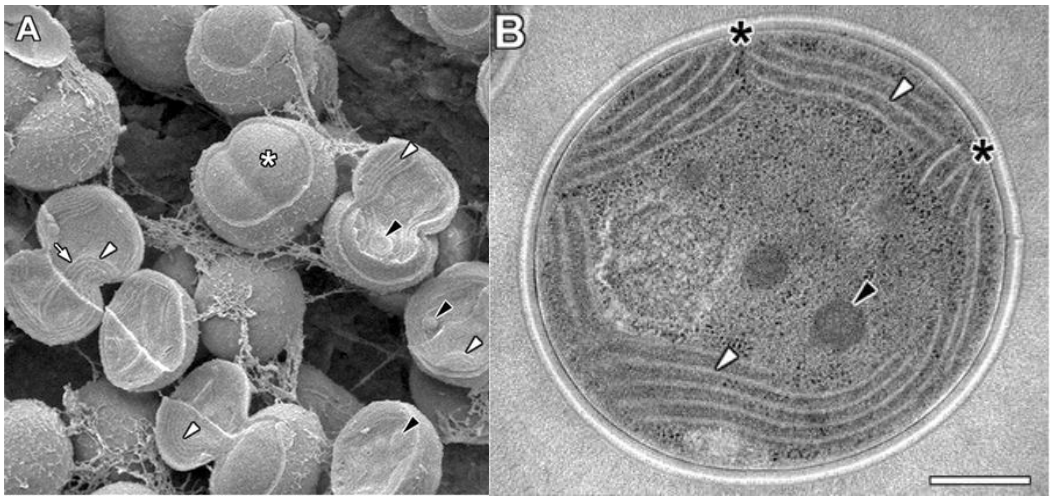

Figure 1-3. Synechocystis sp. PCC 6803 micrographs taken by van de Meene and colleagues [29]. (A) cryoSEM of S. 6803 cells high-pressure frozen and freeze-fractured. (B) TEM image of *S.* 6803. Thylakoid membranes are indicated with white arrowheads. Cytoplasmic membrane is indicated with black asterisks. Carboxysomes are indicated with black arrowheads. Bar=200 nm. Figure modified from Figure 1 in [29].

S. 6803 is naturally competent and performs homologous recombination, meaning DNA from the outside environment is actively transported into the cytoplasm, scanned for similarities to chromosomal DNA, and integrated into the chromosome at large areas of homology. To the engineer, this means that transforming *S.* 6803 with engineered DNA does not require elaborate genetic engineering tricks and instead can be done relatively rapidly and easily.

Synthetic biology tools for cyanobacteria

Transcriptional promoters

Promoter development in *S.* 6803 has received considerable attention. Characterized promoters include the constitutive P_{cpc560} [30] and the J-23 series [31], chemically-inducible P_{sca} series [32], metal ion-inducible $P_{nrsBACD}$ [33], and many others (see [21] for review). In cyanobacteria, the importance of testing a promoter's expression profile in the locus and growth condition of interest cannot be overstated. As has been documented for plants, chromosomal integration locus plays a significant role in the resulting gene expression observed. To

demonstrate this, the activity of several commonly employed *S.* 6803 promoters is reviewed in Table 1-1.

Table 1-1. Review of well-characterized *S.* 6803 promoters from variable expression loci and replicons.

Promoter	Locus/ vector	Replicon/ method	Activity	Light	Ref.
P_{cpc560}	*pta*	Chromosomal gene disruption	Constitutive and strong strength, as measured by *ter* and *dlhD* expression	CL, 100 µE	[30]
	AOY38_ 07470	Chromosomal N.S. integration	Constitutive and weak strength, steadily declining expression with increasing cell density, as measured by EYFP fluorescence	CL, 30 µE	[34]
	pPMQAK1	Non-native expression vector	Constitutive and medium strength, steadily declining expression with increasing cell density, as measured by EYFP fluorescence	CL, 30 µE	[34]
	5137–2	Native plasmid pCC5.2_M integration	Constitutive and strong strength, as measured by EYFP fluorescence	CL, 30 µE	[34]
P_{trc}	pPMQAK1	Non-native expression vector	Strongly IPTG-inducible, as measured by GFP fluorescence (based on *E. coli* $P_{trp,}$ P_{lacZYA})	CL, 50 µE	[35]
	AOY38_ 07470	Chromosomal N.S. integration	Very weak expression, as measured by GFP fluorescence	CL, 100 µE	[34]
	5137–2	Native plasmid pCC5.2_M integration	Strongly ITPG-inducible, as measured by GFP fluorescence	CL, 100 µE	[34]
P_{psbA2}	*psbA2*	Chromosomal gene disruption	Low- to high-light inducible, as measured by *(S)klspS* and *ispS* expression	LL to HL (10 to 500 µE)	[36]
	slr0168	Chromosomal N.S. integration	Constitutive expression, as measured by eYFP fluorescence	LD, 100 µE	[37]
P_{sca} suite	*slr0168*	Chromosomal N.S. integration	IPTG-inducible, as measured by GFP fluorescence (based on *E. coli* P_{tac})	CL, 100 µE	[32]
P_{L03}	pPMQAK1	Non-native expression vector	aTC-inducible, as measured by EYFP fluorescence upon transition to red light (based on P_{tetR})	CL, 30 µE	[38]
P_{nrsB}	pPMQAK1	Non-native expression vector	Nickel-inducible, and weakly copper-inducible, as measured by EYRP fluorescence	CL, 20 µE	[21]
P_{rnpB}	pPMQAK1	Non-native expression vector	Constitutive and weak, as measured by GFP fluorescence	CL, 50 µE	

LL: low-light; HL: high-light; CL: continuous light; LD: light-dark cycles; N.S.: neutral site

P_{cpc560} is considered to be one of the strongest constitutive ("always-on") promoters available for *S.* 6803, or any cyanobacterium for that matter. Derived from the c-phycocyanin gene, P_{cpc560} was discovered by Zhou *et al.* (2013) as a super-strong promoter for the production of two enzymes, crotonyl-CoA-specific trans-enoyl-CoA reductase (*ter*) and D-lactate dehydrogenase (*dlhD)* [30]. In this study, the heterologous gene products were expressed from the chromosome by disrupting phosphotransacetylase (*pta)*, the first gene in the acetate synthesis pathway. Zhou *et al.* (2013) reports that P_{cpc560} produced their heterologous proteins in super-high quantities, comprising up-to 15% of the total soluble protein. However, a systematic study of promoter performance as a function of replicon and locus presents a more complex story. When P_{cpc560} is expressed from a different chromosomal locus (and under a lower light intensity), the expression is markedly weaker and decreases significantly as cell density increases [34].

P_{psbA2} is another great example of locus-dependent promoter performance. Natively, *psbA2* encodes photosystem II D1 protein and is expressed in high quantities during light periods. When a heterologous protein is expressed from the *psbA2* chromosomal locus, the expression increases upon a shift from low- to high-light intensities [36]. However, when the *psbA2* promoter region (P_{psbA2}) region is used to drive heterologous protein expression from other chromosomal loci (*e.g.* the *slr0168* neutral site), the expression no longer has a light-intensity dependency and instead exhibits constitutive expression [37].

Even synthetic inducible promoters exhibit locus-dependence. For example, the P_{trc} promoter—based on the very well-characterized *E. coli* P_{trp} P_{lacZYA} promoters—exhibits different strengths when expressed from a non-native expression vector, a native plasmid neutral site, or a chromosomal neutral site [34,35].

Engineering cyanobacteria for industrial scale production of fuels and chemicals is limited by the genetic tools available. This thesis work addresses this limitation by developing promoters which provide light-entrained gene expression induction under daily LD cycles. These tools function consistently from the *slr0168* neutral site and provide strongly-inducible expression across multiple days of cultivation under daily LD cycles.

Ribosome binding sites

Complicating the story even further, promoter performance and ribosome binding site performance are intertwined: a promoter's performance can be affected by the subsequent RBS, just as a RBS strength can be affected by the preceding promoter sequence. Downstream genetic elements, usually within the 5' UTR, can impact expression by altering mRNA folding/ribosome accessibility and therefore the rate of translation initiation. Howard Salis' research group built a tool to aid in the design of RBSs using thermodynamics of RNA interactions within the ribosome, termed the RBS Calculator [22]. The RBS Calculator has been widely applied to successfully optimize RBSs for engineering efforts in *E. coli* (*e.g.* [39,40]). Englund, Liang and Lindberg (2016) systematically compared several *Synechocystis* sp. PCC 6803 ribosome binding sites using fluorescent reporter and found similar expression levels in *S.* 6803 and *E. coli* [21], which suggests similar performance between the species. However, the success rate of the RBS Calculator in cyanobacteria is less than optimal [35, Peebles Lab unpublished work]. More basic research into the mechanisms of translation initiation will be critical to the development of RBS genetic tools suitable for use in cyanobacteria.

Chromosomal neutral sites and replicative vectors

Genetic engineering in *S.* 6803 is implemented via integration into the chromosome and replicative plasmids. pPMQAK1 is a replicative plasmid built with the RSF1010 replicon and BioBrick-compatible features [35]. Chromosomal loci which provide no detrimental growth defects in insertion/deletion mutants are considered "neutral sites". Pinto *et al.* (2015) applied a systematic genome mapping and genetic validation study to identify neutral sites [42]. Ng, Berla and Pakrasi (2015) identified 21 chromosomal and endogenous plasmid neutral sites in *S.* 6803 by analyzing CL RNA-Seq data and identifying 400-1,100 bp regions to which no RNA transcripts (including ncRNA, rRNA, and tRNA) matched [34,43]. They report drastic expression differences based on integration locus, highlighting the importance and influence of chromosomal locus on gene expression behavior. Another important consideration for chromosomal integration in cyanobacteria is achieving full segregation. *S.* 6803 is highly polyploid, carrying up-to 200 genome copies during exponential growth and ~50 genome copies during linear and stationary growth [44], so multiple rounds of selection are often necessary to achieve full segregation.

Metabolic engineering cyanobacteria biorefineries

Cyanobacteria "biorefineries" couple photosynthetic carbon sequestration of atmospheric carbon dioxide to the production of valuable fuels and chemicals. Analogous to the petroleum refinery which processes crude oil into useful products (*e.g.,* gasoline, diesel fuel, kerosene, etc.), a cyanobacteria biorefinery processes sunlight, CO_2 and water into useful products. Obvious advantages of cyanobacteria biorefineries over petroleum refineries include contributing to independence from finite fossil fuel resources and reducing net greenhouse gas emissions. Metabolic engineering efforts seek to funnel CO_2 into valuable and useful products.

At the most basic level, metabolic engineering is a practice which must (i) make a change to the cell and (ii) detect the effect of that change on metabolism, especially on the amount of product produced. Diverse hypotheses drive the types of changes implemented, but most require a genetic modification. Several cyanobacterial species are genetically tractable, including *S.* 6803, *Synechococcus* sp. 7942 (hereafter *S.* 7942), and *Synechococcus* sp. PCC 7002, and thus dominate the genetic and metabolic engineering research space. Genetic engineering of pathway components—both endogenous and heterologous—have been extensively applied in these cyanobacteria for a host of products including alcohols, fatty metabolites, hydrocarbons, carbohydrates, carboxylic acids, terpenes and isoprenes. Pathway engineering is a common strategy with product-specific approaches. The reader is directed to recent reviews for a summary of the these efforts [45–47]. Here, a review of common engineering strategies in cyanobacteria and algae is presented.

Cultivation strategy

Abiotic factors play a strong role in determining cellar physiology and metabolism, and thereby influencing chemical yield. Important factors influencing cyanobacteria physiology include nutrient ratios, nutrient concentration, pH, temperature, and light. Macronutrient availability, especially the nitrogen to phosphorus ratio (N:P), drastically affects algal biomass composition. For example, phosphorus limitation in *Arthrospira platensis* increased carbohydrate content from 11% to 67% [48]. Nitrogen limitation in *Nannochloropsis* sp. M&M-M24 increased lipid content from 32% to 60% [49]. Micronutrients such as iron, nickel, magnesium, molybdenum, and zinc also affect biomass composition [50]. Under iron starvation, ETC electron carriers are substituted, and new iron scavenging systems are synthesized [51]. Altering light availability and/or intensity affects physiology and biomass. For example,

phycobilisome structure and arrangements are altered in response to changing light quality and are degraded under nitrogen starvation [52]. However, applying environmental stressors—such as nutrient limitation—often come at the expense of decreased biomass productivity; thus, applying environmental stress is considered a non-optimal solution for the generation of photosynthetic biorefineries [50].

Engineering improved carbon fixation

Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) catalyzes first step of carbon fixation. RuBisCO is notoriously slow and promiscuous; thus, much of the engineering efforts for improved carbon fixation have been directed at improving RuBisCO selectivity and/or rate. This has been attempted in several hosts with minimal success [53–55]. *S.* 6803 over-expressing FLAG-tagged RuBisCO exhibited an increased growth rate and photosynthesis rate [56]. In *S.* 7942, a non-native over-expressed RuBisCO resulted in a 1.63-fold increase in photosynthetic rate despite incomplete incorporation into the carboxysome [57]. Another *Synechococcus* sp. 7942 study showed a similar result: over-expression of RuBisCO resulted in a 2-fold increase in isobutyraldehyde [58].

Another common approach to improving carbon fixation is engineering the carbon-concentration mechanism (CCM). Cyanobacterial CCMs actively transport CO_2 and HCO_3^- inside the cell, convert CO_2 to HCO_3^-, concentrate HCO_3^- and RuBisCO inside protein microcompartments called carboxysomes, and convert HCO_3^- back to CO_2 via carbonic anhydrases for RuBisCO fixation [59]. CCM-engineering by way of increased bicarbonate transporters has increased biomass productivity 2-fold in *S.* 6803 [60].

NADPH is required for both activation and utilization for many of the CBB cycle enzymes. Over-expression of NADPH and ATP generation pathways is commonly applied in

addition to product-specific pathway engineering to provide sufficient energy for carbon assimilation and therefor flux to the product of interest [61]. ATP:NADPH balancing is also critical as the ATP:NADPH ratio plays an important regulatory role in photosynthetic processes [62–64]. Cofactor balancing is commonly applied in engineering efforts beyond that specifically targeting improved carbon fixation.

Engineering improved light capture and/or photosynthetic efficiency

Light capture processes and components have also been engineered to improve photosynthetic efficiency. Noy, Moser, and Dutton (2006) consider the engineering guidelines necessary for a *de novo* photosynthetic unit (PSU) and the areas amenable to exploitation for improved efficiency. They propose a PSU composed of units responsible for light absorption, excitation energy transfer (light capture), and electron transfer (convert light energy to chemical energy) [65]. Theoretical light capture alterations for improved photosynthetic efficiency have also been proposed [66,67]. Interestingly, they propose that a decreased light harvesting unit would increase overall photosynthetic efficiency by reducing excess energy dissipation, decreasing cell shading, and thereby increasing the depth of light penetration into a liquid culture or forest.

Progresses in the experimental application of these proposed principles to improve photosynthetic efficiency are unclear. Page, Liberton, and Pakrasi (2012) found that truncation of phycobilisomes in cyanobacterium *S.* 6803 resulted in a significant decrease in photosynthetic O_2 evolution and a slight decrease in biomass productivity [68]. Perriene, Negi, and Sayre (2012) found that an increase in chlorophyll *b* levels in alga *Chlamydomonas reinhardtii* resulted in a significant increase in photosynthetic rate [69]. Further work in this area will be critical to

application of theoretical engineering principles to optimize photosynthetic efficiency in phototrophs.

Protein engineering

Protein engineering of rate-limiting enzymes in cyanobacteria has been successful strategy for terpene synthases. Formighieri and Melis (2015) over-expressed a terpene synthase (β-phellandrene synthase)-phycocyanin (phycocyanin β-subunit) fusion protein in *S.* 6803 to realize a 100-fold increase in β-phellandrene production [70]. Gao *et al.* (2016) expressed plant-derived isoprene synthases with over-expression of IPP synthases in *Synechococcus elongatus* to achieve 1.2 g L^{-1} isoprene [71].

Genome-scale models and flux-balance analysis

Genome-scale metabolic network reconstructions, or genome scale models (GSMs), are mathematical models representing all reactions in a given organism, the enzymes responsible for all reactions, and the genes encoding each enzyme. Genome-scale models (GSMs) have been developed for many species of bacteria, archaea, and eukaryotes and have wide-spread applications including identifying drug targets effective against specific pathogens and designing metabolic engineering targets for improved chemical production. Many different algorithm approaches have been developed to aid in these pursuits [72]. In *S.* 6803, several GSMs have been developed [73–77]. Most recently, *i*SynCJ806 was developed to include 816 genes and 1605 reactions, including photosynthetic reactions previously neglected [77].

GSMs can be used for quantitative approaches to studying metabolic network responses to environmental and/or genetic perturbations and inform engineering approaches. Using a GSM as a starting point, all enzyme-catalyzed metabolite transformations in the metabolic network can

be represented by a stoichiometric matrix (**S**) with columns of reactions (*n*), rows of metabolites (*m*), and stoichiometric coefficients assigned to each metabolite participating in a reaction. Vectors representing reaction flux (**v**) and metabolite concentration (**x**) are constructed of length *n* and *m*, respectively. At steady-state, a system of linear equations can be formulated as:

$$\mathbf{S} \cdot \mathbf{v} = \mathbf{0} \tag{1}$$

Large metabolic networks, such as the *S.* 6803 model *i*SynCJ816, have more reactions than equations and thus are underdetermined. To narrow the solution space, upper- (v_{upper}) and lower (v_{lower}) flux constraints can be applied. Flux balance analysis optimizes a user-defined objective function by predicting network flux distribution which satisfies the constraints in Equations (1) and (2) [78]. Cyanobacterial metabolism is dynamic in nature due to global entrainment to daily fluctuations in light availability over the course of daily day/night cycles (discussed in following section). For engineering efforts aimed at the creation of strains viable in industrially-relevant conditions, standard FBA approaches are inadequate. Instead, dynamic flux balance analysis (DFBA) is more appropriate, as it incorporates kinetic expressions of biochemical processes with FBA to predict network behavior in a dynamic environment, such as daily day/night cycles [79].

Dynamic flux balance analysis (DFBA) predicts metabolic network flux distribution response to dynamic perturbations. Kinetic equations describing metabolite uptake/excretion, dynamic metabolite abundance, and growth rate are incorporated into the model and used as additional constraints. Joshi *et al.* (2017) applied the direct approach to DFBA to *S.* 6803 using the MATLAB-based DFBAlab simulator which embeds a linear program (LP) within the previously mentioned kinetic equations to perform the direct approach to DFBA [80]. However, experimentally-validated data in dynamic light conditions are sparse, and therefore the development of DFBA approaches in *S.* 6803 is limited by the development and curation of

multi-level ‘omics (transcriptomics, proteomics, and metabolomics) data in diurnal light:dark cycles.

This thesis work addresses this limitation by providing data with which improved DFBA can be built for metabolic engineering cyanobacteria for outdoor production facilities. Specifically, this work provides metabolite data to support a dynamic biomass equation across the day/night cycle. Future work developing DFBA for day/night light cycles has tremendous potential to improve the identification of engineering targets for appropriate strain engineering in the industrially-relevant environmental condition of diurnal light:dark cycles.

DIURNAL LIGHT:DARK CYCLES

Paramount to photoautotrophic fitness is the ability to maximize sunlight capture and utilization. In natural conditions, sunlight availability dynamically changes over the course of a diurnal (daily) day/night light:dark (LD) cycle. Gene expression, metabolism, and engineered chemical production are all profoundly influenced by diurnal LD cycles and the underlying biological entrainment to these cycles. A review of studies investigating cyanobacteria and algae behavior in diurnal LD cycles is provided in Table 1-2; each study is discussed further in the following sections.

Table 1-2. Review of algae and cyanobacteria studies in daily LD cycles. Light:dark cycles type indicated by "sqr" for square-wave (ON/OFF) or "sin" for sinusoidal-like (gradual increase/decrease). Max light intensity for sqrLD cycles is light applied for entire period. Max light intensity for sinLD cycles is light intensity reached at peak light. Contributions of this thesis' work are highlighted in blue. Circadian studies conducted in continuous light are not included in this review.

Author (Year)	Organism	L:D cycle (hr:hr)	Maximum light	Molecule/measurement (platform/method)	Ref.
SYSTEMS BIOLOGY & *'OMICS*					
This thesis	*S.* 6803	sin12:12	1,600 μE	Non-polar metabolites (LC-MS), Semi-polar metabolites (LC-MS), Polar metabolites (GC-MS), Insoluble polymers (GC-MS)	Chapter 4
Saha (2016)	*S.* 6803	sqr12:12	50 μE	Transcriptomics (microarray), glycogen, ATP, NADPH	[81]
Angermayr (2016)	*S.* 6803	sqr12:12†	500 μE‡	Excreted metabolites (HPLC), Primary metabolites (NMR), glycogen	[82]
Diamond (2015)	*S.* 7942	sqr12:12	150 μE	Primary metabolites (GC-MS), glycogen	[83]
Willamme (2014)	*C. reinhardtii*	sqr12:12	100 μE	Lipids (GC-MS), amino acids (GC-MS), chlorophyll, starch (specific assays)	[84]
Beck (2014)	*S.* 6803	sqr12:12	80 μE	Transcriptomics (microarray)	[85]
Waldbauer (2012)	*Prochlorococcus*	sin13:11	200 μE	Transcriptomics (RNA-Seq), Proteomics (LC-MS/MS)	[86]
Stockel (2011)	*Cyanothece 51142*	sqr12:12	50 μE	Proteomics (LC-MS/MS)	[87]
PHOTOPHYSIOLOGY & GENETIC PERTUBATIONS					
Cantrell (2017)	*C. reinhardtii*	sin12:12	2,134 μE	Photophysiology, TOC	[88]
Iijima (2015)	*S.* 6803	Light-dark transition	30-70 μE	Select primary metabolites (GC-MS, LC-MS/MS), glycogen	[89]
Hanai (2014)	*S.* 6803	sqr12:12	20 μE	Glycogen, select metabolites (CE-MS)	[90]
McEwen (2013)	*S.* 7942	sqr12:12	25 μE	Glucose and xylose (HPLC)	[91]
CHEMICAL PRODUCTION					
Kanno (2017)	*S.* 7942	sqr12:12	30 μE	2,3-butanediol production (GC-TOF-MS), Glycogen	[92]
Cheah (2015)	*S.* 6803	sqr12:12	200 μE	Free fatty acids (GC-FID)	[93]
Angermayr (2012)	*S.* 6803	sqr16:8	30 μE	Lactic acid (Megazyme assay)	[94]
Oncel (2011)	*C. reinhardtii*	sqr12:12§	70 μE	Hydrogen production (GC-Unspecified)	[95]
BIOMASS PRODUCTIVITY					
Liao (2014)	*C. pyrenoidosa*	sqrLD, 2-10 Hz	50 μE	Cell growth, photosynthetic efficiency	[96]
Tamburic (2012)	*C. reinhardtii*	sqr12:12	25 Wm^{-2}	Cell growth	[97]

Jacob-Lopes (2009)	*A. Nägeli*	sqr12:12[¥]	150 μE	Cell growth, elemental analysis (CHNS/O element analyzer)	[98]
Ogbonna (1998)	*E. gracilis*	sqr10:14	250 μE	Cell growth, elemental analysis (CHN elemental analyzer), glucose	[99]

[†]Nitrogen gas was bubbled during dark phase to maintain anoxic conditions
[‡]Light provided with blue- and red-LEDs.
[§]Photoperiods of 18:6, 14:10, 12:12 (h:h) and 45:15, 35:25, 30:30 (min:min) were investigated
[¥]Photoperiods of 0:24, 2:22, 4:20, 6:18, 8:16, 10:14, 12:12, 14:10, 16:8, 18:6, 20:4, 22:2 and 24:0 were investigated.

Cyanobacteria productivity in diurnal LD cycles

Biomass productivity

Algal and cyanobacterial biomass productivity in diurnal LD cycles is different from that observed in continuous light (CL). Tamburic and colleagues report decreased *Chlamydomonas reinhardtii* (*C. reinhardtii)* biomass productivity under 12:12 LD cycles as compared with CL [97]. Jacob-Lopes *et al.* (2009) systematically studied the effect of photoperiod in two-hour increment increases on the biomass productivity of cyanobacterium *Aphanothece microscopica Nägeli* (*A. Nägeli)* [98]. Interestingly, there was a linear decrease of biomass production with decreasing light period with the exception of 12:12 LD cycles, suggesting an adaptive or evolutionary advantage to growth in the near-natural 12:12 photoperiod.

A trend of decreasing biomass production with decreasing light availability may not be surprising given that light energy is often the limiting cellular resource. This trend is reversed when the LD cycle period is decreased to the order of seconds or minutes. Liao *et al.* (2014) report *Chlorella pyrenoidosa* (*C. pyrenoidosa*) biomass productivity increased under rapid LD cycles as compared with CL [96]. Ogbonna and colleagues engineered a cycling autotrophic/heterotrophic daily LD cycle *Euglena gracilis (E. gracilis*) algae cultivation system where an ethanol carbon source was added during the dark phase; they report increases in biomass and α-tocopherol productivity in this system as compared with CL or non-fed daily LD

cycles [99]. Inconsistencies across experiments could be attributed to differences in light intensity used for the light phase or even the physical arrangement of the light panels [95].

Engineered chemical productivity

Productivity of engineered strains significantly decreases in daily LD cycles in the few cases where performance in diurnal LD cycles has been tested. Oncel and Sukan investigated the effect of LD cycles with varying photoperiod on *Chlamydomonas reinhardtii* hydrogen production in sulfur deplete cultures: they found decreased production during the dark phases resulting in an overall decrease in production in 12:12 LD cycles [95]. Angermayr and colleagues investigated lactic acid production in *S.* 6803 in CL and 16-hour light:8-hour dark (16:8) LD cycles. They report decreased titers of lactic acid in the 16:8 LD strain as compared with the CL strain [94]. Cheah *et al.* (2015) engineered a free fatty acid (FFA) production strain with a simple acyl-ACP synthase knock-out (Δ*aas*) [93]. In CL, the Δ*aas* strain produced twice the FFA as wild-type. In 12:12 LD cycles, the increase in titer conferred by the acyl-ACP synthase knock-out vanished in 12:12 LD cycles rendering FFA production was reduced to that observed in wild-type.

Due to the decreased light availability in LD cycles as compared to CL, a decreased overall titer would be expected due to the decreased photons available to the organism. However, decreased FFA levels to that of wild-type are perplexing. This observation begs the question: why would the FFA titer advantage of Δ*aas* be lost in daily LD cycles? One hypothesis is that dramatic oscillations in NADPH and ATP availability under diurnal LD cycles limit constitutive production of FFAs. FFA production is very energy-intensive: one C18 FFA requires 16 NADPH and 8 ATP. Under this hypothesis, similar issues may be seen for other metabolites. To understand the phenomena underlying the burden experienced by constitutively engineered

production strains in diurnal LD cycles, an understanding of fundamental metabolism and cell biology in diurnal LD cycles is required.

Circadian rhythms

Circadian rhythms synchronize cell biology with the external environment, specifically the light environment on a ~24-hour cycle. Circadian rhythms must persist in CL, be re-set by light/dark cues, and compensate for temperature; thus, circadian studies are conducted in CL after a period of darkness. These rhythms have been discovered in many life forms including plants, bees, flies, mice, humans, and cyanobacteria. Evidence for circadian rhythms in *S.* 6803 was first published by Aoki, Kondo, and Ishiura. Using a LuxAB promoter-trap, they measured bioluminescence in CL after LD entrainment and found many oscillations occurring every ~24 hours [100,101]. In the decades to follow, extensive research has been conducted on cyanobacteria (mostly *S.* 7942) to elucidate the molecular mechanisms which confer circadian rhythms. As circadian rhythms are implicated in many areas of human health, this type of work has been a high scientific priority (see 2017 Nobel Prize in Physiology/Medicine for circadian rhythm work).

The core circadian pacemaker in cyanobacteria is comprised of a KaiABC protein post-translational oscillator (PTO). The proposed mechanism of action is as follows: KaiA phosphorylates KaiC at the expense of ATP; KaiB inhibits this activity; in the absence of KaiA phosphorylation activity, KaiC returns back to a dephosphorylated state. KaiA also binds oxidized quinones, forming an aggregate which is degraded. The clock is re-set by cellular redox status [102,103] and ATP/ADP ratio [104]. During the day, photosynthetic electron transfer confers a reduced cellular state and generates energy cofactors (ATP and NAD(P)H). During the

night, energy cofactors are consumed, and cellular redox markers are oxidized. Several PTO output factors mediate global circadian gene expression changes [105–108], though the details of these protein-DNA or protein-protein interactions remain elusive. Contemporary research seeks to further elucidate the mechanisms by which key PTO output factors—such as CikA, RpaA, and others—mediate global gene expression control.

Alternatively, the oscilloid model asserts that changes in chromosome topology drive gene expression changes via promoter accessibility not promoter sequence. Bacterial chromosomes are circular and therefore are subject to either positive or negative supercoiling. In *S.* 7942, knocking out enzymes key to supercoiling status (DNA topoisomerase and gyrase) abolishes circadian oscillations in gene expression [109]. Yet, there are no apparent chromosomal loci clustering patterns [110], and promoter mutations can alter the circadian phase of expression [18], confounding the oscilloid model and suggesting at least some role of sequence-specific circadian gene expression control.

Sequence-specific gene expression control could be due to trans-acting PTO proteins and/or RNAP sigma subunits. Trans-acting transcription factors bind to specific *cis*-acting "promoter" elements upstream of the TSS to recruit RNAP and initiate transcription. Additionally, the RNAP sigma subunits play a large role in transcription specificity. Based on sequence homology, *S.* 6803 has nine RNAP sigma factors, most of which are not functionally understood [111]. Nevertheless, these sigma factors have been divided into three main groups based on sequence homology: group 1, essential to cell growth (SigA); group 2 (SigB, SigC, SigD, and SigE); and group 3 (SigF, SigG, SigH, and SigI). SigD protein abundance increases significantly under high-light conditions and thus is of interest as a light-responsive gene regulator [112]. Evidence of light-activated [20] and dark-activated [9] sigma protein

accumulation further supports the notion that DNA promoter sequences themselves confer circadian entrainment. Genes/promoters are broadly classified into dusk-peaking (Class 1) and dawn-peaking (Class 2) based on circadian expression [18]. Genetic tools which capitalize on these native temporal gene expression control mechanisms would be immensely useful to metabolic and strain engineers. However, without an understanding of metabolism dynamics in diurnal LD cycles, generating informed hypotheses for strain design is challenging.

Despite a lack of understanding about the specific molecular mechanism which conveys circadian gene control, it is clear that the reach of circadian rhythms on global cellular behavior is broad. To-date, circadian regulation has been identified at nearly every layer of the central dogma (Figure 1-4). To re-cap the work presented above: Changes in chromosome topology may govern global changes in promoter accessibility based on chromosomal loci; Promoter recognition via Sigma factors, PTO factors, or transcription factors may confer sequence-specific temporal transcriptional activation; mRNA stability and therefore translation rate may be controlled by circadian expression of micro- or antisense- RNAs; Lastly, post-translational modifications—both to the core KaiABC pacemaker and other enzymes—may be mediated via circadian rhythms to temporally activate/deactivate enzymatic activities without degradation or resynthesize of peptides.

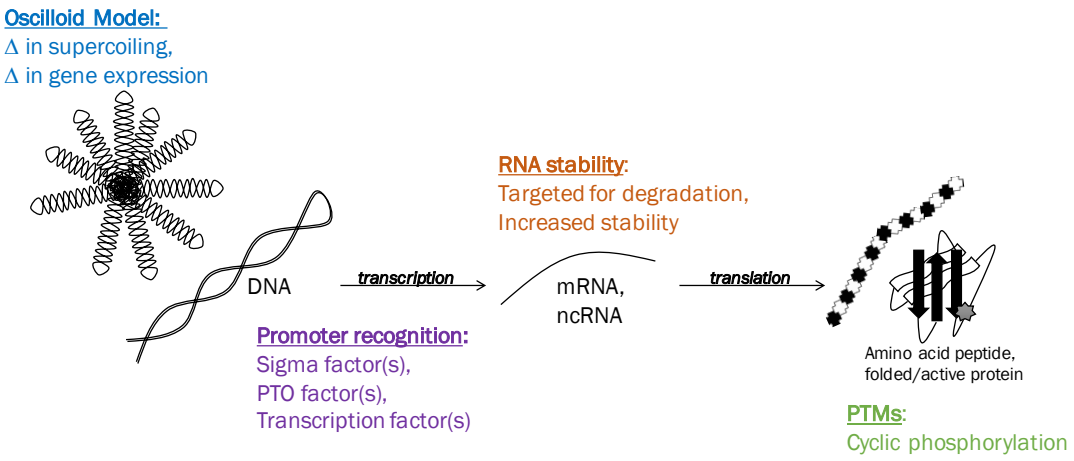

Figure 1-4. Aspects of gene expression regulation controlled via circadian rhythms and/or diurnal LD cycles.

Systems biology approaches in diurnal LD cycles

Studies of all collective constituents of a biological system are referred to as "*-omics"* approaches (*i.e.* genomics, transcriptomics, metabolomics, proteomics, etc.). The integration of *'omics* experiments with disciplines such as computer science, engineering, and bioinformatics is termed "systems biology". Systems biology seeks to understand a system under the premise that system is greater than the sum of all the individual parts. Systems biology and *'omics* approaches have been transformational in nearly every scientific field, including cyanobacteria in diurnal LD cycles.

Transcriptomics studies support wide-spread entrainment to daily LD cycles. In 12:12 LD cycles, functionally-similar genes cluster together with similar oscillation patterns [81,85]. For example, photosynthesis genes peak during the day while ribosomal RNAs peak during the night [85]. Hundreds of antisense RNAs strongly oscillate too, suggesting a layer of gene expression regulation which is widely overlooked [85]. Saha *et al.* (2016) also applied transcriptomics to *S.* 6803 and report that 40% of gene oscillate significantly in 12:12 LD cycles [81]. In addition, they report glycogen, ATP, and NADPH dynamics which provides increased understanding of

dynamic metabolism in LD cycles. However, the correlation between transcript and protein oscillations is complex. Waldbauer *et al.* (2012) studied cyanobacterium *Prochloroccus* and found that while transcript abundance changes significantly over the course of sinLD cycles, the oscillations are dampened at the protein level [86]. Still more elusive are metabolite changes over the course of the day.

Glycogen, ATP, and NADPH abundances oscillate with daily LD cycles (Figure 1-5). Diamond *et al.* (2015) found that *S.* 7942 glycogen pools increase gradually during the day and decrease gradually during the night. ATP dynamics follow the same pattern. NADPH, however, spikes abruptly at the onset of light and remain low until the next morning. These trends were corroborated in a *S.* 6803 study in 2016 [81]. Saha *et al.* demonstrated that glycogen, total ATP, and NADPH/$NADP^+$ in *S.* 6803 are synchronized with diurnal LD cycles, broadly suggesting wide-spread metabolic entrainment. Still, the metabolites studied here were restricted and do not provide details on a global, system response to dynamic light conditions.

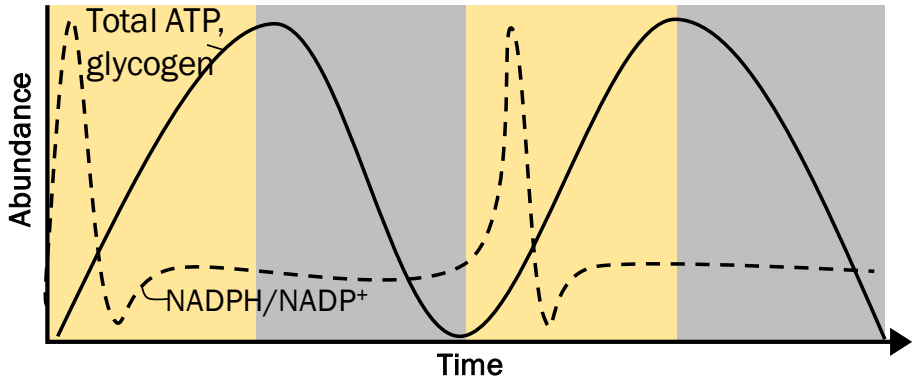

Figure 1-5. Diurnal 12:12 LD cycle energy dynamics. Figure constructed by A. Werner based on data from Diamond *et al.* (2015) [83] and Saha *et al.* (2016) [81].

The state-of-the-art metabolomics studies have been conducted under on/off low-light conditions with a narrow scope of metabolites detected (Table 1-2). Most recently, Kanno *et al.* (2017) engineered 2,3-butanediol (23BD) production and measured productivity in both CL and

12:12 LD cycles. Using GC-MS, they identified several energetic bottlenecks during the dark phase which informed genetic modifications for improved strain design. The breadth of metabolites analyzed is narrow; still, the high-profile nature of this work should raise widespread awareness to the importance of considering daily LD cycles in engineering. Angermayr *et al.* (2016) examined a wider array of metabolites under LD cycles, but supplied N_2 gas during the night. In these unrealistic anoxic conditions, the glycogen dynamics were aberrant and therefore the relevance of these results in realistic conditions is unclear. Furthermore, this study utilized blue- and red-LED panels, which may have resulted in phycobilisome rearrangements at a minimum [52]. Diamond *et al.* (2015) used GC-MS metabolomics to identify metabolite differences between wild-type and circadian clock knock-out (Δ*kaiC* and Δ*rpaA*) *S.* 7942 strains in CL and 12:12 LD cycles. Hanai *et al.* (2014) also studied genetic mutants in daily LD cycles. These studies were successful at uncovering aspects of genetic elements, but do not address system changes over the course of a diurnal LD cycles.

Metabolism under realistic day/night cycles—sinusoidal increases in light intensity peaking at super-saturating high-light—is almost completely unexplored. Photosynthetic fitness depends on the efficient and timely capture and utilization of sunlight, thus the study of temporal metabolite changes under diurnal LD cycles is critical to understanding the regulation of photosynthetic metabolism in natural environments. This represents a significant knowledge gap, with implications spanning the fields of agriculture, photobiology, and biological engineering. This thesis work addresses the knowledge gap with a time-course multi-platform mass spectrometry metabolomics experiment on *S.* 6803 cultivated in sinLD cycles.

METABOLOMICS

MS technology platforms

The goal of this section is to provide a brief introduction to the mass spectrometry (MS) technologies and methods utilized in this work (Table 1-3). Many MS methods and configurations exist, but a general workflow is illustrated in Figure 1-6. An inlet system is utilized to introduce the sample to mass spectrometer without loss of vacuum. Chromatography is commonly employed at the inlet to provide separation of compounds prior to MS analysis. Molecular separation reduces ion suppression as well as competition, *i.e.* more compounds in the sample become ionized and therefore available for detection.

Table 1-3. Summary of MS technology utilized in this work.

Target metabolite class (metabolite examples)	Inlet system (column type)	Ion source	Mass analyzer/ Detector
Non-polar in organic phase (lipids, pigments)	UPLC (Phenyl Hexyl)	ESI	idMS/MS-TOF (non-targeted)
Polar in aqueous phase (pyruvate, acetyl-CoA)	UPLC (HILIC)	ESI	SRM, TQS (targeted)
Polar in aqueous phase (organic acids, small sugars)	GC	EI	Single Q
Pellet contents (amino acids, nucleic acids, polysaccharides)	GC	EI	Single Q

GC: gas chromatography; LC: liquid chromatography; UPLC: ultra-performance LC; EI: electron impact; ESI: electrospray ionization; idMS/MS: indiscriminant MS/MS; SRM: selected reaction monitoring; TQS: triple quadrupole mass spectrometer; TOF: time of flight; Q: quadrupole.

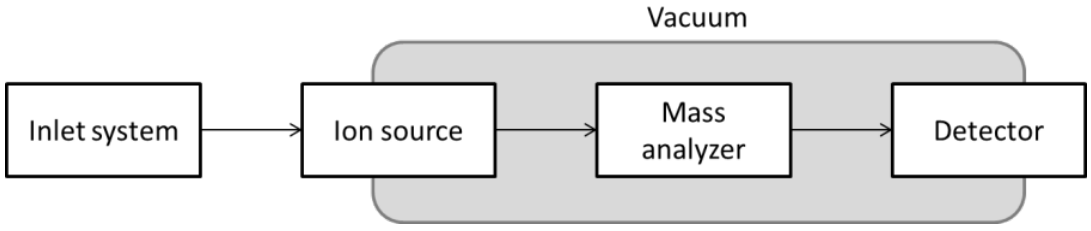

Figure 1-6. Generalized mass spectrometry instrument configuration. Figure made by author.

Gas chromatography (GC) separates compounds based on volatility commonly using a helium or hydrogen mobile phase and a poly(dimethylsiloxane) stationary phase. A temperature

gradient provides gradual volatilization of compounds and subsequent separation down a chromatographic column. Semi- and non-volatile compounds can be identified by GC-MS when paired with a derivatization step to enable volitization. Derivatization consists of removing N-H and O-H groups in a methoximation step, then adding silyl groups in a sialylation step. The added group step varies depending on the application. Due to the ability to volatize many compounds via derivatization, GC-MS is now commonly employed for the identification of sugars, amines, amino acids, organic acids, sterols, free fatty acids, and fatty acid methyl esters, most of which must be derivatized prior to analysis.

Liquid chromatography (LC) is another chromatography separation method utilized at the inlet system of mass spectrometers. LC separates molecules based on polarity using a solvent liquid phase and a column resin stationary phase. LC is more versatile than GC in the compounds able to be detected due to the absence of a volatility requirement. Furthermore, many different column chemistries and mobile phases can be combined to achieve excellent separation of target compounds. Ultra-performance LC (UPLC) is a Waters® variant which utilizes low flow rates and high pressures to enable low sample volumes and good separation. Hydrophilic interaction LC (HILIC) utilizes a hydrophilic stationary phase and organic to polar liquid phase. HILIC enables detection of many semi-polar compounds but is highly pH dependent and generally experiences poorer separation than other methods.

The ion source converts molecules to ions. Many versions exist, including election impact (EI) and electrospray ionization (ESI). EI is used for gas samples and ionizes molecules using electron beams generated from a rhenium or tungsten filament. EI ionization results in strong (“hard”) and predictable fragmentation. ESI is used for liquid samples and ionized

molecules by forcing flow through small steel capillary at high voltage. Under ideal conditions, ESI does not cause fragmentation but only neutral losses (*e.g.* $-H_2O$) and adducts (*e.g.* +Na).

Quadrupole mass analyzers (QMS or Q) utilize rapidly oscillating magnets which subjects molecules to resonance proportional to the mass/charge ratio (*m/z)*. This enables separation of molecules based on *m/z* and/or selection of specific masses. Quadrupoles are commonly paired to improve sensitivity and selectivity. Triple quadrupole mass spectrometers (TQS or QqQ) use a quadrupole to select ions of a particular mass, a second quadrupole to further fragment the ion, and a third to further select ion mass. Targeted analysis via selected reaction monitoring (SRM) enables differentiation between precursor ions of the same mass given that fragment ions have different masses. Given the retention time of the target molecule, the first quadrupole (Q1) is set to $m/z=m_{precursor}$ while the third quadrupole (Q3) scans $m/z=m_{product}$. This method has far superior sensitivity and can be utilized for quantitative metabolite detection [113]. Indiscriminate MS/MS (idMS/MS, MS^E) is an approach to generating molecular fragment information using a TOF tandem quadrupole instrument [114]. In this method, the first quadrupole is not used for mass selection while the collision chamber Q cycles rapidly between low- and high-collision energy. In this way, both precursor and fragment ion information are collected in a single analytical run.

Time of flight (TOF) detectors apply an electric field to accelerate ions at a velocity proportional to the *m/z* and measure the drift time through a vacuum. A so-called pusher initiates ion entry to the linear flight tube and a detector records the electrical signal from the ion. Using the precise time measurement, the *m/z* can be determined. TOF is considered a high-resolution detector, measuring mass accurately to the third- or fourth- decimal place. Even higher resolution can be achieved by pairing a mass analyzer with the detector (Q-TOF or HRMS).

Multiplicity of comparisons

High-throughput methodologies, such as 'omics techniques, test multiple hypotheses simultaneously. In this study, we are interested in testing the null hypothesis (H_0) that a regression model does *not* describe the metabolite behavior as a function of time (the situation we are *not* interested in). The alternative hypothesis (H_a) is that a regression model does describe metabolite behavior as a function of time, meaning that the metabolite is changing as a function of time (the situation we *are* interested in). (Regression analysis is discussed in Appendix C).

The test statistics employed can be assessed with a *p*-value. Formally, the *p*-value is the "probability of obtaining a value of the test statistic that is as likely or more like to reject H_0 as the observed value of the test statistic" (computed assuming H_0 is true) [115]. In this study, a *p*-value of 0.05 means there is a 5% chance that the metabolite variation observed is due to chance. A type I error is committed when a true H_0 is rejected; a type II error is committed when a false H_0 failed to be rejected [115]. The probabilities of type I and type II errors are inversely related, described by alpha and beta, respectively. When multiple comparisons are being made, the rate of a type I error—the case where we reject H_0 when it is true, also called a false-positive—becomes intolerably large [116]. If we run regression analysis on 1,000 metabolites and utilize a confidence threshold (α) of 0.05, a commonly used if not arbitrary value, we are accepting that we may incorrectly conclude at least 50 metabolites are changing when in truth they are not. Therefore, using a multiple testing correction method is essential for the analysis of large datasets such as those generated by high-throughput methods like mass spectrometry.

The tolerable risks of type I and type II errors should depend on the research goals and the likely follow-up studies. In some cases, controlling type I error is of utmost importance (*e.g.* testing the effects of a drug which will go to clinical trials) whereas in other cases controlling

type II errors is more suitable (*e.g.* identifying interesting metabolite changes in a biological process of interest). Thus, the choice of multiple testing corrections ranges across studies.

Multiple hypothesis correction methods

While many methods of multiple testing correction methods exist, the Bonferroni and Benjamini-Hochberg procedures are the most common in the biological sciences [117]. *Q*-values are an alternative to *p*-values which take false discovery rate (FDR) into consideration [118]. Utilization of both *p*-values and *q*-values can be valuable to a given study, such as this one, to assess different aspects or different platforms utilized in the analysis [119].

The Bonferroni correction controls the family wise error rate (FWER), thus controlling probability of *one* false positive among all of the tests (in the whole "family"). In this case, the corrected *p*-value is the original *p*-value divided by the number of tests performed, However, there is much evidence to suggest that FWER correction procedures, such as the Bonferroni procedure, are unsuitable for biological data for the following reasons: (i) the interpretation depends on the number of comparisons, (ii) the likelihood of type II errors increases, and (iii) the correction is too strict when multiple tests are not completely independent, as with many high-throughput experiments [120–122]. Stated simply, Bonferroni corrections are strict and therefore are applied when the risks of type I errors are prioritized above risks of type II errors.

The Benjamini-Hochberg procedure, on the other hand, controls the false discovery rate (FDR), thus controlling the *proportion* of discoveries which are false positives [123]. Because it is a proportion, the distribution of *p*-values remains constant across different numbers of tests. Benjamini-Hochberg ("BH") corrected *p*-values are calculated by ordering the original *p*-values in ascending order, determining the largest *p*-value which is less than (test rank/number of tests)*FDR% [117], where the tolerable FDR is user-defined. Again, the FDR choice should be

determined by the costs associated with both a type I and type II error. An advantage of BH corrections is the decreased sensitivity to the constitution of a family as compared with Bonferroni; BH-correction is likely more suitable for simultaneously assessing metabolites across multiple platforms. As with Bonferroni, however, BH assumes that all tests are independent and therefore is often too-strict for biological applications (*e.g.* when multiple metabolites being independently tested are correlated) [120].

Q-values are an alternative to *p*-value correction methods, appropriate for cases where the false discovery rate (FDR) is of interest and increasingly applied to *'omics* work. In contrast to *p*-values, *q*-values say something about the features actually reported as significant given that multiple tests were conducted simultaneously: *q*-values report the probability that a significant finding will turn out to be a false discovery if further investigated. One of the most concise differentiations between false positive rate and FDR is provided by Storey and Tibshirani [119]:

> *"The false positive rate and FDR are often mistakenly equated, but their difference is actually very important. Given a rule for calling features significant, the false positive rate is the rate that truly null features are called significant. The FDR is the rate that significant features are truly null. For example, a false positive rate of 5% means that on average 5% of the truly null features in the study will be called significant. A FDR of 5% means that among all features called significant, 5% of these are truly null on average."*

Stated in words, the *q*-value is "the minimum FDR incurred when calling a test significant" [119,124]. (Mathematical definition of FDR and *Q*-values is presented in Appendix C). Determination of the *q*-value cut-off can be achieved by assessment of several quantities, including: (i) p-value versus q-value, (ii) number of significant features for a range of *q*-value cut-offs, and (iii) number of expected false positives for a range of significant features.

CONTRIBUTIONS OF THESIS

The overarching goal of this thesis was to improve the industrial feasibility of cyanobacteria bio-production platforms in the industrially relevant condition of daily LD cycles. Each chapter of my thesis makes a unique contribution toward this goal.

Chapter Two presents materials and methods for the cultivation and engineering of S. 6803, a model cyanobacterium, and E. coli. Methods for analyzing the expression of recombinant gene transcripts are also presented. Novel materials and methods which enable a comprehensive time-course metabolomics study of liquid cyanobacterial cultures subjected to diurnal sinusoidal LD (sinLD) cycles peaking at supra-high light intensities (1,600 μE) were developed. Specifically, a light-emitting diode (LED) photobioreactor (LED PBR) is engineered, cultivation of *S.* 6803 under sinLD cycles in the LED PBR is optimized, and a multi-platform MS workflow is developed.

This chapter advances the state-of-the-art for (i) cultivation of phototrophic bacterial cultures, and (ii) metabolomics on bacterial cultures. To our knowledge, this is the first bioreactor which supplies dynamic super-high-light to relatively large volume (< 2 liter) cultures. Further, we develop methods which enable characterization of multiple classes of compounds from both the soluble and insoluble fractions. These contribute to the overall goal of improved understanding of photosynthetic metabolism in industrially-relevant environmental conditions.

Chapter Three presents novel genetic engineering tools which provide light-entrained gene activation in *S.* 6803 cultivated in 12:12 LD cycles. Specifically, the genetic promoters provide repeatable gene induction from the *slr0168* chromosomal neutral site at the onset of light over consecutive days of growth. Three promoters exhibit transcriptional induction corresponding to protein activity induction; one promoter does not which suggests post-transcriptional regulation. These genetic engineering tools provide the first tools for diurnal LD

cycles characterized at the transcript and protein levels. Further work applying these tools toward chemical production will test the hypothesis that heterologous protein expression restriction to periods of light increases engineered yield by decreasing metabolic burden under the environmental condition of daily LD cycles.

This work fills a critical gap in genetic engineering tools suitable for engineering cyanobacteria strains for outdoor cultivation systems. These tools have the potential to impact the industrial feasibility of biofuel production strains by enabling metabolic engineering efforts targeting temporal control of metabolic changes under diurnal LD cycles. Further, the data generated on the entire promoter:probe strains have potential to be utilized by future research efforts to understand the underlying diurnal LD cycle gene expression regulatory mechanisms.

Chapter Four presents a semi-comprehensive profile of *S.* 6803 metabolite abundance dynamics across a diurnal sinLD cycle, utilizing the LED PBR cultivation and MS workflow presented in Chapter 2. Temporal dynamics of amino acids, nucleic acids, insoluble carbohydrates, and central carbon metabolites are discussed. Hypotheses about cellular objectives and global metabolism are presented based on the observed accumulation dynamics. MS data processing steps, overview of data variation, and analyses strategies and are detailed in Appendix C.

To our knowledge, this work is the first to report cyanobacterial metabolism under realistic day/night light cycles. Further, we report comprehensive metabolite dynamics with a high degree of temporal frequency. This work fills a critical knowledge gap in understanding the dynamics of phototrophic metabolism in natural light conditions. Extensive work has been done to describe the light acclimation processes, but very little is known about if and how metabolite pools change over the course of the day in photosynthetic organisms. Together this body of work

vastly improves our understanding of photoautotrophic metabolism in realistic light conditions. These findings have broad implications in the fields of agriculture, photobiology, and phototroph engineering. Integration of the metabolomics data into GSMs and DFBA has potential to improve the identification of engineering targets for production under natural light conditions, which could ultimately result in improved feasibility of industrial cyanobacteria biofuel production strains. Data to support a dynamic biomass equation is presented and summarized in Appendix E.

Chapter Five summarizes the main findings and contributions of this thesis and presents future research directions in this area.

CHAPTER 2 – MATERIALS AND METHODS[1]

SUMMARY

This chapter details the materials, methods, and engineering developments utilized in Chapters 3 and 4. Bacterial cultivation, genetic engineering, and recombinant gene expression assessment are described in the first three sections. Next, materials and methods for the cultivation of *Synechocystis* sp. PCC 6803 (*S.* 6803) under sinusoidal LD (sinLD) cycles are presented. A light-emitting diode (LED) photobioreactor (PBR) which supplies realistic day/night light profiles is engineered and constructed. The PBR was built with dual 4000K white LED arrays which provide 1,600 μmol photons $m^{-2}s^{-1}$ at the peak of the light phase. In-house CO_2 mixing enables a range of carbon delivery percentages and rates. A water batch equipped with a heat exchanger enables temperature control to modular reactor configurations. Custom flat-panel circular glass reactors equipped with custom aerators achieves rapid mixing and high light delivery to relatively dense cultures. Methods for achieving consistent growth of *S.* 6803 under sinLD are provided. A time-course multi-platform mass spectrometry metabolomics workflow was developed for the comprehensive characterization of metabolites at 26 time-points over the course of a 24-hour sinLD cycle. Cells are rapidly quenched in ice-cold phosphate buffered saline and separated from media via centrifugation. Metabolites are extracted from cells with a biphasic MTBE:MeOH:H_2O workflow. Non-polar compounds in the organic phase are detected via non-targeted reverse phase LC-MS. Polar and semi-polar compounds in the aqueous phase are detected via SRM targeted HILIC-TQS and non-targeted GC-MS. Insoluble polymers

[1] A portion of this chapter has been submitted for publication as: Genetic Engineering of Cyanobacteria: Design, implementation, and characterization of recombinant *Synechocystis* sp. PCC 6803.

contained in the pellet post-biphasic extraction are hydrolyzed and the amino acid, nucleic acid, and polysaccharide content is detected via GC-MS.

CULTIVATION AND PRESERVATION OF BACTERIAL STRAINS

E. coli strains and media

E. coli DH5α was used for cloning and was cultivated in Luria broth (LB) at 37°C shaking at 225 rpm. Chemically competent *E. coli* was prepared by inoculating *E. coli* into 5 mL LB and cultivating at 37 degrees C, 220 rpm for ~16 hours. 100 uL of the overnight culture was transferred into two 250 mL flasks with 50 mL LB and cultivated at 37 degrees C, 220 rpm until OD_{600}=0.4-0.6. 22 mL of this culture was aliquoted into cold 50 mL centrifuge tubes and incubated on ice for 30 min, mixing periodically. Tubes were centrifuged at 3,000 rpm for 5 min. at 4 degrees C. The supernatant was discarded and the cell pellet was resuspended in 10 mL ice-cold 100 mM $CaCl_2$ and incubated for 2 hours on ice. The tubes were again centrifuged at 3,000 rpm for 5 min. at 4 degrees C. The supernatant was discarded and the cell pellet was resuspended in 1 mL ice-cold 100 mM $CaCl_2$ with 15% glycerol and incubated on ice ~16 hours. 100 uL aliquots were flash-frozen in liquid nitrogen and stored at -80C until transformation.

Table 2-1. BG-11 media components.

Part Name	***Chemical Formula and Name***	***Amount in 200 mL***
100X Comp. A	$NaNO_3$, Sodium nitrate	30.00 g
	$MgSO_4$-$7H_2O$, Magnesium sulfate heptahydrate	1.498 g
	$CaCl_2$-$2H_2O$, Calcium chloride dehydrate	0.720 g
	$C_6H_8O_7$, Citric acid	0.120 g
	NH_3FeCitrate, Ammonium ferric citrate	0.120 g
	EDTA IDRANAL	0.020 g
600X Comp. B	K_2HPO_4-$3H_2O$, Potassium phosphate trihydrate	4.800 g
100X Comp. C	Na_2CO_3, Sodium carbonate	0.400 g

1000X Trace Metals	H_3BO_3, Boric acid	0.572 g
	$MnCl_2$-$4H_2O$, Manganese II chloride tetrahydrate	0.362 g
	$ZnSO_4$-$7H_2O$, Zinc sulfate heptahydrate	0.044 g
	Na_2MoO_4-$2H_2O$, Sodium molybdate dehydrate	0.078 g
	$CuSO_4$-$5H_2O$, Copper II sulfate pentahydrate	0.016 g
	$Co(NO_3)_2$-$6H_2O$, Cobalt II nitrate hexahydrate	0.010 g

Table 2-2. BG-11 media additions.

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

[a] *N*-Tris(hydroxymethyl)methyl-2-aminoethanesulfonic acid
[b] Dimethyl sulfoxide

Freezer stocks

Plasmids were stored in molecular grade water at -20 degrees C. Additionally, engineered *E. coli* and *S.* 6803 cultures were stored at -80 degrees C. To store *E. coli* strains, a glycerol stock was made: the *E. coli* stain was inoculated into 2 mL of LB and grown at 37 degrees C, 225 rpm for ~16 hours. 600 uL of the culture was mixed with 600 uL 50% glycerol in a cryogenic tube and stored at -80 degrees C. To store *S.* 6803 strains, a DMSO stock was made: the *S.* 6803 strain

was inoculated into 25 mL of BG-11 and grown to an OD_{730} ~0.6. Cells were pelleted at 2,500*g* for 5 minutes at room temperature, the supernatant was discarded, and the pellet was resuspended in 2.5 mL of fresh BG-11. 275 uL of 50% DMSO was added and mixed well. Two-hundred and fifty microliters of the cell-DMSO suspension was aliquoted into 1.7 mL microcentrifuge tubes and stored at -80 degrees C.

GENETIC ENGINEERING OF *SYNECHOCYSTSIS* SP. PCC6803

Plasmid design

Plasmids were designed in key steps: (1) obtain sequence of interest, (2) select expression control elements, (3) select cloning vector and cloning method, and (4) design primers. A brief overview of the actions taken for each step is described.

S. 6803 sequences can be accessed at Cyanobase: http://genome.microbedb.jp/cyanobase/Synechocystis. Gene sequences from many other organisms can be obtained at NCBI GenBank: https://www.ncbi.nlm.nih.gov/genbank/. The reader is referred to Eaton-Rye (2011) for detailed instruction on retrieving sequences from Cyanobase [125].

Promoters and ribosome binding sites are two critical genetic control elements to consider for the genetic engineering of *S.* 6803. Several native *S.* 6803 promoters, as well as modified *E. coli* promoters that perform in *S.* 6803, have been previously characterized [32,35,37,126]. The RBS Calculator is a valuable tool for assessing native RBSs and designing sequences for strong translation [22]. Codon optimization is frequently employed to optimize heterologous gene expression, especially with genes derived from eukaryotes. Commercial gene synthesis companies typically offer codon optimization free of charge when synthetic genes are

ordered. However, codon optimization is not fully understood, and the wide variety of mechanisms by which codon usage impacts gene expression introduces a degree of uncertainty to optimization. A good review of codon usage in this context has been provided by Quax and colleagues [127].

Vector copy number, selection marker, multiple cloning site (MCS), and homologous regions are critical considerations in vector selection for the genetic engineering of *S.* 6803. For example, pIGA4 is a plasmid used in our lab to test promoter strengths. It uses the pUC origin of replication to achieve a high copy number, contains the ampicillin antibiotic resistance marker, and homologous regions (HRs) for the transformation into the *slr0168* neutral site [32]. Between the homologous regions, the plasmid contains the gene sequence for a green fluorescent protein variant mutated to degrade more quickly than the wild-type, facilitating measurement of rapid changes in expression [32]. The plasmid backbone has been mutated to remove an NdeI cut site from the backbone so that KpnI and NdeI cut sites can be used to clone in and test different promoters. If a vector is used that does not have HRs and the objective is to perform homologous recombination, HRs need to be introduced flanking the gene of interest (or MCS) and selection marker.

Cloning method is tied closely with vector selection, as the viability of every option must be assessed by examining the cloning vector. Many options exist, including “cut-and-paste” using restriction enzymes and ligase, GoldenGate [128], ligase cycling [129], Gibson assembly [130] and others. In the simplest cases, a gene can be inserted into the MCS of a vector using a cut-and-paste approach.

In this study, primers were designed using IDT Oligo analyzer (https://www.idtdna.com/calc/analyzer), though many other primer design software are available.

This author utilized the following design criteria: 55-60 Tm, 40-60% GC content, and 15-30 bp in length and synthesized by IDT. Other considerations include primer termination in a G or C at the 3' end and minimal predicted primer dimers and secondary structure. Additionally, the forward and reverse primers melting temperatures should be within 5ºC of each other. If using cut-and-paste into the MCS, add restriction endonuclease recognition sequences to 5' end of both primers. In this case, four bases should be added on the far 5' end to facilitate enzyme binding.

Plasmid construction

Plasmids were constructed using polymerase chain reaction (PCR) amplification of DNA with New England Biolabs (NEB) Taq DNA Polymerase (NEB M0273S) or Phusion High-Fidelity DNA Polymerase (NEB M0530S). PCR reactions were set-up as described in NEB product protocols (https://www.neb.com/protocols/0001/01/01/taq-dna-polymerase-with-standard-taq-buffer-m0273, https://www.neb.com/protocols/0001/01/01/pcr-protocol-m0530), always including the optional DMSO component. PCR reactions were run using 1 minute per kilobase extension time for Taq and 30 second per kilobase extension time for Phusion. PCR products were checked with a 1% agarose gel visualized with SYBR Green. PCR products and plasmids were isolated using PureLink™ PCR Purification Kit (Thermo K310001) and PureLink™ Quick Plasmid Miniprep Kit (Thermo K210010), respectively.

Routine cut-and-paste cloning was employed as follows. Restriction enzyme digests were performed with NEB enzymes, unless noted otherwise. Restriction enzyme digest reactions were set-up as specified by NEB for the enzyme of interest (https://www.neb.com/products/restriction-endonucleases). Digested DNA was treated with the PureLink™ PCR Purification kit. Digested plasmids were treated with Alkaline Phosphatase (CIP, NEB M0290S) as specified on the NEB

protocol (https://www.neb.com/protocols/0001/01/01/protocol-for-dephosphorylating-with-cip). De-phosphorylated backbones were purified with PureLink™ PCR Purification kit. Ligation with NEB T4 Ligase (NEB M0202S) was specified on the NEB website (https://www.neb.com/protocols/0001/01/01/dna-ligation-with-t4-dna-ligase-m0202).

E. coli transformation

Plasmids were transformed into *E. coli* DH5alpha (*E. coli*). *E. coli* transformation by heat shock was completed as follows. 100 uL aliquots of chemically-competent cells were thawed on ice and 1-5 uL of the ligation product was added and incubated on ice for 30 min, flicking the tube ~3 times during this incubation period. A heat shock was applied by submerging the tube in a 37 degree C water bath for 45 seconds. Cells were immediately placed on ice for 2 minutes. 900 uL of LB was added to each tube and incubated at 37 degrees C, 220 rpm for 1 hour. 10-200 uL of the transformation product was spread onto LB plates with the appropriate selection marker and incubated at 37 degrees C for ~16 hours. Colonies were re-streaked and screened via colony PCR. Clones were confirmed with DNA sequencing.

S. 6803 transformation

S. 6803 is naturally competent (*i.e.* naturally uptakes DNA) due to the presence of a type IV-like pilus structure [131] and integrates DNA with regions of homology into the chromosome via homologous recombination. *S.* 6803 transformation efficiency is influenced by variables such as homologous region length and the time of incubation with plasmid DNA on transformation efficiency [132] and premethylation of the plasmid DNA [133]. The protocol outlined here,

while not systematically examined to determine transformation efficiency, resulted in a sufficient number of positive recombinants for successful engineering.

S. 6803 cells were grown to mid-log phase and 10 mL was aliquoted into a 50 mL conical tube on ice. The cell suspension was centrifuged at 4000*g* for 6 minutes at 4 ºC. The supernatant was discarded, the cell pellet was resuspended in 100 µL of fresh BG-11, and the suspension was transferred to a 14 mL round-bottom Falcon tube. A no-DNA control tube was also prepared. 1 µg of the target plasmid was added to the suspension and incubated for 5 hours under low light (20-100 µmol photons $m^{-2}s^{-1}$) at 30°C. Transformation tubes were gently agitated after ~2.5 hours. 100 µL of the transformation mixture was plated on BG-11 agar plates without antibiotics, sealed with surgical tape, and incubated under low light at 30°C for 48 hours (this period is necessary for recovery). After 48 hours, a triple-streak procedure onto BG-11 plates with the selection pressure was performed to achieve single colonies in 5-14 days.

Colony PCR with primers binding outside the homologous region of interest was conducted to check for correct transformants. To achieve full segregation, single colonies were re-streaked on a fresh BG-11 plate with antibiotics and place in 30°C lighted incubator. This process was repeated once a week for 3-4 weeks, or until full segregation is achieved. Segregated mutants were confirmed with DNA sequencing.

RECOMBINANT TRANSCRIPT ANALYSIS VIA RT-QPCR

RNA extraction

RNA was extracted from 15 mL of cultures at OD_{730nm} ~5 harvested by centrifugation at 12,000*g* for 5 min at 4 degrees C. Supernatant was discarded and the pellet was flash-frozen in liquid nitrogen prior to storage at -80 degrees C. RNA was extracted by a modified PGTX

method based on the protocol described by Pinto *et al.* [134]. PGTX was formulated by adding 39.6 g phenol, 6.9 mL glycerol, 0.1 g 8-hydroxyquinoline, 0.58 g EDTA, 0.8 g sodium acetate, 9.5 g guanidine thiocyanate, 4.6 g guanidine hydrochloride, and 2 mL Triton X-100 to a total volume of 100 mL in water and pH adjusted to 4.2. 1 mL of PGTX was added to pelleted cells, transferred to a microcentrifuge tube, and vortexed at speed 6 for 15 seconds. The cells suspension was resuspended well with a pipette, incubated at 95 degrees C for 2 minutes, vortexed at speed 6 for 15 seconds, incubated at 95 degrees C for 5 minutes, and placed on ice for 5 minutes. 100 uL of 1-bromo-3-chloropropane was added to the suspension, vortexed at speed 10 for 10 seconds, and incubated at room temperature for 15 minutes. Tubes were centrifuged at 14,000*g* for 15 minutes at 4 degrees C. The upper aqueous phase was transferred to a fresh tube and equal part isopropanol was added; tubes were inverted 5 times to gently mix and incubated at room temperature for 10 minutes. Samples were centrifuged at 12,000*g* for 10 minutes at 4 degrees C. The supernatant was discarded, 1 mL of 75% ethanol in water was added, vortexed for 20 seconds, and centrifuged at 7,000*g* for 5 minutes at 4 degrees C. The supernatant was discarded and the ethanol wash was repeated. The pellet was air-dried in a laminar flow hood for ~1 hour and resuspended in 50 uL molecular grade water. The sample was incubated at 60 degrees C for 10 minutes, vortexed again, RNA concentration was measured on a NanoDrop, and samples were stored at -80 degrees C and immediately processed by DNase treatment.

DNA removal

DNA was removed by TURBO DNA-Free ™ (Thermo M1907) DNase treatment, as specified by the kit. Briefly, 0.1 volumes 10X TURBO DNase buffer was added to 1 µL RNA

sample, mixed gently, and incubated at 37 degrees C for 30 minutes. DNase was deactivated by adding 2 μL of DNase deactivation reagent and incubated at room temperature for 5 minutes, mixing by flicking every ~1 minute. Samples were centrifuged at 10,000*g* for 1.5 minutes, transferred to a fresh tube, and stored at -80 degrees C for <24 hours.

Attention to this step is critical; poor attention to the removal of DNA will prevent quality transcript analysis in downstream steps. The working volumes in the TURBE DNA-Free™ kit are very low, so sufficient mixing requires a good deal of precision. Without very good mixing, DNA will not be fully removed and will contaminate the RNA, thus preventing good RT-qPCR results.

RNA purification

RNA was purified by the Phenol-Chloroform method, as follows. An equal volume of 25:24:1 phenol:chloroform:isoamyl alcohol pH 5.2 (Fisher BP1753I-100) was added to the sample, vortexed for 20 seconds, and centrifuged at top speed for 1 minute at room temperature. The supernatant was transferred to a fresh tube; filter tips were used from here forward. An equal volume of chloroform was added, vortexed for 20 seconds, centrifuged at top speed for 1 minute at room temperature, and the supernatant was transferred to a fresh tube. A 1/10 volume of cold 3 M NaCl and 2x volume of cold 200 proof ethanol was added, vortexed 10 seconds, placed at -80 degrees C for 20 minutes, and centrifuged for 20 minutes at top speed at 4 degrees C. The supernatant was discarded. The pellet was washed by adding 1 mL of cold 75% ethanol, vortexing for 20 seconds, and centrifuging at top speed for 3 minutes at 4 degrees C. The supernatant was discarded and the pellet was air-dried for ~1 hour. The remaining RNA was

resuspended in 30-40 uL of cold molecular grade water and processed by reverse transcriptase within 24 hours.

cDNA synthesis, DNA contamination check, and RT-qPCR

Complimentary DNA (cDNA) was made using GoScript™ Reverse Transcriptase System (Promega A5003) as specified by the Promega protocol. A no-reverse transcriptase control was also run and used to check for DNA contamination. A check for DNA contamination and primer specificity was performed as follows (Figure 2-1). All PCRs were carried out with the primer designed and selected for use in the qPCR. A PCR was conducted on the cDNA (used reverse transcriptase, "+RT") and the no-reverse transcriptase control ("-RT"). Two reactions were set-up for each template; one was run for 25 cycles and the other was run for 35 cycles. All products were analyzed via gel electrophoresis using a DNA stain such as SYBR-Green. A strong, single band should be visible in the 25+RT sample. Multiple bands indicate non-specific or multiple binding sites of the primer. No band should be visible in the 25-RT. A band in 25-RT indicates the presence of DNA contamination. A strong band should be visible in 35+RT. A faint band can be visible at 35-RT, but should not be strong. A strong band at 35-RT indicates DNA contamination. qPCR was run using SsoAdvanced™ Universal SYBR® Green Supermix as described by the manufacturer (http://www.bio-rad.com/en-ch/product/ssoadvanced-universal-sybr-green-supermix).

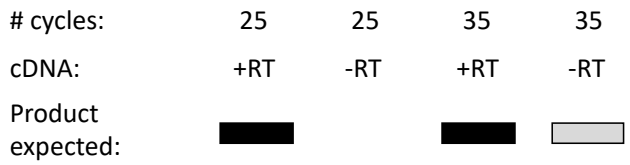

Figure 2-1. Check for DNA contamination set-up and anticipated results.

ENGINEERING AND CONSTRUCTION OF A PHOTOBIOREACTOR

Overview

A custom state-of-the-art light-emitting diode (LED) photobioreactor (PBR) was engineered and constructed for the cultivation of cyanobacteria in diurnal sinusoidal light:dark (sinLD) cycles (Figure 2-2). The LED PBR enables the cultivation of 3 x 1.5 L cultures with illumination of up-to 2,000 μmole photons m^{-2} s^{-1} on both sides of circular flat-panel reactors (FPRs). A temperature-controlled water bath suits multiple reactor sizes and quantities. Controlled carbon dioxide (CO_2) mixing and gas delivery rates are achieved via variable flow meters. Custom aerators provide carbon delivery, pH buffering, and agitation to the cultures. Rapid sampling ports enable ultra-fast sampling times (<5 seconds) for time-sensitive studies, such as the time-course metabolomics work presented in this study. This reactor expands the capabilities of the Peebles' lab and—to our knowledge—represents the first reactor which enables larger volume cultivation under super-high light intensities.

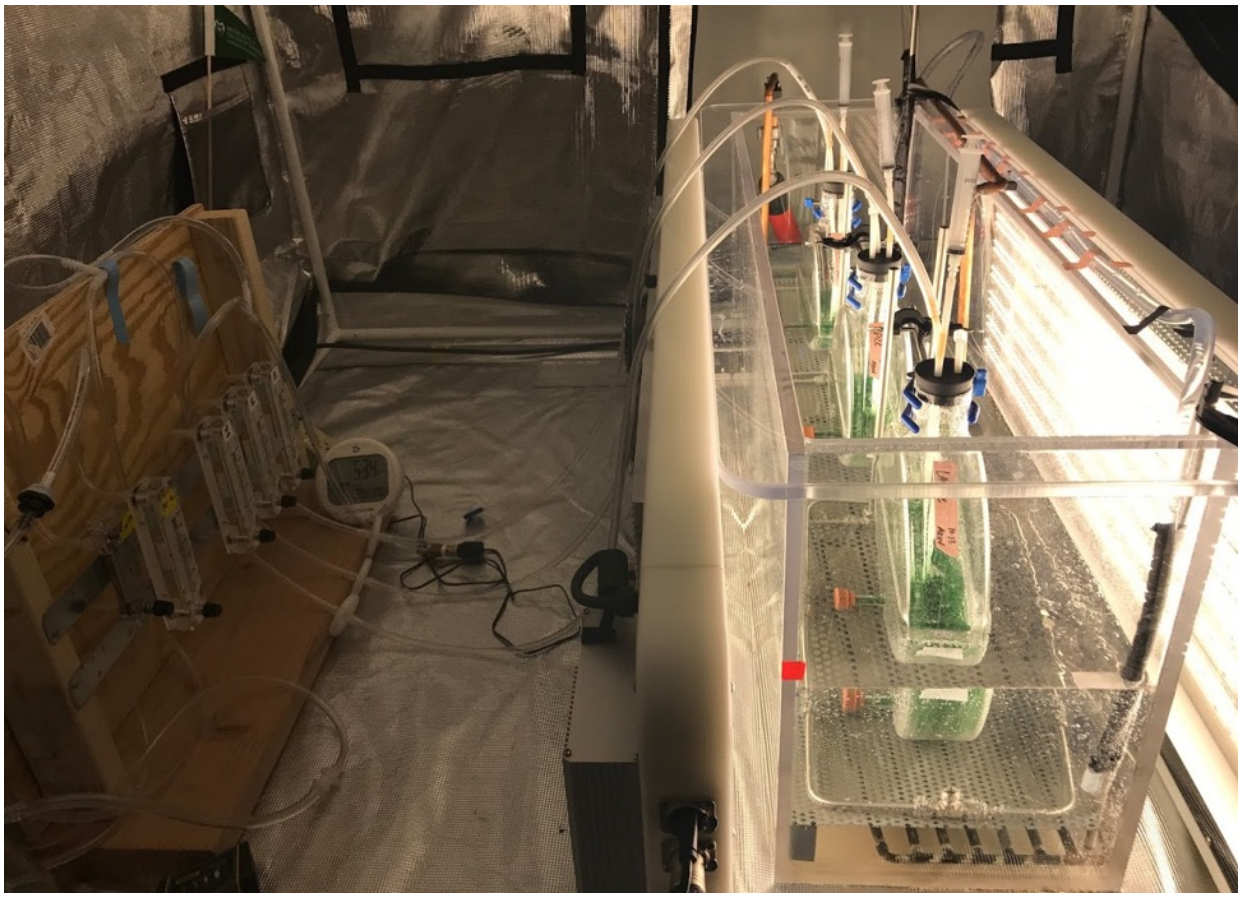

Figure 2-2. LED PBR overview. Three flat panel reactors (1.5 L working volume) are arranged down the center of a water bath and illumined by two 4000K white LED panels on either side. Bath temperature is controlled via circulating thermofluid in a heat exchanger. Uniform water bath temperature is maintained via a circulation pump. House air is mixed with 100% CO_2 to a user-defined concentration; gas delivery rate to each reactor is controlled with variable flow meters. Reactors are fitted with a rapid sampling port, gas delivery port, and a gas vent.

Light delivery

Light delivery was designed to deliver up-to 2,000 micromole photons per m^2s^{-1} (μE) with near-sunlight spectrum to closely mimic natural sunlight quantity and quality. Additional design criteria included dynamic control of irradiance, minimal light source heat production, and uniform delivery. These criteria prohibit the use of common grow-lights which provide targeted wavelength spectra. Fluorescent lights are also prohibited due to intolerably high heat outputs at high light intensities, difficulties associated with rapidly changing light intensities, and unrealistic spectra (Figure 2-3A). White 4000K LEDs provides irradiance across the key

cyanobacterial whole-cell absorbance areas, specifically 400-450 nm and 600-700 nm (Figure 2-3B). While this spectral irradiance does not exactly mimic sunlight, 4000K white LEDs present the most cost-effective technology available at this time.

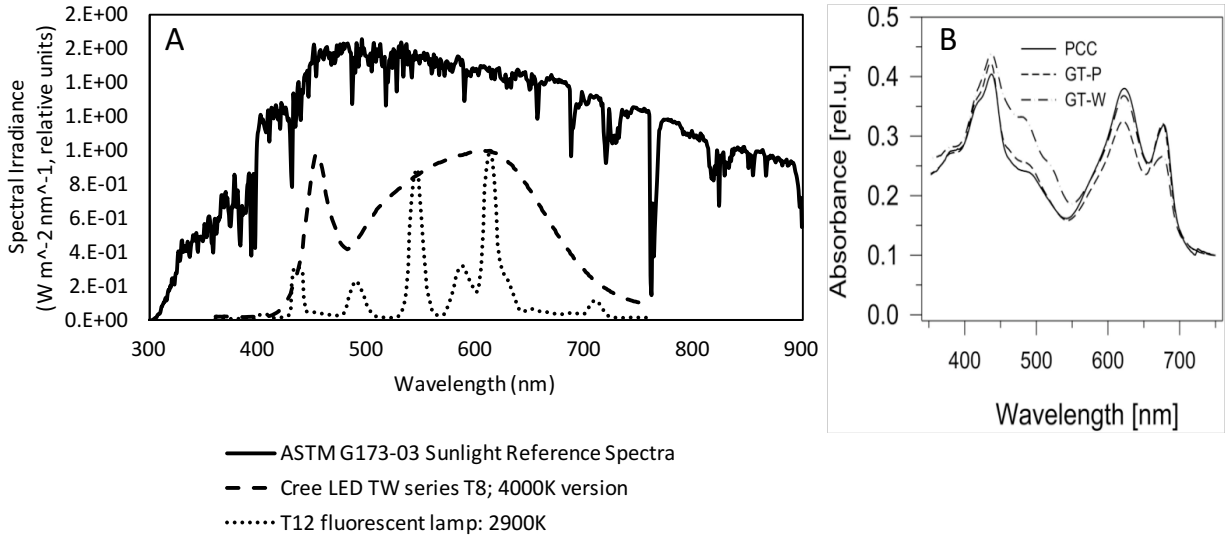

Figure 2-3. Spectral irradiance of light sources and photosynthetic absorbance. (A) Spectral irradiance comparison between sunlight, 4000K White LEDs, and 2900K Fluorescent lamp. Sunlight spectra obtained from National Renewable Energy Laboratory Renewable Resource Data Center, supplied by American Society for Testing and Materials (ASTM) Terrestrial Reference Spectra for Photovoltaic Performance Evaluation (http://rredc.nrel.gov/solar/spectra/am1.5/). LED and fluorescent lamp spectra obtained from Designing with LEDs (http://www.designingwithleds.com/light-spectrum-charts-data/). Spectral irradiance is shown in W $m^{-2}nm^{-1}$. LED and fluorescent spectral irradiance is shown in relative units. Graph created by A. Werner. (B) Whole cell absorption spectra of three Synechocystis strains—PCC, GT-P, and GT-W—as measured by a Shimadzu UV3000 spectrophotometer in Tichy et al. [66].

Two custom-built LED arrays were manufactured by Reliance Laboratories (https://www.reliancelabs.com/) with white 4000K LEDs. The state-of-the-art LEDs produce minimal heat, and on-panel heat sinks dissipate >90% of heat produced. The arrays provide 0-2,000 μE smooth and reproducible light delivery with 3,000 discrete steps from 0-100% output using advanced network control via an on-panel central processing unit. The panels are 45” long by 13” tall by 5” deep. The two LED panels were positioned facing each other and each array was covered by two diffusion sheets (supplied by Reliance Laboratories). Variation in irradiance

across the three reactors was negligible (Figure 2-4A). A light pattern was programmed to deliver irradiance in a sinusoidal manner such that light was delivered for 12 hours peaking at 1600 μmol photons m^2s (μE) (Figure 2-4B).

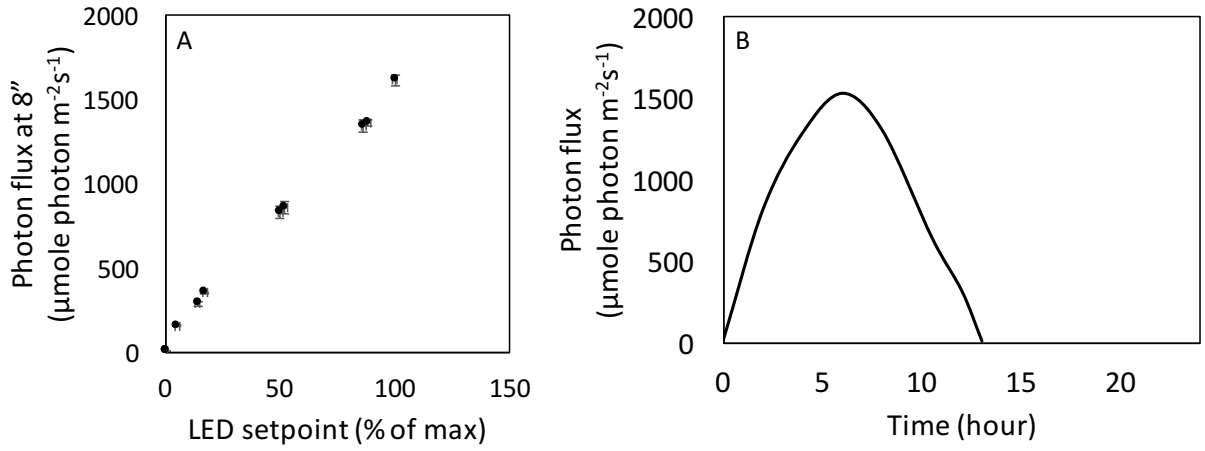

Figure 2-4. LED PBR photon flux delivery. (A) LED set-point versus photon flux between two panels arranged facing each other. Error bars represent standard deviation across the three bioreactors. (B) Sinusoidal 24-hour diurnal LD cycle program.

Gas delivery

Controlled gas mixing and delivery to individual cultures is provided (Figure 2-5). The system was engineered to mix 100% CO_2 with house compressed air to accommodate a range of user-defined CO_2 gas concentrations and avoid the high cost of pre-mixed CO_2. Air and CO_2 flow rates are set with acrylic variable flow meters to achieve the CO_2 concentration set-point and the gases are mixed in a closed chamber. Importantly, check valves are placed between the gas feed and mixing chamber to avoid the obscuring effects of back-pressure on variable flow meter readings. Mixed CO_2-air gas is fed to three reactors via individual acrylic variable flow meters such that each reactor flow can be monitored and adjusted. A pressure release valve is also adjustable to accommodate applications where the total flow rate far exceeds the flow to each reactor. Gas is fed through a 0.2 μM filter and bubbled into liquid culture via a custom-made gas diffuser. Due to the nature of the carbonate buffer system CO_2 delivery provides pH

control: given a fixed flow rate, as the percentage of CO_2 increases the pH will decrease. In the present system, a set percentage of CO_2 is supplied instead of 100% CO_2 to maintain a set pH. This approach was chosen to prevent the periods of super-high CO_2. We do not know the effect this would have on the cells, and do not want to confound our metabolomics experiment with any carbon availability effects.

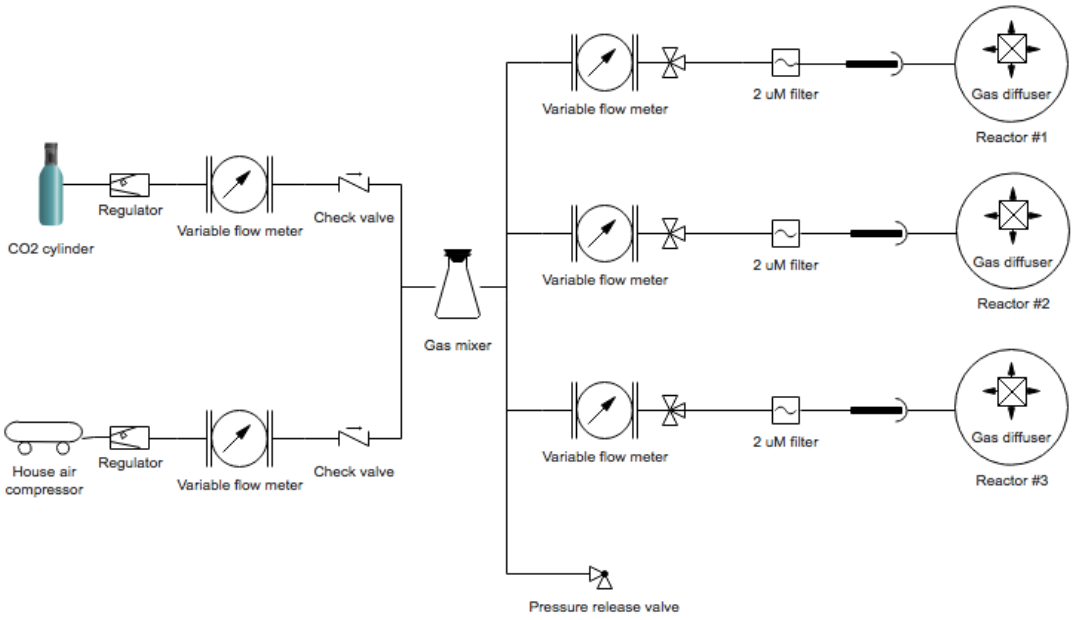

Figure 2-5. CO_2 delivery system flow diagram. 100% CO_2 is mixed with house air, filtered, and delivered to individual reactors at a controlled volumetric flow rate.

Gas delivery into each culture needed to provide (a) mixing of the culture and (b) carbon delivery to the cultures. Commercial aerators with holes on the micron scale were tested, but at low gas delivery rates (<50 mL/min) intolerable amounts of foam were created. Custom bubblers were constructed with 1/8" ID x 1/4" high-density polyethylene (HDPE) tubing with holes created with a 0.8 mm needle. HDPE was chosen as the tubing material because it can be autoclaved and is suitable for submersion into biological samples. Several geometries of bubbler were constructed and tested, including ring-shaped, L-shaped, and Y-shaped (Figure 2-6). Ring-shaped required too high of a flow rate to achieve bubble distribution across the aerator.

Furthermore, deformation cause by squeezing through the bottle mouth never reverted after insertion. L-shaped also required too high of a flow rate to achieve bubble distribution across the aerator. L-shaped geometry was also fragile and subject to frequent displacement which would skew gas delivery in the culture. Y-shaped proved to be the best design. With two holes near the base of each leg, low gas flow rates enabled bubble delivery through each hole. The Y-shaped bubbler was easily inserted and removed and the geometry was stable in the culture. Ends were capped with latex plugs.

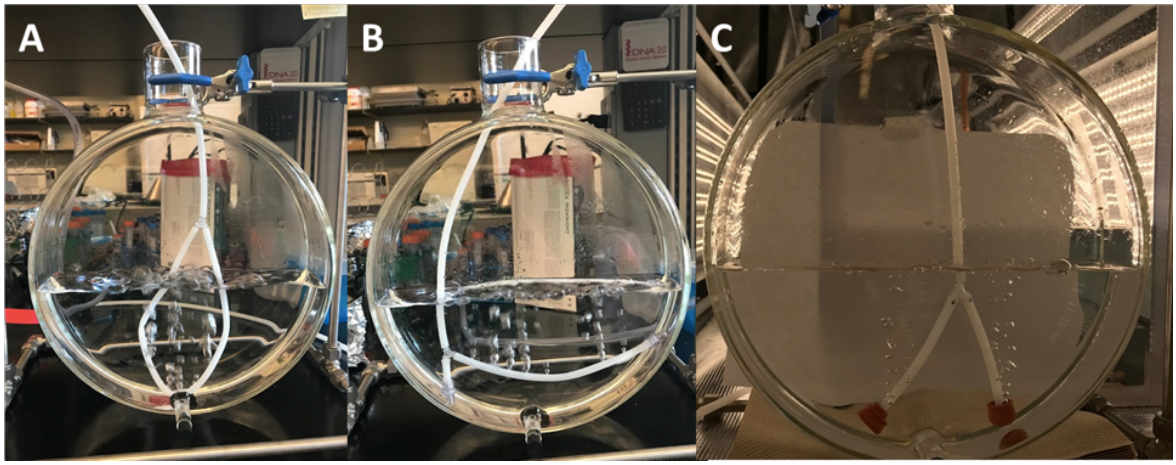

Figure 2-6. Bubbler geometries. (A) Ring-shaped, (B) L-shaped, and (C) Y-shaped. All bubblers were constructed out of HDPE tubing.

The quality of mixing achieved using the Y-shaped bubbler, mixing tests with food dye were completed. For each test, 1.5 L of water was added to a FPR fitted with the Y-shaped bubbler. Air flow rates and dye placements were changed between run. Yellow dye was added first; after complete dispersion, green dye was added. In each case, 0.5 mL of dye was added. The time required for complete mixing was recorded. Observations suggest that the green dye is slightly heavier than the yellow dye, which is reflected in the time recordings.

With dye added to the side wall ~1" from the water surface, mixing rates of less than 2 minutes were achieved (Figure 2-7A). Faster mixing rates were seen with dye added to the center

of the vessel (data not presented). With no gas delivery, uniform dye mixing was not achieved by 45 minutes, at which point the test was ended.

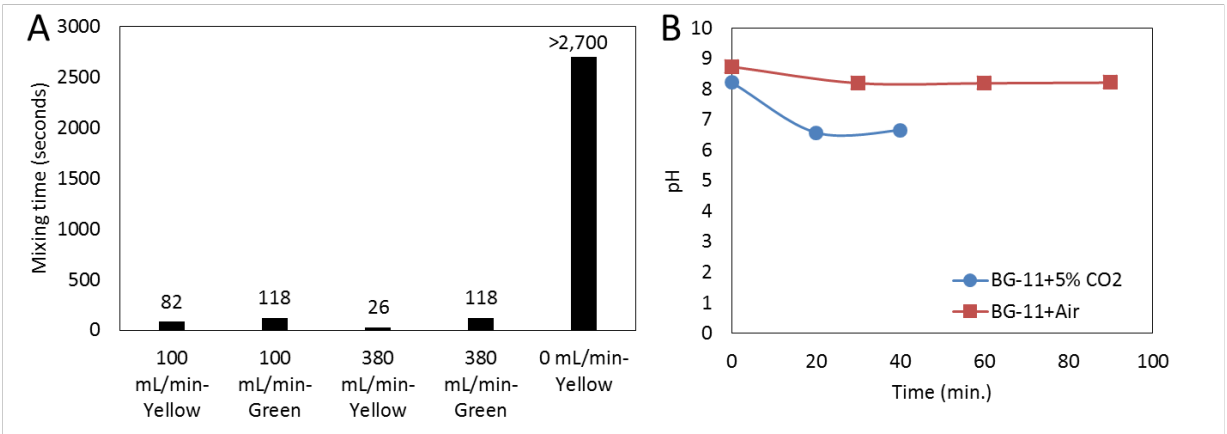

Figure 2-7. Y-shaped bubbler mixing rates and CO_2 delivery. (A) Y-shaped bubbler mixing rates as time required to achieve homogeneous food dye dispersion. Flow rates (mL/min) and dye color are indicated. The 0 mL/min test was ended at 45 minutes (2,700 seconds), but homogenous dye dispersion was not achieved at this point. (B) pH of BG-11 media bubbled with 0% and 5% CO_2 delivery at 15 VVM. No cyanobacteria cultures were present in the media.

The pH of media bubbled with Y-shaped bubblers was assessed using 5% and 0% CO_2 (Figure 2-7B). With air (0% CO_2), media pH was maintained at 8.2. With 5% CO_2, pH dropped to 6.5 after <20 minutes was maintained at 6.5. Also tested were 25-50 μM glass dispersion tubes at 5 VVM and metal tubes with holes at 15 VVM. All tested systems with 5% CO_2 reached a pH of ~6.5-6.6 after ~20 minutes (data not presented). Cyanobacterial growth was expected to affect the pH level significantly, as the process of photosynthesis increases media pH and cyanobacteria are able to tolerate high pH environments. The pH dynamics of *S.* 6803 cultures are presented in the subsequent sections.

Temperature control

Culture temperature is maintained via a water bath system. A custom-built copper tubing heat exchanger with circulating thermofluid controls water bath temperature (Figure 2-8).

Thermofluid enters down the center line and is dispersed outward in both directions along the bottom of the water bath. Thermofluid is heated, cooled, and pumped with a Huber Ministat. Water bath temperature uniformity is ensured with a circulation pump.

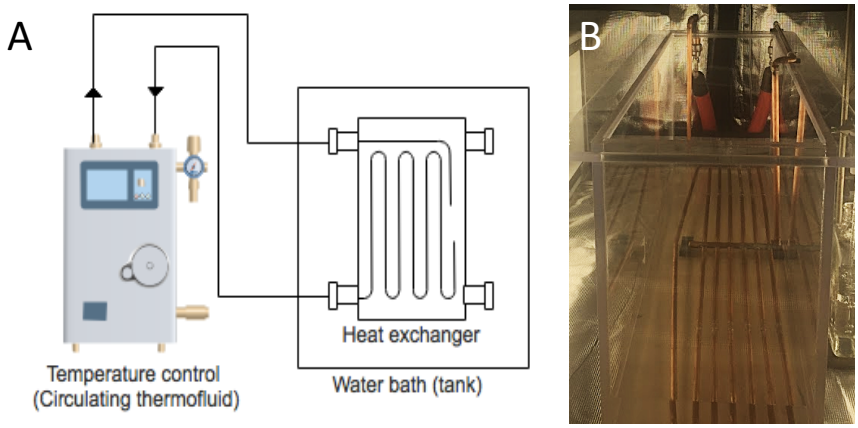

Figure 2-8. Temperature control via water bath with heat exchanger. (A) Temperature control schematic. An acrylic box is filled with water and fitted with a heat exchanger. A Huber Ministat regulates and circulates thermofluid which is fed through a heat exchanger to control the water bath temperature. (B) Top-view of custom-built copper tubing heat exchanger (heat exchanger built by Tim Gonzalez of the CSU Chemical and Biological Engineering Department). Circulation pump not pictured.

Reactors

The water-bath design enables modular utilization of a wide variety of glass reactors. Reactors for *S.* 6803 growth in the FPR developed in this study include 70 mL test tube reactors, 0.5 L tubular reactors, and 1.5-2 L flat-panel circular reactors. To accommodate the biomass requirements for multiple MS platforms, at 1.5 L culture in a flat-panel reactor (FPR) was selected for use in the time-course comprehensive metabolomics experiments.

Glass reactors were built to provide a maximal surface area to volume ratio to achieve the maximal amount of incident light to cells at higher cell densities. Four custom FPRs were constructed by Allen Scientific Glass (http://www.allenglass.com/) with a 3.7 L total volume and 1.0-2.5 L working volume (Figure 2-9). A port for sampling is fitted to the bottom (though it was not used for rapid metabolomics sampling) and a circular top port supports size 10 standard

rubber stoppers. FPR stands were constructed out of stainless steel tubing and Swagelok fittings, and reactors are secured with scientific clamps (Figure 2-9C).

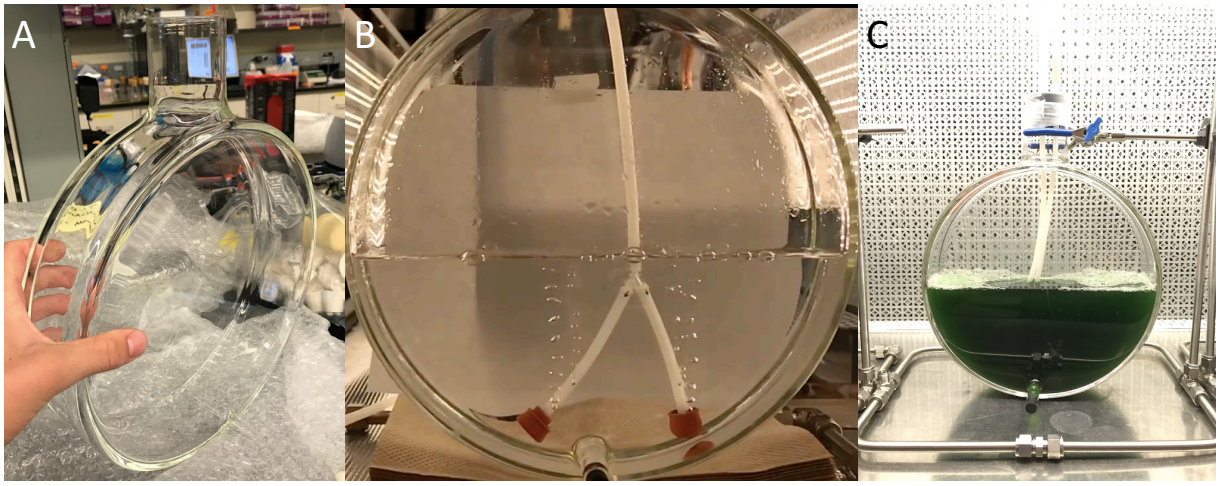

Figure 2-9. Flat-panel glass reactors (FPRs). (A) Side-view of reactor. (B) Front-view of reactor filled with water fitted with aerator. (C) Front-view of reactor secured by a stainless-steel stand and containing 1.5 L of S. 6803 culture.

To prevent contamination, the bubblers were bleached for 24 hours, acid-washed for 24-hours, rinsed with water (through the tubing), and autoclaved inside the FPRs with the stopper not secured into the neck but hanging loosely and covered entirely with aluminum foil. Gravity cycle with 60 minutes of sterilization was used for all part of the FPR and bubbler set-up. Though a filter on the vent was used for the experiments, the absence of a filter did not result in contamination. Modified inoculum train protocols were also essential to maintain clean cultures (discussed in next section).

CULTIVATION OF *S.* 6803 IN SINUSOIDAL DIURNAL LIGHT:DARK CYCLES

Sampling considerations

Many factors were considered when determining culture volume, culture density, sample volume, and number of time-points. The total volume removed from each reactor was restricted to less than 15%, which restricted the total sampling volume to <16 mL for a 1.5 L culture with

14 time-points (Figure 2-10A). The culture density was affected by several factors, including: minimizing cell shading and thereby maximizing incident light to each cell, sampling the biomass requirement at each time-point given the volume restrictions, and enabling adequate cell growth under the pre-determined sinLD cycle regime. Preliminary studies indicated that ~3x10^9 cells are required as feed into the biphasic extraction in order to achieve good signal by GC-MS of the soluble fraction (the least sensitive instrument analyzing the fraction which is split in half). To address the biomass requirement, the number of cells which would feed all MS experiments given three different optical densities (ODs) and four different volumes (7, 8, 9, and 10 mL) were considered (Figure 2-10B). The OD to cell number conversion was calculated using a standard curve created by counting cells via flow cytometry using CountBright™ beads (Figure 2-10). Given this, sampling 10 mL of a culture at an OD~3.0 would provide sufficient biomass for downstream analyses. The cell shading consideration was restricted by the lack of healthy cell growth in sinLD cycles peaking at 1,600 µE to OD>1.5; therefore, higher optical densities higher than 1.5 were required.

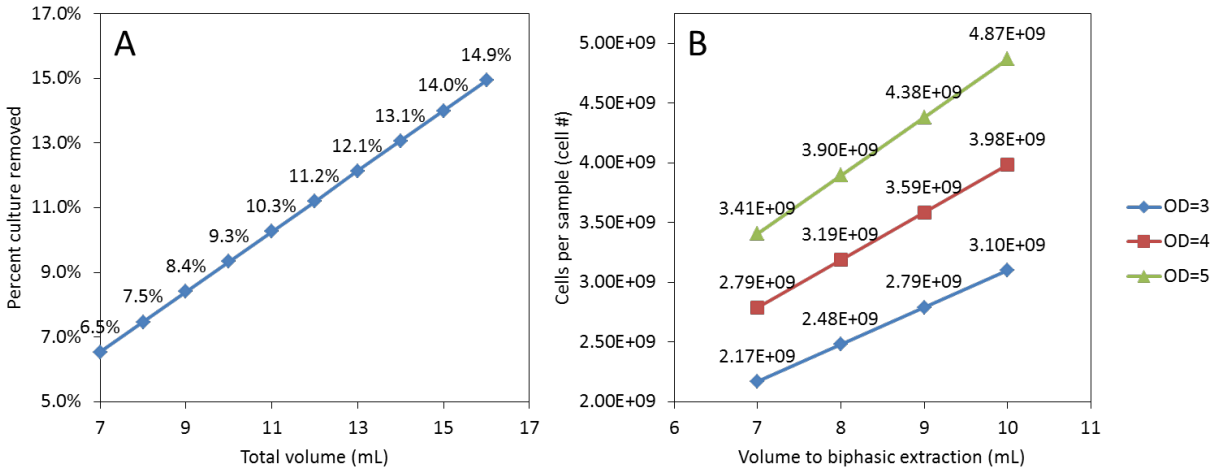

Figure 2-10. Criteria for the determination of sample volume and culture OD. (A) Total sampling volume as a percentage of the total volume for 14 time-points. (B) Volume of the total used for a biphasic extraction and MS analysis for ODs of 3, 4, and 5. The cell number available for MS analysis at each OD and each volume are indicated.

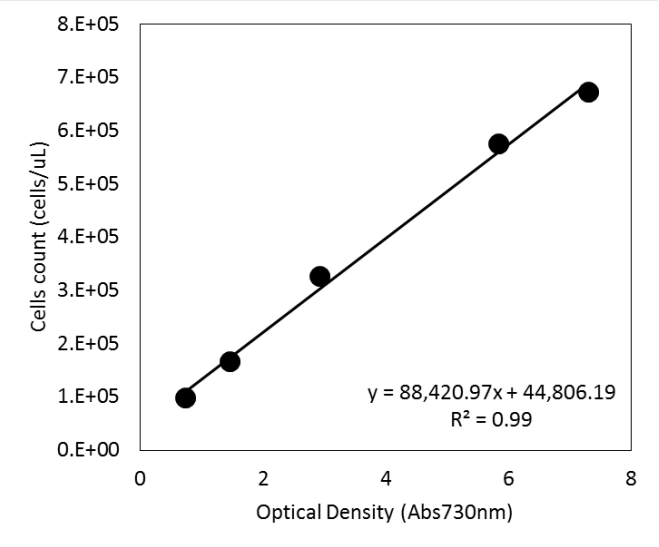

Figure 2-11. S. 6803 optical density (Abs_{730nm}) to cell density (cells per μL). Absorbance was measured on a NanoDrop™ Spectrophotometer. Cells were counted via flow cytometry.

Inoculum train

The inoculum train was developed to serve two key functions: (1) rapidly generate enough *S.* 6803 biomass to inoculate 3 x 1.5 L FPRs, and (2) provide three days of sinLD cycle entrainment prior to sampling for metabolomics. It is well known that *S.* 6803 has relatively poor genetic stability due to the presence of transposons, homologous recombination, and other features [135]. Therefore, a key consideration was to perform the metabolomics experiment quickly after re-suspension from a freezer stock to minimize any potential genetic drift.

Materials preparation and PBR initialization

All inoculum train and FPR growth was done in un-buffered BG-11 media.[2] Ultra-pure 18 Ω water was used for all media components.[3] BG-11 media was prepared fresh for all steps:

[2] Sustained $NaHCO_3$ feeding in 1M Tes-NaOH buffered BG-11 resulted in cell death (yellow cultures). The cause for this growth defect is unknown.
[3] A heavy rain in late September 2017 overwhelmed the building water processing system and the result was small particulate matter contained in the reverse osmosis water supply. This particulate, while alone did not grow on LB plates, caused flocculation of *S.* 6803 cells and the

here, “fresh BG-11” means <1 week old, while “ultra-fresh BG-11” means <36 hours old. Media prepared fresh the day before FPR inoculation resulted in more repeatable growth and higher growth rates in the LED PBR. This effect was not as evident in shake-flasks, although *S.* 6803 growth in media >1 month old is unreliable and often the cultures flocculate within 1 day of growth in old media.

The LED PBR was initialized 2 days prior to inoculation to achieve consistent gas flow rates and uniform water bath temperature (detailed SOPs not provided in this dissertation). Inoculum train schematic (Figure 2-12) and representative growth curves (Figure 2-13) were conducted as follows. Freezer stock wild-type *S.* 6803 is inoculated in 25 mL of fresh BG-11 in quadruplicate 150 mL Erlenmeyer flasks and grown at 150 μE, 30°C, 225 rpm. Cultures are fed with 1M $NaHCO_3$ and pH adjusted to 8.0-8.2 with 6 N HCl twice a day. Only one culture is measured by pH probe directly; the acid addition of the measured flask is applied to the other three. This protocol modification was required to maintain clean *S.* 6803 cultures. Feedings and pH adjustments are spaced ~8-12 hours apart. As soon as cultures reach the *onset* of stationary phase, they are spun-down by centrifugation at 3,000*g* for 5 minutes.[4] The three cultures which did not experience the pH probe directly are resuspended in ~2 mL fresh BG-11, combined, aliquoted into 4 x 150 mL fresh BG-11 in 500 mL Erlenmeyer flasks. All centrifugation steps are done with sterile single-use 50 mL centrifuge tubes and all resuspension steps are done with filter tips.[5]

---

loss of 20 *S.* 6803 cultures (a set-back of ~3.5 weeks). Therefore, it is recommended that 18Ω water is routinely used instead of reverse osmosis water.

[4] Cultivation of *S.* 6803 >2 days into stationary phase resulted in the presence of white debris visible in the cell pellet post-centrifugation step. Restricting growth only to the linear phase was required to achieve healthy and sterile scale-up cultures.

[5] Use of sterile centrifuge tubes (as opposed to 1 L re-useable Nalgene bottles) and filter-tips were required to achieve sterile scale-up cultures.

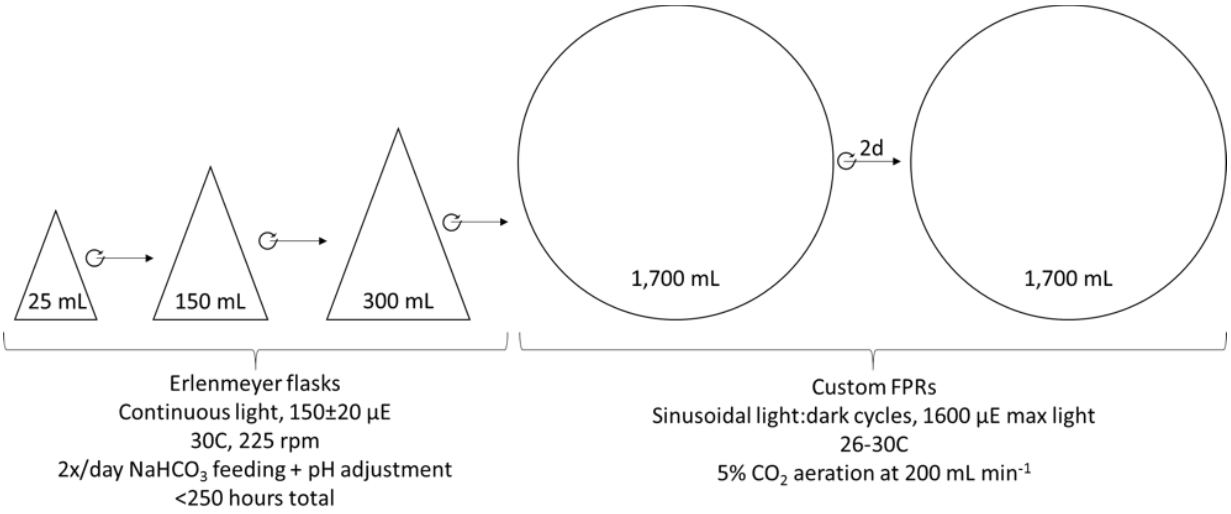

Figure 2-12. S. 6803 inoculum train schematic. Freezer-stocked S. 6803 is inoculated into 25 mL in an Erlenmeyer flask and grown under CL 150 µE with $NaHCO_3$ feeding. Cells are scaled-up into 150 and 300 mL flasks for a total CL cultivation time of <250 hours. Each scale-up step is accompanied by centrifugation and re-suspension in fresh media. Cells are inoculated into FPRs and grown in sinLD cycles for 2 days, centrifuged, re-suspended, grown for a final sinLD cycle, then sampled on the fourth sinLD cycle.

150 mL cultures are grown until the onset of stationary phase, then scaled-up. As with 25 mL cultures, the three cultures which did not experience the pH probe are centrifuged, combined, resuspended, and used to inoculate 4 x 300 mL cultures in 1 L Erlenmeyer flasks. When these cultures reach the onset of stationary phase, the cultures which did not experience the pH probe are centrifuged, combined, and used to inoculate 1 x 1,700 mL FPR with ultra-fresh BG-11. After inoculation, the FPR is incubated in the dark for ~10-12 hours prior to the initialization of the first sinLD cycle. 4.5% CO_2 is provided at 200 mL/min and the water bath is set to 25°C for the entire FPR experiment. OD, pH, and temperature of the LED FPR are measured 5-6 times a day for the first two days of sinLD cycle growth. During the first hour of dark following the second sinLD cycle, the cells were aliquoted into sterile 50 mL centrifuge tubes, centrifuged at 3,000*g* for 5 min. at 25°C, and resuspended in 1-2 mL ultra-fresh BG-11. All resuspended cells were combined, mixed, and aliquoted into 3 x 1.7 L ultra-fresh BG-11 in freshly-autoclaved FPRs. FPRs were entrained in sinLD cycles for a final day prior to sampling for a total of three

days of sinLD cycle entrainment. Three days of entrainment was selected as it is within the standard in the field for LD cycle 'omics experiments: McEwen *et al.* [91], Mettler *et al.* [136], Beck *et al.* [85], Willamme *et al.* [84], and Diamond *et al.* [83] all used 3-days of diurnal LD cycle entrainment prior to sampling. Intermittent centrifugation and resuspension in fresh media ensured that cell growth was still in linear phase despite cultivation as a batch culture.

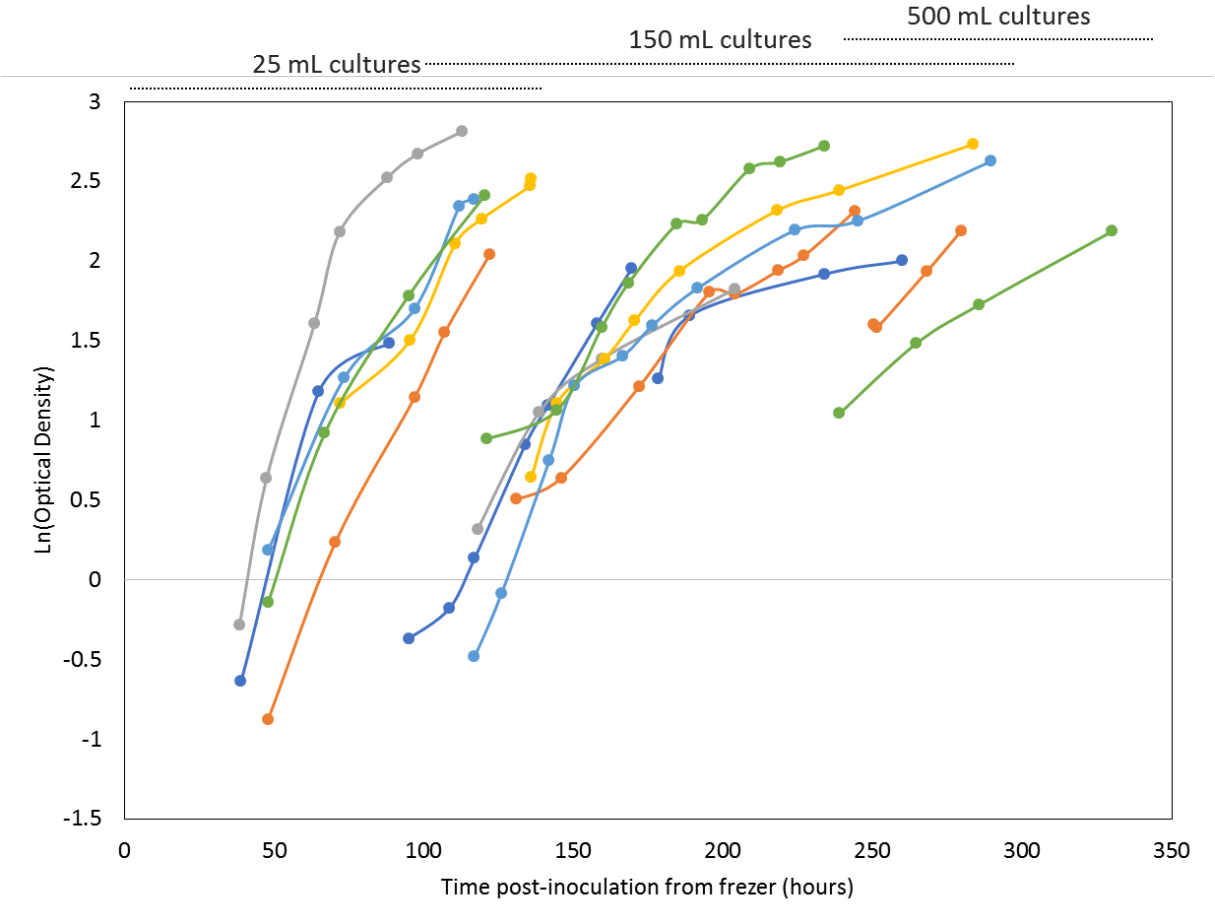

Figure 2-13. Inoculum train growth in 25 mL, 150 mL and 500 mL shake-flask cultures for representative scale-up experiments. Growth is shown as the natural log of optical density. Coloring represents the average of triplicate cultures (five experiments shown).

Contamination checks

Checks for contamination were conducted before inoculation, during inoculation, and at the completion of the rapid sampling experiment to ensure a sterile *S.* 6803 was present. Checks were conducted as follows. Cells were sampled from FPRs (or shake flasks) and plated on LB

plates in 1, 1:10, and 1:100 dilutions and incubated at 37°C in darkness. Plates were checked for contamination after 24 and 48 hours of incubation. Cells were also plated on BG-11+ 10 mM glucose and incubated at 30°C and 50 μE. Plates were checked for contamination after 48 hours and 1 week of incubation. Flow cytometry analysis of cells further confirmed that cultures which passed the plate checks were uniform in cell populations.

sinLD cycle growth curves

A series of *S.* 6803 growth curves in sinLD cycles were conducted to develop the final protocols. A brief description of select experiments which resulted in technical/protocol developments and/or biological insight is provided in this section. A summary of FPR growth in sinLD cycles is also presented.

Findings from a series of developmental sinLD cycle experiments are briefly summarized to aid future researchers in running the LED PBR. LED panels must be plugged-in to different electrical circuits to prevent an electrical mishap. Cultures must be centrifuged and re-suspended in entirely fresh media in the FPR to achieve linear growth on *both* the first and second days of sinLD cycle growth. Without this, growth rate decreases from 0.08 hr^{-1} on day one to 0.04 hr^{-1} on day two. The likely explanation for this is rapid nutrient utilization and therefore nutrient limitation by the second day of rapid growth. Growth in 2x Trace Metals did not significantly improve growth rate. Though never tested, the amount of foaming in cultures under suspected nutrient limitation was far higher than cultures with fresh media every ~2 days. Bubblers and reactors needed acid-washing and stringent sterility procedures prior to inoculation in order to prevent contamination. A constant CO_2 delivery rate was chosen instead of supplying 100% CO_2 to control pH because the effects of sudden transitions between carbon limitation and carbon

saturation are unknown but would likely cause metabolic changes. Such changes would confound the metabolite variations this experiment was designed to assess—those in response to sinusoidal diurnal LD cycles. Thus, the experimental design of constant CO_2 delivery resulted in fluctuations in pH: lower pH during periods of low growth and higher pH during periods of high growth (Figure 2-14). To mitigate the pH drop during the night, a manual decrease in the CO_2 at the onset of darkness and subsequent increase at the onset of light was integrated into the experimental design (Figure 2-14). This resulted in a sharp decrease in pH upon turning the CO_2 back on at the onset of light. During this period in the true metabolomics experiment, a series of rapid samples are scheduled. Therefore, while the pH during the night did not decrease as much, the likelihood of perturbing metabolism with acute changes in carbon availability was deemed unsuitable for the purpose of this experiment and a constant CO_2 delivery rate was chosen for subsequent experiments.

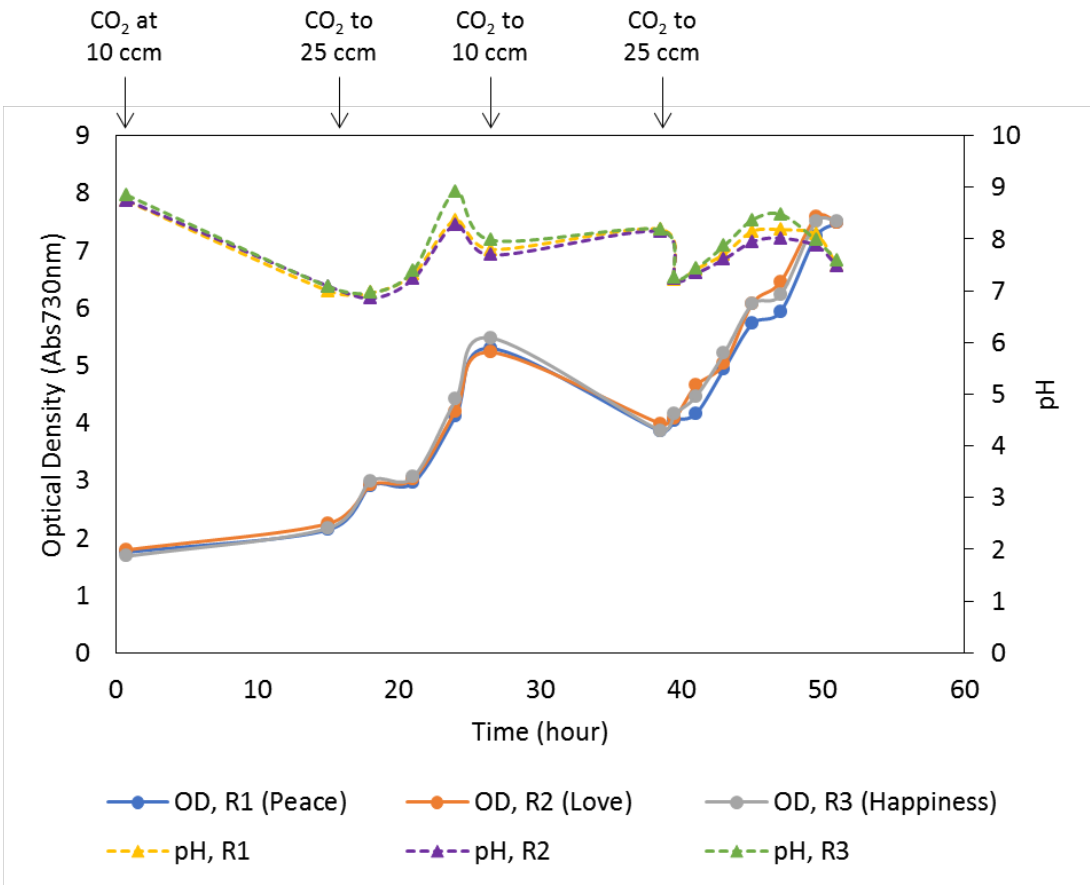

Figure 2-14. sinLD growth curve testing variable CO_2 delivery. CO_2 was decreased to 10 ccm at the onset of darkness and increased to 25 ccm at the onset of light. Growth (optical density, OD) and pH for each of three biological replicates is presented. sinLD cycles applied from *t*=12-24 and *t*=36-48.

Through these developmental experiments, a protocol was developed which enabled consistent FPR cultivation in sinLD cycles. *S.* 6803 growth from eight developmental growth curves are presented in Figure 2-15. Despite variations on the environmental conditions, the growth is relatively consistent across the experiments. Specific growth rate and doubling time for a set of experiments (Figure 2-16) also demonstrates repeatable growth, even under slightly different experimental conditions across the various experiments.

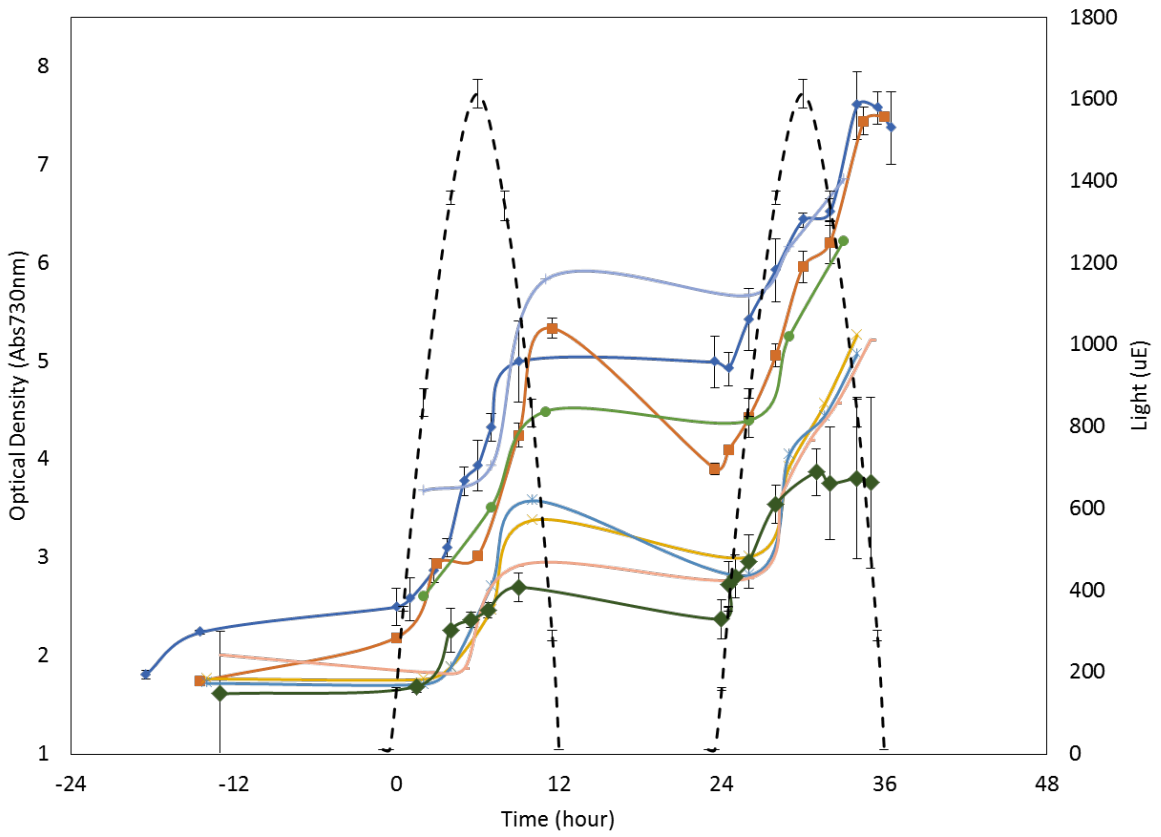

Figure 2-15. Comparison of eight developmental FPR growth curves. Experiments numbers are not presented because they are not all explained in the body of text and therefore are irrelevant. Error bars represent standard deviation across three biological replicates.

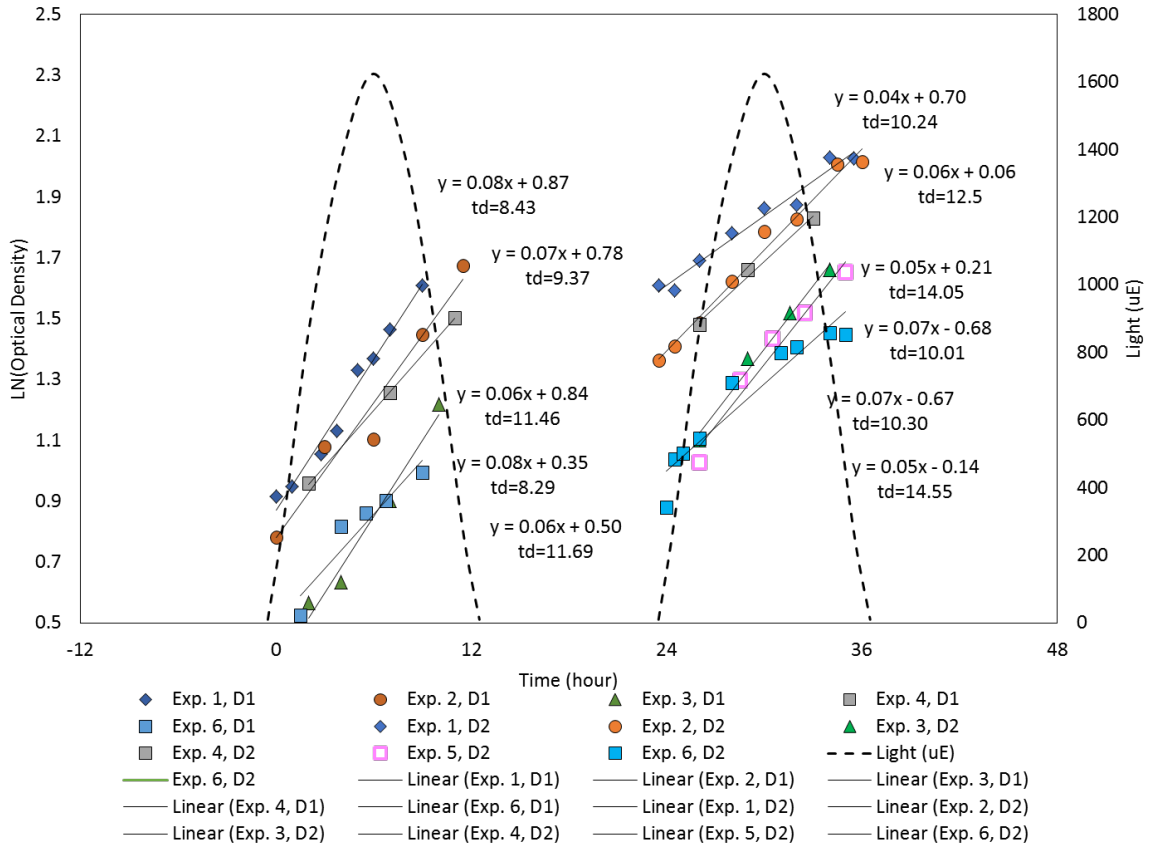

Figure 2-16. S. 6803 specific growth rate and doubling time from six two-day sinLD cycle experiments. Linear regression fits to each day of growth are indicated were m represents μ_{max}. Doubling time for each μ_{max} was calculated and is presented below the linear fit (td).

DEVELOPMENT OF A COMPREHENSIVE METABOLOMICS WORKFLOW

Overview

A workflow for characterizing comprehensive metabolomics over the course of a sinusoidal diurnal light:dark cycle was developed (Figure 2-17). *S.* 6803 are cultivated in the LED PBR for three days prior to sampling. On the fourth day, triplicate cultures are sampled every 0.5-2 hours over a sinLD cycle (Figure 2-17A): 10 mL of each sample is rapidly quenched via ice-cold phosphate buffered saline (PBS) and separated from media via centrifugation (Figure 2-17B); 5 mL of each sample are used for measurements of pH, optical density, cell counts by flow cytometry, and energy cofactor kits. Metabolites are extracted from lyophilized

cell pellets by a biphasic methyl *tert*-butyl ether (MTBE):methanol (MeOH):water(H_2O) workflow (Figure 2-17C). Metabolites of each class are analyzed via multi-platform mass spectrometry (MS, Figure 2-17D). Non-polar metabolites in the MTBE organic phase are detected via non-targeted reverse phase (RP) liquid chromatography (LC) MS. Polar and semi-polar metabolites in the MeOH:H_2O aqueous phase are detected via targeted hydrophilic interaction liquid chromatography (HILIC) MS and non-targeted gas chromatography (GC) MS after derivatization by N-Methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA). Insoluble polymers contained in the pellet post-biphasic extraction are hydrolyzed. Amino acids (AA) are derivatized by N-*tert*-butyldimethylsilyl- N-methyltrifluoroacetamide (MTBSTFA) and detected by GC-MS. Nucleic acids (NA) and sugars are derivatized by MSTFA and detected by GC-MS. The protocol developed here represents a vast improvement in the variety of metabolites detected from a single sample.

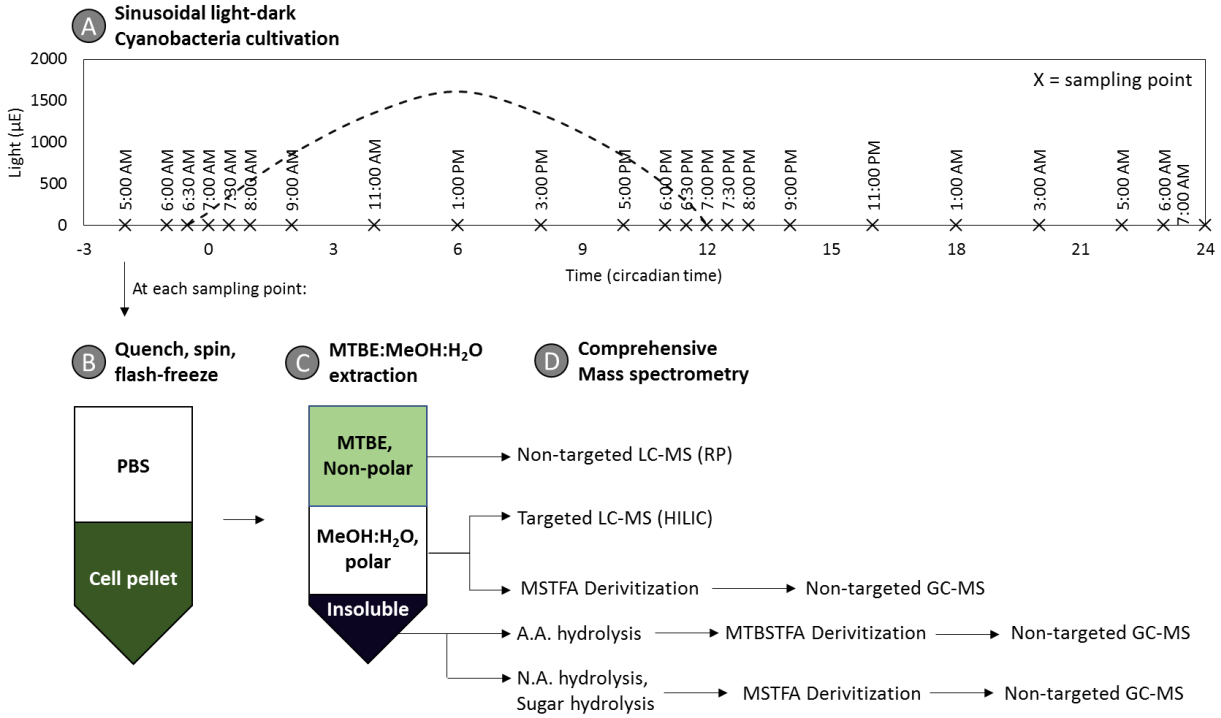

Figure 2-17. Time-course comprehensive metabolomics workflow overview. Abbreviations are defined in the text body.

Metabolite quenching

Development of a rapid-filtration protocol

A filtration-based quenching method was developed. Briefly, a pre-weighed filter was fitted to a vacuum filtration unit and wetted with water. Cyanobacteria samples were filtrated through in <30 seconds, placed in a centrifuge tube, rapidly quenched in liquid nitrogen, stored at -80ºC for 2 days prior to lyophilization. The whole filter was subjected to biphasic extraction for soluble metabolic analysis.

The filtration method was compared to a PBS-quench/centrifugation-separation protocol which is standard in the ^{13}C-MFA community [137]. Both methods produced repeatable data across biological replicates with comparable signals (data not shown). The filtration method, however, did not produce quality pellet-hydrolysis data due to filter interference in the hydrolysis protocols. Therefore, the PBS-Pellet method was selected as the metabolite quenching method for all studies.

Quenching protocol

A 50 mL centrifuge tube for each sample point was labeled and pre-weighted. 10 mL of sterile 1X PBS was aliquoted into each tube and placed at -50°C. 17 minutes prior to the sampling point, the PBS tubes were removed and placed on ice; immediately before the sampling point, tubes were vortexed ~30 seconds or until homogeneous slurry and placed on ice. Nine labeled microcentrifuge tubes and three 15 mL centrifuge tubes were also placed on ice.

At the exact sampling time, 15 mL was pulled from the reactor using a fresh and sterile 20 mL syringe: 10 mL was ejected into the ice-cold PBS slurry; 5 mL was ejected into the 15 mL centrifuge tube. The PBS-cell slurry was immediately capped and vortexed at medium speed for

~3 seconds. A fresh and sterile syringe was replaced on the reactor sampling port. This sampling protocol was repeated for all three reactors.

PBS-quenched cells were spun at 3,000*g* for 5 min. in a centrifuge at -4°C. From sampling to centrifugation, it took <2 minutes. While the centrifugation step ran, 500 uL was aliquoted from the 5 mL tube to 3 x 1.5 mL microcentrifuge tubes for each reactor's sample and tubes were quenched in liquid nitrogen. After the centrifugation step completed, the supernatant was decanted, tube re-capped, and cells quenched in liquid nitrogen. All liquid nitrogen quenched samples were immediately stored at -50°C for 1 day and then transferred to -80°C until further analysis (<3 weeks). 100 uL was taken from the remaining cell suspension and placed in a cuvette to measure optical density and cell count. pH was recorded in the remaining cell suspension prior to flash-freezing in liquid nitrogen.

Cell fixation and biomass measurements

Optical density was measured in a 1:10 dilution on a NanoDrop Spectrophotometer as absorbance at 730 nm in a 1 mm path length cuvette. After measurement, this cell suspension was fixed in a final concentration of 0.25% glutaraldehyde in PBS at room temperature for ~15 minutes prior to flash-freezing in liquid nitrogen and storage at -50°C for <3 weeks. To count cells by flow cytometry, fixed cells were thawed and 200 μL was transferred to a 5 μL flow cytometry tube. 20 μL of CountBright™ beads was added to the cell sample. Cell counts were performed on a CyAn1 flow cytometer. Cells per μL were calculated as follows:

$$(A/B * C/D) * 10 = \text{Cells/uL}$$

where *A*=cell events (singlets), *B*=bead events, *C*=bead lot (beads/uL), *D*=sample volume, and multiplied by a dilution factor of 10. Pellets were lyophilized for 48 hours and the pellet dry cell weight (DCW) was measured.

Biphasic metabolite extraction

Metabolites were extracted from dried pellets via a modified MTBE:MeOH:H_2O biphasic extraction, based on the protocol developed by Salem *et al*. [138]. Briefly, 6 mL of 75% methanol (MeOH) was added to pellets, vortexed, and transferred to glass vials. 9 mL of 100% methyl *tert*-butyl ether (MTBE) was added, vortexed for 30 seconds, placed on automatic shaker for 1.5 hours at 4 ºC, and sonicated for 15 minutes. 3.75 mL of water was added, each extraction was vortexed by hand for 1 minute, and centrifuged for 10 minutes at 3,270*g* at 4ºC. A biphasic solution with a pellet formed: the top green MTBE layer and the bottom clear MeOH:H_2O layer were separated into separate tubes and dried under $N_{2,gas}$ overnight. The pellet was stored at -80 ºC. After drying, the MTBE layer was resuspended in 100 uL 1:1 toluene:MeOH, transferred to a LC-MS vial insert, and stored at -80ºC for <1 month prior to MS analysis. The MeOH:H_2O layer was resuspended in 1 mL of 1:1 H_2O:MeOH, transferred to a 1.7 mL centrifuge tube and spun at 15,000*g* for 2 minutes at 4ºC. The supernatant was split into two 465 µL aliquots—one for GCMS and one for LC(HILIC)MS—in glass vials and dried under $N_{2,gas.}$ The protocol outlined above is suitable for filter-quenched cyanobacteria samples and centrifuged cell pellets.

The polar methanol/water fraction resulting from the biphasic extraction was processed for analysis by hydrophilic interaction liquid chromatography (HILIC) LC-MS. Dried samples were resuspended in 100 µL 1:1 H_2O:MeOH and 10 µL were aliquoted into a pooled QC sample. Samples were stored at -80 ºC until analysis. The pooled QC sample was mixed and aliquoted

into twelve vials. A QC injection was run every tenth injection. The dried polar fraction for analysis by GC-MS was stored at -80 ºC until derivatization, immediately prior to MS analysis. Samples were derivatized in 30 µL methoxyamine HCl and 30 µL MSTFA, as specified in the following section. Ten microliters were removed from each sample to create a pooled QC sample, mixed, and aliquoted into thirteen vials. A QC sample was run after every sixth injection. The non-polar MTBE phase was processed for non-targeted LC-MS analysis. Twenty microliters from each sample were pooled, mixed, and aliquoted into thirteen pooled QC samples. QC injections were placed after every sixth injection.

Derivatization of non- and semi-volatile compounds for GC-MS detection

All samples analyzed by GC-MS were derivatized using a protocol provided by the CSU Proteomics and Metabolomics Facility. Briefly, dried-down samples were resuspended in 30 µL or 50 µL (indicated in sample preparation section) of warm methoxyamine HCl, incubated at 60ºC for 45 minutes, vortexed briefly, and sonicated for 10 minutes at maximum intensity. Samples were again incubated at 60ºC for 45 minutes, 30 µL or 50 µL of derivatization reagent was added (indicated in sample preparation section), vortexed briefly, and incubated at 60ºC for 40 minutes. Samples were centrifuged briefly to collect precipitates and the supernatant was transferred to a GC-MS vial insert for analysis.

Pellet hydrolysis of amino acid, nucleic acid, and carbohydrate detection

An acid hydrolysis protocol was developed for identification of amino acids, nucleosides, and carbohydrates bound in insoluble pellet. Pellets remaining from the biphasic extraction were removed from storage at -80 ºC and residual solvent was evaporated under nitrogen gas. Pellets

were re-suspended in 3 mL of 6 M hydrochloric acid (HCl) using vigorous vortexing and pipette re-suspension. The resulting suspension was a bright teal. The suspension was transferred equally to three separate glass vials for hydrolysis of the separate polymer constituents.

Hydrolysis of protein to amino acids

Hydrolysis of proteins to amino acids was completed with a hydrochloric acid (HCl) hydrolysis, based on previously published protocols (described in Table 2-3). Briefly, vials were incubated at 110ºC for with a loose cap seal. After 4 hours, the acid in each vial was entirely evaporated; 1 mL of 6 M HCl was added to each vial, vortexed, sealed tightly, and returned to 110ºC. After a total of 24 hours, vials were removed and remaining acid was evaporated under nitrogen gas.

Table 2-3. Protein hydrolysis procedures tested for retention of fragile amino acid residues. Experiments performed by K. Olsson. HCl: Hydrochloric acid.

Acid	Temp.	Time	Additive(s)	Result	Ref.
6 M HCl	110 °C	24 hrs.	None	Consistent hydrolysis, loss of Trp, Arg, Cys	[139]
6 M HCl	110 °C	24 hrs.	0.02% Phenol	Inconsistent hydrolysis (intolerable discrepancies among biological replicates)	[139]
6 N HCl	110 °C	24 hrs.	0.4% β-mercaptoethanol	Inconsistent hydrolysis (intolerable discrepancies among biological replicates)	[140]
6 M HCl	166 °C	25 min.	3% Phenol	Incomplete hydrolysis (insoluble protein remaining)	[141]
6 M HCl	145 °C	4 hrs.	3% Phenol	Incomplete hydrolysis (insoluble protein remaining)	[141]

Interpretation of amino acid changes must be done with attention to the short-comings of the hydrolysis method employed. Despite still being the "state-of-the-art" protein hydrolysis method, the 24-hour 110°C 6M HCl hydrolysis method biases the data considerably due to the different chemistries of individual amino acids [139]. Primarily, tryptophan (Trp), cysteine (Cys), and asparagine (Asp) are lost from the analysis due to residue degradation under the harsh conditions. Several additives have been referenced as providing improved recovery of Trp and Cys, including various concentrations of phenol and β-mercaptoethanol. Kristin Olsson, a Master of Education Student in the Peebles' lab over the summer of 2017, assisted in testing whether these additives improved Trp and Cys retention using the experiments outlined in Table 2-3. Proteins hydrolyzed by each method were analyzed for amino acid composition via GC-MS (MTBSTFA derivatization). Olsson found that phenol and *β*-mercaptoethanol improved retention of Trp, but added a layer of inconsistency which we deemed to be intolerable. Specifically, the amino acids identified in these methods

The results of K. Olsson's experiments suggested that a 6 M hydrochloric acid hydrolysis at 110°C for 24-hours was the best method but would bias the data by removal of Trp, Asp, and Cys from the analysis. Additionally, this method is known to result in losses of serine (Ser) and threonine (Thr) due to side-chain chemistries (Figure 2-18) [142]. Therefore, the data derived from this experiment will be interpreted as changes over time, as opposed to abundance, and the relative contribution of each amino acid to the total protein cannot be accurately inferred. This also relies on the assumption that amino acid losses are consistent across the samples. This assumption may be violated if a set of samples have significantly more protein, or significantly more of a particular amino acid, such that hydrolysis is carried to various levels of completion in

the different samples. In an effort to avoid this, additional volume of HCl was added to ensure adequate reagent for high-protein samples.

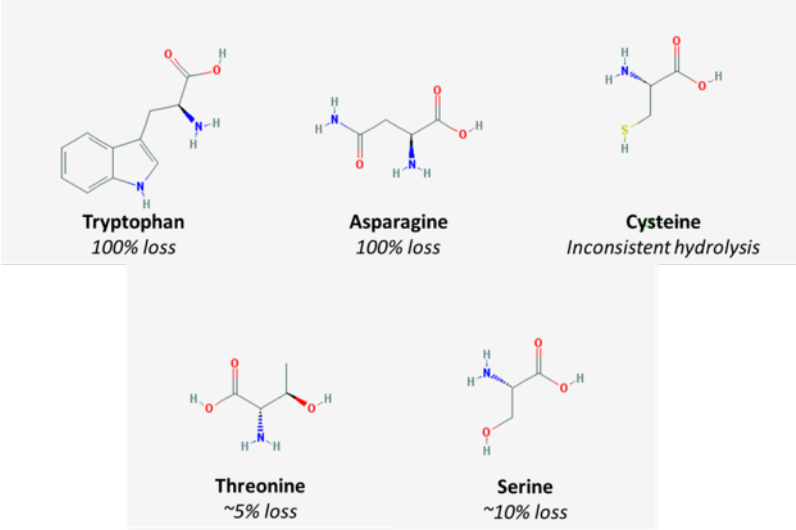

Figure 2-18. Amino acids losses in the protein hydrolysis procedure employed for analysis of the pellet. Images from PubChem (https://pubchem.ncbi.nlm.nih.gov/).

Samples were resuspended in 150 µL of 1:1 MeOH:H2O, 20 µL was removed to create a pooled QC sample, the pooled QC was aliquoted into fourteen vials, and the solvent was evaporated under nitrogen gas. Amino acid samples were derivatized in 30 µL of methoxyamine HCl in pyridine and 30 µL of MTBSTFSA as described in the previous section. The peak integration of each amino acid in each sample was manually checked and curated in the software Chromeleon™. Spectral abundance for each amino acid was exported and normalized to the cell number in each sample and multiplied by $1x10^8$. Amino acids that were not detected include asparagine and tryptophan. Aspartic acid with 2- and 3-derivitization agent modifications were detected and summed for the total spectral abundance. Threonine with 2- and 3-derivitization agent modifications were detected and summed for the total spectral abundance.

Hydrolysis of polysaccharides to sugars

Polysaccharides were hydrolyzed to sugars using a HCl hydrolysis, based on a published protocol developed for algal biomass [143]. Briefly, vials were allowed to evaporate under nitrogen gas for 30 minutes, placed in a 40°C water bath for 1 hour, capped tightly to prevent significant evaporation, and incubated at 110°C for 1 hour. The resulting suspension was a deep purple/blue. 70 µL of 4 M sodium hydroxide (NaOH) was added to neutralize each suspension, and the liquid was transferred to a glass teardrop vial. (Note: There were no/few visible insolubilities in the suspension; everything was transferred on to the next processing step.) The neutralized acid was evaporated under nitrogen gas overnight.

Dried samples were re-suspended in 150 µL 1:1 MeOH:H_2O. Sugar samples were vortexed at the highest speed for 30 seconds. Nucleic acid samples were pipetted five times to achieve a solubilized sample. Twenty milliliters from each sample was transferred to a pooled QC sample for both the sugar and nucleic acid samples. Sugar samples were derivatized in 50 µL of methoxyamine HCl in pyridine and 50 µL of MSTFA. A QC sample was run after every sixth injection.

Hydrolysis of DNA and RNA to nucleosides

Nucleic acid polymers were hydrolyzed to nucleosides using a modified protocol from Huang *et al.* (2012) [144]. Briefly, vials were incubated at 130ºC for 10 minutes, removed and allowed to cool, and 100 µL was transferred to a glass teardrop vial. The remaining pellet was incubated at 160ºC for 40 minutes, removed and allowed to cool, re-suspended in 100 µL LC-MS grade water, vortexed for 15 seconds, centrifuged at 1,500*g* for 2 minutes, and the supernatant was transferred to the glass teardrop vial which contained purines from the 130ºC incubation. (Note: There were significant insoluble chunks in the resulting suspension from the

160ºC incubation that were not transferred on to the next processing step). The acid was evaporated under nitrogen gas.

Samples were removed after the 10 minute 130ºC incubation to preserve purines (guanine and adenine) from degradation during the 40 minute 160ºC incubation. Huang *et al.* (2012) demonstrated that less than ~5-10% of the purines remained after a 40 minute incubation at 160ºC.[6] Therefore, calculations of purine content were assumed to be representative of hydrolysis from half of the sample. However, the presence of pyrimidines in the first hydrolysis product (130ºC for 10 minutes) remains a concern. Huang *et al.* (2012) demonstrated that after 5 minutes at 130ºC less than 5% of the pyrimidines were hydrolyzed, but after 10 minutes at 130ºC nearly 100% of the pyrimidines were hydrolyzed. Huang *et al.* (2012) found that utilizing this protocol resulted in minor deamination of pyrimidines and purines (11% and 16%, respectively). Removal of an amino group from cysteine via deamination (addition of water and release of ammonia) would result in an artificially high abundance of uracil at the expense of cysteine. Deamination of guanine results in the formation of xanthine and deamination of adenine results in the formation of hypoxanthine. Huang *et al.* (2012) describe observing deamination into xanthine, hypoxanthine, and other compounds as a result of the 160ºC hydrolysis step. Nucleic acid samples were derivatized in 30 μL of methoxyamine HCl in pyridine and 30 μL of MSTFA.

In our experiments, the sugar hydrolysis procedure resulted in a great deal of insoluble material, suggesting incomplete hydrolysis. Test runs of both the sugar samples and nucleic acid samples by GC-MS resulted in similar sugar recoveries between the two hydrolysis methods. As

[6] Calculated as [(ending concentration)/(starting concentration)]*100% from Huang *et al.* (2012) data presented in Figure 2.

the nucleic acid method was more repeatable, both nucleic acids and sugars were characterized using the nucleic acid hydrolysis procedure.

MS instrument methods

GC-MS

Metabolites were detected using a Trace 1310 GC coupled to a Thermo ISQ mass spectrometer. Samples (1 µL) were injected at a 10:1 split ratio to a 30 m TG-5MS column (Thermo Scientific, 0.25 mm i.d., 0.25 µm film thickness) with a 1.2 mL/min helium gas flow rate. GC inlet was held at 285°C. The oven program started at 140°C for 1 min, followed by a ramp of 15°C/min to 330°C, and 5 min hold. Masses between 50-650 m/z were scanned at 5 scans/sec under electron impact ionization. Transfer line and ion source were held at 300 and 260°C, respectively. Pooled QC samples were injected after every 6 actual samples.

RP LC-MS

Two microliters of extract were injected onto a Waters Acquity UPLC system in discrete, randomized blocks with a pooled QC injection after every 6 sample injections, and separated using a Waters Acquity UPLC CSH Phenyl Hexyl column (1.7 µM, 1.0 x 100 mm), using a gradient from solvent A (2mM ammonium hydrioxide, 0.1% formic acid) to solvent B (Acetonitrile, 0.1% formic acid). Injections were made in 100% A, held at 100% A for 1 min, ramped to 98% B over 12 minutes, held at 98% B for 3 minutes, and then returned to starting conditions over 0.05 minutes and allowed to re-equilibrate for 3.95 minutes, with a 200 µL/min constant flow rate. The column and samples were held at 65 °C and 6 °C, respectively. The column eluent was infused into a Waters Xevo G2 Q-TOF-MS with an electrospray source in positive mode, scanning 50-2000 m/z at 0.2 seconds per scan, alternating between MS (6 V collision energy) and MSE mode (15-30 V ramp). Calibration was performed using sodium

iodide with 1 ppm mass accuracy. The capillary voltage was held at 2200 V, source temp at 150 °C, and nitrogen desolvation temp at 350 °C with a flow rate of 800 L/hr.

Data pre-processing for non-targeted platforms

Detect peaks

Raw data was converted to *.CSV with Waters® Databridge. For idMS/MS (RP-LC-MS runs), a file was converted for low-collision, high-collision, and LockSpray for each sample. Peaks were detected within the XCMS workflow using the Centwave algorithm [145]. Centwave detects a feature as a bounded, two-dimension signal (*m/z* and retention time) composed of one-dimensional *m/z* peaks (centroids) [146]. In this way, regions of interest (ROIs) are regions where at least p_{min} centroids with a deviation of less than μ_{min} (ppm) occur. This avoids the 'binning' approach which is subject to skewing and disruption of peak detection. The result is a list of mass traces with mass traces of differing lengths, where each ROI may have 0, 1, or >1 distinct chromatographic peaks. Peaks are then detected using a matched filter approach with user-defined peak width ranges (*minpw, maxpw*), mass deviation (μ in ppm), and signal-to-noise ratio (*snthresh*) with continuous wavelet transform (CWT) method, which accommodates peaks of differing widths. Features were detected via XCMS as follows: group peaks with wide bandwidth (3), correct retention time, re-group with narrow bandwidth (1.5), and fill peaks. The output is an XCMS object (*xset*) for each feature (*m/z* and RT) for each sample for both low- and high-collision energy.

Group peaks

Class information (biological replicates) were added to the phenotype slot and XCMS was run again using *minfrac=1,* which specifies that a called peak must be present in all three

biological replicates of at least one time-point.[7] This is a different approach than the standard *minfrac=0.25* where a called peak must be present in 25% of all samples. I adopted the class-based approach because we are specifically interested in peaks which are only present at select time-points, perhaps less than 25% of the samples. The requirement than the peak must be present in all three biological replicates should remove noise.

Normalize per-cell

Each feature's spectral abundance was normalized to the cell number in the sample at that time point. The next step, feature clustering by RAMClustR, did not function properly with very small spectral abundances, such as those which have been divided by ~10^9 cell numbers. Therefore, after biomass normalization each feature was multiplied by $1x10^8$ to achieve numbers compatible with RAMClustR.

Cluster features

Compounds were created by clustering features using RAMClustR [147]. RAMClustR uses a similarity matric which calculates feature correlation across samples and retention time correlation between features. Hierarchical clustering of the similarity matrix was computed via the fastcluter package [148]. The resulting clustered dendrogram is cut using DynamicTreeCut and spectra are created with clusters and features abundances from input data [149]. The abundance for each mass in spectra is a weighted mean of feature intensity. The RAMClustR

[7] The *pmfxcms* function was run with *regroup=TRUE* after the addition of class information; the resulting *xset* was used for all downstream processing.

outputs are compounds (clusters of correlated features) and intensities for each sample; spectral abundance intensities reflect weighted mean of all features within the compound.

Data pre-processing for targeted platforms

For targeted HILIC TQS analysis of metabolites, the integration for spectral abundance of each compound was assessed within the Skyline software package (https://skyline.ms/project/home/begin.view?). Compounds which were not detected and excluded from further analysis, and ions which were removed from spectral integration due to absence in samples, are listed in Table 2-4. Compounds noted as not detected could have been missed due to incorrect retention time windows, abundance below the limit of detection, or absence from our samples. Integration windows for each detected compound and each sample were manually assessed for accuracy within Skyline.

Table 2-4. Compounds or ions removed from analysis in HILIC pre-processing. ND: not detected.

Compound	Note
Methylmalonate	ND
Uracil	ND
5-methyluridine	ND
Xanthine	ND
Cytosine	ND
Trigonelline	ND
Acetylmannosamine	ND
Cysteine	ND
3-Chlorotyrosine	ND
Fructose	ND
Acetylneuraminate	ND
Allantoin	ND
3-hydroxymethylglutarate	ND
Glucose	ND
13C-Glucose	ND
7-methylguanosine	Ion 149 removed
Glycerophosphocholine	Ion 86 removed

Allantoate	ND
beta-alanine	ND
Creatine	ND
Glycine	ND
Gln-6-S	ND
NAD	Ion 448 removed
Xylulose-5-P	ND
Aminoadipate	ND
gamma-aminobutyrate	ND
Phosphocholine	ND
Glucuronate	ND
Carnosine	ND
Myo-inositol	ND
CDP-Choline	ND
Cystine	ND
Saccharopine	ND
Man-6-P	ND
Homocystine	ND
Acetylcholine	ND
Cystathionine	ND
NADPH	Ion 302 removed
Ribulose-1,5-P	ND
Isocitrate	ND
Citrate	ND
Choline	ND
ITP	ND
CTP	Ion 96 removed
Tryptamine	ND
Lysine	Ion 106 removed

For amino acid analysis via GC-MS of the pellet fraction, the peak integration was manually assessed for each detected compound in each sample using the Chromeleon software package (https://www.thermofisher.com/order/catalog/product/CHROMELEON7).

Metabolite annotations

For non-targeted GC-MS analyses, putative metabolites identified via RAMClustR were annotated using the Golm database via the RamSearch GUI (https://vpr.colostate.edu/pmf/metabolomics-resources/). Head-to-tail spectral matches were

manually assessed and annotations were accepted which provided adequate spectral matching (*e.g.* Figure 2-19).

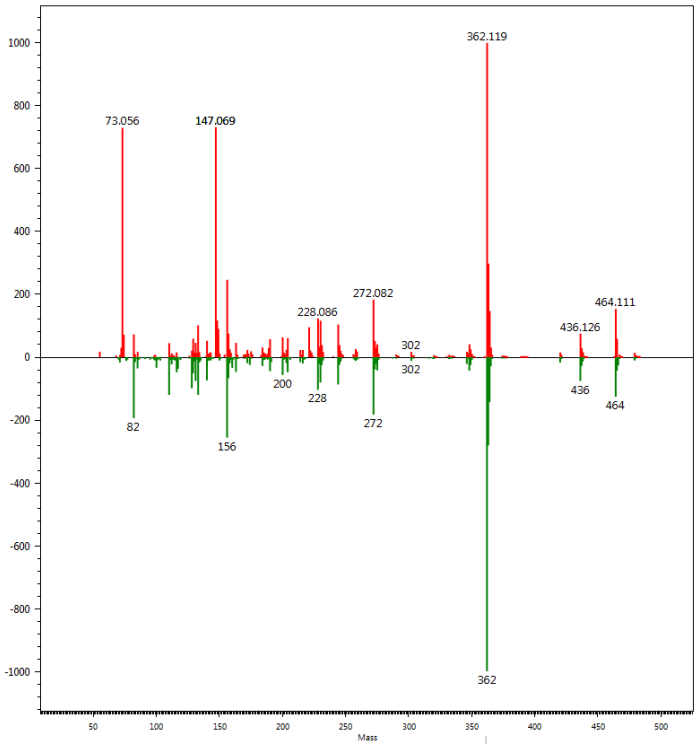

Figure 2-19. Head-to-tail spectral matching in RamSearch. Red spectra are experimentally collected data. Green spectra are reference spectra from the Golm Database.

The Golm Database removes fragments at masses 73 and 147 due to widespread prevalence. However, my data does not have these ions removed and therefore a good spectral match can be achieved with the presence of ions at mass 73 and 147 but absence in the reference spectra, as evidenced in Figure 2-19. Annotations were iteratively assessed using the retention time index for reference spectra/metabolites in the Golm Database. This was achieved by manually annotating metabolites in the RamSearch software, importing annotations into R (into the RC object), and plotting the retention time index from the Golm database against the retention time measured in my experiment (Figure 2-20). Interestingly, the annotations for quantile-normalized (in RamClustR) metabolites were more easily annotated via this workflow.

For non-targeted LC-MS metabolite annotations, molecular formulas and molecular weights were assigned to compounds identified by RP-MS using InterpretMSSpectrum [150]. Putative structures and annotations were assigned by MSFinder [151].

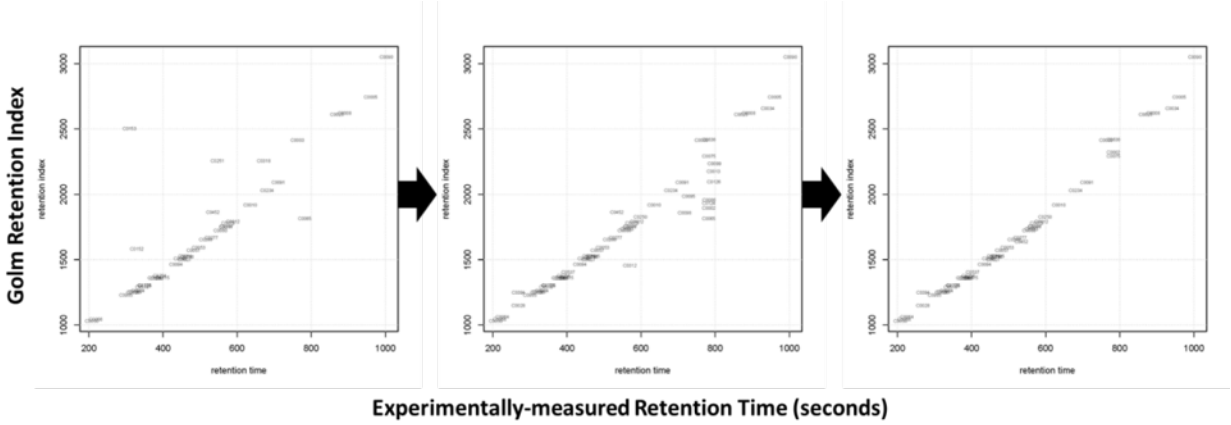

Figure 2-20. Annotation curation via comparison of Golm Retention Index with experimentally measured retention time. Metabolites which did not align with the retention index trend were iteratively re-annotated, if possible, following the left-to-right progression of index vs. experimental retention time plots.

CONTRIBUTIONS

This work was supported in part by the National Science Foundation [Award #1332404 and #1336236]. The photobioreactor development and engineering was made possible by the support of many people and businesses. Tim Gonzales of the CSU Chemical and Biological Engineering (CBE) Department contributed invaluable knowledge, skills, and expertise to many aspects of the PBR engineering, but especially the heat exchanger for the PBR water bath. Tim designed and constructed the heat exchanger from scratch at no-cost to this project! Employees of Ace Hardware in Old Town Fort Collins provided extensive help in selecting appropriate parts and troubleshooting pressure differential issues; Employees of the CSU Chemistry Stock room went out of their way to find the perfect products for the PBR gas delivery; and, Robert Murchinson and Jordan Werner donated stainless steel tubing and fittings which enabled the construction of high-quality bioreactor stands on a tight budget.

The comprehensive metabolomics workflow development was made possible by the continuous support of members of the CSU Proteomics and Metabolomics Facility (PMF) . Linxing Yao aided in the development of the biphasic extraction and GC-MS protocols. Corey Broeckling aided in the development of experimental and data processing workflows, specifically for a time-course study such as this. Emily Davis, Sarah Lyons, and Lisa Wolfe of the CSU PMF ran the MS instruments for the GC-MS, LC-MS (RP), and LC-MS (HILIC), respectively. Instrument methods were provided by PMF. Christopher Allen of the CSU Flow Cytometry Core Facility assisted in the flow cytometry analysis.

CHAPTER 3 – DISCOVERY AND CHARACTERIZATION OF *SYNECHOCYSTIS* SP. PCC 6803 LIGHT-ENTRAINED PROMOTERS IN DIURNAL LIGHT:DARK CYCLES[8]

SUMMARY

Cyanobacteria are photosynthetic bacteria employed for production of valuable chemicals using sunlight and atmospheric carbon dioxide as substrates. Industrial production in outdoor facilities exposes cyanobacteria to daily light:dark (LD) cycles of sunlight availability. Strain engineers need genetic tools suited for diurnal LD cycles to maximize production in these conditions, but tools for engineering in diurnal LD cycles are extremely limited. Here, we discover native *Synechocystis* sp. PCC 6803 promoters which provide light-entrained expression in 12-hour light:12-hour dark (12:12) LD cycles. Promoters were characterized using bacterial luciferase bioluminescent promoter probes and RT-qPCR. P_{hliC}, P_{rbp1}, $P_{slr0006}$, and P_{sigA} provide light-entrained expression in 12:12 LD cycles when expressed from the *slr0168* chromosomal neutral site. None of the promoters provided free-running oscillations in continuous light (CL) following diurnal LD entrainment, but growth in CL resulted in constant mid-level bioluminescence. Transcripts from P_{rbp1}, $P_{slr0006}$, and P_{sigA} increased significantly following the onset of light, whereas transcripts from P_{hliC} exhibited no oscillations in 12:12 LD cycles. Furthermore, P_{hliC} bioluminescence induction at the onset of light increased at higher light intensities. Overall, our work provides additional genetic engineering tools for cyanobacterial strains for chemical production in diurnal LD cycles.

[8] This chapter, with the exclusion of the final results section, has been published as: A. Werner, K. Oliver, A.D. Miller, J. Sebesta, C.A.M. Peebles. Discovery and characterization of Synechocystis sp. PCC 6803 light-entrained promoters in diurnal light:dark cycles, Algal Res. 30 (2018) 121–127. doi:https://doi.org/10.1016/j.algal.2017.12.012.

INTRODUCTION

Engineered microorganisms can be applied for the sustainable production of chemicals, fuels, and pharmaceuticals. Cyanobacteria are a promising chassis for development due to their innate ability to utilize solar energy and carbon dioxide as substrates for growth via photosynthesis. An assortment of chemicals have been produced in cyanobacteria with modest yield improvements from engineering advances [36,152–155], almost all of which have been conducted in continuous light conditions. However, industrial production of commodity chemicals will require outdoor cultivation where cyanobacteria cultures are exposed to natural 24-hour day/night, or diurnal, light:dark (LD) cycles. We recently demonstrated that fatty acid productivity in *Synechocystis* sp. PCC 6803 (hereafter *S.* 6803) is significantly decreased in diurnal LD cycles as opposed to continuous light [93]. This observation is not entirely surprising given that the energy landscape of cyanobacteria changes drastically over the course of a 24-hour LD cycle [81].

Cyanobacteria synchronize global transcription and energy metabolism with diurnal LD cycles. During the day, light is harvested to produce energy cofactors (*e.g.,* ATP, NADPH) and storage molecules (*e.g.,* glycogen); during the night, these energy molecules are nearly depleted [81]. Transcripts within functional categories of translation, ATP synthase, photosynthesis, and CO_2 fixation are up-regulated during the day, while genes related to pentose phosphate pathway, respiration, and DNA replication are up-regulated at night [85]. Extensive work in *Synechococcus elongatus* PCC 7942 (hereafter *S.* 7942) has elucidated several mechanisms of circadian gene regulation. A post-transcriptional oscillator (PTO) comprised of KaiABC proteins mediates global transcriptional changes [105–108] and is re-set by cellular redox status [102,103] and ATP/ADP ratio [104]. Previous work has presented evidence for global transcriptional control via circadian-entrained changes in chromosome topology and thereby

promoter accessibility [109]. However, mutations within transcriptional promoters can alter circadian phase, suggesting sequence specificity of *trans*-acting PTO transcription factors (TFs) and/or RNA polymerase sigma factors [156].

Less is known about *S.* 6803 regulation in daily LD cycles, but several studies describe genetic elements which contribute to daily LD cycle regulation. Deletion of *kaiABC* results in growth defects and differential gene regulation in LD cycles [157]. Of the nine putative sigma factors, many have differential regulation in response to light or dark conditions [112,158–160]. *S.* 6803 promoters are classified into three types based on conserved sigma factor recognition motifs; yet, transcription start points vary drastically even within type 1 promoters with conserved -10 and -35 hexamers [160]. Derivations of these conserved motifs have given rise to inducible non-native promoters for use in *S.* 6803 [32,38]. P*psbAII,* one of the most popular promoters in *S.* 6803, only functions as a light-inducible promoter from the native chromosomal locus and not from neutral site *slr0168* [37]. Engineering viable strains for chemical production in LD cycles requires additional modular genetic tools with diurnal temporal sensitivity.

Here, we evaluate nineteen native *S.* 6803 promoters using the bacterial luciferase reporter, LuxAB, in 12:12 LD cycles by monitoring bioluminescence. We identify P_{hliC}, P_{rbp1}, $P_{slr0006}$, and P_{sigA} as promoters which confer light-entrained expression in diurnal LD cycles from the *slr0168* chromosomal neutral site. Transcript increases were observed from P_{rbp1}, $P_{slr0006}$, and P_{sigA} at the onset of light in diurnal LD cycles. P_{hliC} is interesting due to the observations that mRNA abundance does not change despite a significant bioluminescence increase, and light-induced bioluminescence increased at higher light intensities. Overall, this work provides additional tools for temporal control of gene expression in diurnal LD cycles.

MATERIALS AND METHODS

Plasmids, strains and culture conditions

All strains and plasmids used in this study are listed in Table 3-1. All primers used in this study are listed in Table A-1. Plasmids were constructed in *E. coli* DH5α (New England Biolabs). pAM1580 (Addgene, #40241) was modified by introducing a unique NdeI restriction site immediately upstream of *luxA* using the QuikChange Site Directed Mutagenesis Kit (Agilent, Santa Clara, CA) and flanking *slr0168* homologous regions to create pAZ1-Cam. Promoter regions were excised from *S.* 6803 genomic DNA and inserted into pAZ1 by cut-and-paste with NheI/NdeI restriction endonucleases. Chloramphenicol selection was ineffective as a selection marker in our strain of *Synechocystis* PCC 6803 [N-1] (American Type Culture Collection #27184); thus, all promoter:*luxAB* cassettes were subsequently inserted into pIGA4 [32] with InFusion Cloning (Clontech Labs., Mountain View, CA), resulting in the pAZ promoter probe suite (Figure 3-1). Each of these plasmids were transformed into the *S.* 6803 *slr0168* neutral site via homologous recombination as previously described [32]. *S.* 6803 was cultured in 6X phosphate BG-11, pH adjusted with NaCl to 8.0, buffered with TES, and supplemented with 50 μg/mL kanamycin for selection.

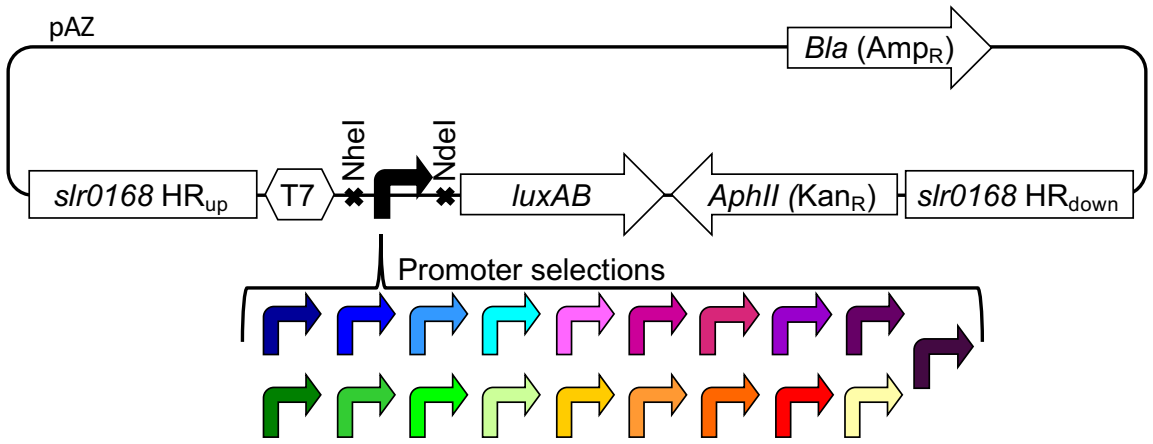

Figure 3-1. pAZ promoter:probe plasmid map. Individual promoters can be inserted/excised by cloning with restriction enzyme NdeI and NheI. E. coli transformants were selected for with ampicillin; S. 6803 transformants were selected for with kanamycin.

Table 3-1. Strains and plasmids.

Name	Description	Source
DH5α	Chemically competent *E. coli*	New England Biolabs
S. 6803	Wild-type (WT) *Synechocystis* sp. PCC6803	ATCC 27184
S. 6803 #2-28	WT *Synechocystis* sp. PCC6803 with *slr0168* replaced with promoter:probe corresponding to pAZ2-28, respectively	This study
pAM1580	Promoterless *luxAB* cloning vector	Addgene
pAZ1	pAM1580 with native NdeI site removed and NdeI site introduced 5' of *luxA,* Chloramphenicol resistance	This study
pIGA4	Cloning vector with *slr0168* homologous regions flanking kanamycin antibiotic resistance (Kan_R)	Albers *et al*. [32]
pAZ2	$P_{slr0749}$:*luxAB* flanked by *slr*0168 homologous regions and Kan_R, as with all pAZ plasmids to follow	This study
pAZ3	$P_{sll0947}$:*luxAB*	This study
pAZ6	$P_{sll1330}$:*luxAB*	This study
pAZ7	$P_{sll0543}$:*luxAB*	This study
pAZ8	$P_{slr1793}$:*luxAB*	This study
pAZ9	$P_{slr0623}$:*luxAB*	This study
pAZ11	$P_{slr1634}$:*luxAB*	This study
pAZ12	$P_{ssl1633}$:*luxAB*	This study
pAZ13	$P_{slr0006}$:*luxAB*	This study
pAZ14	$P_{sll0757}$:*luxAB*	This study
pAZ15	$P_{sll0517}$:*luxAB*	This study
pAZ16	$P_{sll1632}$:*luxAB*	This study

pAZ17	$P_{slr1347}$:*luxAB*	This study
pAZ20	$P_{sll1342}$:*luxAB*	This study
pAZ21	$P_{sll1732}$:*luxAB*	This study
pAZ22	$P_{slr0009}$:*luxAB*	This study
pAZ23	$P_{sll0807}$:*luxAB*	This study
pAZ26	$P_{slr0653}$:*luxAB*	This study
pAZ28	$P_{slr0756}$:*luxAB*	This study

Bioluminescence assays

Pre-cultures were inoculated from *S.* 6803 #2 through *S.* 6803 #28 freezer stocks into 20-50 mL BG-11+Kan_{50} and grown for 2-4 days at 30ºC shaking at 225 rpm with 198 ± 36 µmole photons $m^{-2}s^{-1}$ (µE) light delivered from GE Ecolux High Output 24 Watt T5 High Output fluorescent lights. Cultures were inoculated into 250 mL Erlenmeyer flasks at OD_{730} 0.2-0.5 and entrained to 12 hours light:12 hours dark (12:12) LD cycles or continuous light (CL) schemes for 2-3 days. The initial screen for functional promoters was conducted with biological duplicates; further characterization was conducted with biological triplicates. *In vivo* bioluminescence assays were performed as previously described [161]. Briefly, 150 µL of cell suspension was added to a black-sided flat- and clear-bottom 96-well plate (Corning #3631). Bioluminescence was recorded after >2 minutes dark incubation on a FLOUstar Omega Microplate Reader at 520 nm for 10 seconds immediately following injection of 100 mM decanal in 50% (v/v) methanol:water. *In vitro* bioluminescence assays were performed as previously described [162]. Briefly, 500 µL of cell suspension was mixed with 400 µL cold lux buffer (50 mM sodium phosphate pH 7, 50 mM 2-mercaptoethanol, 2% bovine serum albumin) on ice and sonicated prior to addition of $FMNH_2$ to a final concentration of 5 mM. For both *in vitro* and *in vivo* measurements, average counts per second (cps) over 10 seconds were normalized to absorbance

at 730nm (Abs_{730}) for each technical triplicate; biological replicate averages were normalized to wild-type background, unless otherwise noted.

RNA extraction and RT-qPCR

Fifteen milliliters of cell culture were harvested after 7 days of growth in 12:12 L:D cycles, centrifuged at 10,000*g* for 5 minutes, flash-frozen in liquid N_2, and stored at -80ºC for <2 days before extraction. RNA was isolated using the PGTX method [134], treated with TURBO™ DNase kit (Invitrogen, Carlsbad, CA), and purified with the chloroform:phenol:isoamyl alcohol method [163]. cDNA was synthesized using the GoScript™ Reverse Transcription kit (Promega, Madison, WI) with random primers. Semi-quantitative real-time (q) PCR was performed with SsoAdvanced™ Universal SYBR® Green Supermix (Biorad, Hercules, CA) on a Biorad CFX Connect™ Real-Time PCR Detection System using *rnpB* as a reference gene [164]. Technical triplicates (3 ng of cDNA each) were run with *luxA* and *rnpB* primers. The relative quantification method was used to normalize Ct_{lux} to Ct_{rnpB} (ΔCt) and calculate fold change ($2^{-\Delta\Delta Ct}$) [165].

RESULTS AND DISCUSSION

Selection of promoters for light-entrained expression in daily LD cycles

To identify promoters with strong diurnal entrainment, we re-analyzed a diurnal LD cycle microarray study in *S.* 6803 conducted by Beck and colleagues [85]. Beck *et al.* clustered transcripts based on accumulation patterns into 10 groups: clusters 1-6 peaked at different points during the light phase, clusters 7-8 peaked during the dark phase, and clusters 9-10 did not oscillate [85]. Nineteen genes were selected for promoter investigation from a broad range of clusters (Table 3-2and Table A-2). Gene selections were made based on literature validation (*e.g.*

previously indicated to have *cis*-acting light-inducible elements), large accumulation amplitude (*e.g.* strongly "on/off" in Beck *et al.* LD microarray data*)*, and/or *S.* 7942 validation (*e.g.* same expression pattern in Vijayan *et al. S.* 7942 LD microarray data [109]). Approximately 500 nucleotides in the 5' direction from the start codon, or until the adjacent translation start/stop codon, were selected assuming that most positive regulatory elements are location between -500 and the translation initiation codon [30].

Table 3-2. Promoter selections. The native loci, annotated gene name/function, promoter length, and Beck et al. expression cluster is indicated for each promoter analyzed.

Native loci	Annotated gene name, function	Promoter length (bp)	Beck et al. Cluster
sll0807	rpe/cfxE, pentose-5-phosphate-3-epimerase	504	
sll1342	gap2, NAD(P)-dependent glyceraldehyde-3-phosphate dehydrogenase	542	1
slr0653	sigA/rpoDI, principal RNA polymerase sigma factor	518	1
slr0756	kaiA, circadian clock protein KaiA homolog	500	1
sll0947	lrtA, light repressed protein A homolog	445	2
sll0757	purF, amidophosphoribosyltransferase	456	2
sll1632	ftsQ, hypothetical protein	519	2
slr1347	ccaA/icfA/cab/cca, beta-type carbonic anhydrase localized in the carboxysome	542	2
sll0543	Hypothetical protein	447	3
sll0517	rbp1/rbpA, putative RNA binding protein	503	3
slr0009	rbcL, ribulose bisphosphate carboxylase large subunit	277	3
slr0749	chlL, light-independent protochlorophyllide reductase iron protein subunit ChlL	463	4
sll1330	rre37/crr37, Two-component system response regulation OmpR subfamily	329	6
slr1793	talB, transaldolase	236	6
slr0623	trxA, thioredoxin	500	6
slr1634	Hypothetical protein	302	6
slr0006	unknown protein	539	6
sll1732	ndhF3, NADH dehydrogenase subunit 5 (low CO2-inducible, high affinity CO2 uptake)	397	6
ssl1633	hliC/scpB, high light-inducible polypeptide HliC, CAB/ELIP/HLIP superfamily	523	9

Bioluminescence screen of nineteen *S.* 6803 promoters in 12:12 LD cycles

Bacterial luciferase from *Vibrio harveyi* was chosen as a probe for promoter activity due to widespread application in cyanobacteria to monitor circadian rhythms [100,101,166] and increased temporal sensitivity as compared with fluorescent protein reporters [161]. Nineteen promoters were cloned in front of a *luxAB* bacterial luciferase probe and inserted into *S.* 6803 *slr0168* chromosomal neutral site (Figure 3-1). *LuxAB* integration into *slr0168* did not affect the growth rate in 12:12 LD cycles relative to wild-type (Figure A-5-1Comparison of *in vitro* versus *in vivo* bioluminescence recordings in 12:12 LD cycles determined that bioluminescence assays were not limited by $FMNH_2$ substrate availability (Figure A-5-2).

Promoter:probe *S.* 6803 cultures were entrained to 12:12 LD cycles for two days in biological duplicate. Bioluminescence was measured every 2-3 hours on day 3 (linear phase, low OD_{730}) and day 8 (stationary phase, high OD_{730}) post-inoculation (Figure 3-2 and Table A-3). Twelve promoters produced no measurable bioluminescence above the wild-type background. Many hypotheses could explain the high negative rate for promoters tested here. Genes natively controlled by polycistronic operons or regulator elements in upstream genes would be missed by this screen. Genes with regulator regions within the native gene ORF or outside the selected promoter region would also be missed in this screen.

Three promoters produced low-level bioluminescence on day 3 but not day 8. Four promoters produced light-entrained bioluminescence induction on both days 3 and 8. These four promoters—P_{hliC}, P_{rbp1}, P_{sigA}, and $P_{slr0006}$—were selected for further characterization.

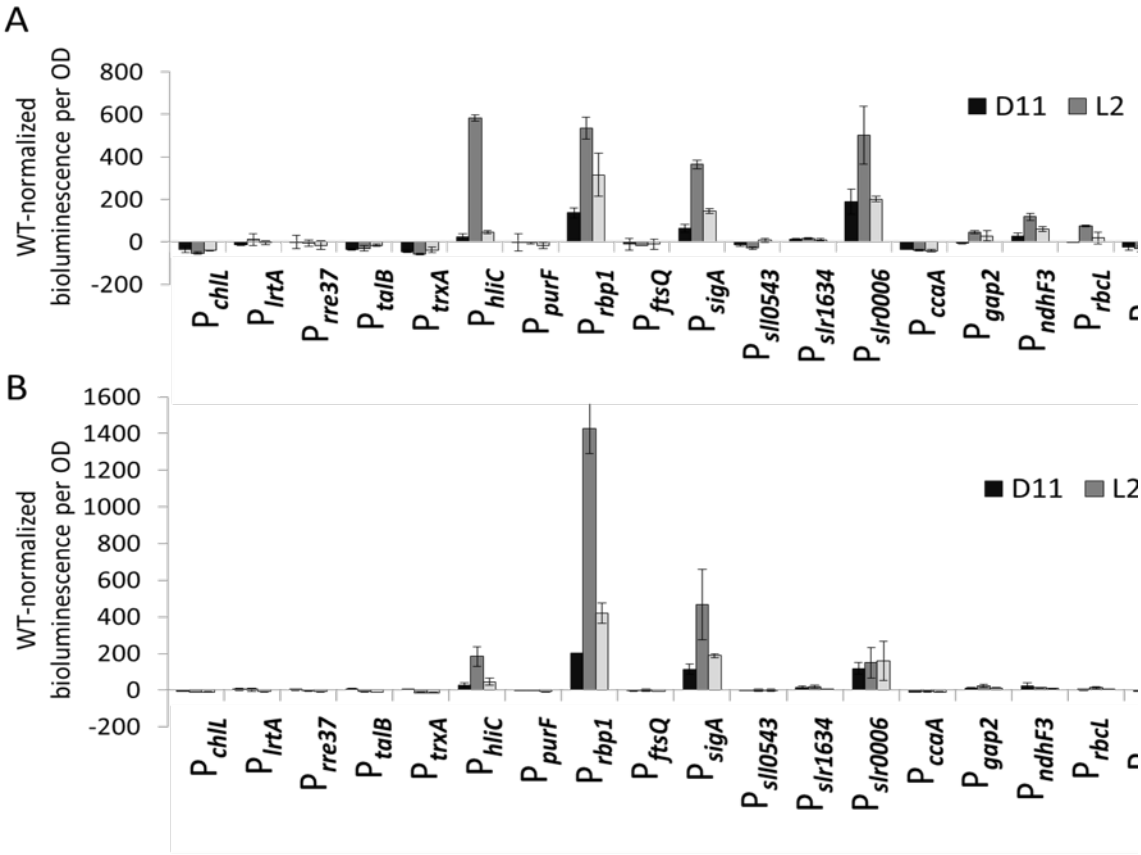

Figure 3-2. Initial screen for functional promoters in 12:12 LD cycles. Bioluminescence from promoter:luxAB strains were measured and normalized to WT background on (A) day 3 and (B) day 8 post-inoculation. Error bars represent standard deviation from biological duplicates and technical triplicates. D11: 11th hour of darkness, L2: 2nd hour of light; L5: 5th hour of light.

We were initially surprised to see light-induced activity from P_{sigA} as it is widely viewed as a housekeeping gene with a constitutive promoter. However, other housekeeping genes *secA, petB,* and *rpoA* have cycling transcripts in diurnal LD cycles [164], and transcriptomics data suggest a global increase in transcription at the onset of day [85]. This is in contrast to our reference gene (*rnpB*) which had uniformly expressed transcripts in our RT-qPCR data (Table A-4), consistent with previously published microarray data [85].

Characterization of light-entrained promoters P_{rbp1}, $P_{slr0006}$, P_{hliC}, and P_{sigA} in 12:12 LD cycles

To further characterize light-entrained promoters P_{rbp1}, $P_{slr0006}$, P_{hliC}, and P_{sigA}, we measured bioluminescence activity in biological triplicate over three consecutive days of growth in 12:12 LD cycles. Five milliliters of fresh BG-11+Kan_{50} was added daily at L11; this maintained a constant culture volume and kept cultures in linear growth phase (Figure A-5-3). Bioluminescence was measured on days 4, 5, and 6 post-inoculation (Figure 3-3). The fold-change bioluminescence induction at the onset of light was generally consistent across days 4-6. Averaged across the three days, P_{rbp1}, $P_{slr0006}$, P_{hliC}, and P_{sigA} exhibited a 5.0, 3.9, 5.0, and 4.0 fold-change increases in bioluminescence at the onset of light, respectively.

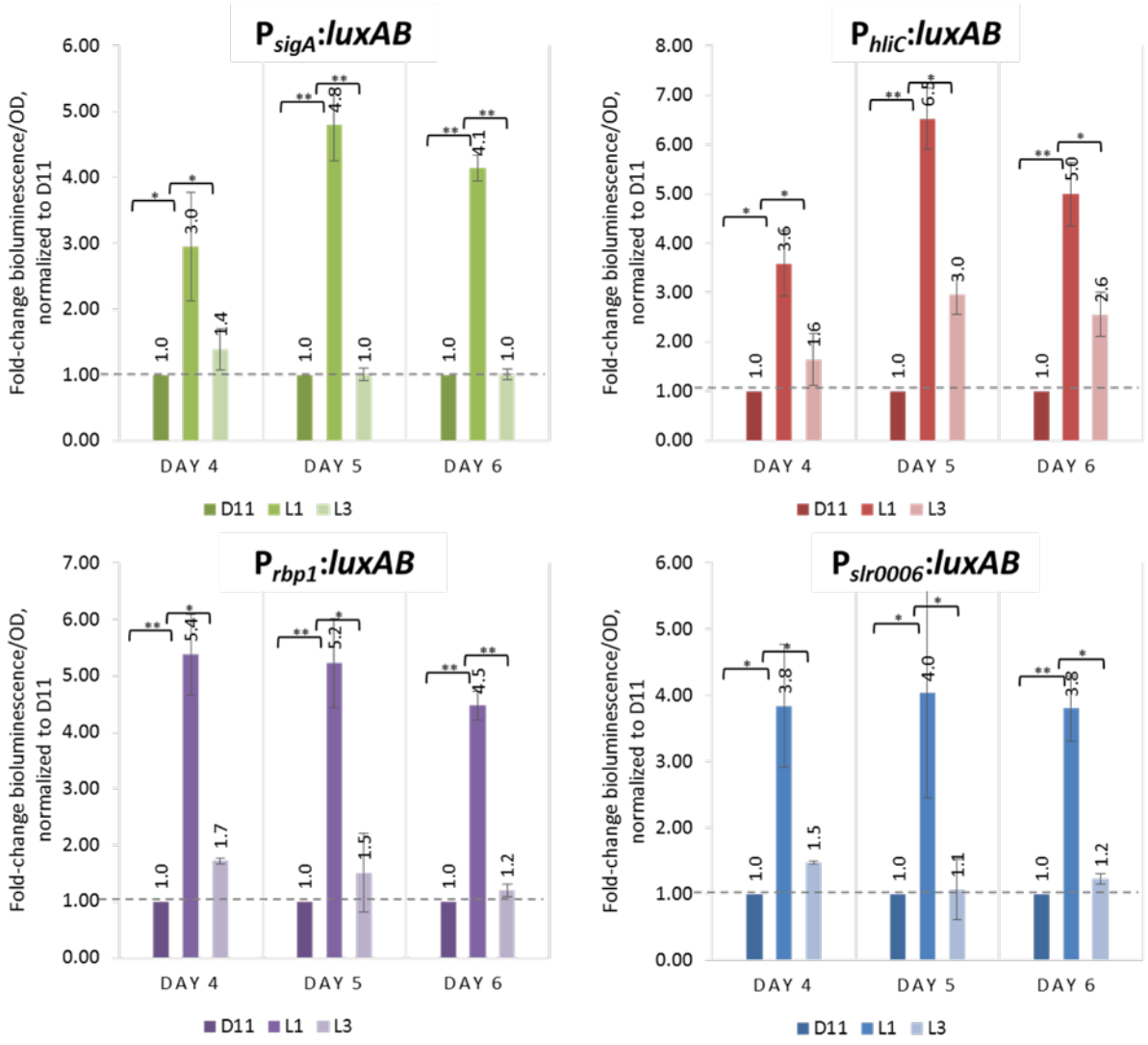

Figure 3-3. Characterization of P_{sigA}, P_{hliC}, P_{rbp1}, and $P_{slr0006}$ as light-induced promoters in 12:12 LD cycles. Error bars represent standard deviation of three biological replicates. Paired t-tests for statistical significance indicated by * for $p<0.05$ and ** for $p<0.005$.

Multiple sequence alignment of P_{sigA}, P_{rbp1}, and $P_{slr0006}$ with CLUSTALW [167] revealed two potential conserved sequences at -10 and -35 of each respective gene (Figure A-5-4). Inclusion of P_{hliC} disrupted alignment patterns. Virtual Footprint [20] analysis identified 50+ prospective prokaryotic TFs or sigma factor binding sites in each promoter. None of these motifs are from cyanobacteria databases, so it is difficult to identify true promoter and/or enhancer

motifs. Further studies are necessary to determine what motifs may play a role in light-induced transcriptional activation.

LuxAB transcript accumulation was analyzed with reverse-transcriptase real-time polymerase chain reaction (RT-qPCR) in 12:12 LD cycles. *LuxAB* mRNA from P_{sigA}, P_{rbp1}, and $P_{slr0006}$ was significantly induced at the onset of light (Table 3-3), corresponding to the observed increase in bioluminescence at this time-point. However, *luxAB* mRNA driven by P_{hliC} did not significantly increase at the onset of light (Table 3-3). The observation is not entirely surprising as transcript and protein accumulation do not always correlate [81] and 5' UTRs can regulate expression post-transcriptionally [168]. Furthermore, this observation agrees with Beck *et al.* microarray data: *hliC* (*ssl1633*) transcript abundance did not change significantly in 12:12 LD cycles [85]. *RnpB* transcripts from WT *S.* 6803 did not increase from dark to light indicating that our reference gene was suitable for these conditions (Table A-4).

Our results reveal two distinct groups with different gene expression control. P_{sigA}, P_{rbp1}, and $P_{slr0006}$ had similar characteristics: bioluminescence was strongly induced at L1 but returned to baseline by L3; mRNA levels increased at L1, similar to the levels of protein activity; and promoter sequence alignment revealed putative -10 and -35 elements. P_{hliC} had differing characteristics: bioluminescence was strongly induced at L1 and more slowly returned to baseline; mRNA levels did not significantly change at the onset of light; and promoter sequence was not similar to the other three analyzed. These observations suggest P_{sigA}, P_{rbp1}, and $P_{slr0006}$ may be controlled transcriptionally, while P_{hliC} may be controlled post-transcriptionally.

Table 3-3. Transcript abundance in 12:12 LD cycles as measured by RT-qPCR. The target gene was normalized to rnpB (ΔCt). Data is presented from four time points. Light-induced mRNA accumulation was calculated as fold-change ($2^{-\Delta\Delta Ct}$) from D11 to L1. Standard deviation (STDEV) is from three biological replicates.

Threshold cycle, ΔCt ± STDEV						**Fold-change ($2^{-\Delta\Delta Ct}$)**
***S.* 6803 strain**	**Target gene**	**D9**	**D11**	**L1**	**L3**	**D11 to L1**
P_{hliC}:*luxAB*	*luxAB*	10.9 ± 1.1	12.8 ± 1.4	12.4 ± 0.5	12.0 ± 0.7	**1.3**
P_{rbp1}:*luxAB*	*luxAB*	12.8 ± 0.2	13.5 ± 0.2	9.0 ± 0.3	8.9 ± 0.7	**23.3**
$P_{slr0006}$:*luxAB*	*luxAB*	15.1 ± 0.8	16.1 ± 0.2	13.5 ± 0.7	13.3 ± 0.9	**6.0**
P_{sigA}:*luxAB*	*luxAB*	13.4 ± 0.4	13.9 ± 0.6	10.0 ± 1.9	12.0 ± 0.6	**15.5**

P*hliC* light-induced bioluminescence increases at higher light intensities

We were intrigued by the observation that P_{hliC}:*luxAB* light-induced bioluminescence decreased from 26.4-fold on day 3 to 6.0-fold on day 8 (Figure 3-2). On day 3, cultures were dilute (Abs_{730}=0.4-0.5) compared to day 8 (Abs_{730}=1.0-1.1). Dilute cultures were likely receiving a higher amount of incident light than dense cultures affected by cell-to-cell shading. Thus, the decreased response on day 8 could be due to decreased incident light or decreased light stress. *S.* 6803 has several annotated high-light inducible (Hli) proteins which help mitigate high-light stress by non-photochemical quenching [169,170] or facilitating photosystem stability or repair [171,172].

We tested if P_{hliC} light-induced bioluminescence increases with increasing light by measuring bioluminescence in 12:12 LD cycles with mid (196 μE) and high (500 μE) light. Both sets of cells were inoculated from the same pre-culture entrained in 12:12 LD cycles with 196 μE light. We observed a 27.7-fold increase in bioluminescence at the onset of light in the high-light exposed cells as compared to a 2.93-fold change increase in the low-light exposed cells (Figure

3-4). Further studies are necessary to understand how P_{hliC} regulates expression in changing light intensities.

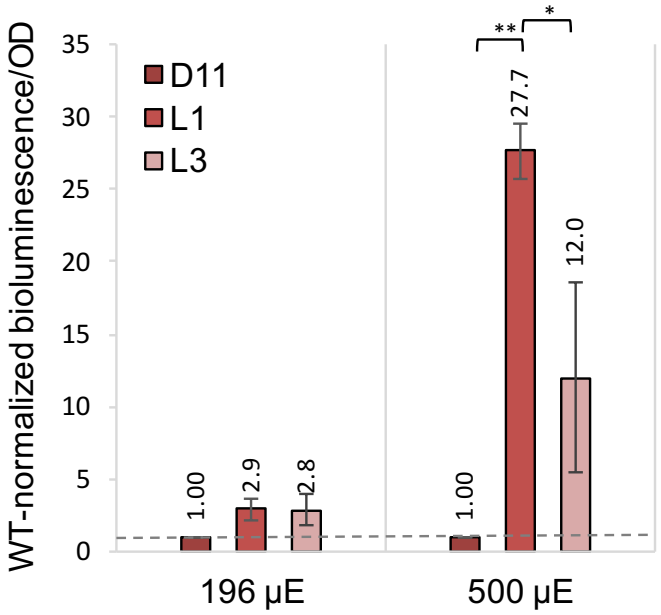

Figure 3-4. P_{hliC} light-induced response increases at higher light intensities. Error bars represent standard deviation from three biological replicates. Paired *t*-tests for statistical significance indicated by * for $p<0.05$ and ** for $p<0.005$.

Light-entrained promoters do not cycle in continuous light

To be classified as a true circadian rhythm, oscillations must persist in continuous light. We assessed bioluminescence in continuous light (CL) after both 12:12 LD and CL entrainment. After two days of 12:12 LD cycle entrainment, *S.* 6803 promoter:*luxAB* strains were transferred to CL. We did not observe free-running rhythms of bioluminescence in continuous light (Figure 3-5). This result is consistent with previous studies indicating that the majority of *S.* 6803 transcript oscillations are lost in CL [85,173]. However, in CL-entrained strains we observed constant bioluminescence from all four promoters at varying intensities (Figure A-5-5).

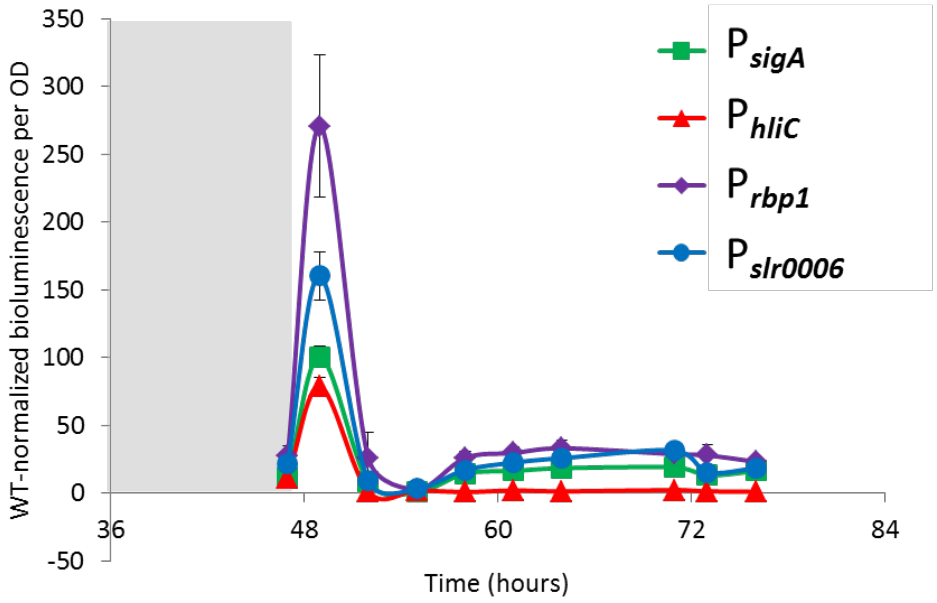

Figure 3-5. Promoter:probe characterization in continuous light. P_{sigA}, P_{hliC}, P_{rbp1}, and $P_{slr0006}$ do not exhibit sustained light-entrained bioluminescence oscillations in continuous light. Error bars represent standard deviation from three biological replicates.

Transcripts increase at the onset of dawn in non-bioluminescent strains P_{kaiA} and P_{rbcL} (unpublished data)

Fifteen of the nineteen promoter:*luxAB* strains produced little or no bioluminescence in diurnal LD cycles. Several hypotheses could explain this observation, including: (1) *cis*-acting "promoter" elements beyond the selected DNA are required to initiate transcription; (2) *trans*-acting elements (*e.g.* ncRNA, transcription factors) interact within the native ORF or 3' UTR and are required for transcript control/stability; and/or (3) *luxAB* mRNA was stably expressed but not translated due to the presence/absence of regulatory elements (*e.g.* ribosome binding sites, unfavorable secondary structure).

To test these hypotheses, we investigated mRNA accumulation in $P_{rbcL:luxAB}$ and $P_{kaiA:luxAB}$, corresponding to circadian clock protein A homolog and Rubisco large subunit, respectively. Both genes had dramatic transcript oscillations in Beck *et al.* (2014) 12:12 LD cycle microarray data [85], but their respective promoters produced no measureable

bioluminescence in our 12:12 LD bioluminescence study. We used RT-qPCR to measure *luxAB* mRNA as well as native gene mRNA (*kaiA* and *rbcL*) at D9, D11, L1, and L3 in 12:12 LD cycles (Table 3-4).

Table 3-4. Transcript abundance from non-bioluminescent P*kaiA* and P*rbcL* strains in 12:12 LD cycles. *LuxAB* mRNA and the native gene driven by each promoter (*kaiA* and *rbcL*) were measured by RT-qPCR. *LuxAB* transcripts increase at the onset of light (D11 to L1), but not to the extent that native gene transcripts increase. Standard deviation is shown across three biological replicates.

Threshold cycle, ΔCt ± STDEV						Fold-change ($2^{-\Delta\Delta Ct}$)
***S.* 6803 strain**	**Target gene**	**D9**	**D11**	**L1**	**L3**	**D11 to L1**
$P_{kaiA:luxAB}$	*luxAB*	17.0 ± 1.1	18.2 ± 0.2	15.4 ± 0.4	16.9 ± 0.9	**2.9**
$P_{kaiA:luxAB}$	*kaiA*	22.1 ± 1.3	20.7 ± 0.9	17.4 ± 0.8	17.4 ± 1.7	**10.5**
$P_{rbcL:luxAB}$	*luxAB*	15.7 ± 0.4	17.0 ± 0.7	14.3 ± 0.3	15.4 ± 0.1	**6.3**
$P_{rbcL:luxAB}$	*rbcL*	13.7 ± 0.5	14.7 ± 0.6	10.3 ± 0.3	14.4 ± 0.8	**22.1**

We observed significant *luxAB* transcript increases at the onset of light in P_{kaiA} and P_{rbcL} strains. Interestingly, native *kaiA* and *rbcL* transcripts had even higher induction at the onset of light. These observations support the hypothesis that *luxAB* mRNA accumulates at the onset of light, but this message is not translated into functional bacterial luciferase protein due. Furthermore, *luxAB* mRNA induction is diminished from the native gene mRNA induction in both cases, supporting decreased mRNA induction or stability. Elucidating the disconnect between native gene expression and engineered *luxAB* expression, as well as the disconnect between engineered bacterial luciferase transcript and protein expression, could provide broad insights into gene expression and work toward more accurate engineering design.

CONCLUSIONS

This study sought to develop genetic engineering tools useful for cyanobacterial strains for the industrially-relevant condition of daily light:dark cycles. We discover and characterize four promoters which provided light-entrained expression from the *slr0168* neutral site using the bioluminescence reporter *luxAB*. $P_{hliC,}$ $P_{slr0006}$, P_{sigA}, and P_{rbp1} provide 5.0, 3.9, 4.0, and 5.0 fold-change increases in bioluminescence at the onset of light, respectively. *LuxAB* transcript abundance increases in $P_{slr0006}$, P_{sigA}, and P_{rbp1} by 6.0, 15.5, and 23.3 fold-change, respectively, from the *slr0168* chromosomal neutral site. Interestingly, *luxAB* transcripts from P_{hliC} did not significantly change at the onset of light, suggesting a potential role of post-transcriptional regulation. P_{hliC} bioluminescence induction increased significantly with increasing light intensity. None of the strains exhibited free-running bioluminescence rhythms in continuous light following 12:12 LD cycle entrainment, but all produced constant bioluminescence in continuous light following continuous light entrainment. Ineffective expression from the majority of tested promoters (81%) suggests mechanisms of gene control outside the 5' untranslated region. RT-qPCR analysis of two of these promoters, P*kaiA* and P*rbcL*, revealed that both *luxAB* and native (*kaiA* and *rbcL*) transcripts increase at the onset of light in diurnal LD cycles. The transcript-protein disconnect is interesting and demonstrate that promoters alone are not sufficient to control protein expression in many cases. Together, this work expands the toolset available for cyanobacteria strain engineering in the industrially-relevant condition of daily LD cycles.

CONTRIBUTIONS

This work was supported by National Science Foundation [Award #1332404 and #1336236]. A.W. and C.P. conceived, planned and designed the research. A.W., K.O., and D.O performed promoter:probe cloning. A.W. performed experiments, analyzed data, and drafted the

manuscript. C.P., J.S., revised the manuscript for intellectual content. All authors have read and approved the final manuscript.

CHAPTER 4 – METABOLOME CHARACTERIZATION OF THE PHOTOSYNTHETIC CYANOBACTERIUM *SYNECHOCYSTIS* SP. PCC 6803 ACROSS A DAY/NIGHT LIGHT CYCLE[9]

SUMMARY

Cyanobacteria are studied and engineered as a model photoautotroph and a chassis for the sustainable production of fuels and chemicals. Knowledge of photoautotrophic metabolism in the natural environment of day/night cycles is lacking yet has implications for improved yield from plants, algae, and cyanobacteria. Model cyanobacterium *Synechocystis* sp. PCC 6803 was cultivated under day/night light cycles with sinusoidal light patterns (sinLD) peaking at 1,600 μmol photons m^{-2} s^{-1}. Metabolite accumulation was profiled every 30-120 minutes over a 24-hour diurnal sinLD cycle. Growth lagged for ~2 hours after transitioning to day followed by a ~4 hour period of rapid biomass accumulation then cell division. During the lag phase, amino acids (AA) and nucleic acids (NA) accumulated to high levels. The biomass accumulation phase was characterized by AA incorporation into proteins and decreased levels of free NAs. Insoluble carbohydrates displayed sharp oscillations at the day-to-night transition. Insoluble galactosamine and xylose display a sharp decrease in accumulation at the night-to-day transition, evocative of a 're-set', in addition to the day-to-night oscillation. Pyruvate, acetyl-CoA, and alpha-ketoglutarate—key intermediates of central carbon metabolism—maintained stable levels across the day/night cycle. High phosphoenolpyruvate accumulation during the early day suggests a

[9] This chapter has been submitted for publication as: A. Werner, C. Broeckling, C.A.M Peebles. Metabolome characterization of the photosynthetic bacterium *Synechocystis* sp. PCC 6803 across a day/night light cycle.

potential bottleneck in central carbon metabolism. Together these data illustrate widespread changes in metabolite availability and cellular objective across a day/night light cycle.

SIGNIFICANCE STATEMENT

Photosynthetic organisms have evolved under day/night cycles of sunlight energy availability where fitness depends on maximally utilizing light energy during the day. Cyanobacteria are the progenitors of the chloroplast and serve as model organisms for understanding fundamental aspects of photosynthesis and photoautotrophic metabolism. Here, we present a comprehensive time-course characterization of metabolite accumulation across a realistic day/night light cycle. We discovered a diverse range of metabolite classes which display cyclic metabolite accumulation patterns under day/night light cycles, including striking carbohydrate oscillations at the day/night transition. We present a timeline of observations and hypothesized cellular objectives. Our findings provide insight into the temporal dynamics of day/night photosynthetic metabolism with implications in agricultural productivity and cyanobacterial bioengineering.

INTRODUCTION

Photosynthetic organisms—including plants, algae, and cyanobacteria—utilize sunlight as an energy source to fix atmospheric CO_2 into organic materials. Photosynthetic organisms are an attractive alternative to heterotrophs due to the promise of sustainable production of food, fuels and chemicals. Cyanobacteria have been widely engineered for the production of a variety of products [174–176] under continuous light (CL) laboratory conditions. Yet, industrial production will require outdoor cultivation exposing strains to natural day/night cycles of sunlight availability [177]. Productivity of engineered *Synechocystis* sp. PCC 6803 (hereafter *S.*

6803) is decreased in diurnal light:dark (LD) cycles as compared with CL [93]. Thus, it is important to understand how metabolism changes under realistic light conditions to appropriately engineer industrial production strains [178].

Phototrophs have evolved under natural day/night LD cycles of sunlight availability where fitness depends on the efficient and timely capture of photons. Acclimation to changing light quality [179] and quantity [180] have been studied extensively, as have circadian rhythms in plants [181,182] and cyanobacteria [100,101]. The cyanobacterial circadian clock has been elucidated in *Synechococcus elongatus* sp. 7942 (hereafter *S.* 7942) [183,102,184,185], entraining cell behavior to the external light environment. In *S.* 6803, the accumulation of many transcripts [81,85] and glycogen [81,82,89,90] are synchronized to diurnal LD cycles. Global metabolism across diurnal LD cycles is relatively unexplored in any phototroph, though evidence of circadian control was recently presented in *S.* 7942 [83].

Here, we advance the study of photosynthetic metabolism under diurnal LD cycles in two areas. First, we cultivate cyanobacteria under realistic LD cycles with sinusoidal LD (sinLD) cycles which peak at super-high light (1,600 μE) in a custom photobioreactor. To our knowledge, this is the first report of cyanobacteria growth under realistic diurnal sinLD cycles. Second, we expand the scope of metabolites studied, utilizing a comprehensive multi-platform mass spectrometry (MS) workflow which enabled characterization of polar and non-polar metabolites from the soluble and insoluble phases. We employed very high temporal frequency of sampling, characterizing all metabolites every 30-120 minutes for 26 hours. To our knowledge, this is the first comprehensive metabolite report for any photosynthetic organism under diurnal LD cycles. We report daily oscillations in abundance for metabolites spanning a range of metabolic pathways, including amino acid biosynthesis, nucleotide and cofactor

metabolism, insoluble carbohydrates, and central carbon metabolism. The hypotheses generated in this study have important implications for both engineering and understanding the metabolism of photosynthetic organisms.

RESULTS AND DISCUSSION

Cell growth in sinLD cycles is characterized by lag, mass accumulation, and cell division phases

S. 6803 was cultivated under diurnal sinLD cycles which peaked at 1,600 μmol photons $m^{-2}s^{-1}$ (μE) by white LEDs (Figure 4-1A) and bubbled with 5% CO_2 in a custom photobioreactor (Figure B-5-6). Cultures were entrained to sinLD cycles for three days prior to sampling for metabolomics to synchronize the cell population (Figure B-5-7). Cell growth was measured by optical density (OD, Abs_{730nm}) and cell density (cells per microliter), where OD provides a rough estimate of biomass independent of photosynthetic pigments [186]. We report specific growth rate of 0.06 ± 0.01 hr^{-1} by OD and 0.07 ± 0.001 hr^{-1} by cell count on the fourth day of sinLD entrainment (Table B-5).

Growth measurements across the sinLD cycle suggest the following sequence of cell phases: a lag phase (Zeitgeber time "ZT" 0-2), biomass accumulation (ZT 2-6), cell division (ZT 6-12), and biomass loss in the early night likely due to respired CO_2 (ZT 12-18) (Figure 4-1B). Previous reports of cyanobacteria growth in sinLD cycles are limited as much of the work has been in a chemostat-like manner [85,82,83], or were not reported with enough temporal sensitivity to discern the timing of cell division [93,81,90]. Still, several studies report increased OD during the day in diurnal LD cycles [90,92,94,187] and circadian studies of *Synechococcus elongatus* PCC7942 report cell division during the circadian day [188]. These reports are in

agreement with our observations. In contrast, the marine cyanobacterium *Prochloroccus* has been reported to divide during the night under sinLD (~200 μE peak) cycles [86,189].

We extracted soluble non-polar, soluble polar, and insoluble polar compounds using a biphasic extraction workflow (Figure 4-1C). Five mass spectrometry (MS) platforms were used to detect metabolites in each phase including both targeted and non-targeted analyses (Figure 4-1C). Across all five MS platforms, we detected 4,900 compound signatures and annotated 311 metabolites (SI_file1). Of the annotated compounds, 91%, 94%, and 40% of the aqueous, organic, and insoluble pellet, respectively, were identified as significantly changing by linear or quadratic regression (SI_file3, Table B-6, Table B-7). Unbiased principal component analysis (Figure B-5-8) and hierarchical cluster analysis (Figure B-5-9) confirm the observation that time across the sinLD cycle contributes to the majority of underlying variation. Annotated metabolites represented in chemical databases (KEGG, PubChem, and HMBD) were analyzed further and discussed by metabolic pathway in the following sections.

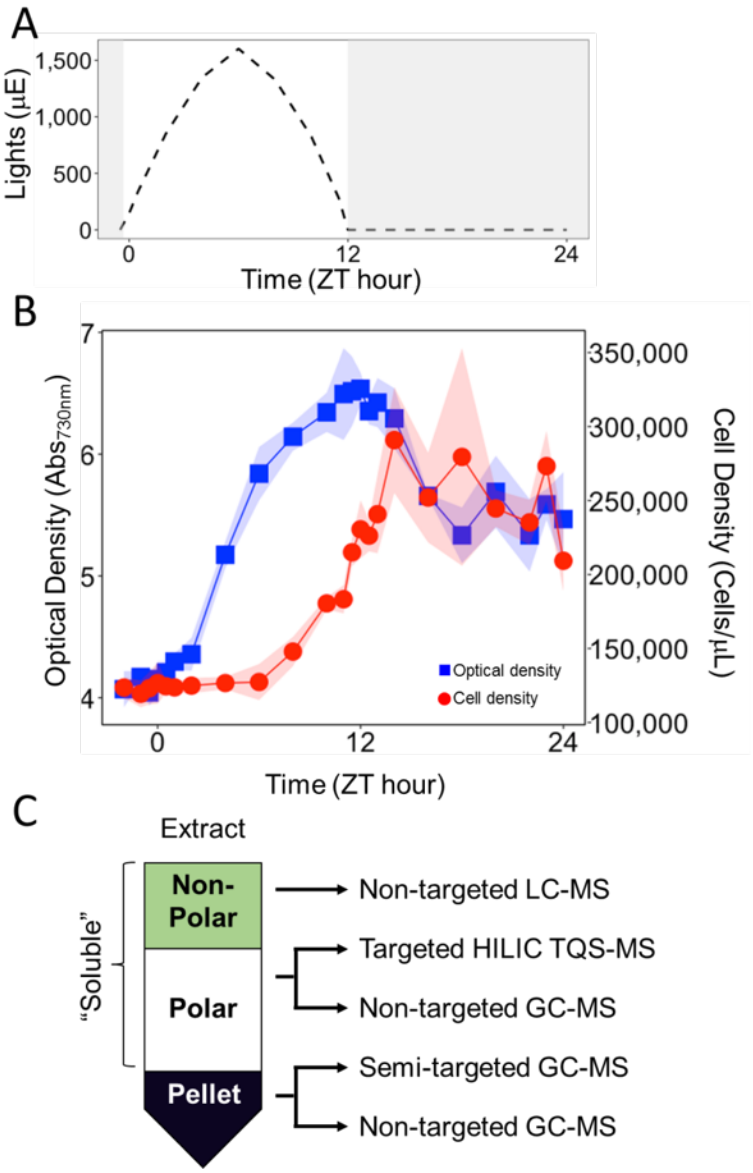

Figure 4-1. S. 6803 growth over the course of a sinusoidal day/night light cycle. (A) Diurnal sinusoidal photon flux pattern (micromoles photons per m^2s). (B) Cell growth as measured by optical density (blue squares) and cell density (red circles). Shaded regions represent standard deviation among three biological replicates. (C) Comprehensive MS detection of metabolites.

Soluble amino acids rapidly accumulate at the onset of day, followed by protein incorporation

Soluble and pellet-bound amino acids (AA) were measured every 0.5-2 hours by non-targeted GC-MS (Figure 4-2, Figure B-5-10). Free AAs increase within the first 30-120 minutes of the day. Insoluble (pellet-bound) AAs gradually increase during the early day, decline in

tandem with cell division during the late day, and remain steady throughout the night. We hypothesize the decline in free AAs per-cell from ZT 2-4 is due to incorporation into proteins, and the decline in protein-bound AAs during ZT 6-12 is due to cell division.

The dramatic increase in free AAs which would require substantial resources to synthesize is surprising. Transcriptomic studies of cyanobacteria under daily LD cycles report peak accumulation of amino acid metabolism genes from ZT -1-3 [81], and ribosomal genes from ZT 0-6 [85]. These observations support the hypothesis that AA synthesis is a primary objective during the first hours of the day followed by protein synthesis. Liberation from proteins via a salvage pathway is an alternative hypothesis to explain the observations; protein-bound AAs only partially support this hypothesis. Pellet-bound aspartic acid and arginine accumulation follows the same pattern as all other protein-bound AAs (Figure 4-2). This observation suggests cyanophycin does not accumulate under the nutrient-replete linear-growth conditions, which is in agreement with the current understanding of cyanophycin as a nitrogen reserve synthesized in response to nutrient limitation or stagnant growth [190,191].

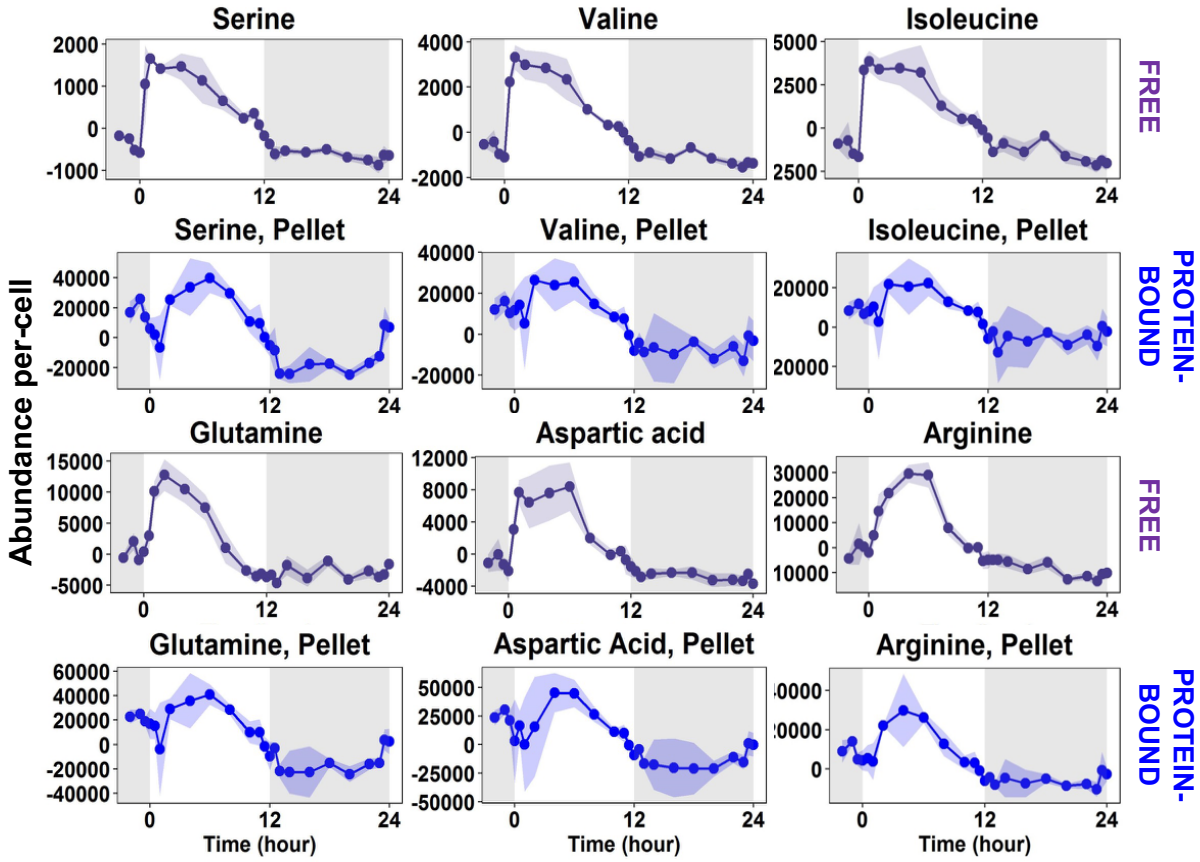

Figure 4-2. Free and protein bound amino acid accumulation over the day/night light cycle. Amino acid abundance per cell is pareto scaled. Periods of darkness are represented with a shaded background. The sinusoidal profile in Figure 4-1A is applied during the light phase. Soluble amino acids are shown in purple and protein-bound amino acids detected from the hydrolyzed pellet are shown in blue (denoted as 'Pellet' in title). Shaded regions represent the standard deviation across three biological replicates.

Insoluble carbohydrate content displays sharp oscillations at the day/night transition

We detected a variety of carbohydrates from the insoluble pellet which display temporal oscillations, including sharp oscillations at the day/night transitions (Figure 4-3). Insoluble C6 carbohydrates display a pattern of gradual increase in accumulation during the day and decrease in accumulation during the night, with a sharp oscillation immediately before/after the day-to-night transition. Insoluble C5 carbohydrates and galactosamine display a slightly different pattern of consistent levels over the day/night cycle with sharp oscillations at both the day-to-night and night-to-day transitions. Soluble carbohydrates do not display these oscillations (Figure B-5-12).

Insoluble carbohydrates in cyanobacteria could be associated with intracellular "capsular" polysaccharides (CPS, *e.g.* glycogen), extracellular polysaccharides (EPS) [192,193], glycosylated lipids [194,195], and glycosylated proteins [196]. Due to the pooled and hydrolyzed methods employed in this study, associated entities can only be hypothesized; further, MS does not enable accurate structural determination. Development of a method which separates CPS, EPS, lipids, and protein from the pellet, paired with an analytical technique which can determine linkages and structures, will be necessary to test the hypotheses presented. Nonetheless, we present some likely hypotheses based on our observations and previous knowledge.

Insoluble glucose content is likely attributable to glycogen. The temporal dynamic we observe is similar to previous reports of glycogen increasing during the day and decreasing during the night [81,83]. Glycogen dynamics as measured by fluorescent kit displayed (). Insoluble galactose is likely attributable to glycerolipid polar heads [194,195] or EPS [192], which contributes to increased cell buoyancy and iron sequestration [193].

Glycoproteins are a possible but unexplored sink for insoluble carbohydrates. Glycoproteins have been reported in cyanobacteria [196] among other bacteria [197]. *S.* 6803 harbors many putative glycosyltransferases including a GlcNAc transferase (AOY38_15715), which suggest glycosylation of proteins is biochemically feasible. Galactosamine is a glycoprotein modification in bacteria [198], though it has not been studied in *S.* 6803. The sharp oscillations in insoluble C5 carbohydrates and galactosamine at the night-to-day transition could suggest a glycoprotein regulatory event. Investigating this hypothesis would be an interesting area for future work.

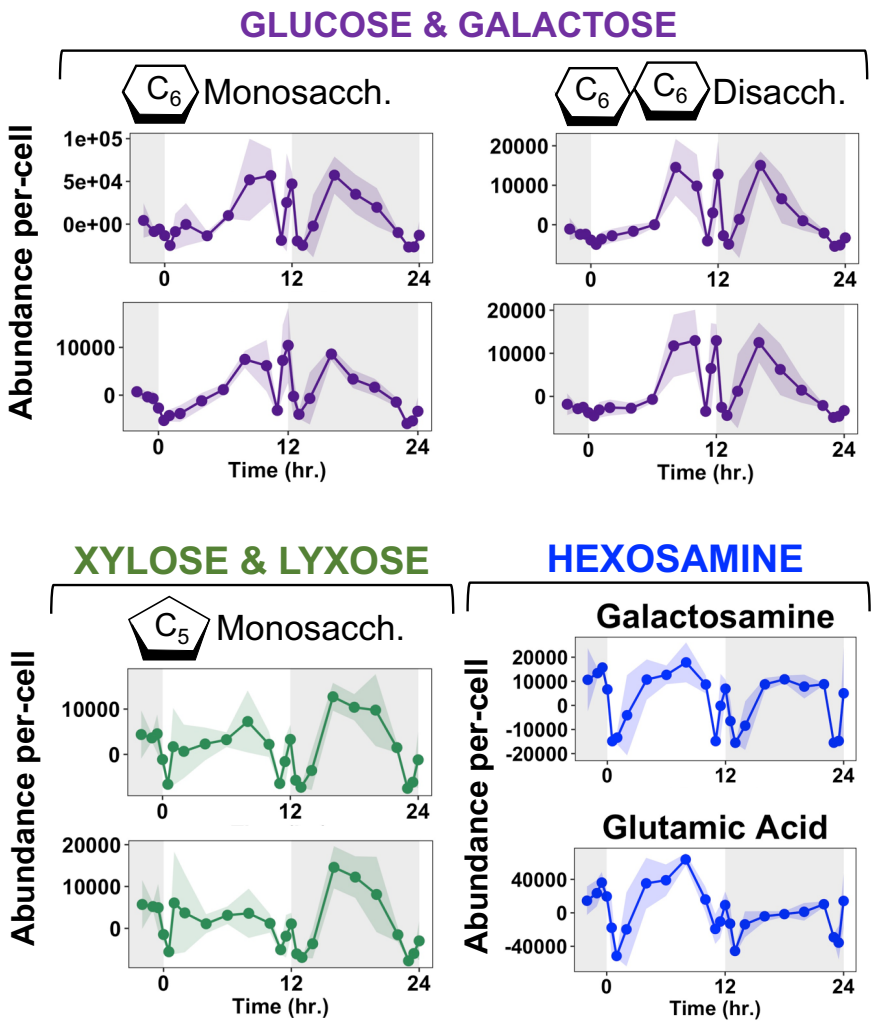

Figure 4-3. Insoluble carbohydrate oscillations over the day/night light cycle. Abundance per cell is pareto scaled. Periods of darkness are represented with a shaded background. The sinusoidal profile in Figure 4-1A is applied during the light phase.

Nucleic acid accumulation peaks early in the day

Nucleosides, nucleotides, NDPs, and NTPs were detected from the pellet and soluble phase by non-targeted GC-MS and targeted LC-TQS-MS, respectively. We observe an early-day spike in accumulation of C, U, G, and I nucleic acids (Figure 4-4). Pellet-bound purines displayed increased accumulation peaks near mid-day, while pellet-bound pyrimidines displayed relatively constant accumulation (Figure B-5-11).

Our results suggest immediate production of NTPs (ZT 0-2) in preparation for early-day DNA replication (ZT 2-6) prior to cell division (ZT 6-12). Reports of *S.* 7942 replication during the lag phase upon transition from dark to light support this hypothesis [199,200]. We observe sharp accumulation of IMP at the onset of light, which has also been reported in *S.* 7942 [199]. IMP is synthesized by NO• or enzymatic deamination [201,202] and is involved in RNA editing, including RNAi and viral defense [203,204]. These pathways are unknown in *S.* 6803, though the presence of deaminases (AOY38_14455, AOY38_04905) and a non-canonical purine NTP pyrophosphatase (AOY38_10185) suggest controlled enzymatic IMP synthesis is biochemically feasible in *S.* 6803.

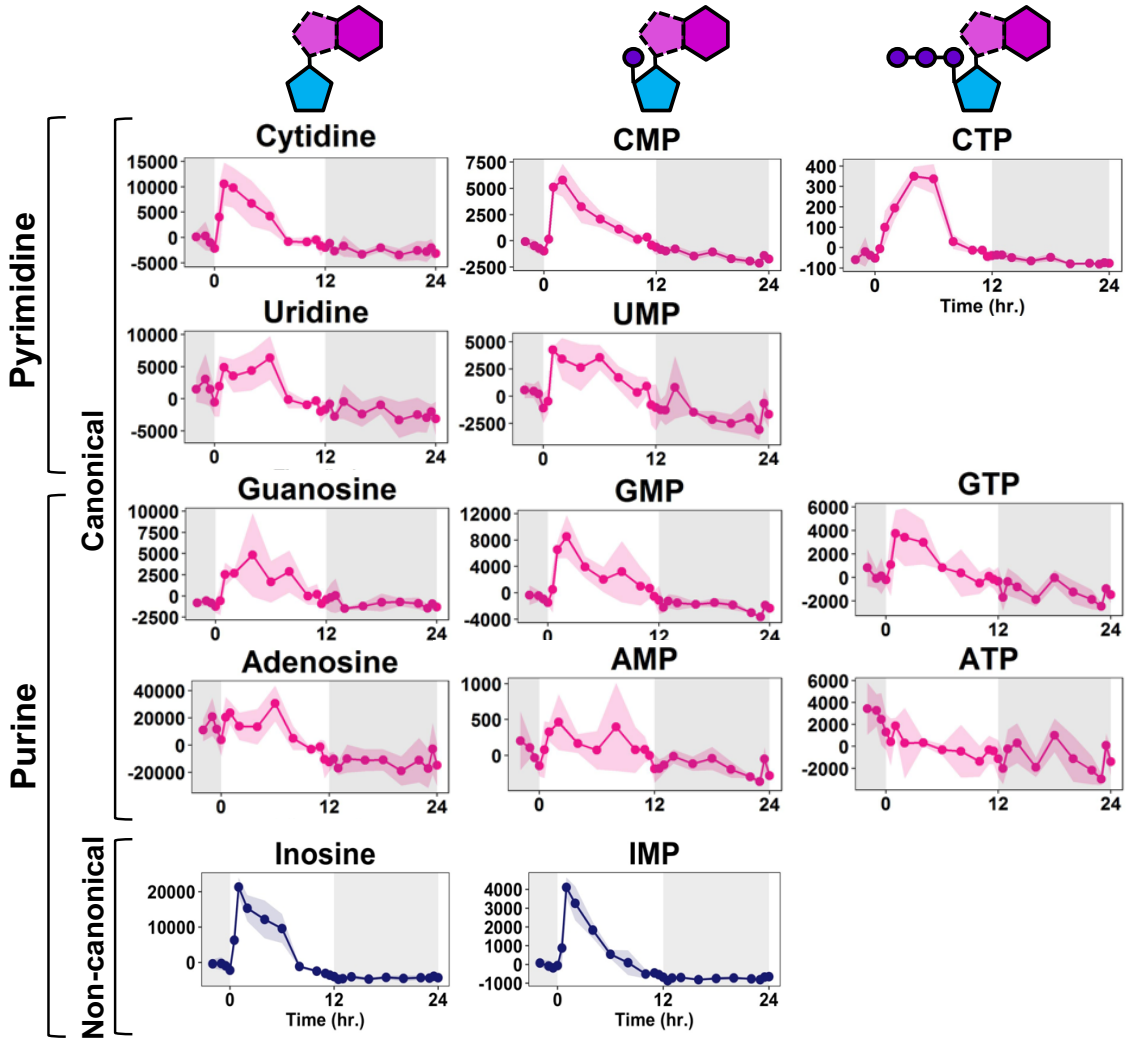

Figure 4-4. Nucleic acid day/night dynamics. Amino acid abundance per cell is pareto scaled. Periods of darkness are represented with a shaded background. The sinusoidal profile in Figure 4-1A is applied during the light phase.

Other metabolites display dampened or less consensus diurnal accumulation patterns

Generally, we observe dampened metabolite accumulation within the Calvin Benson Bassham (CBB) pathway, photorespiration (RESP), and the tricarboxylic acid (TCA) cycle (Figure 4-5). Surprisingly, pyruvate (PYR), acetyl-CoA (ACA), and 2-ketoglutarate (AKG) display relatively consistent accumulation across the day/night cycle. Accumulation of PEP, with no accumulation of PYR, could suggest a bottleneck at this conversion; a similar bottleneck from CIT/Cis-Aconitate to AKG is also observed.

Monogalactosyldiacylglycerol (MGDG), digalactosyldiacylglycerol (DGDG), sulfoquinovosyldiacylglycoerol (SQDG) and phosphatidylglycerol (PG)—the main constituents of thylakoid and cell membranes—displayed varying oscillations with no consensus pattern by class or alkyl chain length (Figure B-5-14). We observe diurnal accumulation oscillations from a variety of pigments (Figure B-5-15) and cofactors (Figure B-5-16). Flavin mononucleotide (FMN) accumulation peaks during mid-day, consistent with a hypothesis of increased demand during periods of high oxidative stress [205]. GTP, a precursor for the synthesis of FMN and other cofactors, peaks early in the day to support increased demand. Adonixanthin ($C_{40}H_{54}O_3$) and alloxanthin ($C_{40}H_{52}O_2$) species displayed accumulation patterns which were similar to the sinusoidal light profile; however, not all pigment metabolites displayed this pattern.

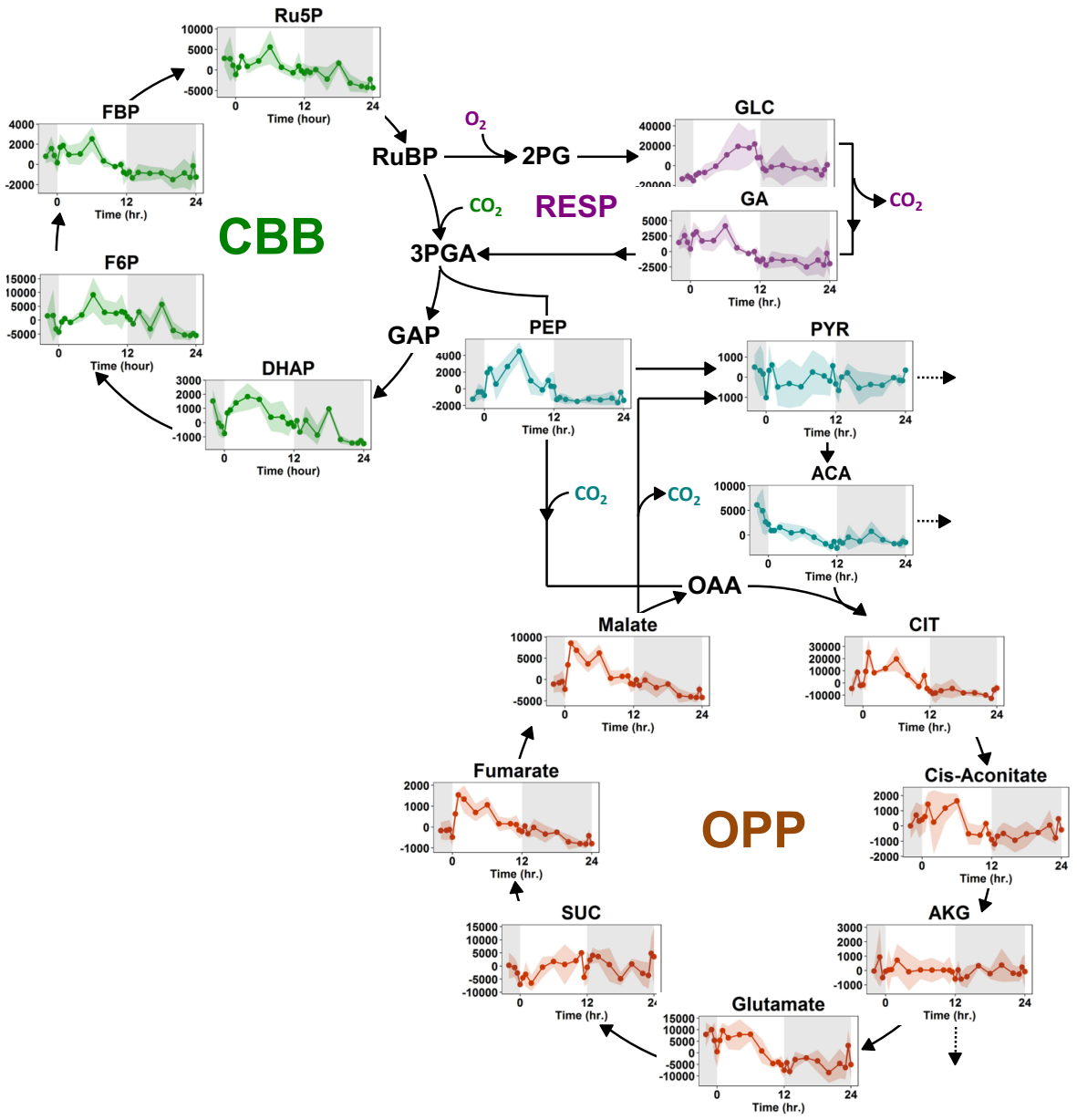

Figure 4-5. Carbon fixation (Calvin Benson Bassham, CBB) and oxidative phosphorylation (OPP) pathway dynamics. Photorespiration (RESP) intermediates are also shown. Amino acid abundance per cell is pareto scaled. Periods of darkness are represented with a shaded background. A sinusoidal profile is applied during the light phase. Dotted arrows represent utilization by pathways not displayed in the figure.

CONCLUDING REMARKS

The findings of this study can be summarized quite simply: metabolite abundances are dynamic across a day/night light cycle in cyanobacteria. Oscillations vary in pattern and

magnitude by metabolite class and biochemical pathway. Consensus patterns emerge for some, but not all, metabolic pathways. While metabolite accumulation is not a perfect proxy for metabolic activity, we formulate hypotheses for pathways which displayed consensus patterns across many metabolites. This study highlights the need for further investigation of photosynthetic metabolism under realistic day/night light cycles, with major implications spanning the fields of plant biology, agriculture, and biological engineering of photosynthetic organisms.

MATERIALS AND METHODS

Synechocystis growth conditions and measurements

Synechocystis sp. PCC 6803 WT triplicate cultures were cultivated in 1.5 L BG-11 medium (6XP) under sinusoidal diurnal LD cycles peaking at 1,600 μmol photons $m^{-2}s^{-1}$ in glass flat-panel reactors submerged in a 27-30°C water bath. Cultures were agitated and supplied with carbon via 5% CO_2 at 200 mL min^{-1}. Optical density was recorded on a NanoDrop™ at 730nm. Cell counts were performed on cells fixed in 0.25% glutaraldehyde via flow cytometry with CountBright™ beads. Cells were entrained to sinLD cycles for three days prior to metabolomics sampling (Table B-5).

Metabolite sampling, extraction, and mass spectral detection

For each metabolomics time-point: 10 mL culture were rapidly sampled and quenched in -4°C 1X PBS, spun at 3,000*g* for 5 min., decanted, quenched in liquid nitrogen, and lyophilized at -50°C. Metabolites were extracted with a 2:1:1.6 MTBE:MeOH:H_2O workflow resulting in a biphasic liquid layer and an insoluble pellet. A fraction of each phase from each sample was pooled into a quality control (QC) sample. The MTBE layer was analyzed via Q-TOF-MS with a

UPLC Phenyl Hexyl column ("RP-MS"). The MeOH:H_2O layer was split evenly and analyzed via GC-ISQ-MS and via targeted SRM analysis on TQS-MS equipped with HILIC. The insoluble pellet was split evenly, and acid hydrolyzed for (a) amino acids and (b) polysaccharides/nucleic acids; both fractions were analyzed via GC-ISQ-MS. Samples for GC-MS were derivatized by either MTBSTFA or MSTFA prior to MS runs. QC samples were injected after every 6^{th} actual sample.

MS data processing and analysis

MS peak detection, grouping, and RT correction was performed via XCMS. Quality control samples signal did not drift across injection order (Figure B-5-17), thus no instrument normalization scheme was necessary. Spectral abundances were normalized per-cell (Figure B-5-18) and clustered into putative compounds via RamClustR. GC-MS compounds were annotated using the Golm database with a requirement for both spectral and RT index matches. RP-MS compounds were assigned a molecular formula and structure using InterpretMSSpectrum and MSFinder, respectively. All presented data is pareto-scaled per-cell.

CONCLUSIONS AND FUTURE WORK

Photosynthetic organisms support life on Earth, harnessing sunlight to capture carbon dioxide and release oxygen. While plants provide the primary food source for humans and animals, algae, cyanobacteria, and other phytoplankton perform the vast majority of photosynthesis. Cyanobacteria biotechnology holds great promise as a sustainable method of producing petroleum replacements and has been engineered for the production of biofuels, polymers, pigments, and other products. Industrial production strains will be subjected to natural cycles of sunlight availability, a very different environment than standard continuous light (CL) laboratory conditions. Thus, engineering cyanobacteria under diurnal light:dark (LD) cycles is a key hurdle to overcome in order to achieve industrial viability of cyanobacteria biotechnology. Further, understanding photosynthetic metabolism in diurnal LD cycles has implications in a broad range of fields, including agricultural productivity. This thesis contributes improved genetic tools, novel materials, improved methods, and significant knowledge advancements toward the ultimate goal of realizing industrial production of commodity chemicals in cyanobacteria.

It was previously demonstrated that energy molecules ATP, NADPH, and glycogen oscillated in abundance across 12:12 LD cycles [81,83]. Further, we previously observed that engineered free fatty acids in *Synechocystis* sp. PCC 6803 (hereafter *S. 6803*) were diminished to the levels produced by wild-type in 12:12 LD cycles [93]. Together, these observations led us to hypothesize that restricting chemical production to periods of high energy availability (*i.e.* during the day) would improve yield by decreasing the metabolic burden experienced by the cell during

periods of high chemical production. One method of controlling chemical production is by inducible promoters: genetic elements which initiate transcription of the gene of interest. In the first portion of this thesis we discover and characterize native *S.* 6803 promoters which synchronize gene expression to diurnal LD cycles. Specifically, we show P_{hliC}, $P_{slr0006,}$ P_{rbp1}, and P_{sigA} provide consistent light-activated expression of the bacterial luciferase LuxAB reporter as measured by bioluminescence at the onset of day under 12:12 LD cycles. $P_{hliC,}$ unlike the others, did not display *luxAB* transcript increases at the onset of light, suggesting a potential role of post-translational modifications in gene expression regulation.

The next logical direction to take this work is to apply these light-entrained promoters toward chemical production and compare yield under CL versus LD cycles on a per-photon basis (*i.e.* normalizing for total light received). I have developed experimental design perform this study for the production of heptadecane, an alkane with octane rating superior to many conventional fuels [206]. I propose utilizing the light-entrained promoters to drive the production of acyl-ACP via fatty acyl-ACP reductase (*sll0209, aar,* Aar) and aldehyde-deformylating oxygenase (*sll0208*, *ado,* Ado) [207] to enable light-entrained production of heptadecane; experimental details and further proposed modifications are provided in Appendix C.

Further study investigating the disconnect between transcript and protein could provide broad insight into genetic expression regulation and improve genetic engineering efforts in cyanobacteria and other related organisms. Toward this end, we investigated P_{kaiA} and P_{rbcL} which are known to be circadian-controlled and upregulated during the light phase [85]. Native P_{kaiA}*:kaiA* and P_{rbcL}*:rbcL* exhibited very-high transcript induction at the onset of light, engineered P_{kaiA}*:luxAB* and P_{rbcL}*:luxAB* exhibited slightly dampened transcript induction at the onset of light, and neither engineered strains exhibited any measureable bioluminescence above the wild-type

background. Two hypotheses can be made from these observations: element(s) responsible for high levels of transcript activation and/or stability were lost from the native to engineered constructs, and/or element(s) responsible for translation were lost assuming native transcripts give rise to high levels of protein. The P_{kaiA}:*luxAB* and P_{rbcL}:*luxAB* strains provide a nice platform to investigate the regulatory disconnects with further genetic studies.

Investigating the mechanism by which P_{hliC} activates expression is another impactful future direction from this work. Two experiments of interest would be (i) investigating whether P_{hliC} activates in response to high incident light, or in response to related signals (*i.e.* reactive oxygen species (ROS) [169,170]); and (ii) investigating whether P_{hliC} is bound by proteins or pigments [208] sequestering and/or protecting transcripts. Experiment (i) could be easily addressed by comparing bioluminescence under high-light versus artificially supplying ROS; the protein hypothesis in experiment (ii) could be addressed with chromatin immunoprecipitation experiment. Hli proteins are studied in cyanobacteria as the ancestors to plant light harvesting complexes (LHCs) [209], so the results of this study should be of interest to a wide audience.

The conclusion from the second portion of this thesis can be stated quite simply: metabolite abundances oscillate dramatically over a diurnal LD cycle in cyanobacteria. We observe sharp and immediate increases in the abundance of free amino acids (AAs) and nucleic acids (NAs) within the first 30-120 minutes of the day. No cell growth was observed during this period. Around two hours into the day, the free AAs and NAs began to decrease as protein-bound AAs and optical density increase. Near noon, optical density reached a peak and cell division began, continuing until dusk. Just before dusk, insoluble carbohydrates oscillated sharply. This oscillation was observed in all insoluble C and C6 carbohydrates, glucosamine, and glutamic acid but not free C5 and C6 carbohydrates, free amino acids, or other compounds with

similar polarity and molecular weight detected from the same fraction. During the early night, cell mass decreased; this observation was highly repeatable.

These observations suggest the following sequence of events: free amino acids and nucleic acid are synthesized early in the day followed by incorporation into protein and DNA, respectively, during the late day. This period of protein synthesis and DNA replication corresponds to the period of maximal cell mass (optical density) increase but no cell number increase. Together these support the hypothesis that a major cellular objective is protein synthesis and chromosomal DNA replication during the first six hours of the day to ready the cell for cell division during the second six hours of the day.

Metabolite accumulation data does not provide information on metabolite flux and thus the conclusions we can draw from this study are limited. Supplemental 'omics work—fluxomics, proteomics, transcriptomics, etc.—are necessary to corroborate our observations. Still, our findings serve to high the widespread dynamic nature of photosynthetic biology under natural light environments.

A primary motivation for this study was to enable improvement of genome scale modelling of cellular metabolism under realistic light conditions. Specifically, developing a dynamic biomass equation would enable dynamic flux balance analysis (DFBA). This is an area of interest for our lab group and the group of Ashok Prasad, a Peebles' lab collaborator and committee member for this dissertation work. Toward this goal, a summary of the metabolites identified which are classified as biomass equation components in Nogales *et al.* (2012) [76] is provided in Appendix E. This appendix should serve as a useful reference for future modelling studies which seek to integrate a dynamic biomass equation into DFBA approaches.

The clear next-step for investigating the insoluble carbohydrates is to determine the entities to which these carbohydrates associate. Isolation of each particular entity prior to analysis would enable determination of the associations. Further, analysis by NMR would enable determination of the polymeric linkages.

Varying degrees of EPS association based on growth conditions and isolation methods leave the question of whether we captured EPS in our separation method unanswered [193,210,211]. The strong insoluble carbohydrate signals we observe suggests we are observing extracellular polysaccharide (EPS) content. Future work should determine whether EPS was captured by Alcain Blue analysis [210] and development of a protocol to separate the polymeric entities. After this development, investigation to when EPS is produced/secreted/consumed would be fascinating. Feeding ^{13}C-EPS is an option to investigate the role of EPS consumption patterns under sinLD cycles.

Analysis of the insoluble pellet by an analytical chemistry technique suitable for non-polar compounds, such as lipids and pigments, would improve the study. Cyanobacteria thylakoid membranes are dominated by galactoglycerolipids, which are likely held together by Van der Walls interactions between non-polar tails and attractive forces, such as hydrogen bonds between polar head groups [212]. These interactions should be easily disrupted by the chemical (MTBE:MeOH:H2O) and physical (vortexing, sonication) methods applied in the biphasic extraction. Phycocyanobilin pigments are covalently bound to phycobiliproteins [213]. The insoluble post-extraction pellet was blue in color, suggesting the phycobilisome was not completely disassociated. Subsequent acid hydrolysis almost certainly disrupted these bonds; however, the pigment constituents were not assessed. Analyzing the pellet by RP-MS would

improve future experiments by providing a measure of phycocyanobilin pigments and determine whether significant membrane components remain post-extraction.

Many experimental modifications and additions would improve the metabolomics study. Analysis of metabolites during different stages of growth would assess whether the metabolite accumulation patterns we observe exhibit growth-phase dependency. PicoGreen and RiboGreen analysis of DNA and RNA content, respectively would enable the calculation of RNA:DNA ratio across a day/night cycle. Proteomics examination of *S.* 6803 under sinLD cycles—especially at the day/night transitions—would be helpful toward investigation of potential glycoprotein events. An improved method of quantifying protein-bound amino acids which does not bias the data toward certain residues would improve the interpretability of protein content.

Rapid quenching and separation of samples in microbial 'omics experiments remain an area in need of development. Traditional cold-shock methods of quenching metabolites suffer from bias by the 'cold shock phenomena' [214] and, in photosynthetic organisms, metabolite loss to traditional organic quenching solvents [215]. In this thesis, a filtration method was developed which provided improved quenching and separation minimizing from this bias but was not compatible with analysis of the insoluble pellet and thus was not used. Development of the filtration protocol to enable pellet analysis would improve future study by improving the quenching aspect as well as enabling analysis of excreted metabolites in the flow-through without salt contaminants introduced by salt quenching solutions (such as PBS utilized in this study).

Currently, absolute quantification of metabolites by MS requires standards making it prohibitively expensive or impossible for compounds without available standards. Metabolite quantification by GC-MS is complicated matrix-dependent and thus biased derivatization

chemistry [216]. Absolute quantification by LC-MS is also plagued by matrix-effects (*e.g.* ion suppression) [217]. Fully ^{13}C-labeled cells can be used as an internal standard for quantification [218] and has been applied in cyanobacteria [219]. Utilization of this method in future studies would enable quantification of a broad range of metabolites. High instrumentation costs and the cost of creating fully labeled cells limit the widespread application of this method for quantification, as was the case for the experiments conducted in this thesis. Still, the importance of overcoming this hurdle cannot be overstated, as it opens the door to metabolic network integration for improved prediction of reaction thermodynamics [220].

A major advantage of untargeted mass spectrometry (MS) based metabolomics is an unbiased approach to metabolite identification. This attribute is also a drawback, as many of the compounds identified by MS cannot be annotated. Annotation is limited by reference spectra availability and computational tools which can accurately predict spectra. Nuclear magnetic resonance (NMR) is an alternative which provides structural information but lacks sensitivity. Sensitivity improvements to NMR [221] or the recently emerging NMR/MS hybrid technologies [222] promise to propel the field of metabolomics forward.

Lastly, mapping compounds identified by mass spectrometry to biochemical pathways is another hurdle toward the usefulness of metabolomics to iterative strain design processes. Systems biology tools such as KBase (https://kbase.us/) utilized for building metabolic models do not use the same the same compound identifiers as mass spectral databases (*e.g.* Golm Metabolome Database, http://gmd.mpimp-golm.mpg.de/; Human Metabolome Database, http://www.hmdb.ca/) or chemical databases (*e.g.* PubChem, https://pubchem.ncbi.nlm.nih.gov/; KEGG COMPOUND, http://www.genome.jp/kegg/compound/). Consistent identifiers—such as

the IUPAC International Chemical Identifier ("InChI Key")—could enable mapping of pathways *onto* mass spectral results, simplifying and streamlining the data analysis process.

REFERENCES

[1] A.S. Verma, S. Agrahari, S. Rastogi, A. Singh, Biotechnology in the Realm of History, J. Pharm. Bioallied Sci. 3 (2011) 321–323. doi:10.4103/0975-7406.84430.

[2] M. Salque, P.I. Bogucki, J. Pyzel, I. Sobkowiak-Tabaka, R. Grygiel, M. Szmyt, R.P. Evershed, Earliest evidence for cheese making in the sixth millennium bc in northern Europe, Nature. 493 (2013) 522–525. http://dx.doi.org/10.1038/nature11698.

[3] L. Solieri, P. Giudici, Vinegars of the World, in: L. Solieri, P. Giudici (Eds.), Vinegars of the World, Springer Milan, Milano, 2009: pp. 1–16. doi:10.1007/978-88-470-0866-3_1.

[4] P.E. McGovern, J. Zhang, J. Tang, Z. Zhang, G.R. Hall, R.A. Moreau, A. Nuñez, E.D. Butrym, M.P. Richards, C. Wang, G. Cheng, Z. Zhao, C. Wang, Fermented beverages of pre- and proto-historic China, Proc. Natl. Acad. Sci. United States Am. . 101 (2004) 17593–17598. doi:10.1073/pnas.0407921102 .

[5] S. Malik, R.M. Cusidó, M.H. Mirjalili, E. Moyano, J. Palazón, M. Bonfill, Production of the anticancer drug taxol in Taxus baccata suspension cultures: A review, Process Biochem. 46 (2011) 23–34. doi:https://doi.org/10.1016/j.procbio.2010.09.004.

[6] C.-J. Huang, H. Lin, X. Yang, Industrial production of recombinant therapeutics in Escherichia coli and its recent advancements, J. Ind. Microbiol. {&} Biotechnol. 39 (2012) 383–399. doi:10.1007/s10295-011-1082-9.

[7] I. Borodina, J. Nielsen, Advances in metabolic engineering of yeast Saccharomyces cerevisiae for production of chemicals, Biotechnol. J. 9 (2014) 609–620. doi:10.1002/biot.201300445.

[8] B.M. Berla, R. Saha, C.M. Immethun, C.D. Maranas, T.S. Moon, H.B. Pakrasi, Synthetic biology of cyanobacteria: unique challenges and opportunities, Front. Microbiol. 4 (2013) 246. doi:10.3389/fmicb.2013.00246.

[9] S.H. Mohd Azhar, R. Abdulla, S.A. Jambo, H. Marbawi, J.A. Gansau, A.A. Mohd Faik, K.F. Rodrigues, Yeasts in sustainable bioethanol production: A review, Biochem. Biophys. Reports. 10 (2017) 52–61. doi:https://doi.org/10.1016/j.bbrep.2017.03.003.

[10] T.L. Hamilton, D.A. Bryant, J.L. Macalady, The role of biology in planetary evolution: cyanobacterial primary production in low-oxygen Proterozoic oceans, Environ. Microbiol. 18 (2016) 325–340. doi:10.1111/1462-2920.13118.

[11] G.R. Wolfe, F.X. Cunningham, D. Durnfordt, B.R. Green, E. Gantt, Evidence for a common origin of chloroplasts with light-harvesting complexes of different pigmentation, Nature. 367 (1994) 566–568. http://dx.doi.org/10.1038/367566a0.

[12] J.D. Delwiche, C. F. & Palmer, Origins of the Algae and Their Plastids, in: 1997: pp. 53–96.

[13] C.B. Field, M.J. Behrenfeld, J.T. Randerson, P. Falkowski, Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components, Science (80-.). 281 (1998) 237 LP-240. http://science.sciencemag.org/content/281/5374/237.abstract.

[14] B. Büdel, Cyanobacteria: Habitats and Species, in: U. Lüttge, E. Beck, D. Bartels (Eds.), Plant Desiccation Toler., Springer Berlin Heidelberg, Berlin, Heidelberg, 2011: pp. 11–21. doi:10.1007/978-3-642-19106-0_2.

[15] S.M. Singh, J. Elster, Cyanobacteria in Antarctic Lake Environments, in: J. Seckbach (Ed.), Algae Cyanobacteria Extrem. Environ., Springer Netherlands, Dordrecht, 2007: pp.

303–320. doi:10.1007/978-1-4020-6112-7_16.
[16] T. Hauer, R. Mühlsteinová, M. Bohunická, J. Kaštovský, J. Mareš, Diversity of cyanobacteria on rock surfaces, Biodivers. Conserv. 24 (2015) 759–779. doi:10.1007/s10531-015-0890-z.
[17] F. Montero, Photosynthesis, in: M. Gargaud, R. Amils, J.C. Quintanilla, H.J. (Jim) Cleaves, W.M. Irvine, D.L. Pinti, M. Viso (Eds.), Encycl. Astrobiol., Springer Berlin Heidelberg, Berlin, Heidelberg, 2011: pp. 1245–1249. doi:10.1007/978-3-642-11274-4_1154.
[18] D.J. Lea-Smith, P. Bombelli, R. Vasudevan, C.J. Howe, Photosynthetic, respiratory and extracellular electron transport pathways in cyanobacteria, Biochim. Biophys. Acta - Bioenerg. 1857 (2016) 247–255. doi:https://doi.org/10.1016/j.bbabio.2015.10.007.
[19] Y.K.& Y.K. Kazuyoshi Murata, Sayuri Hagiwara, Ultrastructure of compacted DNA in cyanobacteria by high-voltage cryo-electron tomography, Sci. Rep. 6 (2016) 34934.
[20] R. Münch, K. Hiller, A. Grote, M. Scheer, J. Klein, M. Schobert, D. Jahn, Virtual Footprint and PRODORIC: an integrative framework for regulon prediction in prokaryotes, Bioinformatics. 21 (2005) 4187–4189. http://dx.doi.org/10.1093/bioinformatics/bti635.
[21] E. Englund, F. Liang, P. Lindberg, Evaluation of promoters and ribosome binding sites for biotechnological applications in the unicellular cyanobacterium Synechocystis sp. PCC 6803, 6 (2016) 36640.
[22] H.M. Salis, Chapter two - The Ribosome Binding Site Calculator, in: C.V.B.T.-M. in Enzymology (Ed.), Synth. Biol. Part BComputer Aided Des. DNA Assem., Academic Press, 2011: pp. 19–42. doi:http://dx.doi.org/10.1016/B978-0-12-385120-8.00002-4.
[23] K. Thiel, E. Mulaku, H. Dandapani, C. Nagy, E.-M. Aro, P. Kallio, Translation efficiency of heterologous proteins is significantly affected by the genetic context of RBS sequences in engineered cyanobacterium Synechocystis sp. PCC 6803., Microb. Cell Fact. 17 (2018) 34. doi:10.1186/s12934-018-0882-2.
[24] I.A. Osterman, S.A. Evfratov, P. V Sergiev, O.A. Dontsova, Comparison of mRNA features affecting translation initiation and reinitiation., Nucleic Acids Res. 41 (2013) 474–486. doi:10.1093/nar/gks989.
[25] R. Grantham, Working of the genetic code, Trends Biochem. Sci. 5 (1980) 327–331. doi:https://doi.org/10.1016/0968-0004(80)90143-7.
[26] Q. Xiong, Z. Chen, F. Ge, Proteomic analysis of post translational modifications in cyanobacteria, J. Proteomics. 134 (2016) 57–64. doi:https://doi.org/10.1016/j.jprot.2015.07.037.
[27] T. Kaneko, S. Tabata, Complete Genome Structure of the Unicellular Cyanobacterium Synechocystis sp. PCC6803, Plant Cell Physiol. . 38 (1997) 1171–1176. http://pcp.oxfordjournals.org/content/38/11/1171.abstract.
[28] M. Spannagl, T. Nussbaumer, K.C. Bader, M.M. Martis, M. Seidel, K.G. Kugler, H. Gundlach, K.F.X. Mayer, PGSB PlantsDB: updates to the database framework for comparative plant genome research, Nucleic Acids Res. 44 (2016) D1141–D1147. doi:10.1093/nar/gkv1130.
[29] A.M.L. van de Meene, M.F. Hohmann-Marriott, W.F.J. Vermaas, R.W. Roberson, The three-dimensional structure of the cyanobacterium Synechocystis sp. PCC 6803, Arch. Microbiol. 184 (2006) 259–270. doi:10.1007/s00203-005-0027-y.
[30] J. Zhou, H. Zhang, H. Meng, Y. Zhu, G. Bao, Y. Zhang, Y. Li, Y. Ma, Discovery of a

super-strong promoter enables efficient production of heterologous proteins in cyanobacteria, Sci. Rep. 4 (2014). http://dx.doi.org/10.1038/srep04500.
[31] D. Camsund, T. Heidorn, P. Lindblad, Design and analysis of LacI-repressed promoters and DNA-looping in a cyanobacterium, J. Biol. Eng. 8 (2014) 4. doi:10.1186/1754-1611-8-4.
[32] S.C. Albers, V.A. Gallegos, C.A.M. Peebles, Engineering of genetic control tools in Synechocystis sp. PCC 6803 using rational design techniques, J. Biotechnol. 216 (2015) 36–46. doi:http://dx.doi.org/10.1016/j.jbiotec.2015.09.042.
[33] L. Peca, P.B. Kos, Z. Mate, A. Farsang, I. Vass, Construction of bioluminescent cyanobacterial reporter strains for detection of nickel, cobalt and zinc., FEMS Microbiol. Lett. 289 (2008) 258–264. doi:10.1111/j.1574-6968.2008.01393.x.
[34] A.H. Ng, B.M. Berla, H.B. Pakrasi, Fine-Tuning of Photoautotrophic Protein Production by Combining Promoters and Neutral Sites in the Cyanobacterium Synechocystis sp. Strain PCC 6803, Appl. Environ. Microbiol. . 81 (2015) 6857–6863. doi:10.1128/AEM.01349-15 .
[35] H.H. Huang, D. Camsund, P. Lindblad, T. Heidorn, Design and characterization of molecular tools for a Synthetic Biology approach towards developing cyanobacterial biotechnology, Nucleic Acids Res. 38 (2010). doi:10.1093/nar/gkq164.
[36] P. Lindberg, S. Park, A. Melis, Engineering a platform for photosynthetic isoprene production in cyanobacteria, using Synechocystis as the model organism, Metab. Eng. 12 (2010) 70–79. doi:http://dx.doi.org/10.1016/j.ymben.2009.10.001.
[37] S.C. Albers, C.A.M. Peebles, Evaluating Light-Induced Promoters for the Control of Heterologous Gene Expression in Synechocystis sp. PCC 6803, Biotechnol. Prog. (2016) n/a-n/a. doi:10.1002/btpr.2396.
[38] H.-H. Huang, P. Lindblad, Wide-dynamic-range promoters engineered for cyanobacteria., J. Biol. Eng. 7 (2013) 10. doi:10.1186/1754-1611-7-10.
[39] Y. Zheng, Q. Yuan, X. Yang, H. Ma, Engineering Escherichia coli for poly-(3-hydroxybutyrate) production guided by genome-scale metabolic network analysis, Enzyme Microb. Technol. 106 (2017) 60–66. doi:https://doi.org/10.1016/j.enzmictec.2017.07.003.
[40] T. Walther, C.M. Topham, R. Irague, C. Auriol, A. Baylac, H. Cordier, C. Dressaire, L. Lozano-Huguet, N. Tarrat, N. Martineau, M. Stodel, Y. Malbert, M. Maestracci, R. Huet, I. André, M. Remaud-Siméon, J.M. François, Construction of a synthetic metabolic pathway for biosynthesis of the non-natural methionine precursor 2,4-dihydroxybutyric acid, Nat. Commun. 8 (2017) 15828. doi:10.1038/ncomms15828.
[41] B. Wang, C. Eckert, P.-C. Maness, J. Yu, A Genetic Toolbox for Modulating the Expression of Heterologous Genes in the Cyanobacterium Synechocystis sp. PCC 6803, ACS Synth. Biol. 7 (2018) 276–286. doi:10.1021/acssynbio.7b00297.
[42] F. Pinto, C.C. Pacheco, P. Oliveira, A. Montagud, A. Landels, N. Couto, P.C. Wright, J.F. Urchueguía, P. Tamagnini, Improving a Synechocystis-based photoautotrophic chassis through systematic genome mapping and validation of neutral sites, DNA Res. . (2015). doi:10.1093/dnares/dsv024.
[43] J. Mitschke, J. Georg, I. Scholz, C.M. Sharma, D. Dienst, J. Bantscheff, B. Voß, C. Steglich, A. Wilde, J. Vogel, W.R. Hess, An experimentally anchored map of transcriptional start sites in the model cyanobacterium Synechocystis sp. PCC6803, Proc. Natl. Acad. Sci. . 108 (2011) 2124–2129. doi:10.1073/pnas.1015154108 .

[44] M. Griese, C. Lange, J. Soppa, Ploidy in cyanobacteria, FEMS Microbiol. Lett. 323 (2011) 124–131. doi:10.1111/j.1574-6968.2011.02368.x.
[45] J.K. Oliver, S. Atsumi, Metabolic design for cyanobacterial chemical synthesis, Photosynth. Res. 120 (2014) 249–261. doi:10.1007/s11120-014-9997-4.
[46] H.M. Woo, Solar-to-chemical and solar-to-fuel production from CO2 by metabolically engineered microorganisms, Curr. Opin. Biotechnol. 45 (2017) 1–7. doi:https://doi.org/10.1016/j.copbio.2016.11.017.
[47] A.E. Case, S. Atsumi, Cyanobacterial chemical production, J. Biotechnol. 231 (2016) 106–114. doi:https://doi.org/10.1016/j.jbiotec.2016.05.023.
[48] G. Markou, I. Chatzipavlidis, D. Georgakakis, Carbohydrates Production and Bio-flocculation Characteristics in Cultures of Arthrospira (Spirulina) platensis: Improvements Through Phosphorus Limitation Process, BioEnergy Res. 5 (2012) 915–925. doi:10.1007/s12155-012-9205-3.
[49] L. Rodolfi, G. Chini Zittelli, N. Bassi, G. Padovani, N. Biondi, G. Bonini, M.R. Tredici, Microalgae for oil: Strain selection, induction of lipid synthesis and outdoor mass cultivation in a low-cost photobioreactor, Biotechnol. Bioeng. 102 (2009) 100–112. doi:10.1002/bit.22033.
[50] D. Cheng, Q. He, Assessment of Environmental Stresses for Enhanced Microalgal Biofuel Production – An Overview , Front. Energy Res. . 2 (2014) 26. https://www.frontiersin.org/article/10.3389/fenrg.2014.00026.
[51] F. Ferreira, N.A. Straus, Iron deprivation in cyanobacteria, J. Appl. Phycol. 6 (1994) 199–210. doi:10.1007/BF02186073.
[52] a R. Grossman, M.R. Schaefer, G.G. Chiang, J.L. Collier, The phycobilisome, a light-harvesting complex responsive to environmental conditions., Microbiol. Rev. 57 (1993) 725–749.
[53] S.M. Whitney, R.L. Houtz, H. Alonso, Advancing Our Understanding and Capacity to Engineer Nature's CO₂-Sequestering Enzyme, Rubisco, Plant Physiol. 155 (2011) 27 LP-35. http://www.plantphysiol.org/content/155/1/27.abstract.
[54] M.R. Parikh, D.N. Greene, K.K. Woods, I. Matsumura, Directed evolution of RuBisCO hypermorphs through genetic selection in engineered E.coli, Protein Eng. Des. Sel. 19 (2006) 113–119. http://dx.doi.org/10.1093/protein/gzj010.
[55] D.N. Greene, S.M. Whitney, I. Matsumura, Artificially evolved Synechococcus PCC6301 Rubisco variants exhibit improvements in folding and catalytic efficiency, Biochem. J. 404 (2007) 517 LP-524. http://www.biochemj.org/content/404/3/517.abstract.
[56] F. Liang, P. Lindblad, Synechocystis PCC 6803 overexpressing RuBisCO grow faster with increased photosynthesis, Metab. Eng. Commun. 4 (2017) 29–36. doi:https://doi.org/10.1016/j.meteno.2017.02.002.
[57] T. Iwaki, K. Haranoh, N. Inoue, K. Kojima, R. Satoh, T. Nishino, S. Wada, H. Ihara, S. Tsuyama, H. Kobayashi, A. Wadano, Expression of foreign type I ribulose-1,5-bisphosphate carboxylase/ oxygenase (EC 4.1.1.39) stimulates photosynthesis in cyanobacterium Synechococcus PCC7942 cells, Photosynth. Res. 88 (2006) 287. doi:10.1007/s11120-006-9048-x.
[58] S. Atsumi, W. Higashide, J.C. Liao, Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde, Nat Biotechnol. 27 (2009). doi:10.1038/nbt.1586.
[59] M.R. Badger, G.D. Price, CO2 concentrating mechanisms in cyanobacteria: molecular

components, their diversity and evolution, J. Exp. Bot. 54 (2003) 609–622. http://dx.doi.org/10.1093/jxb/erg076.
[60] N.A. Kamennaya, S. Ahn, H. Park, R. Bartal, K.A. Sasaki, H.-Y. Holman, C. Jansson, Installing extra bicarbonate transporters in the cyanobacterium Synechocystis sp. PCC6803 enhances biomass production, Metab. Eng. 29 (2015) 76–85. doi:http://dx.doi.org/10.1016/j.ymben.2015.03.002.
[61] L. Rosgaard, A.J. de Porcellinis, J.H. Jacobsen, N.-U. Frigaard, Y. Sakuragi, Bioengineering of carbon fixation, biofuels, and biochemicals in cyanobacteria and plants, J. Biotechnol. 162 (2012) 134–147. doi:https://doi.org/10.1016/j.jbiotec.2012.05.006.
[62] J.F. ALLEN, Oxygen reduction and optimum production of ATP in photosynthesis, Nature. 256 (1975) 599. http://dx.doi.org/10.1038/256599a0.
[63] D.M. Kramer, J.R. Evans, The importance of energy balance in improving photosynthetic productivity., Plant Physiol. 155 (2011) 70–78. doi:10.1104/pp.110.166652.
[64] J.F. Allen, Cyclic, pseudocyclic and noncyclic photophosphorylation: new links in the chain, Trends Plant Sci. 8 (2003) 15–19. doi:https://doi.org/10.1016/S1360-1385(02)00006-7.
[65] D. Noy, C.C. Moser, P.L. Dutton, Design and engineering of photosynthetic light-harvesting and electron transfer using length, time, and energy scales, Biochim. Biophys. Acta - Bioenerg. 1757 (2006) 90–105. doi:https://doi.org/10.1016/j.bbabio.2005.11.010.
[66] D.R. Ort, A. Melis, Optimizing Antenna Size to Maximize Photosynthetic Efficiency, Plant Physiol. 155 (2011) 79 LP-85. http://www.plantphysiol.org/content/155/1/79.abstract.
[67] A. Melis, Solar energy conversion efficiencies in photosynthesis: Minimizing the chlorophyll antennae to maximize efficiency, Plant Sci. 177 (2009) 272–280. doi:https://doi.org/10.1016/j.plantsci.2009.06.005.
[68] L.E. Page, M. Liberton, H.B. Pakrasi, Reduction of Photoautotrophic Productivity in the Cyanobacterium Synechocystis sp. Strain PCC 6803 by Phycobilisome Antenna Truncation, Appl. Environ. Microbiol. . 78 (2012) 6349–6351. doi:10.1128/AEM.00499-12 .
[69] Z. Perrine, S. Negi, R.T. Sayre, Optimization of photosynthetic light energy utilization by microalgae, Algal Res. 1 (2012) 134–142. doi:https://doi.org/10.1016/j.algal.2012.07.002.
[70] C. Formighieri, A. Melis, A phycocyanin·phellandrene synthase fusion enhances recombinant protein expression and β-phellandrene (monoterpene) hydrocarbons production in Synechocystis (cyanobacteria), Metab. Eng. 32 (2015) 116–124. doi:https://doi.org/10.1016/j.ymben.2015.09.010.
[71] X. Gao, F. Gao, D. Liu, H. Zhang, X. Nie, C. Yang, Engineering the methylerythritol phosphate pathway in cyanobacteria for photosynthetic isoprene production from CO2, Energy Environ. Sci. 9 (2016) 1400–1411. doi:10.1039/C5EE03102H.
[72] T.Y. Kim, S.B. Sohn, Y. Bin Kim, W.J. Kim, S.Y. Lee, Recent advances in reconstruction and applications of genome-scale metabolic models, Curr. Opin. Biotechnol. 23 (2012) 617–623. doi:https://doi.org/10.1016/j.copbio.2011.10.007.
[73] P. Fu, Genome-scale modeling of Synechocystis sp. PCC 6803 and prediction of pathway insertion, J. Chem. Technol. Biotechnol. 84 (2009) 473–483. doi:10.1002/jctb.2065.
[74] A. Montagud, E. Navarro, P. Fernández de Córdoba, J.F. Urchueguía, K.R. Patil, Reconstruction and analysis of genome-scale metabolic model of a photosynthetic bacterium, BMC Syst. Biol. 4 (2010) 156. doi:10.1186/1752-0509-4-156.

[75] A. Montagud, A. Zelezniak, E. Navarro, P.F. de Córdoba, J.F. Urchueguía, K.R. Patil, Flux coupling and transcriptional regulation within the metabolic network of the photosynthetic bacterium Synechocystis sp. PCC6803, Biotechnol. J. 6 (2011) 330–342. doi:10.1002/biot.201000109.
[76] J. Nogales, S. Gudmundsson, E.M. Knight, B.O. Palsson, I. Thiele, Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis, Proc. Natl. Acad. Sci. . 109 (2012) 2678–2683. doi:10.1073/pnas.1117907109.
[77] C.J. Joshi, C.A.M. Peebles, A. Prasad, Modeling and analysis of flux distribution and bioproduct formation in Synechocystis sp. PCC 6803 using a new genome-scale metabolic reconstruction, Algal Res. 27 (2017) 295–310. doi:https://doi.org/10.1016/j.algal.2017.09.013.
[78] A. Varma, B.O. Palsson, Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110., Appl. Environ. Microbiol. 60 (1994) 3724–3731. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC201879/.
[79] R. Mahadevan, J.S. Edwards, F.J. Doyle, Dynamic flux balance analysis of diauxic growth inEscherichia coli, Biophys J. 83 (2002). doi:10.1016/S0006-3495(02)73903-9.
[80] J.A. Gomez, K. Höffner, P.I. Barton, DFBAlab: a fast and reliable MATLAB code for dynamic flux balance analysis, BMC Bioinformatics. 15 (2014) 1–10. doi:10.1186/s12859-014-0409-8.
[81] R. Saha, D. Liu, A. Hoynes-O'Connor, M. Liberton, J. Yu, M. Bhattacharyya-Pakrasi, A. Balassy, F. Zhang, T.S. Moon, C.D. Maranas, H.B. Pakrasi, Diurnal Regulation of Cellular Processes in the Cyanobacterium Synechocystis sp. Strain PCC 6803: Insights from Transcriptomic, Fluxomic, and Physiological Analyses, MBio. 7 (2016) e00464-16. doi:10.1128/mBio.00464-16.
[82] S.A. Angermayr, P. van Alphen, D. Hasdemir, G. Kramer, M. Iqbal, W. van Grondelle, H.C. Hoefsloot, Y.H. Choi, K.J. Hellingwerf, Culturing Synechocystis sp. Strain PCC 6803 with N2 and CO2 in a Diel Regime Reveals Multiphase Glycogen Dynamics with Low Maintenance Costs, Appl. Environ. Microbiol. . 82 (2016) 4180–4189. doi:10.1128/AEM.00256-16 .
[83] S. Diamond, D. Jun, B.E. Rubin, S.S. Golden, The circadian oscillator in Synechococcus elongatus controls metabolite partitioning during diurnal growth, Proc. Natl. Acad. Sci. . 112 (2015) E1916–E1925. doi:10.1073/pnas.1504576112.
[84] R. Willamme, Z. Alsafra, R. Arumugam, G. Eppe, F. Remacle, R.D. Levine, C. Remacle, Metabolomic analysis of the green microalga Chlamydomonas reinhardtii cultivated under day/night conditions, J. Biotechnol. 215 (2015) 20–26. doi:http://dx.doi.org/10.1016/j.jbiotec.2015.04.013.
[85] C. Beck, S. Hertel, A. Rediger, R. Lehmann, A. Wiegard, A. Kölsch, B. Heilmann, J. Georg, W.R. Hess, I.M. Axmann, Daily Expression Pattern of Protein-Encoding Genes and Small Noncoding RNAs in Synechocystis sp. Strain PCC 6803, Appl. Environ. Microbiol. . 80 (2014) 5195–5206. doi:10.1128/AEM.01086-14.
[86] J.R. Waldbauer, S. Rodrigue, M.L. Coleman, S.W. Chisholm, Transcriptome and Proteome Dynamics of a Light-Dark Synchronized Bacterial Cell Cycle, PLoS One. 7 (2012). doi:10.1371/journal.pone.0043432.
[87] J. Stöckel, J.M. Jacobs, T.R. Elvitigala, M. Liberton, E.A. Welsh, A.D. Polpitiya, M.A. Gritsenko, C.D. Nicora, D.W. Koppenaal, R.D. Smith, H.B. Pakrasi, Diurnal Rhythms

Result in Significant Changes in the Cellular Protein Complement in the Cyanobacterium Cyanothece 51142, PLoS One. 6 (2011) e16680. doi:10.1371/journal.pone.0016680.
[88] M. Cantrell, G. Peers, A mutant of Chlamydomonas without LHCSR maintains high rates of photosynthesis, but has reduced cell division rates in sinusoidal light conditions., PLoS One. 12 (2017) e0179395. doi:10.1371/journal.pone.0179395.
[89] H. Iijima, T. Shirai, M. Okamoto, A. Kondo, M.Y. Hirai, T. Osanai, Changes in primary metabolism under light and dark conditions in response to overproduction of a response regulator RpaA in the unicellular cyanobacterium Synechocystis sp. PCC 6803, Front. Microbiol. 6 (2015) 888. doi:10.3389/fmicb.2015.00888.
[90] M. Hanai, Y. Sato, A. Miyagi, M. Kawai-Yamada, K. Tanaka, Y. Kaneko, Y. Nishiyama, Y. Hihara, The Effects of Dark Incubation on Cellular Metabolism of the Wild Type Cyanobacterium Synechocystis sp. PCC 6803 and a Mutant Lacking the Transcriptional Regulator cyAbrB2, Life. 4 (2014) 770–787. doi:10.3390/life4040770.
[91] J.T. McEwen, I.M.P. Machado, M.R. Connor, S. Atsumi, Engineering Synechococcus elongatus PCC 7942 for Continuous Growth under Diurnal Conditions, Appl. Environ. Microbiol. . 79 (2013) 1668–1675. doi:10.1128/AEM.03326-12 .
[92] M. Kanno, A.L. Carroll, S. Atsumi, Global metabolic rewiring for improved CO2 fixation and chemical production in cyanobacteria, Nat. Commun. 8 (2017) 14724.
[93] Y.E. Cheah, A.J. Zimont, S.K. Lunka, S.C. Albers, S.J. Park, K.F. Reardon, C.A.M. Peebles, Diel light:dark cycles significantly reduce FFA accumulation in FFA producing mutants of Synechocystis sp. PCC 6803 compared to continuous light, Algal Res. 12 (2015) 487–496. doi:http://dx.doi.org/10.1016/j.algal.2015.10.014.
[94] S.A. Angermayr, M. Paszota, K.J. Hellingwerf, Engineering a Cyanobacterial Cell Factory for Production of Lactic Acid, Appl. Environ. Microbiol. 78 (2012) 7098–7106. doi:10.1128/AEM.01587-12.
[95] S. Oncel, F.V. Sukan, "Effect of light intensity and the light: dark cycles on the long term hydrogen production of Chlamydomonas reinhardtii by batch cultures," Biomass and Bioenergy. 35 (2011) 1066–1074. doi:https://doi.org/10.1016/j.biombioe.2010.11.017.
[96] Q. Liao, L. Li, R. Chen, X. Zhu, A novel photobioreactor generating the light/dark cycle to improve microalgae cultivation, Bioresour. Technol. 161 (2014) 186–191. doi:https://doi.org/10.1016/j.biortech.2014.02.119.
[97] B. Tamburic, F.W. Zemichael, G.C. Maitland, K. Hellgardt, Effect of the Light Regime and Phototrophic Conditions on Growth of the H2-producing Green Alga Chlamydomonas Reinhardtii, Energy Procedia. 29 (2012) 710–719. doi:https://doi.org/10.1016/j.egypro.2012.09.083.
[98] E. Jacob-Lopes, C.H.G. Scoparo, L.M.C.F. Lacerda, T.T. Franco, Effect of light cycles (night/day) on CO2 fixation and biomass production by microalgae in photobioreactors, Chem. Eng. Process. Process Intensif. 48 (2009) 306–310. doi:https://doi.org/10.1016/j.cep.2008.04.007.
[99] J.C. Ogbonna, H. Tanaka, Cyclic autotrophic/heterotrophic cultivation of photosynthetic cells: A method of achieving continuous cell growth under light/dark cycles, Bioresour. Technol. 65 (1998) 65–72. doi:https://doi.org/10.1016/S0960-8524(98)00018-2.
[100] S. Aoki, T. Kondo, H. Wada, M. Ishiura, Circadian rhythm of the cyanobacterium Synechocystis sp. strain PCC 6803 in the dark, J. Bacteriol. 179 (1997) 5751–5755. http://jb.asm.org/cgi/content/long/179/18/5751 (accessed June 20, 2016).
[101] S. Aoki, T. Kondo, M. Ishiura, Circadian expression of the dnaK gene in the

cyanobacterium Synechocystis sp . strain PCC These include : Circadian Expression of the dnaK Gene in the Cyanobacterium Synechocystis sp . Strain PCC 6803, J. Bacteriol. 177 (1995) 5606–5611.
[102] T.L. Wood, J. Bridwell-Rabb, Y.-I. Kim, T. Gao, Y.-G. Chang, A. LiWang, D.P. Barondeau, S.S. Golden, The KaiA protein of the cyanobacterial circadian oscillator is modulated by a redox-active cofactor., Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 5804–5809. doi:10.1073/pnas.0910141107.
[103] Y.-I. Kim, D.J. Vinyard, G.M. Ananyev, G.C. Dismukes, S.S. Golden, Inaugural Article: Oxidized quinones signal onset of darkness directly to the cyanobacterial circadian oscillator, Proc. Natl. Acad. Sci. 109 (2012) 17765–17769. doi:10.1073/pnas.1216401109.
[104] M.J. Rust, S.S. Golden, E.K. O'Shea, Light-driven changes in energy metabolism directly entrain the cyanobacterial circadian oscillator, Science. 331 (2011) 220–223. doi:10.1126/science.1197243.
[105] J.S. Markson, J.R. Piechura, A.M. Puszynska, E.K. O'Shea, Circadian Control of Global Gene Expression by the Cyanobacterial Master Regulator RpaA, Cell. 155 (2013) 1396–1408. doi:http://dx.doi.org/10.1016/j.cell.2013.11.005.
[106] M.L. Paddock, J.S. Boyd, D.M. Adin, S.S. Golden, Active output state of the Synechococcus Kai circadian oscillator, Proc. Natl. Acad. Sci. . 110 (2013) E3849–E3857. doi:10.1073/pnas.1315170110.
[107] T. Osanai, T. Shirai, H. Iijima, A. Kuwahara, I. Suzuki, A. Kondo, M.Y. Hirai, Alteration of cyanobacterial sugar and amino acid metabolism by overexpression hik 8 , encoding a KaiC-associated histidine kinase, 2015. doi:10.1111/1462-2920.12715.
[108] Y. Xu, P.D. Weyman, M. Umetani, J. Xiong, X. Qin, Q. Xu, H. Iwasaki, C.H. Johnson, Circadian Yin-Yang Regulation and Its Manipulation to Globally Reprogram Gene Expression, Curr. Biol. 23 (2015) 2365–2374. doi:10.1016/j.cub.2013.10.011.
[109] V. Vijayan, R. Zuzow, E.K. O'Shea, Oscillations in supercoiling drive circadian gene expression in cyanobacteria., Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 22564–22568. doi:10.1073/pnas.0912673106.
[110] Y. Xu, P. Ma, P. Shah, A. Rokas, Y. Liu, C.H. Johnson, Non-optimal codon usage is a mechanism to achieve circadian clock conditionality., Nature. 495 (2013) 116–120. doi:10.1038/nature11942.
[111] T. Kaneko, S. Sato, H. Kotani, A. Tanaka, E. Asamizu, Y. Nakamura, N. Miyajima, M. Hirosawa, M. Sugiura, S. Sasamoto, T. Kimura, T. Hosouchi, A. Matsuno, A. Muraki, N. Nakazaki, K. Naruo, S. Okumura, S. Shimpo, C. Takeuchi, T. Wada, A. Watanabe, M. Yamada, M. Yasuda, S. Tabata, 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 (1996) 109–136. doi:10.1093/dnares/3.3.109.
[112] S. Imamura, S. Yoshihara, S. Nakano, N. Shiozaki, A. Yamada, K. Tanaka, H. Takahashi, M. Asayama, M. Shirai, Purification, Characterization, and Gene Expression of All Sigma Factors of RNA Polymerase in a Cyanobacterium, J. Mol. Biol. 325 (2003) 857–872. doi:http://dx.doi.org/10.1016/S0022-2836(02)01242-1.
[113] L. Anderson, C.L. Hunter, Quantitative Mass Spectrometric Multiple Reaction Monitoring Assays for Major Plasma Proteins, Mol. Cell. Proteomics . 5 (2006) 573–588. doi:10.1074/mcp.M500331-MCP200.

[114] R.S. Plumb, K.A. Johnson, P. Rainville, B.W. Smith, I.D. Wilson, J.M. Castro-Perez, J.K. Nicholson, UPLC/MS(E); a new approach for generating molecular fragment information for biomarker structure elucidation., Rapid Commun. Mass Spectrom. 20 (2006) 1989–1994. doi:10.1002/rcm.2550.
[115] P. Turk, Experimental design and data anlalysis for researchers I: with applicatoins using SAS university edition, in: Course Notes Stat. 511, n.d.: p. 91.
[116] A. Banerjee, U.B. Chitnis, S.L. Jadhav, J.S. Bhawalkar, S. Chaudhury, Hypothesis testing, type I and type II errors, Ind. Psychiatry J. 18 (2009) 127–131. doi:10.4103/0972-6748.62274.
[117] W.S. Noble, How does multiple testing correction work?, Nat. Biotechnol. 27 (2009) 1135–1137. doi:10.1038/nbt1209-1135.
[118] J.D. Storey, A direct approach to false discovery rates, J. R. Stat. Soc. Ser. B (Statistical Methodol. 64 (2002) 479–498. doi:10.1111/1467-9868.00346.
[119] J.D. Storey, R. Tibshirani, Statistical significance for genomewide studies, Proc. Natl. Acad. Sci. 100 (2003) 9440 LP-9445. http://www.pnas.org/content/100/16/9440.abstract.
[120] D. Pascovici, D.C.L. Handler, J.X. Wu, P.A. Haynes, Multiple testing corrections in quantitative proteomics: A useful but blunt tool, Proteomics. 16 (2016) 2448–2453. doi:10.1002/pmic.201600044.
[121] T. V Perneger, What's wrong with Bonferroni adjustments, BMJ Br. Med. J. 316 (1998) 1236–1238.
[122] P.C. Sham, S.M. Purcell, Statistical power and significance testing in large-scale genetic studies, Nat. Rev. Genet. 15 (2014) 335.
[123] Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, J. R. Stat. Soc. Ser. B. 57 (1995) 289–300.
[124] J. Storey, A. Bass, qvalue: Q-value estimation for false discovery rate control., R Packag. Version 2.10.0. (2015).
[125] J.J. Eaton-Rye, Construction of Gene Interruptions and Gene Deletions in the Cyanobacterium Synechocystis sp. Strain PCC 6803, in: R. Carpentier (Ed.), Photosynth. Res. Protoc., Humana Press, Totowa, NJ, 2011: pp. 295–312. doi:10.1007/978-1-60761-925-3_22.
[126] D. Camsund, P. Lindblad, Engineered transcriptional systems for cyanobacterial biotechnology , Front. Bioeng. Biotechnol. . 2 (2014). http://www.frontiersin.org/Journal/Abstract.aspx?s=1217&name=synthetic_biology&ART_DOI=10.3389/fbioe.2014.00040.
[127] T.E.F. Quax, N.J. Claassens, D. Söll, J. van der Oost, Codon Bias as a Means to Fine-Tune Gene Expression, Mol. Cell. 59 (2015) 149–161. doi:10.1016/j.molcel.2015.05.035.
[128] C. Engler, S. Marillonnet, Golden Gate Cloning, in: S. Valla, R. Lale (Eds.), DNA Cloning Assem. Methods, Humana Press, Totowa, NJ, 2014: pp. 119–131. doi:10.1007/978-1-62703-764-8_9.
[129] S. de Kok, L.H. Stanton, T. Slaby, M. Durot, V.F. Holmes, K.G. Patel, D. Platt, E.B. Shapland, Z. Serber, J. Dean, J.D. Newman, S.S. Chandran, Rapid and Reliable DNA Assembly via Ligase Cycling Reaction, ACS Synth. Biol. 3 (2014) 97–106. doi:10.1021/sb4001992.
[130] D.G. Gibson, L. Young, R.-Y. Chuang, J.C. Venter, C.A. Hutchison, H.O. Smith, Enzymatic assembly of DNA molecules up to several hundred kilobases, Nat Meth. 6 (2009) 343–345. http://dx.doi.org/10.1038/nmeth.1318.

[131] S. Yoshihara, X. Geng, S. Okamoto, K. Yura, T. Murata, M. Go, M. Ohmori, M. Ikeuchi, Mutational Analysis of Genes Involved in Pilus Structure, Motilityand Transformation Competency in the Unicellular Motile CyanobacteriumSynechocystis sp. PCC6803, Plant Cell Physiol. . 42 (2001) 63–73. doi:10.1093/pcp/pce007.
[132] X. Zang, B. Liu, S. Liu, K.K.I.U. Arunakumara, X. Zhang, Optimum conditions for transformation of Synechocystis sp. PCC 6803, J. Microbiol. 45 (2007) 241–245.
[133] B. Wang, J. Yu, W. Zhang, D.R. Meldrum, Premethylation of Foreign DNA Improves Integrative Transformation Efficiency in Synechocystis sp. Strain PCC 6803, Appl. Environ. Microbiol. (2015) AEM.02575-15. doi:10.1128/AEM.02575-15.
[134] F.L. Pinto, A. Thapper, W. Sontheim, P. Lindblad, Analysis of current and alternative phenol based RNA extraction methodologies for cyanobacteria, BMC Mol. Biol. 10 (2009) 79. doi:10.1186/1471-2199-10-79.
[135] P.R. Jones, Genetic Instability in Cyanobacteria – An Elephant in the Room?, Front. Bioeng. Biotechnol. 2 (2014) 12. doi:10.3389/fbioe.2014.00012.
[136] T. Mettler, T. Mühlhaus, D. Hemme, M.-A. Schöttler, J. Rupprecht, A. Idoine, D. Veyel, S.K. Pal, L. Yaneva-Roder, F.V. Winck, F. Sommer, D. Vosloh, B. Seiwert, A. Erban, A. Burgos, S. Arvidsson, S. Schönfelder, A. Arnold, M. Günther, U. Krause, M. Lohse, J. Kopka, Z. Nikoloski, B. Mueller-Roeber, L. Willmitzer, R. Bock, M. Schroda, M. Stitt, Systems Analysis of the Response of Photosynthesis, Metabolism, and Growth to an Increase in Irradiance in the Photosynthetic Model Organism Chlamydomonas reinhardtii., Plant Cell. 26 (2014) 2310–2350. doi:10.1105/tpc.114.124537.
[137] L.J. Jazmin, Y. Xu, Y.E. Cheah, A.O. Adebiyi, C.H. Johnson, J.D. Young, Isotopically nonstationary (13)C flux analysis of cyanobacterial isobutyraldehyde production, Metab. Eng. 42 (2017) 9–18. doi:10.1016/j.ymben.2017.05.001.
[138] M.A. Salem, J. Jüppner, K. Bajdzienko, P. Giavalisco, Protocol: a fast, comprehensive and reproducible one-step extraction method for the rapid preparation of polar and semi-polar metabolites, lipids, proteins, starch and cell wall polymers from a single sample, Plant Methods. 12 (2016) 45. doi:10.1186/s13007-016-0146-2.
[139] M. Fountoulakis, H.-W. Lahm, Hydrolysis and amino acid composition analysis of proteins, J. Chromatogr. A. 826 (1998) 109–134. doi:http://dx.doi.org/10.1016/S0021-9673(98)00721-3.
[140] L.T. Ng, A. Pascaud, M. Pascaud, Hydrochloric acid hydrolysis of proteins and determination of tryptophan by reversed-phase high-performance liquid chromatography, Anal. Biochem. 167 (1987) 47–52. doi:https://doi.org/10.1016/0003-2697(87)90132-1.
[141] K. Muramoto, H. Kamiya, Recovery of tryptophan in peptides and proteins by high-temperature and short-term acid hydrolysis in the presence of phenol, Anal. Biochem. 189 (1990) 223–230. doi:https://doi.org/10.1016/0003-2697(90)90112-M.
[142] J. Ozols, [44] Amino acid analysis, in: M.P.B.T.-M. in E. Deutscher (Ed.), Guid. to Protein Purif., Academic Press, 1990: pp. 587–601. doi:https://doi.org/10.1016/0076-6879(90)82046-5.
[143] B.O. McConnell, M.R. Antoniewicz, Measuring the Composition and Stable-Isotope Labeling of Algal Biomass Carbohydrates via Gas Chromatography/Mass Spectrometry, Anal. Chem. 88 (2016) 4624–4628. doi:10.1021/acs.analchem.6b00779.
[144] Q. Huang, K. Kaiser, R. Benner, A simple high performance liquid chromatography method for the measurement of nucleobases and the RNA and DNA content of cellular material, Limnol. Oceanogr. Methods. 10 (2012) 608–616. doi:10.4319/lom.2012.10.608.

[145] C.A. Smith, E.J. Want, G. O'Maille, R. Abagyan, G. Siuzdak, XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification, Anal. Chem. 78 (2006) 779–787. doi:10.1021/ac051437y.
[146] R. Tautenhahn, C. Böttcher, S. Neumann, Highly sensitive feature detection for high resolution LC/MS, BMC Bioinformatics. 9 (2008) 504. doi:10.1186/1471-2105-9-504.
[147] C.D. Broeckling, F.A. Afsar, S. Neumann, A. Ben-Hur, J.E. Prenni, RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data., Anal. Chem. 86 (2014) 6812–6817. doi:10.1021/ac501530d.
[148] D. Müllner, Fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for R and Python, 2013. doi:10.18637/jss.v053.i09.
[149] P. Langfelder, B. Zhang, S. Horvath, Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R., Bioinformatics. 24 (2008) 719–720. doi:10.1093/bioinformatics/btm563.
[150] C. Jaeger, F. Hoffmann, C.A. Schmitt, J. Lisec, Automated Annotation and Evaluation of In-Source Mass Spectra in GC/Atmospheric Pressure Chemical Ionization-MS-Based Metabolomics, Anal. Chem. 88 (2016) 9386–9390. doi:10.1021/acs.analchem.6b02743.
[151] H. Tsugawa, T. Kind, R. Nakabayashi, D. Yukihira, W. Tanaka, T. Cajka, K. Saito, O. Fiehn, M. Arita, Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software, Anal. Chem. 88 (2016) 7946–7958. doi:10.1021/acs.analchem.6b00770.
[152] D. Lagarde, L. Beuf, W. Vermaas, Increased Production of Zeaxanthin and Other Pigments by Application of Genetic Engineering Techniques to Synechocystis sp. Strain PCC 6803, Appl. Environ. Microbiol. 66 (2000) 64–72. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC91786/.
[153] X. Liu, J. Sheng, R. Curtiss III, Fatty acid production in genetically modified cyanobacteria, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 6899–6904. doi:10.1073/pnas.1103014108.
[154] J. Zhou, H. Zhang, Y. Zhang, Y. Li, Y. Ma, Designing and creating a modularized synthetic pathway in cyanobacterium Synechocystis enables production of acetone from carbon dioxide, Metab. Eng. 14 (2012) 394–400. doi:http://dx.doi.org/10.1016/j.ymben.2012.03.005.
[155] J. Ungerer, L. Tao, M. Davis, M. Ghirardi, P.C. Maness, J.P. Yu, Sustained photosynthetic conversion of CO2 to ethylene in recombinant cyanobacterium Synechocystis 6803, Energ Env. Sci. 5 (2012). doi:10.1039/c2ee22555g.
[156] V. Vijayan, E.K. O'Shea, Sequence Determinants of Circadian Gene Expression Phase in Cyanobacteria, J. Bacteriol. . 195 (2013) 665–671. doi:10.1128/JB.02012-12.
[157] A.K. Dörrich, J. Mitschke, O. Siadat, A. Wilde, Deletion of the Synechocystis sp. PCC 6803 kaiAB1C1 gene cluster causes impaired cell growth under light–dark conditions, Microbiology. 160 (2014) 2538–2550. http://mic.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.081695-0.
[158] T. Osanai, Y. Kanesaki, T. Nakano, H. Takahashi, M. Asayama, M. Shirai, M. Kanehisa, I. Suzuki, N. Murata, K. Tanaka, Positive Regulation of Sugar Catabolic Pathways in the Cyanobacterium Synechocystis sp. PCC 6803 by the Group 2 σ Factor SigE, J. Biol. Chem. . 280 (2005) 30653–30659. doi:10.1074/jbc.M505043200.
[159] S. Imamura, M. Asayama, H. Takahashi, K. Tanaka, H. Takahashi, M. Shirai, Antagonistic dark/light-induced SigB/SigD, group 2 sigma factors, expression through

redox potential and their roles in cyanobacteria, FEBS Lett. 554 (2003) 357–362. doi:10.1016/S0014-5793(03)01188-8.
[160] S. Imamura, M. Asayama, Sigma factors for cyanobacterial transcription., Gene Regul. Syst. Biol. . 3 (2009) 65–87.
[161] A. Kunert, M. Hagemann, N. Erdmann, Microbiological Methods Construction of promoter probe vectors for Synechocystis sp . PCC 6803 using the light-emitting reporter systems Gfp and LuxAB, 41 (2000) 185–194.
[162] O. Olsson, C. Koncz, A.A. Szalay, The use of the luxA gene of the bacterial luciferase operon as a reporter gene, Mol. Gen. Genet. MGG. 215 (1988) 1–9. doi:10.1007/BF00331295.
[163] S. Shalel-Levanon, K.-Y. San, G.N. Bennett, Effect of ArcA and FNR on the expression of genes related to the oxygen regulation and the glycolysis pathway in Escherichia coli under microaerobic growth conditions, Biotechnol. Bioeng. 92 (2005) 147–159. doi:10.1002/bit.20583.
[164] F. Pinto, C.C. Pacheco, D. Ferreira, P. Moradas-Ferreira, P. Tamagnini, Selection of Suitable Reference Genes for RT-qPCR Analyses in Cyanobacteria, PLoS One. 7 (2012) e34983. http://dx.doi.org/10.1371%2Fjournal.pone.0034983.
[165] S. Taylor, M. Wakem, G. Dijkman, M. Alsarraj, M. Nguyen, A practical approach to RT-qPCR—Publishing data that conform to the MIQE guidelines, Methods. 50 (2010) S1–S5. doi:http://dx.doi.org/10.1016/j.ymeth.2010.01.005.
[166] T. Kondo, C. a Strayer, R.D. Kulkarni, W. Taylor, M. Ishiura, S.S. Golden, C.H. Johnson, Circadian rhythms in prokaryotes: luciferase as a reporter of circadian gene expression in cyanobacteria., Proc. Natl. Acad. Sci. U. S. A. 90 (1993) 5672–5676. doi:10.1073/pnas.90.12.5672.
[167] H. McWilliam, W. Li, M. Uludag, S. Squizzato, Y.M. Park, N. Buso, A.P. Cowley, R. Lopez, Analysis Tool Web Services from the EMBL-EBI, Nucleic Acids Res. 41 (2013) W597–W600. http://dx.doi.org/10.1093/nar/gkt376.
[168] H. Samartzidou, W.R. Widger, Transcriptional and Posttranscriptional Control of mRNA from lrtA, a Light-Repressed Transcript in Synechococcus sp. PCC 7002, Plant Physiol. 117 (1998) 225–234. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC35007/.
[169] Q. He, N. Dolganov, O. Björkman, A.R. Grossman, The High Light-inducible Polypeptides inSynechocystis PCC6803: EXPRESSION AND FUNCTION IN HIGH LIGHT , J. Biol. Chem. . 276 (2001) 306–314. doi:10.1074/jbc.M008686200.
[170] M. Havaux, G. Guedeney, Q. He, A.R. Grossman, Elimination of high-light-inducible polypeptides related to eukaryotic chlorophyll a/b-binding proteins results in aberrant photoacclimation in Synechocystis PCC6803, Biochim. Biophys. Acta - Bioenerg. 1557 (2003) 21–33. doi:http://dx.doi.org/10.1016/S0005-2728(02)00391-2.
[171] K. Promnares, J. Komenda, L. Bumba, J. Nebesarova, F. Vacha, M. Tichy, Cyanobacterial Small Chlorophyll-binding Protein ScpD (HliB) Is Located on the Periphery of Photosystem II in the Vicinity of PsbH and CP47 Subunits, J. Biol. Chem. . 281 (2006) 32705–32713. doi:10.1074/jbc.M606360200.
[172] Q. Wang, S. Jantaro, B. Lu, W. Majeed, M. Bailey, Q. He, The High Light-Inducible Polypeptides Stabilize Trimeric Photosystem I Complex under High Light Conditions in Synechocystis PCC 6803, Plant Physiol. 147 (2008) 1239–1250. doi:10.1104/pp.108.121087.
[173] K. Kucho, K. Okamoto, Y. Tsuchiya, S. Nomura, M. Nango, M. Kanehisa, M. Ishiura,

Global Analysis of Circadian Expression in the Cyanobacterium Synechocystis sp. Strain PCC 6803 , J. Bacteriol. 187 (2005) 2190–2199. doi:10.1128/JB.187.6.2190-2199.2005.
[174] I.M.P. Machado, S. Atsumi, Cyanobacterial biofuel production, J. Biotechnol. 162 (2012) 50–56. doi:http://dx.doi.org/10.1016/j.jbiotec.2012.03.005.
[175] C.J. Knoot, J. Ungerer, P.P. Wangikar, H.B. Pakrasi, Cyanobacteria: Promising biocatalysts for sustainable chemical production., J. Biol. Chem. 293 (2018) 5044–5052. doi:10.1074/jbc.R117.815886.
[176] A.L. Carroll, A.E. Case, A. Zhang, S. Atsumi, Metabolic engineering tools in model cyanobacteria., Metab. Eng. (2018). doi:10.1016/j.ymben.2018.03.014.
[177] R.H. Wijffels, O. Kruse, K.J. Hellingwerf, Potential of industrial biotechnology with cyanobacteria and eukaryotic microalgae., Curr. Opin. Biotechnol. 24 (2013) 405–413. doi:10.1016/j.copbio.2013.04.004.
[178] G. Peers, Increasing algal photosynthetic productivity by integrating ecophysiology with systems biology, Trends Biotechnol. 32 (2014) 551–555. doi:https://doi.org/10.1016/j.tibtech.2014.09.007.
[179] M.-Y. Ho, N.T. Soulier, D.P. Canniffe, G. Shen, D.A. Bryant, Light regulation of pigment and photosystem biosynthesis in cyanobacteria, Curr. Opin. Plant Biol. 37 (2017) 24–33. doi:https://doi.org/10.1016/j.pbi.2017.03.006.
[180] M. Muramatsu, Y. Hihara, Acclimation to high-light conditions in cyanobacteria: from gene expression to physiological responses., J. Plant Res. 125 (2012) 11–39. doi:10.1007/s10265-011-0454-6.
[181] M. Johansson, T. Koster, On the move through time - a historical review of plant clock research., Plant Biol. (Stuttg). (2018). doi:10.1111/plb.12729.
[182] J.A. Kim, H.-S. Kim, S.-H. Choi, J.-Y. Jang, M.-J. Jeong, S.I. Lee, The Importance of the Circadian Clock in Regulating Plant Metabolism., Int. J. Mol. Sci. 18 (2017). doi:10.3390/ijms18122680.
[183] T. Kondo, C.A. Strayert, R.D. Kulkarnit, W. Taylor, M. Ishiura, S.S. Goldent, C. Hirschie, J. Ii, Circadian rhythms in prokaryotes : Luciferase as a reporter of circadian gene expression in cyanobacteria, 90 (1993) 5672–5676.
[184] Q. Yang, B.F. Pando, G. Dong, S.S. Golden, A. van Oudenaarden, Circadian Gating of the Cell Cycle Revealed in Single Cyanobacterial Cells, Science (80-.). 327 (2010) 1522–1526. http://science.sciencemag.org/content/327/5972/1522.abstract.
[185] M.L. Paddock, J.S. Boyd, D.M. Adin, S.S. Golden, Active output state of the Synechococcus Kai circadian oscillator., Proc. Natl. Acad. Sci. U. S. A. 110 (2013) E3849-57. doi:10.1073/pnas.1315170110.
[186] R.E. Blankenship, Photosynthetic Pigments: Structure and Spectroscopy, in: Mol. Mech. Photosynth., Blackwell Science Ltd, 2008: pp. 42–60. doi:10.1002/9780470758472.ch4.
[187] A. Werner, K. Oliver, A.D. Miller, J. Sebesta, C.A.M. Peebles, Discovery and characterization of Synechocystis sp. PCC 6803 light-entrained promoters in diurnal light:dark cycles, Algal Res. 30 (2018) 121–127. doi:https://doi.org/10.1016/j.algal.2017.12.012.
[188] T. Mori, B. Binder, C.H. Johnson, Circadian gating of cell division in cyanobacteria growing with average doubling times of less than 24 hours, Proc. Natl. Acad. Sci. . 93 (1996) 10183–10188. http://www.pnas.org/content/93/19/10183.abstract.
[189] E.R. Zinser, D. Lindell, Z.I. Johnson, M.E. Futschik, C. Steglich, M.L. Coleman, M. a. Wright, T. Rector, R. Steen, N. McNulty, L.R. Thompson, S.W. Chisholm, Choreography

of the transcriptome, photophysiology, and cell cycle of a minimal photoautotroph, Prochlorococcus, PLoS One. 4 (2009). doi:10.1371/journal.pone.0005135.
[190] D.P. Stephan, H.G. Ruppel, E.K. Pistorius, Interrelation between cyanophycin synthesis, L-arginine catabolism and photosynthesis in the cyanobacterium Synechocystis sp. strain PCC 6803., Z. Naturforsch. C. 55 (2000) 927–942.
[191] M. Maheswaran, K. Ziegler, W. Lockau, M. Hagemann, K. Forchhammer, PII-regulated arginine synthesis controls accumulation of cyanophycin in Synechocystis sp. strain PCC 6803., J. Bacteriol. 188 (2006) 2730–2734. doi:10.1128/JB.188.7.2730-2734.2006.
[192] J.-M. Panoff, B. Priem, H. Morvan, F. Joset, Sulphated exopolysaccharides produced by two unicellular strains of cyanobacteria, Synechocystis PCC 6803 and 6714, Arch. Microbiol. 150 (1988) 558–563. doi:10.1007/BF00408249.
[193] T. Jittawuttipoka, M. Planchon, O. Spalla, K. Benzerara, F. Guyot, C. Cassier-Chauvat, F. Chauvat, Multidisciplinary Evidences that Synechocystis PCC6803 Exopolysaccharides Operate in Cell Sedimentation and Protection against Salt and Metal Stresses, PLoS One. 8 (2013) e55564.
[194] L. Boudiere, M. Michaud, D. Petroutsos, F. Rebeille, D. Falconet, O. Bastien, S. Roy, G. Finazzi, N. Rolland, J. Jouhet, M.A. Block, E. Marechal, Glycerolipids in photosynthesis: composition, synthesis and trafficking., Biochim. Biophys. Acta. 1837 (2014) 470–480. doi:10.1016/j.bbabio.2013.09.007.
[195] K. Awai, T. Kakimoto, C. Awai, T. Kaneko, Y. Nakamura, K. Takamiya, H. Wada, H. Ohta, Comparative Genomic Analysis Revealed a Gene for Monoglucosyldiacylglycerol Synthase, an Enzyme for Photosynthetic Membrane Lipid Synthesis in Cyanobacteria, Plant Physiol. 141 (2006) 1120–1127. doi:10.1104/pp.106.082859.
[196] Y. Zilliges, J.-C. Kehr, S. Mikkat, C. Bouchier, N.T. de Marsac, T. Börner, E. Dittmann, An Extracellular Glycoprotein Is Implicated in Cell-Cell Contacts in the Toxic Cyanobacterium Microcystis aeruginosa PCC 7806 , J. Bacteriol. 190 (2008) 2871–2879. doi:10.1128/JB.01867-07.
[197] R.K. Upreti, M. Kumar, V. Shankar, Bacterial glycoproteins: functions, biosynthesis and applications., Proteomics. 3 (2003) 363–379. doi:10.1002/pmic.200390052.
[198] K. Manoj, M. Nivedita, U.R. K., A novel membrane glycoprotein of Escherichia coli, J. Basic Microbiol. 43 (2003) 28–35. doi:doi:10.1002/jobm.200390001.
[199] S. Watanabe, R. Ohbayashi, Y. Kanesaki, N. Saito, T. Chibazakura, T. Soga, H. Yoshikawa, Intensive DNA Replication and Metabolism during the Lag Phase in Cyanobacteria, PLoS One. 10 (2015) e0136800.
[200] S. Watanabe, R. Ohbayashi, Y. Shiwa, A. Noda, Y. Kanesaki, T. Chibazakura, H. Yoshikawa, Light-dependent and asynchronous replication of cyanobacterial multi-copy chromosomes, 2012. doi:10.1111/j.1365-2958.2012.07971.x.
[201] I. Alseth, B. Dalhus, M. Bjørås, Inosine in DNA and RNA, Curr. Opin. Genet. Dev. 26 (2014) 116–123. doi:https://doi.org/10.1016/j.gde.2014.07.008.
[202] E. Grunebaum, A. Cohen, C.M. Roifman, Recent advances in understanding and managing adenosine deaminase and purine nucleoside phosphorylase deficiencies, Curr. Opin. Allergy Clin. Immunol. 13 (2013).
[203] M.W. Gray, Evolutionary Origin of RNA Editing, Biochemistry. 51 (2012) 5235–5242. doi:10.1021/bi300419r.
[204] B.-E. Wulff, K. Nishikura, Substitutional A-to-I RNA editing., Wiley Interdiscip. Rev. RNA. 1 (2010) 90–101. doi:10.1002/wrna.10.

[205] V.L. Gonçalves, J.B. Vicente, L.M. Saraiva, M. Teixeira, Flavodiiron Proteins and Their Role in Cyanobacteria BT - Bioenergetic Processes of Cyanobacteria: From Evolutionary Singularity to Ecological Diversity, in: G.A. Peschek, C. Obinger, G. Renger (Eds.), Springer Netherlands, Dordrecht, 2011: pp. 631–653. doi:10.1007/978-94-007-0388-9_22.
[206] B. Creton, C. Dartiguelongue, T. de Bruin, H. Toulhoat, Prediction of the Cetane Number of Diesel Compounds Using the Quantitative Structure Property Relationship, Energy & Fuels. 24 (2010) 5396–5403. doi:10.1021/ef1008456.
[207] A. Schirmer, M.A. Rude, X.Z. Li, E. Popova, S.B. del Cardayre, Microbial Biosynthesis of Alkanes, Science (80-.). 329 (2010). doi:10.1126/science.1187936.
[208] M.K. Shukla, M.J. Llansola-Portoles, M. Tichy, A.A. Pascal, B. Robert, R. Sobotka, Binding of pigments to the cyanobacterial high-light-inducible protein HliC., Photosynth. Res. (2017). doi:10.1007/s11120-017-0475-7.
[209] J.A.D. Neilson, D.G. Durnford, Structural and functional diversification of the light-harvesting complexes in photosynthetic eukaryotes., Photosynth. Res. 106 (2010) 57–71. doi:10.1007/s11120-010-9576-2.
[210] J.F.C. DE BROUWER, K. WOLFSTEIN, L.J. Stal, Physical characterization and diel dynamics of different fractions of extracellular polysaccharides in an axenic culture of a benthic diatom, 2002. doi:10.1017/S0967026201003419.
[211] A. Pannard, J. Pédrono, M. Bormans, E. Briand, P. Claquin, Y. Lagadeuc, Production of exopolymers (EPS) by cyanobacteria: impact on the carbon-to-nutrient ratio of the particulate organic matter, Aquat. Ecol. 50 (2016) 29–44. doi:10.1007/s10452-015-9550-3.
[212] P. Attard, D.J. Mitchell, B.W. Ninham, The attractive forces between polar lipid bilayers., Biophys. J. 53 (1988) 457–460. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1330213/.
[213] A.N. Glazer, Phycobilisome a macromolecular complex optimized for light energy transfer, Biochim. Biophys. Acta - Rev. Bioenerg. 768 (1984) 29–51. doi:https://doi.org/10.1016/0304-4173(84)90006-5.
[214] C. Wittmann, J.O. Kromer, P. Kiefer, T. Binz, E. Heinzle, Impact of the cold shock phenomenon on quantification of intracellular metabolites in bacteria., Anal. Biochem. 327 (2004) 135–139. doi:10.1016/j.ab.2004.01.002.
[215] L. Krall, J. Huege, G. Catchpole, D. Steinhauser, L. Willmitzer, Assessment of sampling strategies for gas chromatography–mass spectrometry (GC–MS) based metabolomics of cyanobacteria, J. Chromatogr. B. 877 (2009) 2952–2960. doi:http://dx.doi.org/10.1016/j.jchromb.2009.07.006.
[216] H.H. Kanani, M.I. Klapa, Data correction strategy for metabolomics analysis using gas chromatography-mass spectrometry., Metab. Eng. 9 (2007) 39–51. doi:10.1016/j.ymben.2006.08.001.
[217] B.K. Matuszewski, M.L. Constanzer, C.M. Chavez-Eng, Matrix effect in quantitative LC/MS/MS analyses of biological fluids: a method for determination of finasteride in human plasma at picogram per milliliter concentrations., Anal. Chem. 70 (1998) 882–889.
[218] L. Wu, M.R. Mashego, J.C. van Dam, A.M. Proell, J.L. Vinke, C. Ras, W.A. van Winden, W.M. van Gulik, J.J. Heijnen, Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly 13C-labeled cell extracts as internal standards., Anal. Biochem. 336 (2005) 164–171. doi:10.1016/j.ab.2004.09.001.
[219] S. Noguchi, S.P. Putri, E.I. Lan, W.A. Lavina, Y. Dempo, T. Bamba, J.C. Liao, E. Fukusaki, Quantitative target analysis and kinetic profiling of acyl-CoAs reveal the rate-

limiting step in cyanobacterial 1-butanol production., Metabolomics. 12 (2016) 26. doi:10.1007/s11306-015-0940-2.
[220] S.V. Martínez, O.J. Krömer, Quantification of Microbial Phenotypes, Metab. . 6 (2016). doi:10.3390/metabo6040045.
[221] J.L. Markley, NMR analysis goes nano, Nat. Biotechnol. 25 (2007) 750. http://dx.doi.org/10.1038/nbt0707-750.
[222] K. Bingol, Recent Advances in Targeted and Untargeted Metabolomics by NMR and MS/NMR Methods., High-Throughput. 7 (2018). doi:10.3390/ht7020009.
[223] E. Layne, Spectrophotometric and Turbidimetric Methods for Measuring Proteins, 1957. doi:10.1016/S0076-6879(57)03413-8.
[224] V.A.R. Huss, H. Festl, K.H. Schleifer, Studies on the spectrophotometric determination of DNA hybridization from renaturation rates, Syst. Appl. Microbiol. 4 (1983) 184–192. doi:https://doi.org/10.1016/S0723-2020(83)80048-4.
[225] S.C. Hicks, R.A. Irizarry, quantro: a data-driven approach to guide the choice of an appropriate normalization method, Genome Biol. 16 (2015) 117. doi:10.1186/s13059-015-0679-0.
[226] M. Calderon-Santiago, M.A. Lopez-Bascon, A. Peralbo-Molina, F. Priego-Capote, MetaboQC: A tool for correcting untargeted metabolomics data with mass spectrometry detection using quality controls., Talanta. 174 (2017) 29–37. doi:10.1016/j.talanta.2017.05.076.
[227] B.M. Bolstad, R.A. Irizarry, M. Astrand, T.P. Speed, A comparison of normalization methods for high density oligonucleotide array data based on variance and bias., Bioinformatics. 19 (2003) 185–193.
[228] J.E. Jackson, Scaling of Data, in: A User's Guid. to Princ. Components, John Wiley & Sons, Inc., 1991: pp. 63–79. doi:10.1002/0471725331.ch3.
[229] R.A. van den Berg, H.C.J. Hoefsloot, J.A. Westerhuis, A.K. Smilde, M.J. van der Werf, Centering, scaling, and transformations: improving the biological information content of metabolomics data, BMC Genomics. 7 (2006) 142. doi:10.1186/1471-2164-7-142.
[230] J. Osborne, Estimating the False Discovery Rate Using SAS, in: SUGI 31 Proc., San Francisco, USA, 2006: p. Paper 190-31.
[231] D.M. Hawkins, Identification of oultiers, 1980. doi:10.1007/978-94-015-3994-4.
[232] T. Ebbels, Non-linear Methods for the Analysis of Metabolic Profiles, 2007. doi:10.1016/B978-044452841-4/50008-4.
[233] J. Xia, I. V Sinelnikov, B. Han, D.S. Wishart, MetaboAnalyst 3.0—making metabolomics more meaningful, Nucleic Acids Res. 43 (2015) W251–W257. doi:10.1093/nar/gkv380.
[234] F. Li, J. Wang, L. Nie, W. Zhang, Computational Methods to Interpret and Integrate Metabolomic Data, 2012. doi:10.5772/32517.
[235] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chemom. Intell. Lab. Syst. 58 (2001) 109–130. doi:https://doi.org/10.1016/S0169-7439(01)00155-1.
[236] J. Xia, D.S. Wishart, MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data, Nucleic Acids Res. 38 (2010) W71–W77. doi:10.1093/nar/gkq329.
[237] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, J.P. Mesirov, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles,

Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 15545–15550. doi:10.1073/pnas.0506580102.
[238] J.J. Goeman, S.A. van de Geer, F. de Kort, H.C. van Houwelingen, A global test for groups of genes: testing association with a clinical outcome., Bioinformatics. 20 (2004) 93–99.
[239] N. Kinsley, F. Sandrine, B. Emilie, J. Christophe, C. Corinne, C. Franck, Oxidative-stress detoxification and signalling in cyanobacteria: the crucial glutathione synthesis pathway supports the production of ergothioneine and ophthalmate, Mol. Microbiol. 100 (2016) 15–24. doi:10.1111/mmi.13296.
[240] O. ZITKA, S. SKALICKOVA, J. GUMULEC, M. MASARIK, V. ADAM, J. HUBALEK, L. TRNKOVA, J. KRUSEOVA, T. ECKSCHLAGER, R. KIZEK, Redox status expressed as GSH:GSSG ratio as a marker for oxidative stress in paediatric tumour patients, Oncol. Lett. 4 (2012) 1247–1253. doi:10.3892/ol.2012.931.
[241] C.R. Andersson, N.F. Tsinoremas, J. Shelton, N. V Lebedeva, J. Yarrow, H. Min, S.S. Golden, Application of bioluminescence to the study of circadian rhythms in cyanobacteria, Methods Enzymol. 305 (2000) 527–542.
[242] A. Taton, F. Unglaub, N.E. Wright, W.Y. Zeng, J. Paz-Yepes, B. Brahamsha, B. Palenik, T.C. Peterson, F. Haerizadeh, S.S. Golden, J.W. Golden, Broad-host-range vector system for synthetic biology and biotechnology in cyanobacteria, Nucleic Acids Res. 42 (2014) e136–e136. doi:10.1093/nar/gku673.
[243] S. Klähn, D. Baumgartner, U. Pfreundt, K. Voigt, V. Schön, C. Steglich, W.R. Hess, Alkane Biosynthesis Genes in Cyanobacteria and Their Transcriptional Organization, Front. Bioeng. Biotechnol. 2 (2014) 24. doi:10.3389/fbioe.2014.00024.
[244] D.M. Warui, M.-E. Pandelia, L.J. Rajakovich, C. Krebs, J.M. Bollinger, S.J. Booker, Efficient Delivery of Long-Chain Fatty Aldehydes from the Nostoc punctiforme Acyl–Acyl Carrier Protein Reductase to Its Cognate Aldehyde-Deformylating Oxygenase, Biochemistry. 54 (2015) 1006–1015. doi:10.1021/bi500847u.
[245] E.N.G. Marsh, M.W. Waugh, Aldehyde Decarbonylases: Enigmatic Enzymes of Hydrocarbon Biosynthesis, ACS Catal. 3 (2013) 2515–2521. doi:10.1021/cs400637t.
[246] B.E. Eser, D. Das, J. Han, P.R. Jones, E.N.G. Marsh, Oxygen-Independent Alkane Formation by Non-Heme Iron-Dependent Cyanobacterial Aldehyde Decarbonylase: Investigation of Kinetics and Requirement for an External Electron Donor, Biochemistry-US. 50 (2011). doi:10.1021/bi2012417.
[247] N. Li, W.C. Chang, D.M. Warui, S.J. Booker, C. Krebs, J.M. Bollinger, Evidence for Only Oxygenative Cleavage of Aldehydes to Alk(a/e)nes and Formate by Cyanobacterial Aldehyde Decarbonylases, Biochemistry-US. 51 (2012). doi:10.1021/bi300912n.
[248] P. Hu, S. Borglin, N.A. Kamennaya, L. Chen, H. Park, L. Mahoney, A. Kijac, G. Shan, K.L. Chavarría, C. Zhang, N.W.T. Quinn, D. Wemmer, H.-Y. Holman, C. Jansson, Metabolic phenotyping of the cyanobacterium Synechocystis 6803 engineered for production of alkanes and free fatty acids, Appl. Energy. 102 (2013) 850–859. doi:http://dx.doi.org/10.1016/j.apenergy.2012.08.047.
[249] Y.-X. Cao, W.-H. Xiao, J.-L. Zhang, Z.-X. Xie, M.-Z. Ding, Y.-J. Yuan, Heterologous biosynthesis and manipulation of alkanes in Escherichia coli, Metab. Eng. 38 (2016) 19–28. doi:http://dx.doi.org/10.1016/j.ymben.2016.06.002.
[250] X. Song, H. Yu, K. Zhu, Improving alkane synthesis in Escherichia coli via metabolic engineering, Appl. Microbiol. Biotechnol. 100 (2016) 757–767. doi:10.1007/s00253-015-7026-y.

[251] W. Wang, X. Liu, X. Lu, Engineering cyanobacteria to improve photosynthetic production of alka(e)nes, Biotechnol. Biofuels. 6 (2013) 1–9. doi:10.1186/1754-6834-6-69.
[252] C. Andre, S.W. Kim, X.-H. Yu, J. Shanklin, Fusing catalase to an alkane-producing enzyme maintains enzymatic activity by converting the inhibitory byproduct H2O2 to the cosubstrate O2, Proc. Natl. Acad. Sci. . 110 (2013) 3191–3196. doi:10.1073/pnas.1218769110.
[253] Y.E. Cheah, S.C. Albers, C.A.M. Peebles, A novel counter-selection method for markerless genetic modification in Synechocystis sp. PCC 6803, Biotechnol. Prog. 29 (2013) 23–30. doi:10.1002/btpr.1661.
[254] A.M. Kunjapur, K.L.J. Prather, Microbial Engineering for Aldehyde Synthesis, Appl. Environ. Microbiol. . 81 (2015) 1892–1901. doi:10.1128/AEM.03319-14.
[255] Y.-X. Cao, W.-H. Xiao, J.-L. Zhang, Z.-X. Xie, M.-Z. Ding, Y.-J. Yuan, Heterologous biosynthesis and manipulation of alkanes in Escherichia coli, Metab. Eng. 38 (2016) 19–28. doi:http://dx.doi.org/10.1016/j.ymben.2016.06.002.
[256] X. Song, H. Yu, K. Zhu, Improving alkane synthesis in Escherichia coli via metabolic engineering, Appl. Microbiol. Biotechnol. 100 (2016) 757–767. doi:10.1007/s00253-015-7026-y.
[257] B.K. Kaiser, M. Carleton, J.W. Hickman, C. Miller, D. Lawson, M. Budde, P. Warrener, A. Paredes, S. Mullapudi, P. Navarro, F. Cross, J.M. Roberts, Fatty Aldehydes in Cyanobacteria Are a Metabolically Flexible Precursor for a Diversity of Biofuel Products, PLoS One. 8 (2013) e58307. doi:10.1371/journal.pone.0058307.
[258] J. Zvezdanović, S. Petrovic, D. Marković, T. D. ANDJELKOVIĆ, D. H. ANDJELKOVIĆ, Electrospray ionization mass spectrometry combined with ultra high performance liquid chromatography in the analysis of in vitro formation of chlorophyll complexes with copper and zinc, 2014. doi:10.2298/JSC130918009Z.

SUPPLEMENTAL FIGURES

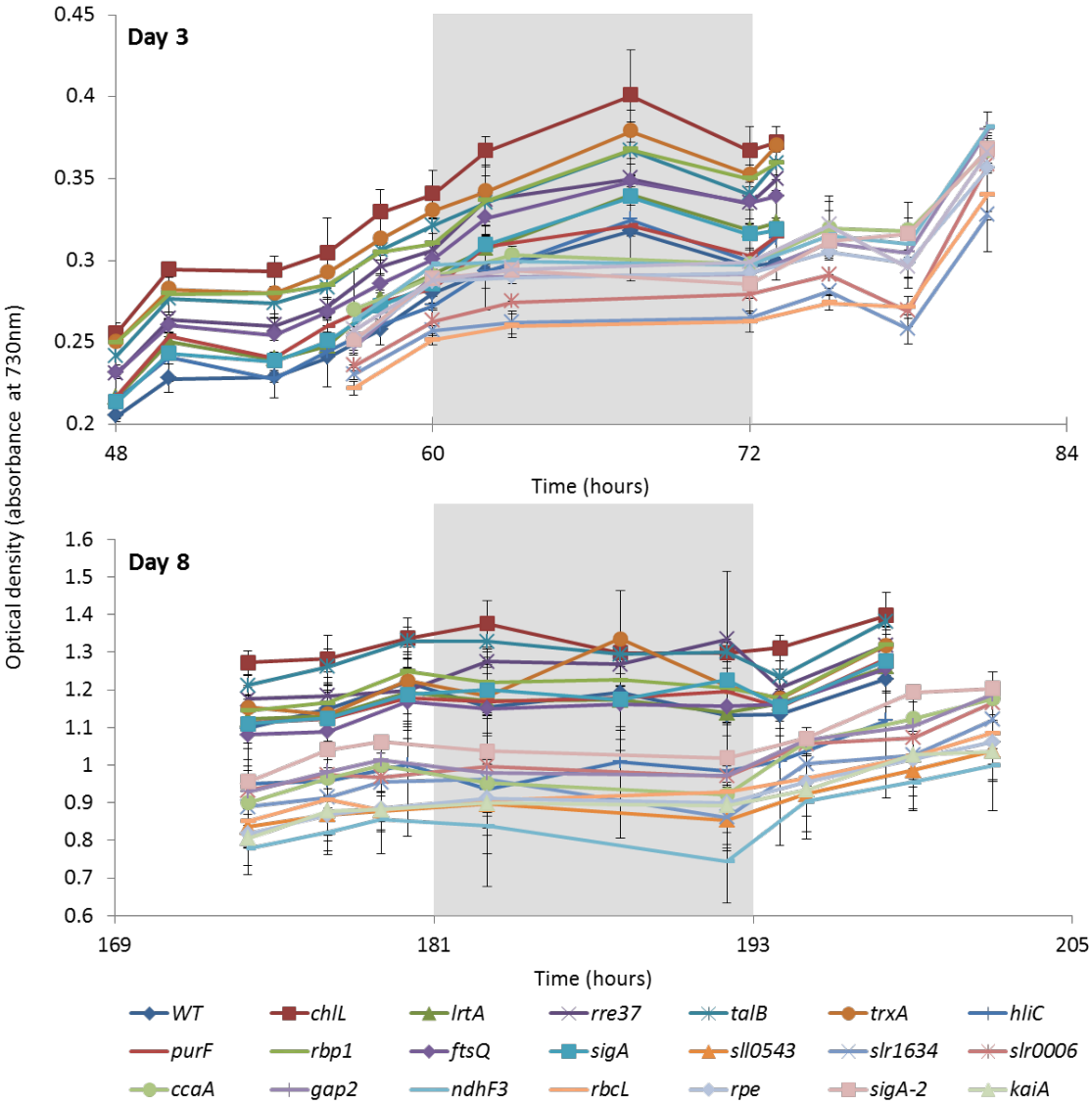

Figure A-5-1. Growth measurements of promoter:luxAB strains in LD cycles on day 3 and day 8 post-inoculation. Error bars represent standard deviation from biological duplicates and technical triplicates. Shaded areas indicate periods of darkness.

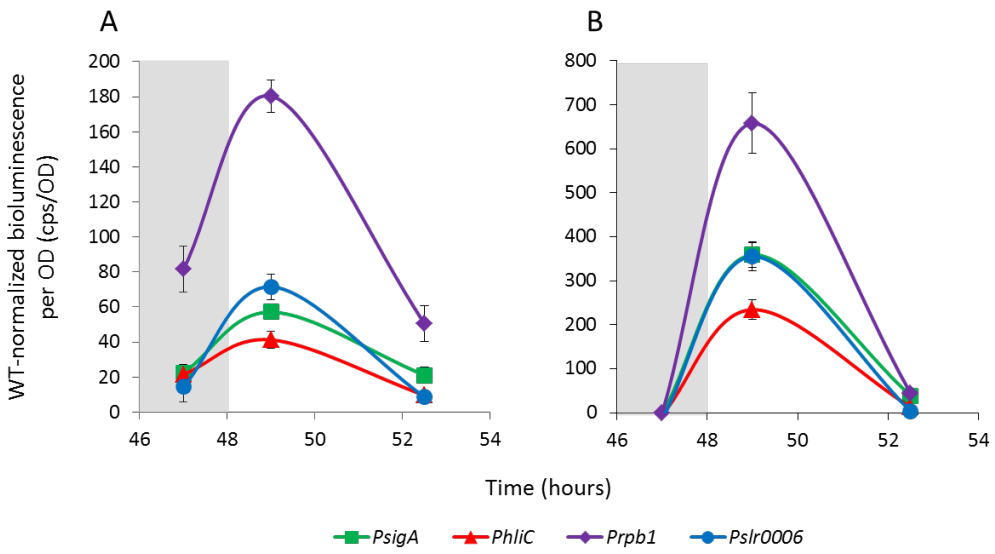

Figure A-5-2. (A) In vitro versus (B) in vivo bioluminescence assays. Cells were grown in 12:12 LD cycles for two days prior to measurements. Error bars represent standard deviation from biological triplicates. Shaded areas indicate periods of darkness.

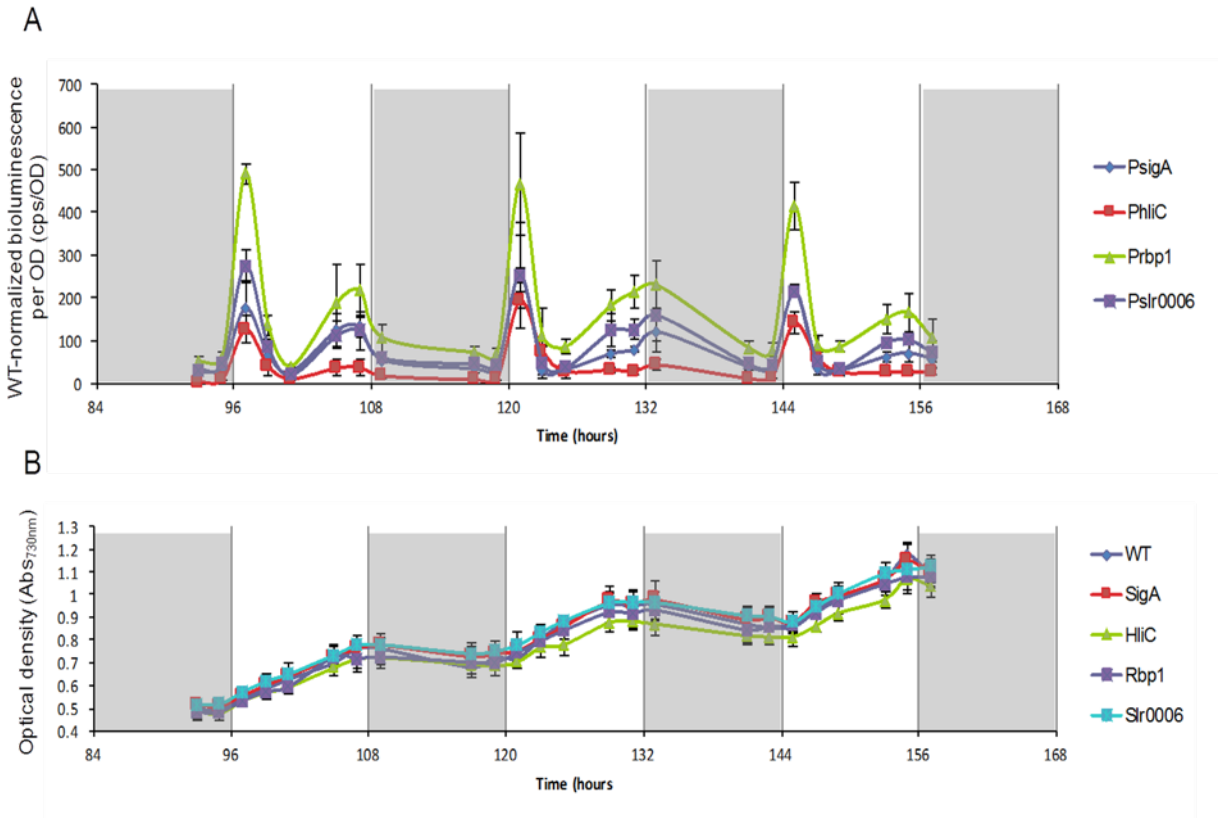

Figure A-5-3. PsigA, PhliC, Prbp1, and Pslr0006 provide light-entrained expression in 12:12 LD cycles. (A) Wild-type normalized bioluminescence per OD over three consecutive days of growth. (B) Optical density (absorbance at 730nm) as measured by the FLOUstar Omega Microplate Reader. Error bars represent standard deviation from biological triplicates with technical triplicates. Shaded areas indicate periods of darkness.

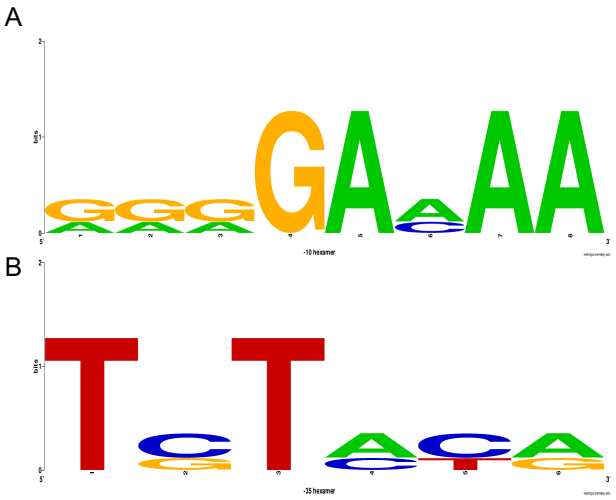

Figure A-5-4. Potential consensus (A) -10 and (B) -35 hexamers for PsigA, Prbp1, and Pslr0006. Conserved regions were identified by CLUSTALw multiple sequence alignment [167] and figures were generated by Weblogo (http://weblogo.berkeley.edu/logo.cgi).

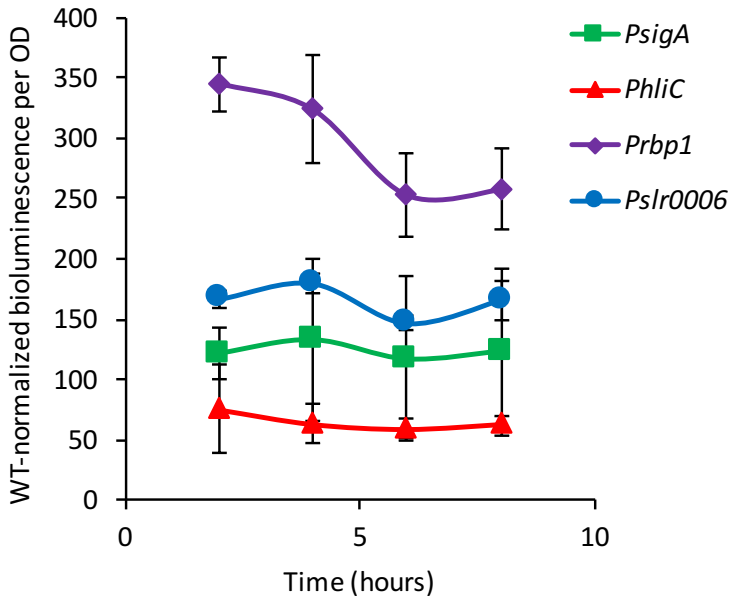

Figure A-5-5. Bioluminescence in continuous light following two days of cultivation in continuous light. Error bars represent standard deviation from biological triplicates.

SUPPLEMENTAL TABLES

Table A-1. Primers used in this study. Restriction endonuclease recognition sequences are underlined.

Name	Purpose/Description	Sequence (5'→3')
27F	pAM1580 site-directed mutagenesis 3754 A->T fw	gtg aga agg aag ttt cca aat ttc ata tga ttt cct tat ttg ttg gta tta cga t
27R	pAM1580 site-directed mutagenesis 3754 A->T rv	atc gta ata cca aca aat aag gaa atc ata tga aat ttg gaa act tcc ttc tca c
28F	pAM1580 site-directed mutagenesis 8744 C->T fw	gat tgt act gag agt gca ctg tat gcg gtg tga aat acc g
28R	pAM1580 site-directed mutagenesis 8744 C->T rv	cgg tat ttc aca ccg cat aca gtg cac tct cag tac aat c
44F	P*slr0749:luxAB (chlL)* from *Synechocystis* sp. PCC6803 genome fw	TAT AAG CTA GCG TTT GAT TGC GCA ATT ATG
44R	P*slr0749:luxAB (chlL)* from *Synechocystis* sp. PCC6803 genome rv	TAT AAC ATA TGT ATT AAC GAG GTT TGG GG
43F	P*sll0947:luxAB (lrtA)* from *Synechocystis* sp. PCC6803 genome fw	TAT AAG CTA GCG CGC TAC AAA GGA AAC
43R	P*sll0947:luxAB (lrtA)* from *Synechocystis* sp. PCC6803 genome rv	TAT AAC ATA TGA CTG AAT TCT CCC TTC G
45F	P*ssr2062:luxAB* (hypothetical protein) from *Synechocystis* sp. PCC6803 genome fw	TAT AAG CTA GCC CTG GGA TCA AGA CAG
45R	P*ssr2062:luxAB* (hypothetical protein) from *Synechocystis* sp. PCC6803 genome rv	TAT AAC ATA TGG GGG GTG CGT GGG G
47F	P*sll1330:luxAB (rre37) from Synechocystis sp. PCC6803 genome* fw	TAT AAG CTA GCA GAT GAC CCG GAA TAG
47R	P*sll1330:luxAB (rre37) from Synechocystis sp. PCC6803 genome* rv	TAT AAC ATA TGC GAT CTA GAC CCC TGA C
48F	P*sll0543:luxAB* (hypothetical protein) from *Synechocystis* sp. PCC60803 genome fw	TAT AAG CTA GCT TAA AGT TAC ATT TTT TG
48R	P*sll0543:luxAB* (hypothetical protein) from *Synechocystis* sp. PCC6803 genome rv	TAT AAC ATA TGC TAC GGT GCC
68F	P*slr1793:luxAB (talB)* from *Synechocystis sp. PCC6803* genome fw	TAT AAG CTA GCG GCA AAA GTT ACA GAG GTT G
68R	P*slr1793:luxAB (talB)* from *Synechocystis sp. PCC6803* genome *rv*	TTA TAC ATA TGT GTA AAT TGG CGC AGT TGT TC
82F	P*slr0623:luxAB (trxA)* from *Synechocystis* sp. PCC6803 genome fw	TAT AAG CTA GCG AAA TGC GGA CTT TAG CCA GGA AT
82R	P*slr0623:luxAB (trxA)* from *Synechocystis* sp. PCC6803 genome rv	TTA TAC ATA TGA CTG GAA ATC CTT CTA AAC CTC GTG ATG GGC TGA T
86F	P*sll0616:luxAB (secA)from Synechocystis sp. PCC6803 genome fw*	TAT AAG CTA GCG CAG AAA AAG CCA TTG
86R	P*sll0616:luxAB (secA)from Synechocystis sp. PCC6803 genome rv*	TAT AAC ATA TGG GGA CTT GGG ATA GGG
85F	P*slr1634:luxAB* (hypothetical protein) from *Synechocystis* sp. PCC6803 genome fw	TAT AAG CTA GCG ATC AAA GCT AAA CGA C
85R	P*slr1634:luxAB* (hypothetical protein) from *Synechocystis* sp. PCC6803 genome rv	TAT AAC ATA TGT TCC TAC TCC TTC CCT TGA C
87F	P*ssl1633:luxAB (hliC)* from *Synechocystis* sp. PCC6803 genome fw	TAT AAG CTA GCC AGA GAA GCT GCT AG
87R	P*ssl1633:luxAB (hliC)* from *Synechocystis* sp. PCC6803 genome rv	TAT AAC ATA TGG GTT AAG AAT GTA ATT ACG AGC TG

88F	P*slr0006:luxAB* (unknown protein) from *Synechocystis* sp. PCC6803 genome fw	TAT AAG CTA GCC AAC TGT TCA TCA AAA CC
88R	P*slr0006:luxAB* (unknown protein) from *Synechocystis* sp. PCC803 genome rv	TAT AAC ATA TGT GAA TTG TCC CTT GAA ACT GC
89F	P*sll0757:luxAB (purF)* from *Synechocystis* sp. PCC6803 genome fw	TAT AAG CTA GCC CAC GGT TGA AAC
89R	P*sll0757:luxAB (purF)* from *Synechocystis* sp. PCC803 genome rv	TAT AAC ATA TGG GCG GCA ATC TTC C
90F	P*sll0517:luxAB (rbp1)* from *Synechocystis* sp. PCC6803 genome fw	TAT AAG CTA GCC ACA GAC TAA CAA GC
90R	P*sll0517:luxAB (rbp1)* from *Synechocystis* sp. PCC6803 genome rv	TAT AAC ATA TGG GAT TTT CTC CAA AAA CTG AGG
91F	P*sll1632:luxAB (ftsQ)* from *Synechocystis* sp. PCC6803 genome fw	TAT AAG CTA GCC ACA ATT TTC TTC GTA TCC
91R	P*sll1632:luxAB (ftsQ)* from *Synechocystis* sp. PCC6803 genome rv	TAT AAC ATA TGG GGA GTT CAA TAA CCA TCA TC
92F	P*slr1347:luxAB (ccaA)* from *Synechocystis* sp. PCC6803 genome fw	TAT AAG CTA GCG ACG AAT AAT AGC CAA TTG
92R	P*slr1347:luxAB (ccaA)* from *Synechocystis* sp. PCC6803 genome rv	TAT AAC ATA TGG GAA AGG TTT TAA TGA CCT CC
94F	P*sll1525:luxAB (prk)* from *Synechocystis* sp. PCC6803 genome fw	TAT AAG CTA GCG TTA ATT GCT CCC TAC TC
94R	P*sll1525:luxAB (prk)* from *Synechocystis* sp. PCC6803 genome rv	TAT AAC ATA TGA ACA TCC CTC AAA AAT GCA AAA C
95F	P*sll1342:luxAB (gap2)* from *Synechocystis* sp. PCC6803 genome fw	TAT AAG CTA GCC TCC AGG GTT AAA CC
95R	P*sll1342:luxAB (gap2)* from *Synechocystis* sp. PCC6803 genome rv	TAT AAC ATA TGG TTC GTC TTG CCC TCT C
96F	P*sll1732:luxAB (ndhF4)* from *Synechocystis* sp. PCC6803 genome fw	TAT AAG CTA GCT TCG GAT ATT TAT GTA ATT CGG ATC
96R	P*sll1732:luxAB (ndhF4) from Synechocystis sp. PCC6803 genome* rv	TAT AAC ATA TGG GGG GTT GTG GGA AAT TTT G
97F	P*slr0009:luxAB (rbcL)* from *Synechocystis* sp. PCC6803 genome fw	TAT AAG CTA GCT CAC CAT TTG GAC AAA AC
97R	P*slr0009:luxAB (rbcL)* from *Synechocystis* sp. PCC6803 genome rv	TAT AAC ATA TGC TAG GTC AGT CCT CCA TAA AC
98F	P*sll0807:luxAB (rpe) from Synechocystis sp. PCC6803 genome* fw	TAT AAG CTA GCC GTT CAG GAT AAC AAT G
98R	P*sll0807:luxAB (rpe)* from *Synechocystis* sp. PCC6803 genome rv	TAT AAC ATA TGG GGA AGG GTC TCT AAG TAT TTC
100F	P*slr0653:luxAB (sigA) from Synechocystis sp. PCC6803 genome* fw	TAT AAG CTA GCC CGT CAT CGA TTC
100R	P*slr0653:luxAB (sigA) from Synechocystis sp. PCC6803 genome* rv	TAT AAC ATA TGG CCG TTT TCC TCG TTA AC
84F	P*slr0756:luxAB (kaiA) from Synechocystis sp. PCC6803 genome* fw	TAT AAG CTA GCA TTT TCC GGG TCC GG
84R	P*slr0756:luxAB (kaiA)* from *Synechocystis* sp. PCC6803 genome rv	TTA TAC ATA TGA GGG GAT CAA TAA TCA GTC GAA GGT TAA TC
32F	pAM1580 promoter insertion verification fw	GTT GCG AAG CAA CGG C
32R	pAM1580 promoter insertion verification rv	CAT GTG CCT GAG CAA CTT C
115F	InFusion overlap for inserting *promoter:luxAB into pIGA* fw	AAT GGA ATC AAA GTT CCG GAT CCG GCC GGA TC
115R	InFusion overlap for inserting *promoter:luxAB into pIGA* rv	GGT AAA CAG AGG GCG CAT CGC TGC CCC TTC AGC
117F	pAZ-IF sequencing fv	GTT GTC AAA GGC AAT CTG TTG GG

117R	pAZ-IF sequencing rv	GCA GAG CAT TAC GCT GAC TTG
114F	*slr0168* homolgous region sequencing fw	CCT CAC CGC CAG TAA GGT CAC
114R	*slr0168* homolgous region sequencing rv	CTG CAC CGT TAT CGT CCA GAA GTC
122F	*luxAB* qPCR fv	TGG TGC TCC TGT TTA TGT CG
122R	*luxAB* qPCR rv	CGT TGT AAA GAT CAA GCT GCG
123F	*rnpB* qPCR fw	GTG AGG ACA GTG CCA CAG AA
123R	*rnpB* qPCR rv	GAT ACT GCT GGT GCG CTC TT
138F	*sigA* qPCR fw	GCC GGA TTC GTC TGC TC
138R	*sigA* qPCR rv	CCA ACT TCC AGA CCT GTT TAC C

Table A-2. Promoter sequences.

Promoter	Sequence (5'→3')
PchlL, slr0749	Gtttgattgcgcaattatgctaatttcccgctgtttggcaaaaattcacatttattcatctttccttggtcttttctggtcaatcggtcattttctca gttgcaactctggtcgtaaccacggcaatcggacaaaatctagttagttcactgactgcttattttcgcctactaatttctgcatagctctgcc agtaagagtttaaagtctaattctttttagtacaactctcatttttgtcttaatcccatctttggcgatcgccataagtaactatacccagacaag ttattttattttttgtctcaatttaataaaaataaaaacacaaaatttgtttttaggcgttttaaagactagaagcagttttaaaagcaattaaaacg aatttaagttttaaaaataagatttaagccagattaataccccaaactatttctccaagaccccaaacctcgttaata
PlrtA, sll0947	gcgctacaaaggaaacggaatcaaactaaagttcaaagttggcagaaattaagaaacgtaaagagatgcaaaggaaagtcaaaatca cctgaccgattaggtcttattcaatacatagtgctaatctgaagatagtcttaggagttaattatttaccaccacaattttctggaaaactttac ctctaccctagggatgattaaaagtaaactagagaataacaaggttgggtttataattcatcaccaagctcaaatttatggtgttttttcaatg atccatgcttttgatatctttagcagaaaggcattttaagtaatgattccacctcactgtttctcggaaaaattgcccaatctaacttagtttttat aacttaagtttagatctgcggaaaaccaaccattgctcatttttttattaattttacgaagggagaattcagt
Prre37, sll1330	AGATGACCCGGAATAGATTTTTAAAGCACCCAGACTCCCCAGATTTGGGGCGAAC ATCAAAGTCTCCTGGTATTGAGGGAGCTTGATTGGGTAAATTTGTGCTCTACACC AAAACCTTGCTAAGGGATTTTCAAATTTTTTCATGGCCCATTGGTCTGACAAAAGT TAACCGTTGAGCGACGGAAATGGTACGAACAATTAACATTATGGCGAGCAAATT CTTTATATTTTTTGAAGATTTGCCCCCCCATGCAGGTAACTGTTGTTACAAAGCCT TGACATTGACTTTGTTAGATTAACAGGGAACTTTCAGTCAGGGGTCTAGATCG
Psll0543	ATGTTTAGTTTGAAAAATTATTCCCCATCTCCATTACTATCCGCTAAGGCCAAGGA ATATTTAGTATTCACTTTGGTAACCCTAACCATTGTCGTTGCTGGCACTCTTCTGTT CCTGTTTGATCCTGCTTCATCGAAATTATTTCCCCCCAGCCCCTTTCGCAGTTTAA CTGGGCTGTACTGCCCCGGTTGCGGCACTTTACGGGCGTTGCATCAATTATTGCAT GGCAACTTAGTGGGAGCTTTTGGCCTAAATCCTTTGATGATGATTTCCCTCCCATT TATGATTTATTCCTACGTTGCCTATGGCTTAAAAACTATTAATGGGCGGACTTTAT TAGTTTGGTTTGTTCCCCCCAAGGTAATCTGGTTCATACTACAGGCGATTTTAGCT TATTGGGTGGCCAGAAATATTCCCTTTGCACCCTTCTCTTGGTTGGCACCGTAG
PtalB, slr1793	GCTAGCggcaaaagttacagaggttgAAAAATAGTGAACAGTAATTGGATGTCAGTCCGAAG CCAATTTGCCCCCTAGGGTTCGGCAACGCATCGGCGTTCCCTAGCTTAAATTATCT TGATGTCCAAAGAAGGCTCGCTTTCGGGAATCACCATTACCATTGGAGGGAAAGA CGTTGAACTTGTTTAATCGCCATTATGGGTAAGAATCTCCTCgaacaactgcgccaatttaca
PtrxA, slr0623	GAAATGCGGACTTTAGCCAGGAATTTCGACCCCGCCTCCAACAAAATTACTCTAC CCTCCGGCAGATTTCTGGCACAGTTAATCGCCCCAAAAAAGCCAGCTGCTCCTCC CCCGACCACAATCACAGTTCTTGTCATAGTTTTGACCAACCTTTATCTCTGGATTT CACTGGAAAATGGATCTAATCACCCCAAAAATCCCTTTAAAAAACTTAACAAATA CGGAACTCCCCACCGGCAAAAACCCTATGCCCCCCGTCCCAACCTGTACAATGAA GAGGGCGGAGACGTAAGTTTCCGTTCACTCCTCACACCACACTCCGCCTGGATGA TGTTCGGGCGGTTTCTTCTTATCTGCTCCCCAGGGGGAAAAGTGTGACGCCAACT GTGACAAAAGATGAATAAATTCTAAGTTTCACGATATTTTTCCATACAGGGGTCA ACAATTGGTTATGGTAGTATTCTAATCAGCCCATCACGAGGTTTAGAAGGATTTC CAGT
Pslr1634	GATCAAAGCTAAACGACAAACTAGCCATGAAATCAGGCTTTAAACTGAGATAAA TAGAGTTGAGGTTAATTTTTACCTATGGCTGGGAAACAGCAGAGTGTAAGAATAA TGAATAAATGTCAATGATTAAATGTAAATAATTATAAAGAAATCTCAGTTTCTCG AAGTCGTTGCCTATACTGGACTCAACAGCCCCCAGTCGGAGCCTTCAATTTGGGC ACCGTTGTTTGGGGTCTGCCTCACCCGATTGTTGTTGTTGGAGTTAATAGTTATGG AGCTTTCGTCAAGGGAAGGAGTAGGAA
PhliC, ssl1633	CAGAGAAGCTGCTAGGGGGAATTATTGGCGGGTTAGTTTCCAGTGCCTGCCATTG CCCTAATAATGGCGATCGCGCTCCAGGTCAAAAATTGCCTTTTCCCAATACCAGT GCTCCGGGCGATCAGGATGTTTGGCCCGTAGATTGGTTACCAGTCTTTCCGCTACC TGGGCATTGCCCGCAATACGAATTAGATCCCGCCGTAACTTACCGGGCACCCTGG GTGATGTTGGGCGGGAAGCTTGGATGGCCTGCTTGAAACCTGATGAATGACGACG GCCTCTGCTCCAGGCAATTTTCCGACCTGATCTACGCCTAGCCCACCAGATCAAG CCCCACAACAGGGCGATCGCCAAAATTAGGGCCAACATTGCCTTGTTCTCCGAAG GTTAGTCACTTCTGTTCTGTTTGAGCCTAACATTATCCTCCCGTTGCCTTTTTCACC

	ACAGACTTGCCATGGGCGCAATACTCTGTTACATTTATTTACATAGACAACCCTCC CAGCTCGTAATTACATTCTTAACC
Pslr0006	CAACTGTTCATCAAAACCCACCGTTGGCGTTAGATTTGCCCCCGTTACTGGCATTA ATGGTCGCGACTGGGCCCTAAGACCACTGCCATGGAAGCCAAAAACCAACAGTA AAGCCAATAACGGGAAAAATTGCTTAATCATTTATTGCTTTGTTGAAATTACCTA CCGCTGACGAAAGTTAAGTTTACCTGCTCCGTCGGGAATAGGGGATGGGAAACC ACCAGCGCACAGACTAGCCCAGGGCCATATTTAGCTAAATACCATCATTAGGGAC TTTTATTTATTAATTTTTAAGAAGAAAAATTGAGCGATTTTAATGGCTATGTCTGG GTATTTTTACCGAGATAAATTGTCCCAAAATCAAGAATGATAAGTTTATATAATG GTAGTCAAAACTATTTTAAAAATTTCCCCAAGTCTTGCTCCCTACCTAGATTACGT AAAAAATTAACTGTTTAGAAGCGTTTTTATACCTAGAGTCTAAATTGAGTGAACT AAGAATCTCTGTTAATCTGGAGGCAGTTTCAAGGGACAATTCA
PpurF, *sll0757*	GCTAGCCCACGGTTGAAACAGAGGTAAAAACTGGGGTTTTAAGGGAGTGTCCAG GCAAAAAACACCTGGAAAAAAGACCCTTTAATTGGACTATTTTTCGATGCGGGGG GGTTCCCGAAAGGCGATCGCAAAGAAAAGGACCGCAAGGGCCATGGTCAGCACC AGAATGTAAGCAACACTTTCCATAATCTAAAAAATTCCTAAAAGTTTATCCTTAC ACACTAATCTATACATTCCCCTGTCCCCTGACAACTAGACCGACTTGGCCAAGAA CCAAACACTGCGGTGTGGGCACTGCTTCTTTGCTAGTATGGGTTAACGATTAACA ATTGATTAAGGGGTCTAGACCCGTTCGTCATTGCCTTCCCTGTCTGGCGGGAATCC ACCAATTCCCATAGTCTAAGATTTAGTTTCTGTACCCGTTTTTCTCTACCCTGATCC AGGAAGATTGCCGCC
Prbp1, *sll0517*	CACAGACTAACAAGCCAGGTAAATTTCCGTCCGCCACTAACCAGGCCACCACCGA GTCAACGGTTTGTTGATGCCCCACCCGGATGTCTTTTTGCACGAAATTGACTTGGC CGTCACCATAGATCCGGCGCCAACGGTCAAAGCCGGTGGAATTAAAATAGTCAC GAACGATGGTTTTATCGTCTAGGGCGGCGTTGGTCATGGAAAGCTAAGCTCGAAA ATTTGAGTTAGGTAAAGGTTGGTATAAGCTATAGGCACGAGCCGTCTAAGCAATG GTACCATTCTTTTCTGGTTACTCTGGGGGAGGAACTTTCGTGGTTCCATATCAGGA ATTCAAGACCTAATATCCCCAGCCGAGGGGAGAACGTGTTACTATGCCCCCATAG AGTAATGATCGGTTGGTAGCATTTGTTTCTAGTATTAACGGTTTTTCGCGTTTTTC CATTGACAGGCATTTCTCCGAAATCACCCTCTACATATCCCTCAGTTTTTGGAGAA AATCC
PftsQ, *sll1632*	CACAATTTTCTTCGTATCCTGGTCCTGGGTGATTTGAATACCCACCCCGGTCAATT CACCGGAAGTGTCGATACGCATTGATTGAAACTCATCGGGGGACATGAAGCGAG TGTAGGGGTCATTTAGCTTCTCCAGCATTTCCCGAATGGCGGTGTAGGCTTCTTCC TGGTTTTTGTAGTCCCGTGTTAGATAATCCTGGCGAACCGCTACCCAATCCTCACC GTTGAAAGTACCATCTACATAGGTGCGGTTGACAATCTGCCAAACTTCGTCAACC AGCTCCTTGGGATTATCCTGTAGGTAGCCCTGGGAGCGGGCCAATCGCAACCCAA CCCCCGTCACCGCCACTGTTGTCAATAACAGAGCTGTAGTTCCCAAAATCAGACT GCGTTTTTGTTTCAACATGGGATTTACCGGACAGTTGGTGAAGATTAGTTAAGAA AATAAACCCTAGGGCGGTGCTTAGAGTCCTTGATTCCTAGGTTATAACATTTCAG AGATGATGGTTATTGAACTCCC
PccaA, *slr1347*	GACGAATAATAGCCAATTGTTGGGCATCTTTGGCACCGAAATAAGCCACATCGGG ACAAACAATGGTGAACAATTTAGTGACGATAGTAGCTACCCCCTGGAAATGGCC GGGACGATGGGCACCGCAGAGAACTTCCGTCATGGCTGGAGGAGGCAGTACCGT AGTTTGATCTCCTTTGGCATCAATGCCCAAATCCGTCACCGTCGGGGCAAAAATT ACCGCCACCCCCAAAGATTCTGCCCATTGGCGATCGCCGTCGAAGTCCCGGGGAT ATTTTTCCAAATCTTCACCGGGGCCAAATTGCAGGGGATTAACGAAAATAGATAA CACGACTAAATCCATTTCCGCCACCGCCCGTTTGAGCAGACTGCCATGGCCGGCG TGGAGAGATCCCATGGTAGGAACTAAGCCCACGGTTTTGCCCCGACCAGCTTGGC GGAGATAGGTTTGTAACCCTGCGATTGTGCGAAAAACCTGCACCATTACTTCTTT GTTATTCTTTGGATTGTGTGCAAACGATGGAGGTCATTAAAACCTTTCC
Pgap2, *sll1342*	CTCCAGGGTTAAACCTGAACCGGGGGAGCCAGTGGGTAAAGCCACCGCTGGGTT AACTTGCAGGGATTGAAAATAGTCTAAATTCTCCGGCACCTGATCTTGGAAGATT TTTGCCCCCACGGACTCCAGGCGATCGGTAATATGGGTGCGGCGTAGGTCAGAGC CCGACACAGGCAGATGGCGTTTTGCGAGGACGTAGGCCAAAGCCGACATACCAA TGCCCCCAATTCCCAAGAAATGGAAAGGACGACCACTAAAATCCACAGCACCCA

	CAGCAACGCTCCTCACAAGCCATAAATGTTAACGTCTACAATCACCTGCCGGTTC ACCGCACCTGTAATTGAGCGACATCATAGCAAAAACTCTACCCCCAGGCATAGAC AAAACCCAGAGGCCACGGTAAACAGCGACGGAGAGAAAGGGAAAAAGATTCCC ACCCCCCCTTTGAAATTCGGGATCTAAGTGTTCAGGCAATAGGATATGATTGGTG TCATTAATAATCCTTAAATGTACTGAGCTAAAACAGAGAGGGCAAGACGAAC
PndhF3, sll1732	TTCGGATATTTATGTAATTCGGATCAAACTCTCAAATTTTCCCTTGGAGTGGACTG CCACCCCGTAGCCTGCCTCTTCCTCCAATTACCCAGACTTATCATTCACTTAAGCA AATGGTTGTTGATCTGGAATTTTATTTCCCCTATGATATGGTGTAACAAATAACAC TCTTTCTATGCATCTAGGTTATACCATTGTCGCTTGACTATGGTAATGCCCCTTTG GGGCCTGGATAAGATCAGCCATTTGGGCCGTCACTTTGGATGTATTAGCTTTCCCT GTCTGGGTTCTGGCAGTCCCAGATTCCTAGCACCGAGTAAGCCGGAAAATTCTTC GGTGACCGGGATGTTTGGACGGCGAAAGTGGAATCTTGTTAACAAAATTTCCCAC AACCCCC
PrbcL, slr0009	TCACCATTTGGACAAAACATCAGGAATTCTAATTAGAAAGTCCAAAAATTGTAAT TTAAAAAACAGTCAATGGAGAGCATTGCCATAAGTAAAGGCATCCCCTGCGTGAT AAGATTACCTTCAGAAAACAGATAGTTGCTGGGTTATCGCAGATTTTTCTCGCAA CCAAATAACTGTAAATAATAACTGTCTCTGGGGCGACGGTAGGCTTTATATTGCC AAATTTCGCCCGTGGGAGAAAGCTAGGCTATTCAATGTTTATGGAGGACTGACCT AG
Prpe, sll0807	CGTTCAGGATAACAATGGCAAAAATGGCGATCGTGTCCTTGAAGGGAAAACCCT CCATGGTTTCTTCCCCCAACCTTAAGCTGATCAGATCTAATGCTCCAGAAATAATT GCCACCCCGATCAACATCAAGAGCATGACATTGGTGAACTGATCCAGCAAAATTT GCCAATTATTACGACCGGCGATCGTGGTAATTTCATTTAGTCCATAGGTGCGTCTT CGTTCCATCACAGTTTTTTCTGCCAACCCCCGGTGGGGAGCAGTGGTCAACCGAT GCAGGGTTTTATCGATGCCATAGCTATACCAAGCCTCGGCCTTAACTGGGGAGAC GGACACATCAGGGTCAGACATTAAAAAGTAGGCTTAAATTTAGGTTCCCATTGGT TCTAATAGGCTATTTCCTATTTTAAAGGCCGGGTTTGGGGGCCTACTACAATTGAC CGTAATTGTTTACACTGAATTTCAAGCTTGTTTATGAGTTGAAATACTTAGAGACC CTTCCC
PsigA, slr0653	CCGTCATCGATTCCGTTAACTTTTTGGCGATCGATGTCCCCTTAGTCCACTAATTA GTAATGATTAGTAATGAAAGCTTGACTAGGCAAAATTCCCCACACCAGTTGCAAA AAAAGATTGAAAAAGCCACTATCACGCAGAAAATTCCGTTGAGTCAAAAGCTCC AACCCCAGTTAGGTCAATCATCGGCAATAATTTTAAGTCTTTCTGTCCCCGCTCAC CAGCAATGTGAAAAAAAGTCGTGGTCAGACCACAATTAAGCCCCATTTTTTTGTT ATCGAGGCTACAATCAATGAAGTCTTTTTTCACTGATTCATCAAGCTCTACCGCTG ACTAGCCTTCTCTAACACTTTTACGCCTGAATCCTTGCATTCTGTTTTTCCCCAACG AAGAAGTTAGAGAACCCCGAATCAATAGACTGACCGAGCGCACATCATCTAGCA ACGCCTAGGGGTGGCCTCAACTTTGACTGACCAGCTAATTTTGTACACGACTTAG GAGTTAACGAGGAAAACGGC
PkaiA, slr0756	ATTTTCCGGGTCCGGTTGCTGACGGTAATAGTCGTCTAAAAGTTTGGCCACATCC AAAAGGCTGTCGGCGGGGGGATGCTGGCCGGCGAGGGGATTAATTCTGCTTGTC ATATACAAAAATTGTAAAAAATGGAGGGCGGCGATCAGGGGCTTAGACACCCAA ATCCTAGCCAAAAAGGGTTAACTAGCCAAGGGCTATCCATGGGCAAAGAGATAA AAGAAAAAGTCTCCAAATCCCTGGTCATAGAGAAAAAATTGCCAAAGTTACCCC AGGCCATACACGGCCCAGCGCCAAGATGGGGAGCACAAATTCAAACTTTGTAAA CAGGCCGGAAGCTATCCGGCCAAGGAGCACTCAGATTGTGTTAACGTTCAGGGG AGTTGCTTAACACAATTTTCCAATTAATAGTATTAATATTTTCTTAACTTGCACCG TACCATGGTGAGAAAGCCTATCTGAGCCCTTATTTGATTAACCTTCGACTGATTAT TGATCCCCT

Table A-3. Initial characterization in 12:12 LD cycles.

Abs730 (Group A)											
Time (hours)	**WT**	**PchlL**	**PlrtA**	**Prre37**	**PtalB**	**PtrxA**	**PhliC**	**PpurF**	**Prbp1**	**PftsQ**	**PsigA**
0	0.20	0.26	0.22	0.23	0.24	0.25	0.22	0.22	0.25	0.23	0.21
2	0.23	0.29	0.25	0.26	0.28	0.28	0.24	0.25	0.28	0.26	0.24
4	0.23	0.29	0.24	0.26	0.27	0.28	0.23	0.24	0.28	0.25	0.24
6	0.24	0.30	0.25	0.27	0.28	0.29	0.24	0.26	0.28	0.27	0.25
8	0.26	0.33	0.28	0.30	0.31	0.31	0.26	0.27	0.30	0.29	0.27
10	0.28	0.34	0.29	0.31	0.32	0.33	0.27	0.28	0.31	0.30	0.29
12	0.29	0.37	0.31	0.34	0.34	0.34	0.29	0.31	0.34	0.33	0.31
14	0.32	0.40	0.34	0.35	0.37	0.38	0.32	0.32	0.37	0.35	0.34
20	0.30	0.37	0.32	0.33	0.34	0.35	0.30	0.30	0.35	0.34	0.32
24	0.30	0.37	0.32	0.35	0.36	0.37	0.31	0.31	0.36	0.34	0.32
126	1.10	1.27	1.12	1.18	1.21	1.15	0.95	1.11	1.15	1.08	1.11
129	1.15	1.28	1.13	1.18	1.26	1.13	0.96	1.12	1.17	1.09	1.12
132	1.22	1.34	1.19	1.20	1.33	1.22	1.00	1.18	1.25	1.17	1.19
135	1.15	1.38	1.17	1.27	1.33	1.18	0.93	1.17	1.22	1.15	1.20
140	1.19	1.30	1.17	1.27	1.29	1.34	1.01	1.18	1.23	1.16	1.17
144	1.13	1.30	1.14	1.33	1.30	1.21	0.98	1.20	1.21	1.16	1.23
146	1.13	1.31	1.16	1.20	1.24	1.18	1.01	1.15	1.18	1.16	1.16
150	1.23	1.40	1.26	1.32	1.38	1.32	1.12	1.28	1.32	1.26	1.28
Abs730 STDEV (Group A)											
Time (hours)	**WT**	**PchlL**	**PlrtA**	**Prre37**	**PtalB**	**PtrxA**	**PhliC**	**PpurF**	**Prbp1**	**PftsQ**	**PsigA**
0	0.00	0.01	0.01	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.00
2	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.00	0.01	0.00	0.01
4	0.00	0.01	0.00	0.01	0.01	0.00	0.01	0.00	0.01	0.00	0.01
6	0.00	0.02	0.00	0.01	0.01	0.01	0.02	0.00	0.02	0.01	0.01
8	0.00	0.01	0.01	0.00	0.00	0.01	0.01	0.01	0.02	0.00	0.01
10	0.00	0.01	0.01	0.01	0.00	0.01	0.02	0.01	0.02	0.00	0.02
12	0.01	0.01	0.01	0.02	0.01	0.02	0.02	0.01	0.02	0.00	0.01
14	0.00	0.03	0.01	0.00	0.02	0.02	0.04	0.01	0.02	0.00	0.02
20	0.01	0.02	0.02	0.01	0.02	0.02	0.03	0.01	0.01	0.01	0.02
24	0.00	0.01	0.02	0.00	0.01	0.02	0.03	0.01	0.02	0.01	0.02
126	0.00	0.03	0.14	0.01	0.03	0.11	0.17	0.10	0.06	0.02	0.07
129	0.03	0.06	0.12	0.02	0.05	0.08	0.16	0.06	0.07	0.05	0.07
132	0.04	0.05	0.12	0.06	0.04	0.12	0.19	0.06	0.05	0.06	0.09
135	0.03	0.06	0.12	0.03	0.05	0.04	0.26	0.09	0.08	0.02	0.07
140	0.03	0.03	0.14	0.02	0.02	0.13	0.20	0.08	0.05	0.06	0.11
144	0.03	0.02	0.14	0.18	0.03	0.10	0.16	0.06	0.13	0.04	0.08
146	0.00	0.03	0.04	0.02	0.07	0.08	0.22	0.02	0.06	0.04	0.08
150	0.01	0.06	0.11	0.00	0.01	0.10	0.20	0.06	0.08	0.06	0.08
Abs730 (Group B)											
Time (hrs.)	**WT**	**sll0543**	**slr1634**	**slr0006**	**ccaA**	**gap2**	**ndhF3**	**rbcL**	**rpe**	**sigA**	**kaiA**
57	0.24	0.23	0.24	0.27	0.25	0.26	0.22	0.25	0.25	0.27	0.26
60	0.27	0.26	0.26	0.29	0.29	0.30	0.25	0.29	0.29	0.31	0.29
63	0.27	0.26	0.27	0.30	0.29	0.30	0.26	0.29	0.29	0.32	0.29
72	0.27	0.27	0.28	0.30	0.29	0.30	0.26	0.29	0.29	0.32	0.30
75	0.29	0.28	0.29	0.32	0.31	0.32	0.27	0.31	0.31	0.36	0.32
78	0.27	0.26	0.27	0.32	0.30	0.31	0.27	0.30	0.32	0.33	0.30
81	0.33	0.33	0.36	0.37	0.38	0.38	0.34	0.36	0.37	0.41	0.37
174	0.85	0.84	0.89	0.93	0.90	0.93	0.78	0.85	0.82	0.96	0.81
177	0.89	0.87	0.91	0.97	0.96	0.98	0.82	0.91	0.87	1.04	0.88
179	0.89	0.88	0.95	0.97	1.00	1.01	0.86	0.88	0.88	1.06	0.88
183	0.87	0.90	0.96	1.00	0.95	0.98	0.84	0.91	0.91	1.04	0.90
192	0.89	0.85	0.86	0.97	0.92	0.97	0.74	0.93	0.90	1.02	0.89

195	0.96	0.92	1.00	1.06	1.06	1.07	0.90	0.96	0.95	1.07	0.93
199	1.01	0.99	1.03	1.07	1.12	1.10	0.95	1.02	1.02	1.19	1.03
202	1.08	1.04	1.12	1.17	1.18	1.19	1.00	1.09	1.06	1.20	1.04
Abs730 STDEV (Group B)											
Time (hrs.)	**WT**	**sll0543**	**slr1634**	**slr0006**	**ccaA**	**gap2**	**ndhF3**	**rbcL**	**rpe**	**sigA**	**kaiA**
57	0.001	0.003	0.001	0.025	0.004	0.004	0.004	0.007	0.002	0.003	0.004
60	0.009	0.003	0.008	0.002	0.002	0.003	0.003	0.001	0.006	0.002	0.005
63	0.001	0.007	0.001	0.000	0.004	0.009	0.007	0.001	0.005	0.001	0.005
72	0.008	0.000	0.002	0.004	0.007	0.011	0.006	0.010	0.006	0.006	0.007
75	0.006	0.002	0.000	0.017	0.000	0.015	0.004	0.003	0.005	0.018	0.018
78	0.004	0.009	0.009	0.017	0.021	0.004	0.007	0.009	0.000	0.005	0.006
81	0.018	0.003	0.001	0.003	0.011	0.001	0.034	0.002	0.008	0.016	0.012
174	0.017	0.049	0.050	0.064	0.008	0.013	0.045	0.031	0.040	0.058	0.097
177	0.001	0.048	0.051	0.089	0.028	0.001	0.058	0.044	0.053	0.071	0.107
179	0.011	0.050	0.025	0.101	0.045	0.018	0.033	0.046	0.033	0.065	0.118
183	0.055	0.064	0.017	0.087	0.063	0.023	0.026	0.051	0.038	0.087	0.133
192	0.036	0.064	0.081	0.061	0.046	0.025	0.112	0.043	0.053	0.081	0.121
195	0.021	0.100	0.060	0.028	0.038	0.011	0.040	0.064	0.042	0.056	0.131
199	0.014	0.103	0.085	0.096	0.016	0.020	0.036	0.048	0.054	0.072	0.148
202	0.023	0.077	0.087	0.082	0.045	0.015	0.044	0.080	0.058	0.050	0.157
WT-Normalized maximum bioluminescence per OD (Group A)											
Time (hrs.)	**PchlL**	**PlrtA**	**Prre37**	**PtalB**	**PtrxA**	**PhliC**	**PpurF**	**Prbp1**	**PftsQ**	**PsigA**	
0	-37.1	-14.4	-1.6	-36.4	-45.4	21.3	-2.1	135.6	-10.8	62.0	
2	-53.7	9.9	-7.4	-31.6	-58.0	582.7	-3.2	534.5	-19.2	364.0	
6	-41.5	-4.0	-15.9	-16.2	-39.1	43.2	-17.4	315.0	-11.8	144.8	
8	-42.8	-7.6	-5.6	-22.5	-29.1	126.9	-5.1	796.5	3.2	271.4	
10	-56.9	-23.2	-23.1	-52.7	-69.5	45.8	-43.5	361.5	-54.4	140.9	
12	-40.2	-31.4	-42.7	-38.3	-51.0	35.3	-46.6	348.1	-45.6	161.0	
14	-36.5	-21.4	-25.7	-28.9	-41.9	14.8	-29.8	205.9	-33.7	97.9	
20	-3.1	-16.6	-13.4	-14.1	-34.9	28.4	-25.5	214.1	-29.3	104.4	
24	-48.9	-32.0	-41.9	-49.2	-63.1	18.5	-27.5	249.5	-49.9	120.8	
126	-6.0	-1.9	-0.7	-3.7	-5.9	36.3	5.4	239.6	-1.5	95.8	
129	-4.3	-1.8	2.3	2.7	-4.7	54.0	-4.0	550.9	5.6	200.9	
132	-10.8	-4.5	-5.8	-10.8	-11.7	30.4	-8.1	794.8	-8.0	270.1	
135	-2.7	-2.6	2.3	-5.0	-5.7	28.8	-2.2	382.3	-3.5	148.3	
140	-4.7	-2.0	-3.9	-7.0	-5.7	22.4	-5.4	194.3	-7.0	93.1	
144	-6.1	3.4	0.5	5.9	2.8	26.0	-1.9	201.6	-1.4	114.6	
146	-12.4	2.8	-1.8	-9.1	-13.7	183.7	-3.3	1429.7	-0.1	467.4	
150	-9.9	-8.6	-8.6	-13.0	-13.5	45.2	-7.6	420.5	-7.3	188.4	
Standard deviation of WT-Normalized maximum bioluminescence per OD (Group A)											
Time (hrs.)	**PchlL**	**PlrtA**	**Prre37**	**PtalB**	**PtrxA**	**PhliC**	**PpurF**	**Prbp1**	**PftsQ**	**PsigA**	
0	14.7	5.2	32.7	4.0	7.1	14.6	39.8	24.5	28.0	18.1	
2	5.4	27.4	15.3	13.3	2.2	14.8	5.4	52.7	0.1	20.2	
6	2.9	8.3	19.4	6.6	12.9	7.3	13.5	101.8	25.2	12.0	
8	2.1	10.9	35.9	7.1	19.7	11.9	6.8	278.4	18.4	43.0	
10	19.5	11.9	20.2	0.9	5.0	34.7	1.1	1.9	5.1	10.2	
12	4.2	4.8	6.1	5.8	1.8	21.0	5.2	57.4	5.7	58.8	
14	4.0	0.7	5.5	0.7	0.1	8.0	0.7	20.1	13.0	12.6	
20	35.5	11.4	15.2	8.4	10.3	1.5	5.0	25.9	10.7	20.9	
24	3.3	13.6	6.6	5.2	3.0	21.3	13.1	6.8	0.7	43.7	
126	0.4	3.6	2.0	1.5	3.8	10.7	12.6	74.7	0.7	10.1	
129	7.4	5.7	5.9	10.5	0.8	20.4	3.2	73.5	14.5	40.3	
132	0.9	1.0	2.7	1.5	2.9	1.2	3.1	8.6	0.4	82.2	
135	3.5	2.5	1.1	4.4	1.8	19.8	0.1	7.2	0.3	3.7	
140	0.5	4.7	0.1	2.1	0.4	12.5	2.1	3.7	2.3	0.6	
144	0.4	4.9	2.4	4.5	2.3	12.4	1.0	1.9	4.2	28.4	

146	0.5	8.6	3.2	0.5	2.2	52.7	1.3	139.4	5.4	191.2	
150	1.8	2.4	3.3	0.1	0.4	19.2	1.8	54.2	0.0	11.2	
WT-Normalized maximum bioluminescence per OD (Group B)											
Time (hrs.)	**sll0543**	**slr1634**	**slr0006**	**ccaA**	**gap2**	**ndhF3**	**rbcL**	**rpe**	**sigA-2**	**kaiA**	
57	35.1	27.6	379.9	-21.0	1.5	-0.7	16.3	-33.7	89.3	-41.5	
60	3.8	39.1	464.5	-19.9	6.0	97.3	47.4	-13.2	365.6	-20.7	
63	-53.6	-46.1	313.4	-73.8	-57.7	27.1	-30.8	-77.9	251.3	-92.4	
72	-12.8	12.4	188.1	-34.9	-6.4	27.4	-0.5	-24.6	286.2	-37.0	
75	-29.2	16.1	501.3	-38.7	45.2	117.4	74.8	-34.0	485.4	-51.6	
78	7.4	9.1	201.1	-41.8	27.7	58.7	18.0	-37.2	181.8	-27.0	
81	-8.3	12.0	347.5	-33.1	72.1	45.0	38.6	-8.5	227.1	2.8	
174	1.7	7.3	122.6	-10.0	43.3	17.0	21.2	4.4	118.2	-5.6	
177	4.1	14.6	124.5	-4.1	28.3	26.8	30.4	9.3	99.5	-6.1	
179	-0.6	13.6	150.8	-2.8	23.6	31.9	32.6	11.7	96.7	2.0	
183	-10.0	4.2	112.1	-14.1	10.3	5.5	4.2	-6.4	96.6	-15.7	
192	-0.5	15.4	118.6	-11.1	11.3	23.7	2.3	-5.1	96.4	-10.7	
195	-1.5	18.3	150.0	-7.2	23.1	12.9	13.6	0.3	197.9	-8.1	
199	-2.2	0.7	160.5	-9.6	12.1	8.4	3.9	-1.5	121.2	-7.7	
202	-3.4	8.7	180.5	-9.1	31.2	11.3	5.1	-1.2	203.2	-5.7	
Standard deviation of WT-Normalized maximum bioluminescence per OD (Group B)											
Time (hrs.)	**sll0543**	**slr1634**	**slr0006**	**ccaA**	**gap2**	**ndhF3**	**rbcL**	**rpe**	**sigA-2**	**kaiA**	
57	6.1	2.8	35.6	13.9	5.0	19.5	30.4	5.8	1.7	10.7	
60	3.1	15.2	22.9	7.3	4.5	15.9	34.3	7.1	52.1	14.9	
63	20.2	6.1	62.9	10.8	7.7	29.2	11.8	1.4	23.8	2.2	
72	8.0	4.2	60.4	2.8	3.7	12.2	1.8	13.6	90.2	9.0	
75	5.4	3.3	136.8	3.9	8.7	16.9	4.6	11.9	175.7	3.4	
78	9.3	5.4	13.9	5.9	24.1	10.3	26.5	3.5	168.9	22.9	
81	5.0	21.7	51.7	0.8	7.7	13.7	31.0	1.2	114.4	39.5	
174	0.4	4.4	37.5	0.3	3.8	1.6	9.0	5.5	88.4	3.5	
177	6.8	0.4	10.8	8.0	3.0	1.6	8.3	4.9	16.0	0.0	
179	2.1	7.1	43.3	2.7	9.5	0.4	5.2	4.0	3.7	1.1	
183	2.5	7.2	27.4	0.5	11.2	1.2	3.0	1.3	5.8	6.8	
192	2.6	8.0	30.6	1.0	1.8	16.5	3.2	0.8	8.3	1.0	
195	6.7	7.8	84.0	4.5	9.5	0.7	6.8	3.5	82.1	7.4	
199	6.1	0.8	107.6	0.8	0.4	2.3	1.9	4.6	52.3	6.8	
202	3.2	4.2	99.2	4.3	12.1	0.6	1.1	3.3	106.7	5.6	
150	-9.9	-8.6	-8.6	-13.0	-13.5	45.2	-7.6	420.5	-7.3	188.4	

Table A-4. RnpB transcript accumulation in 12:12 LD cycles as measured by RT-q-PCR.

Threshold cycle, Ct ± STDEV						Fold-change ($2^{-\Delta\Delta Ct}$)
S. 6803 strain	Target gene	D9	D11	L1	L3	D11 to L1
WT *S.* 6803	*rnpB*	13.06 ± 0.66	13.40 ± 0.46	13.58 ± 0.49	12.99 ± 0.82	**0.89**

APPENDIX B – SUPPLEMENTAL INFORMATION FOR CHAPTER 4

SUPPLEMENTAL FIGURES

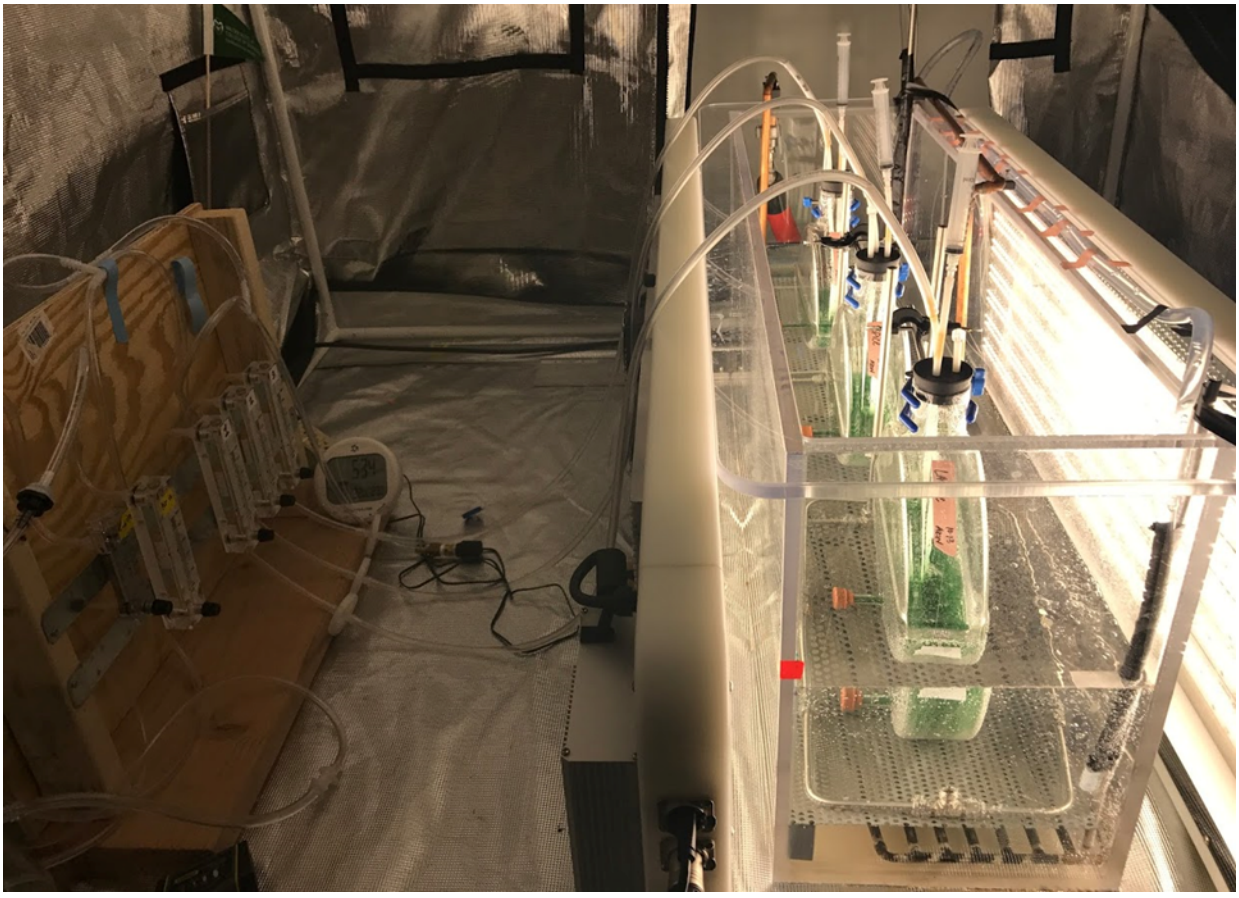

Figure B-5-6. Light-emitting diode photobioreactor. Three flat panel reactors (1.5 L working volume) are arranged down the center of a water bath and illumined by two 4000K white LED panels on either side. Bath temperature is controlled via circulating thermofluid in a heat exchanger. Uniform water bath temperature is maintained via a circulation pump. House air is mixed with 100% CO2 to a user-defined concentration; gas delivery rate to each reactor is controlled with variable flow meters. Reactors are fitted with a rapid sampling port, gas delivery port, and a gas vent.

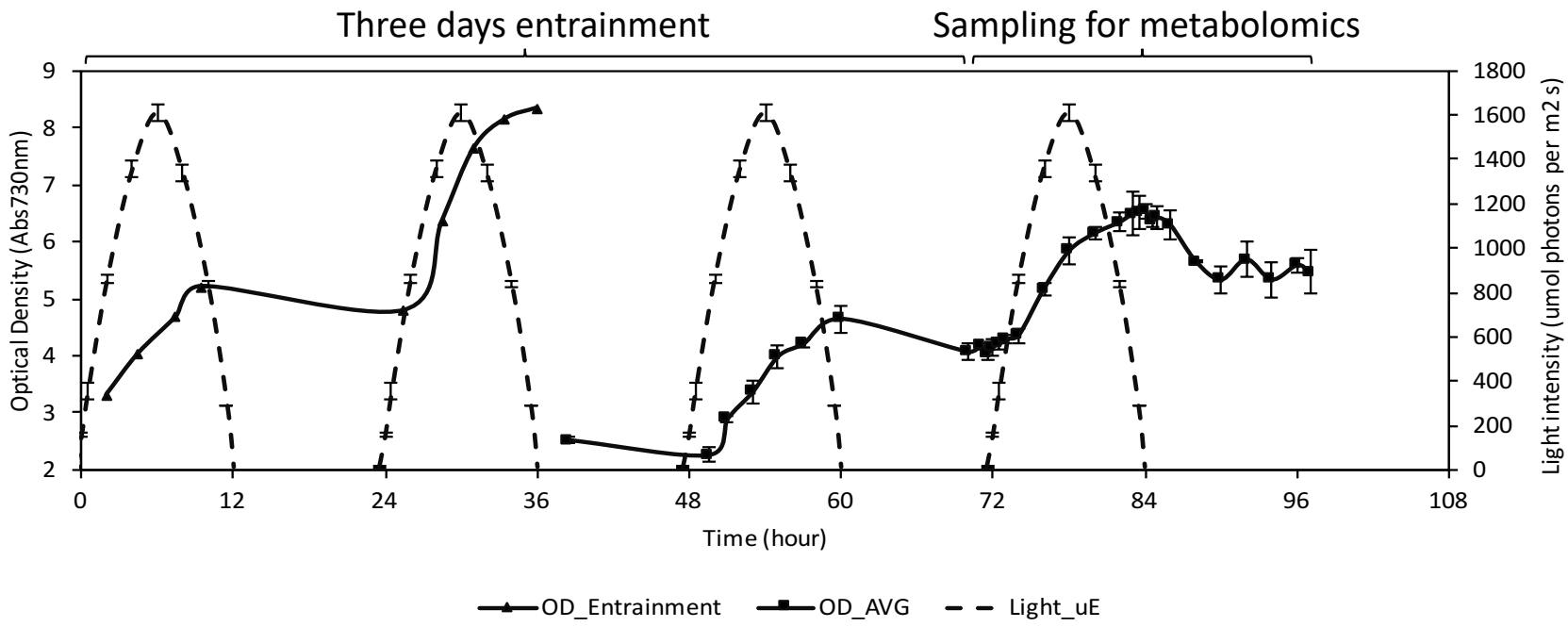

Figure B-5-7. S. 6803 growth under sinLD cycles.

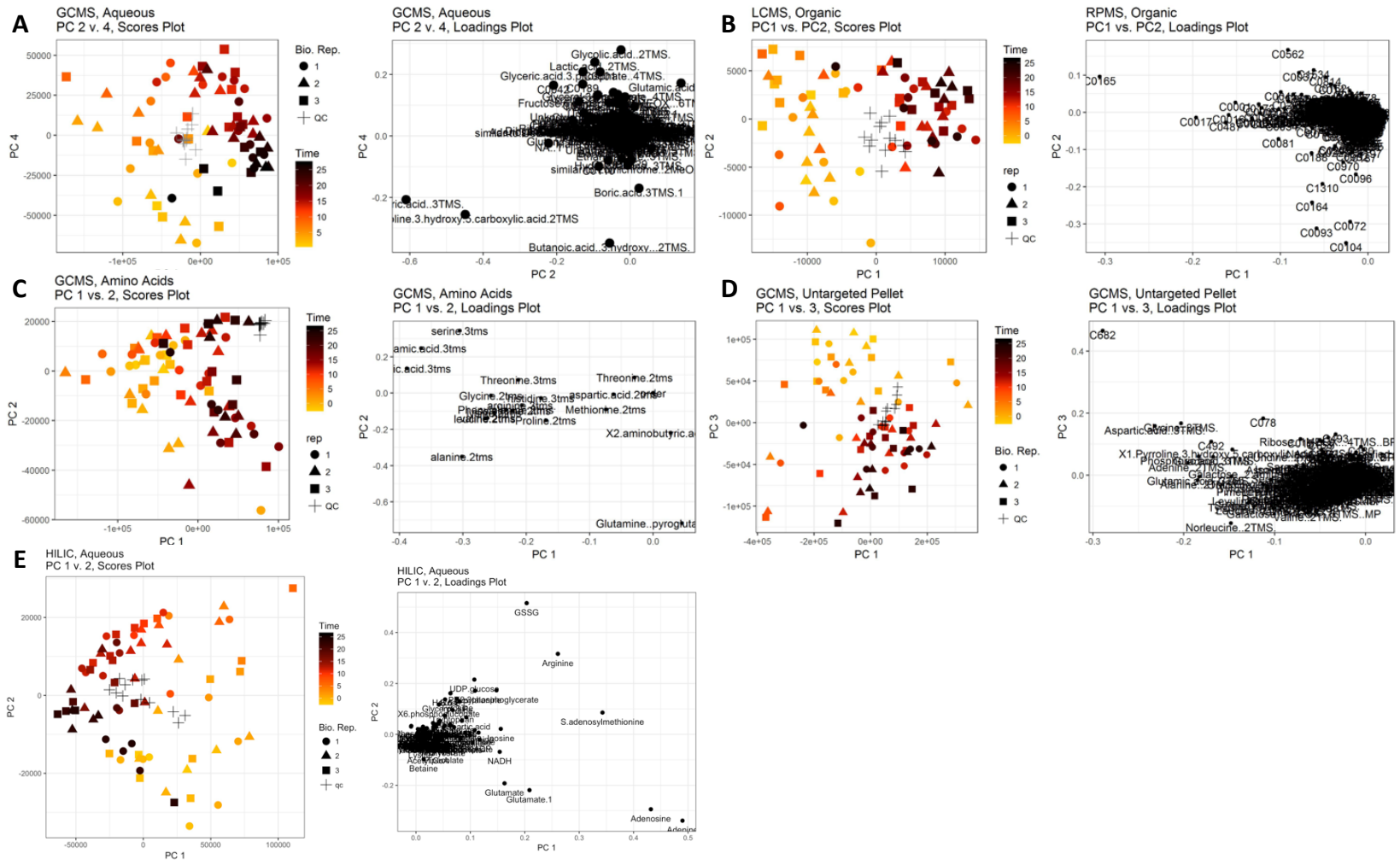

Figure B-5-8. Unbiased PCA by-platform. (A) Non-targeted GC-MS analysis of the aqueous phase. (B) Non-targeted LC-TOF-MS (RP-MS) analysis of the organic phase. (C) Semi-targeted GC-MS analysis of the insoluble pellet amino acid content. (D) Non-targeted GC-MS analysis of the insoluble pellet. (E) Targeted TQS-MS analysis of the aqueous phase. All data pareto-scaled per-cell.

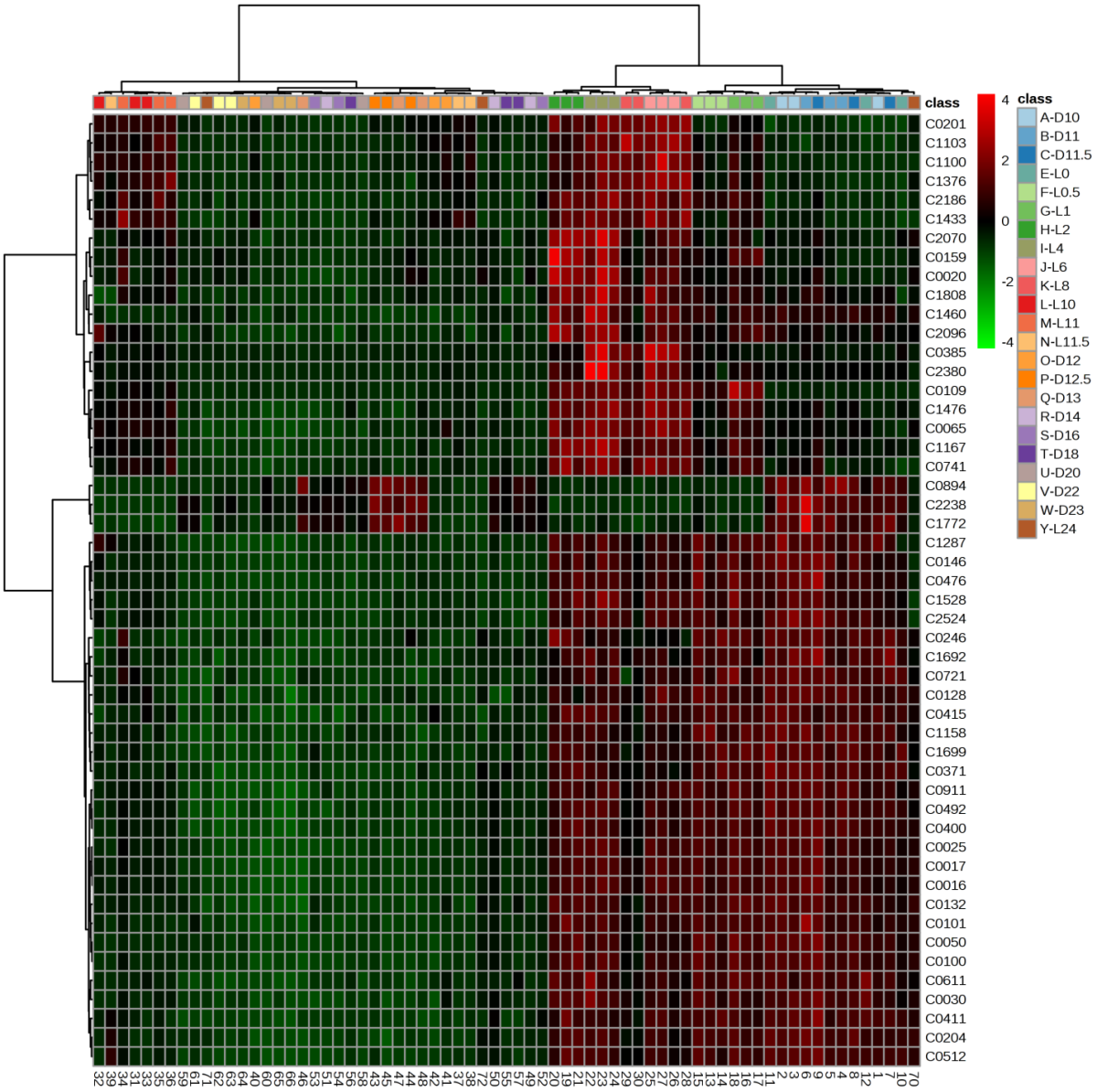

Figure B-5-9. Unbiased HCA and heat-map of changing non-polar metabolites identified from the organic phase by non-targeted RP-MS. Both the samples (columns) and metabolites (rows) were clustered without bias. Metabolite up-regulation is depicted in red, while down-regulation is depicted in green. Class A, B, and C correspond to last hours of darkness before transition to light. Classes E-N correspond to hours during the day. Classes O-Y corresponds to hours during the nig

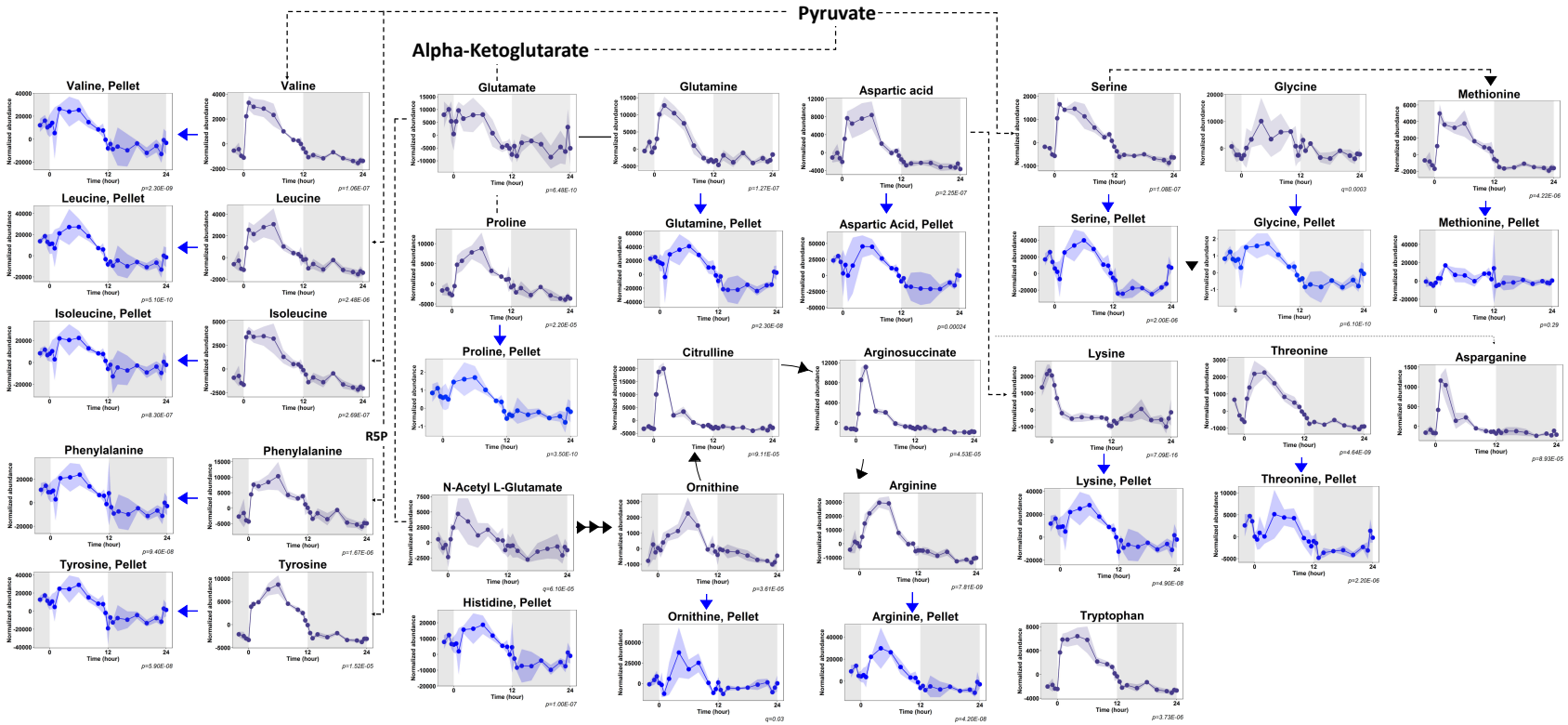

Figure B-5-10. Amino acid biosynthesis pathway(s) with changing metabolites.

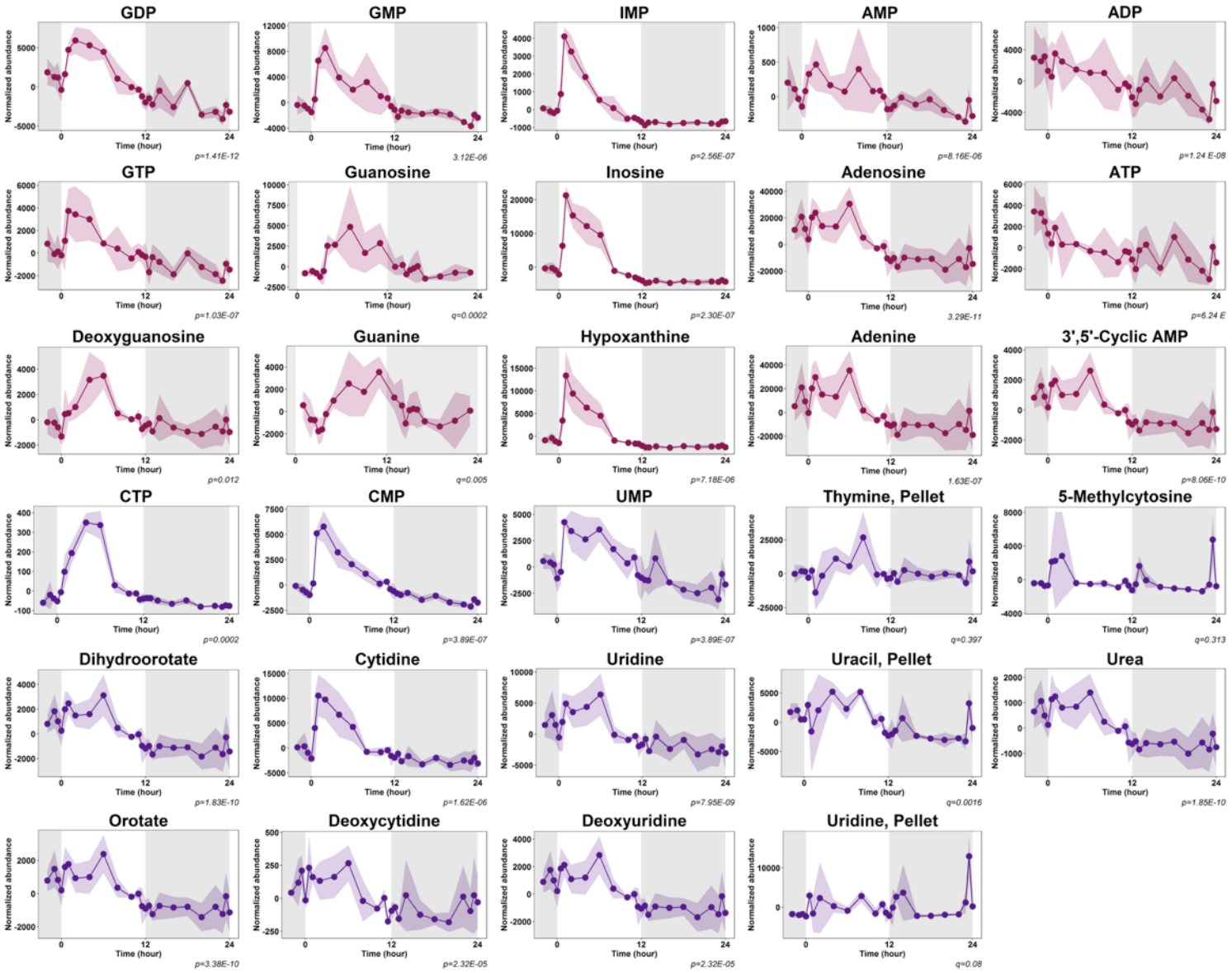

Figure B-5-11. Nucleic acid biosynthesis metabolites.

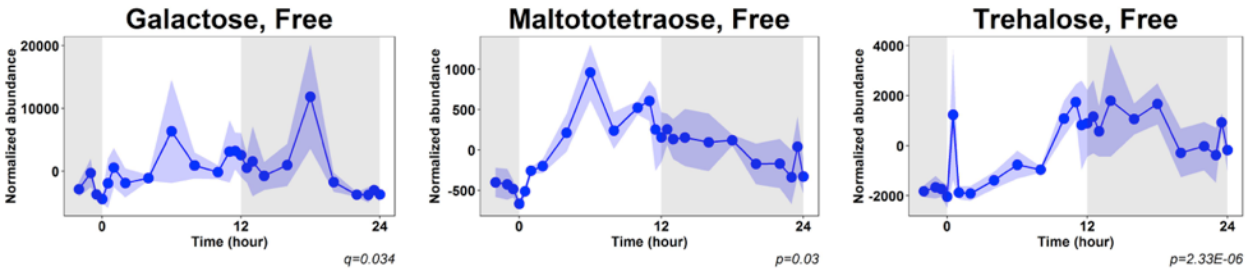

Figure B-5-12. Free carbohydrates.

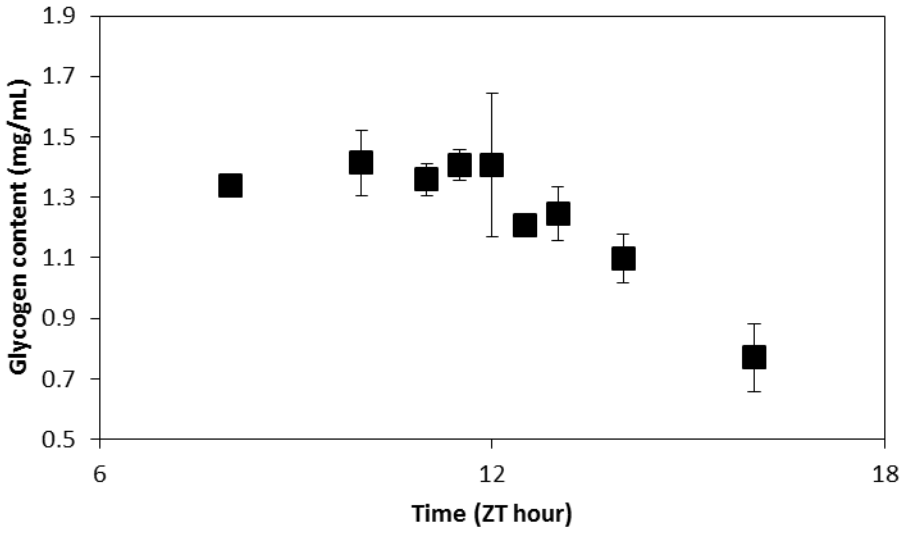

Figure B5-13. Glycogen content as measured by a commercially available fluorescence kit. Error bars represent standard deviation across three biological replicates each with technical triplicates.

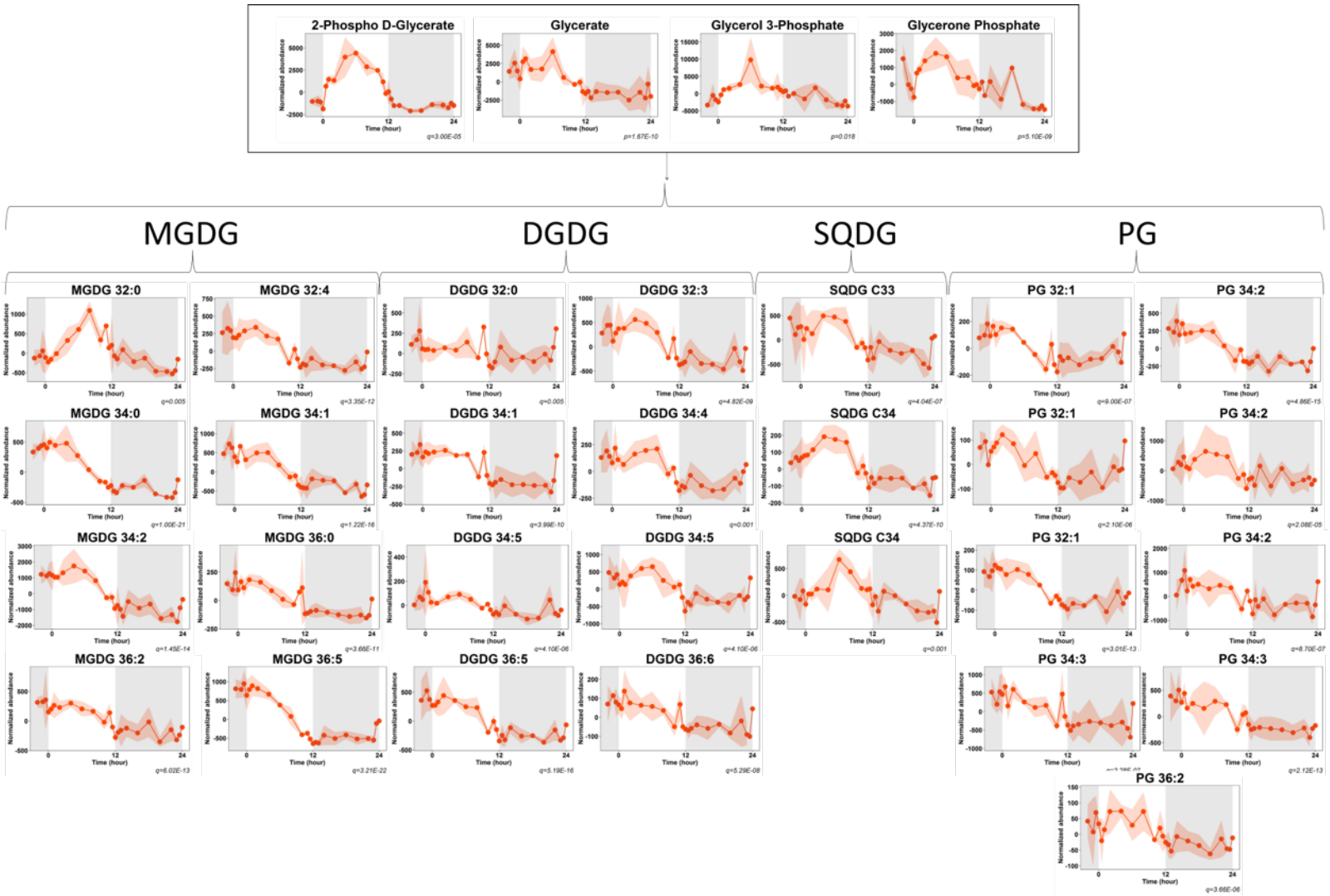

Figure B-5-14. MGDG, DGDG, SQDG and PG dynamics. Metabolite abundance is pareto-scaled per-cell. Lipid biosynthesis precursors shown in black box. Monogalactosyldiacylglycerol (MGDG), Digalactosyldiacylglycerol (DGDG), Sulfoquinovosyldiacylglycerol (SQDG), and Phosphatidylglycerol (PG) lipid species. Shaded regions represent standard deviation across three biological replicates.

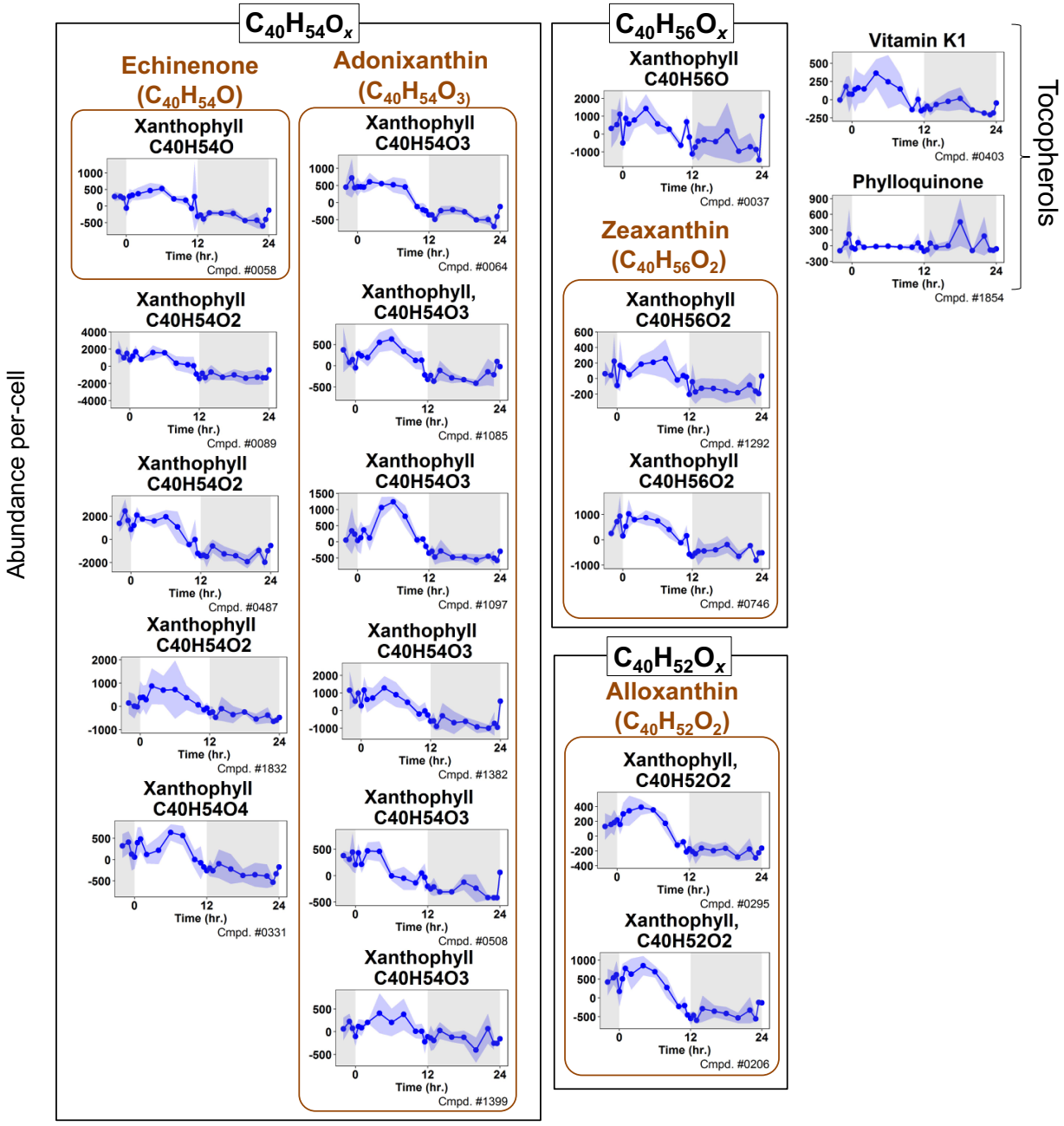

Figure B-5-15. Pigment dynamics. Structural information is difficult to assess via mass spectrometry. Compounds annotated as carotenoids, xanthophylls, and tocopherols are presented with the class and molecular formulas presented. Key photosynthetic pigments and metabolites with matching molecular formulas are identified with brown boxes. All compounds were identified by non-targeted LC-TOF-MS and the corresponding compound number is provided for each metabolite in the bottom right corner of the plot.

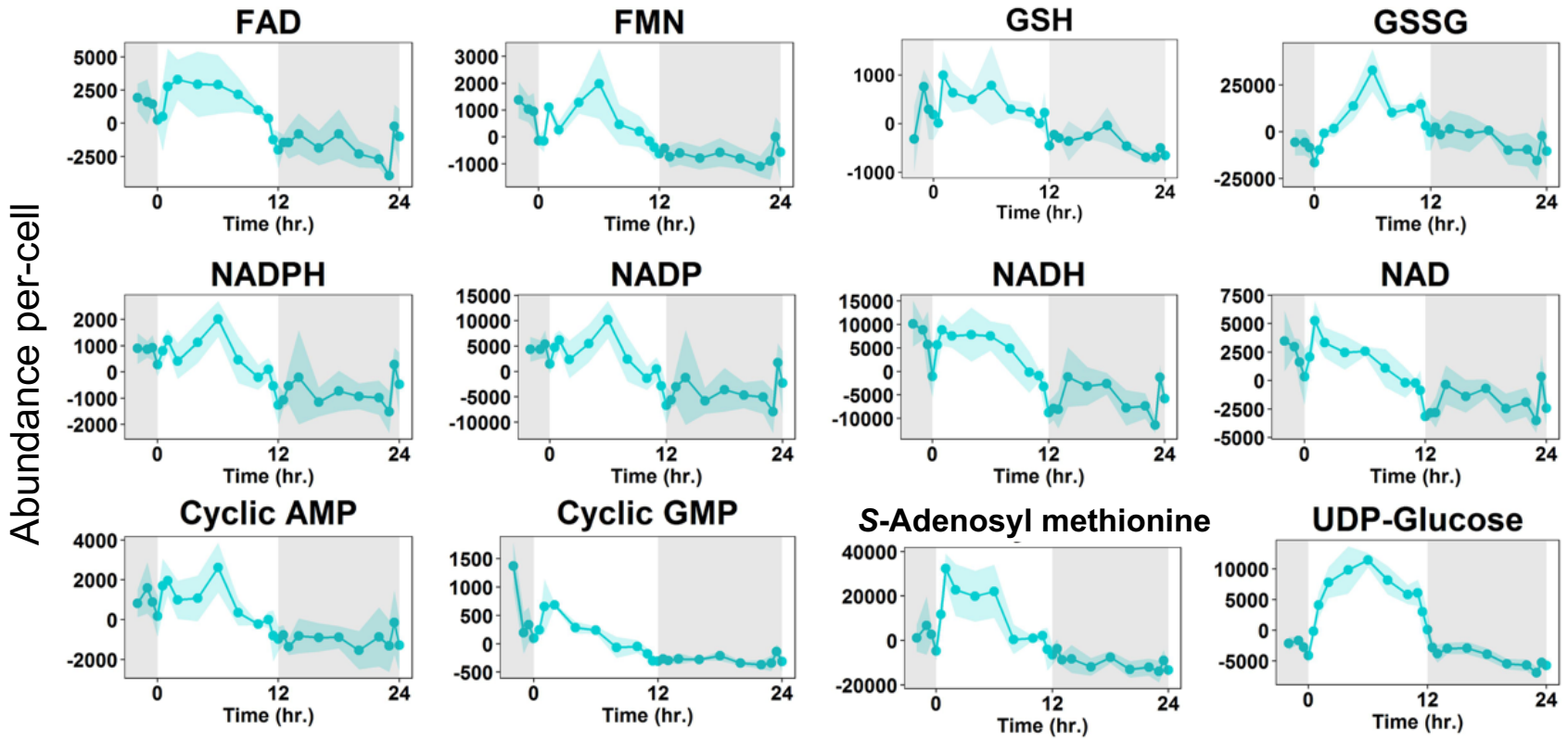

Figure B-5-16. Cofactor and cosubstrate dynamics.

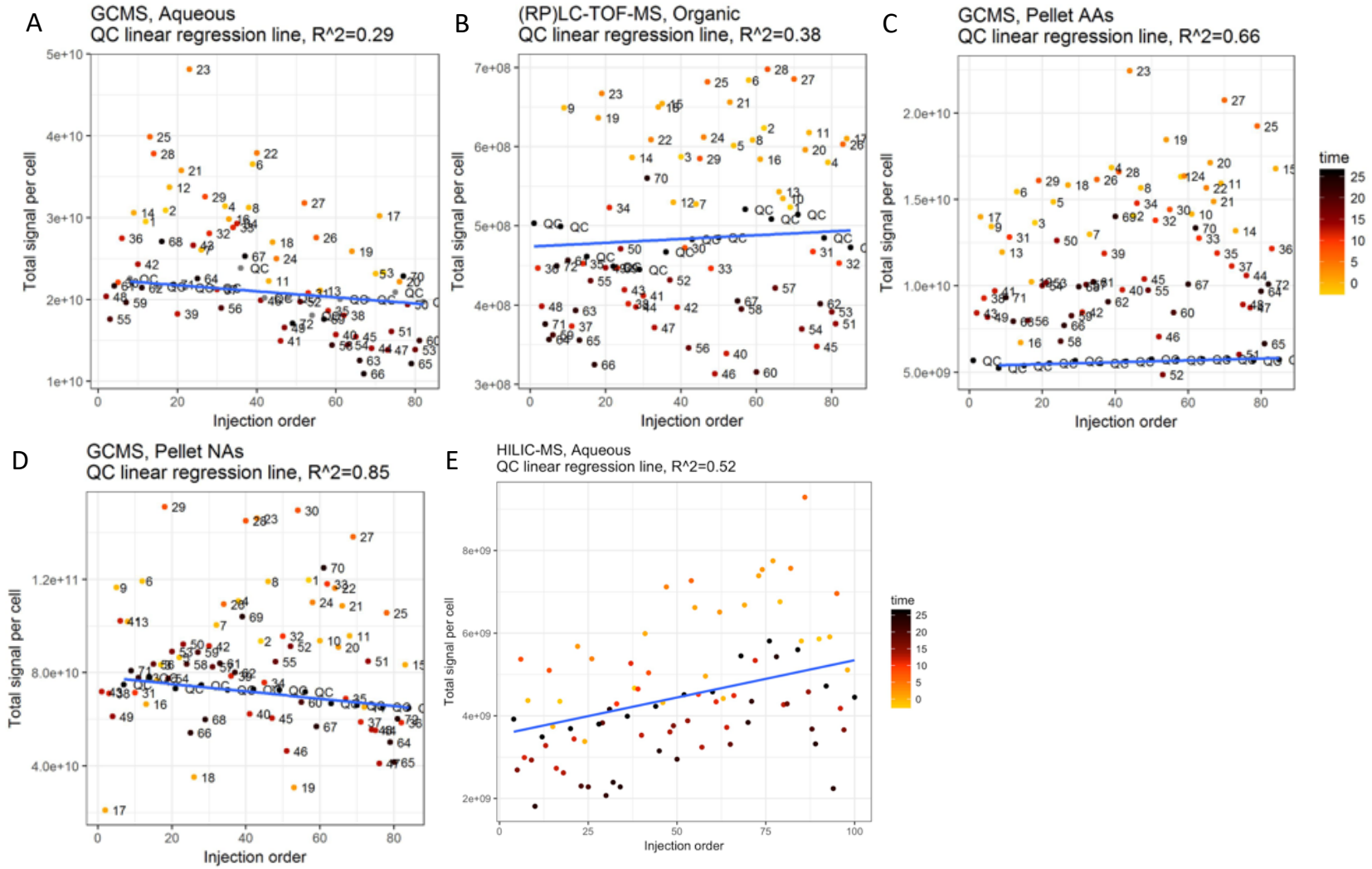

Figure B-5-17. Total signal per-cell versus randomized injection order. (A) Non-targeted GC-MS analysis of the aqueous phase. (B) Non-targeted LC-TOF-MS (RP-MS) analysis of the organic phase. (C) Semi-targeted GC-MS analysis of the insoluble pellet amino acid content. (D) Non-targeted GC-MS analysis of the insoluble pellet. (E) Targeted TQS-MS analysis of the aqueous phase. Linear regression lines and corresponding R^2 are displayed for QC injections.

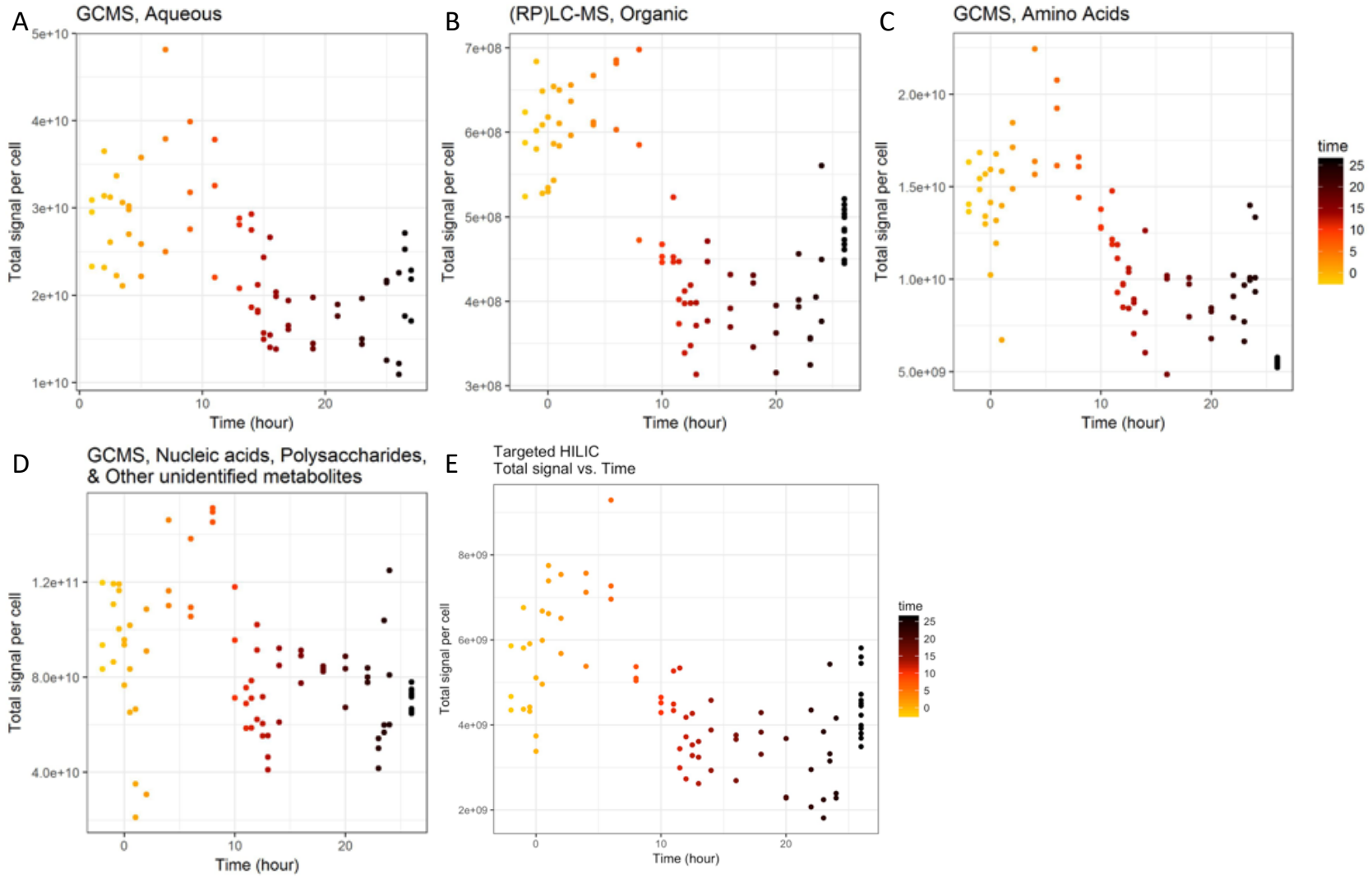

Figure B-5-18. Total signal per-cell versus time in day/night cycle (ZT). (A) Non-targeted GC-MS analysis of the aqueous phase. (B) Non-targeted LC-TOF-MS (RP-MS) analysis of the organic phase. (C) Semi-targeted GC-MS analysis of the insoluble pellet amino acid content. (D) Non-targeted GC-MS analysis of the insoluble pellet. (E) Targeted TQS-MS analysis of the aqueous phase. QC injections are displayed arbitrarily at t=26.

SUPPLEMENTAL TABLES

Table B-5. *S.* 6803 cultivation measurements under sinLD cycles. Consecutive time (Consec. Time) is shown as hours post-inoculation. Zeitgeber time (ZT) is shown as hour post-light in the diurnal sinLD cycle. Sampling volume is the volume of culture removed at that time-point. Only one reactor (R1) was utilized for the first two days of entrainment (Day 1, Day 2). Three reactors (R1-R3) were utilized for the third day of entrainment (Day 3) and the day of sampling (Day4). Specific growth rate was calculated for each day of growth. The time frame utilized for calculation and μ_{max} is presented for each day using the average natural log of optical density for that day. For Days 1-2, only one reactor's optical density was utilized.

	Consec Time	ZT	Sampp Vol (mL)	OD			pH			Light (uE)	Temp (°C)	Avg Ln(OD)	Spec. Growth Rate (1/hr)		Cell Count		
				R1	R2	R3	R1	R2	R3				Time frame	μ_{max}	Avg. (cell/uL)	Stdev.	μ_{max}
Day 0	-12.5	N/A	INOCULATE, OD=3.53										N/A	N/A	N/A		
	-12	N/A	4	3.62	N/A	N/A	7.5	N/A	N/A	N/A	N/A	N/A	N/A	N/A			
Day 1	2	2	4	3.31	N/A	N/A	10.3	N/A	N/A	895	25	1.20	2 hrs. - 9.5 hrs.	0.06			
	4.5	4.5	5	4.03	N/A	N/A	7.6	N/A	N/A	1562	28	1.39					
	7.5	7.5	4	4.68	N/A	N/A	10.0	N/A	N/A	1597	30	1.54					
	9.5	9.5	4	5.21	N/A	N/A	10.1	N/A	N/A	1090	30	1.65					
Day 2	25.5	1.5	4	4.79	N/A	N/A	7.9	N/A	N/A	725	26	1.57	1.5 hrs. - 9.5 hrs.	0.07			
	28.5	4.5	4	6.35	N/A	N/A	8.0	N/A	N/A	1504	29	1.85					
	31	7	4	7.63	N/A	N/A	8.2	N/A	N/A	1578	30	2.03					
	33.5	9.5	4	8.15	N/A	N/A	8.1	N/A	N/A	1052	29	2.10					
	36	12	4	8.34	N/A	N/A	7.6	N/A	N/A	133	28	2.12					
	38	14	INOCULATE, OD=2.52														
Day 3	1.5	1.5	1.5	2.12	2.37	2.29	6.7	6.8	6.7	735	25	0.81	1.5 hrs. - 9 hrs.	0.08			
	3	3	1.5	2.91	2.84	2.93	6.8	6.8	6.8	1204	27	1.06					
	5	5	1.5	3.19	3.34	3.58	6.9	7.0	7.0	1615	29	1.21					
	7	7	1.5	3.77	4.08	4.12	7.7	7.6	7.7	1615	30	1.38					
	9	9	1.5	4.13	4.27	4.23	9.0	9.0	9.2	1200	30	1.44					

	12	12	1.5	4.38	4.8	4.76	9.5	9.6	9.6	0	29	1.54					
Day 4	22	-2	15	3.97	4	4.24	6.6	6.6	6.7	0	24	1.40	2 hrs. - 8 hrs.	0.06± 0.01	1.24E+05	7.01E+03	0.07± 0.02
	23	-1	15	4.14	4.14	4.23	6.5	6.6	6.7	0	25	1.43			1.19E+05	9.09E+03	
	23.5	-0.5	15	3.92	4.09	4.12	6.6	6.6	6.6	2	25	1.40			1.23E+05	1.10E+04	
	24	0	15	3.99	4.27	4.2	6.6	6.7	6.7	139	25	1.42			1.27E+05	1.39E+04	
	24.5	0.5	15	4.26	4.1	4.26	6.6	6.7	6.7	202	25	1.44			1.24E+05	4.18E+03	
	25	1	15	4.24	4.28	4.37	6.8	6.8	6.8	409	25	1.46			1.24E+05	3.97E+03	
	26	2	15	4.22	4.36	4.49	6.9	6.9	6.9	796	25	1.47			1.25E+05	5.18E+03	
	28	4	15	5.06	5.28	5.18	7.2	7.2	7.2	1413	29	1.64			1.27E+05	4.24E+03	
	30	6	15	5.79	6.08	5.65	7.7	7.7	7.6	1732	30	1.76			1.27E+05	1.22E+04	
	32	8	15	6.2	6.22	6.01	7.8	7.8	7.7	1509	30	1.82			1.48E+05	9.94E+03	
	34	10	15	6.24	6.53	6.26	7.6	7.7	7.7	961	30	1.85			1.80E+05	4.14E+03	
	35	11	15	6.446	6.886	6.138	7.5	7.5	7.5	583	30	1.87			1.83E+05	9.57E+03	
	35.5	11.5	15	6.3	6.84	6.41	7.3	7.3	7.4	420	29	1.87			2.15E+05	9.32E+03	
	36	12	15	6.45	6.69	6.47	7.3	7.3	7.3	0	29	1.88			2.31E+05	1.96E+04	
	36.5	12.5	15	6.29	6.48	6.29	7.0	7.1	7.2	0	28	1.85			2.26E+05	1.06E+04	
	37	13	15	6.24	6.39	6.64	6.8	6.9	6.9	0	27	1.86			2.41E+05	2.64E+04	
	38	14	15	6.15	6.58	6.15	7.0	7.0	7.0	0	27	1.84			2.91E+05	3.57E+04	
	40	16	15	5.67	5.64	5.66	7.8	7.7	7.6	0	26	1.73			2.52E+05	3.04E+04	
	42	18	15	5.48	5.45	5.07	7.7	7.6	7.6	0	26	1.67			2.79E+05	7.35E+04	
	44	20	15	5.56	6.03	5.48	7.0	7.0	7.0	0	26	1.74			2.45E+05	1.24E+04	
	46	22	15	5.49	5.52	4.99	7.7	7.7	7.6	0	25	1.67			2.35E+05	1.53E+04	
	48	24	15	5.57	5.72	5.47	7.0	7.0	7.0	0	25	1.72			2.73E+05	2.38E+04	
	49	25	15	5.79	5.57	5.04	7.8	7.7	7.7	140	25	1.70			2.09E+05	2.09E+04	

Table B-6. Summary of regression analysis multiplicity of testing measures of significance and cut-off value determination.

MS platform	# tests	Null raw *p*-value distribution	Null BH adj. *p*-value distribution	Null *q*-value distribution	Fit for λ	Est. π_0	Measure, cut-off
LINEAR REGRESSION							
GC-MS, aqueous	1,329	Uniform[0,1]	Uniform[0,1]	Uniform[0,1]	Good	0.043	*q*-value, 0.01
HILIC-MS, aqueous	103	N/A	N/A	N/A	Poor	0.119	*p*-value, 0.01
RP-MS, organic	2,683	Uniform[0,1]	Uniform[0,1]	Uniform[0,1]	Good	0.113	*q*-value, 0.001
GC-MS, Pellet Aas	20	N/A	N/A	N/A	N/A	N/A	*p*-value, 0.05
GC-MS, Pellet NAs, Sugars, etc.	862	Uniform[0,1]	Uniform[0,1]	Uniform[0,1]	Good	0.424	*q*-value, 0.01
QUADRATIC REGRESSION							
GC-MS, aqueous	1,329	Uniform[0,1]	Non-uniform, shifted right	Uniform[0,1]	Poor	0.609	*q*-value, 0.05
HILIC-MS, aqueous	103	N/A	N/A	N/A	Poor	0.298	*p*-value, 0.01
RP-MS, organic	2,683	Uniform[0,1]	Uniform[0,1]	Uniform[0,1]	Good	0.262	*q*-value 0.001,
GC-MS, Pellet Aas	20	N/A	N/A	N/A	N/A	N/A	raw *p*-value, 0.05
GC-MS, Pellet NAs, Sugars, etc.	862	Uniform[0,1]	Non-uniform, shifted right	Uniform[0,1]	Poor	0.648	*q*-value, 0.01

Table B-7. Summary of significantly changing annotated metabolites by extraction phase and platform. The total number of metabolites identified as significantly changing was determined as the number of annotated metabolites which had *p* or *q* values less than the specific cut-off for either linear or quadratic regression analysis (Table B-6); compounds which had significant values for both were only counted once ("non-redundant"). The percentage of total was calculated as the percentage of total significant metabolites of all annotated metabolites for that phase.

Extraction Phase	**MS Platform**	**Total # of annotated metabolites**	**Identified as significantly changing by:**		**Total metabolites significant by linear or quadratic regression:**	
			Linear regression analysis	*Quadratic regression analysis*	*Total # (linear or quadratic, non-redundant)*	*% of total*
Aqueous	Non-targeted GC-MS	77	61	25	69	**91%**
	SRM-Targeted HILIC-TQS-MS	96	84	33	89	
Organic	Non-targeted LC(RP)-TOF-MS	48	43	11	45	**94%**
Insoluble Pellet	Semi-targeted GC-MS (Amino Acids)	20	17	1	17	**40%**
	Non-targeted GC-MS	70	14	5	19	

SI MATERIALS AND METHODS

Cyanobacteria cultivation

The inoculum train was developed to serve two key functions: (1) rapidly generate enough *S.* 6803 biomass to inoculate 3 x 1.5 L FPRs, and (2) provide three days of sinLD cycle entrainment prior to sampling for metabolomics. It is well known that *S.* 6803 has relatively poor genetic stability due to the presence of transposons, homologous recombination, and other features [135]. Therefore, a key consideration was to perform the metabolomics experiment as quickly after re-suspension from a freezer stock to minimize any potential genetic drift.

All inoculum train and FPR growth was done in un-buffered BG-11 media. Ultra-pure 18 Ω water was used for all media components. BG-11 media was prepared fresh for all steps: here,

"fresh BG-11" means <1 week old, while "ultra-fresh BG-11" means <36 hours old. Media prepared fresh the day before FPR inoculation resulted in more repeatable growth and higher growth rates in the LED PBR. This effect was not as evident in shake-flasks, although *S.* 6803 growth in media >1 month old is unreliable and often the cultures flocculate within 1 day of growth in old media (data not shown).

The LED PBR was initialized 2 days prior to inoculation to achieve consistent gas flow rates and uniform water bath temperature. The LED FPR is initialized by setting input air flow rate to 600 ccm, CO_2 regulator open, input CO_2 flow rate to 30 ccm, vent set-point to 100 psi, reactor flow rates to 200 ccm, water bath filled with RO to fill-line, Huber Ministat set-point to 25°C, and LED arrays to sinLD cycles, and circulation pump ON.

Freezer stock wild-type *S.* 6803 is inoculated in 25 mL of fresh BG-11 in quadruplicate 150 mL Erlenmeyer flasks and grown at 150 μE, 30°C, 225 rpm. Cultures are fed to 1M $NaHCO_3$ and pH adjusted to 8.0-8.2 with 6 N HCl twice a day. Only one culture is measured by pH probe directly; the acid addition of the measured flask is applied to the other three. This protocol modification was required to achieve pure (un-contaminated) *S.* 6803 seed cultures. Feedings and pH adjustments are spaced ~8-12 hours apart. As soon as cultures reach the *onset* of stationary phase, they are spun-down by centrifugation at 3,000*g* for 5 minutes.[10] The three cultures which did not experience the pH probe directly are resuspended in ~2 mL fresh BG-11, combined, aliquoted into 4 x 150 mL fresh BG-11 in 500 mL Erlenmeyer flasks. All centrifugation steps are done with sterile single-use 50 mL centrifuge tubes and all resuspension

[10] Cultivation of *S.* 6803 >2 days into stationary phase resulted in the presence of white debris visible in the cell pellet post-centrifugation step. Restricting growth only to the linear phase was required to achieve healthy and sterile scale-up cultures.

steps are done with filter tips.[11] 150 mL cultures are grown until the onset of stationary phase, then scaled-up. As with 25 mL cultures, the three cultures which did not experience the pH probe are centrifuged, combined, resuspended, and used to inoculate 4 x 300 mL cultures in 1 L Erlenmeyer flasks. When these cultures reach the onset of stationary phase, the cultures which did not experience the pH probe are centrifuged, combined, and used to inoculate 1 x 1,700 mL FPR with ultra-fresh BG-11. After inoculation, the FPR is incubated in the dark for ~10-12 hours prior to the initialization of the first sinLD cycle. 4.5% CO_2 is provided at 200 mL/min and the water bath is set to 25°C for the entire FPR experiment. OD, pH, and temperature of the LED FPR are measured ~5-6 times a day for the first two days of sinLD cycle growth. During the first hour of dark following the second sinLD cycle, the cells were aliquoted into sterile 50 mL centrifuge tubes, centrifuged at 3,000*g* for 5 min. at 25°C, and resuspended in 1-2 mL ultra-fresh BG-11. All resuspended cells were combined, mixed, and aliquoted into 3 x 1.7 L ultra-fresh BG-11 in freshly-autoclaved FPRs. FPRs were entrained in sinLD cycles for a final day prior to sampling for a total of three days of sinLD cycle entrainment. Three days of entrainment was selected as it is within the standard in the field for LD cycle 'omics experiments: McEwen *et al.* [91], Mettler *et al.* [136], Beck *et al.* [85], Willamme *et al.* [84], and Diamond *et al.* [83] all used 3-days of diurnal LD cycle entrainment prior to sampling. Intermittent centrifugation and resuspension in fresh media ensured that cell growth was still in linear phase despite cultivation as a batch culture.

Checks for contamination were conducted before inoculation, during inoculation, and at the completion of the rapid sampling experiment to ensure a sterile *S.* 6803 was present. Checks

[11] Use of sterile centrifuge tubes (as opposed to 1 L re-useable Nalgene bottles) and filter-tips were required to achieve sterile scale-up cultures.

were conducted as follows. Cells were sampled from FPRs (or shake flasks) and plated on LB plates in 1, 1:10, and 1:100 dilutions and incubated at 37°C in darkness. Plates were checked for contamination after 24 and 48 hours of incubation. Cells were also plated on BG-11+ 10 mM glucose and incubated at 30°C and 50 μE. Plates were checked for contamination after 48 hours and 1 week of incubation. Flow cytometry analysis of cells further confirmed that cultures which passed the plate checks were uniform in cell populations.

Rapid quenching of metabolites

A 50 mL centrifuge tube for each sample point was labeled and pre-weighted. 10 mL of sterile 1X PBS was aliquoted into each tube and placed at -50°C. 30 minutes prior to the sampling point, the PBS tubes were removed from the -50°C and placed on ice. 10 minutes prior to the sampling point, the PBS tubes were vortexed ~30 seconds or until homogeneous slurry and placed on ice. Nine labeled microcentrifuge tubes and three 15 mL centrifuge tubes were also placed on ice. At the exact sampling time, 15 mL was pulled from the reactor using a fresh and sterile 20 mL syringe: 10 mL was ejected into the ice-cold PBS slurry; 5 mL was ejected into the 15 mL centrifuge tube. The PBS-cell slurry was immediately capped and vortexed at medium speed for ~3 seconds. A fresh and sterile syringe was replaced on the reactor sampling port. This sampling protocol was repeated for all three reactors. PBS-quenched cells were spun at 3,000*g* for 5 min. in a centrifuge at -4°C. From sampling to centrifugation, it took >2 minutes. While the centrifugation step run, 500 uL was aliquoted from the 15 mL tube to 3 x 1.5 mL microcentrifuge tubes for each reactor's sample, and tubes were quenched in liquid nitrogen. After the centrifugation step completed, the supernatant was decanted, tube re-capped, and cells quenched in liquid nitrogen. All liquid nitrogen quenched samples were immediately stored at -50°C for 1 day and then transferred to -80°C until further analysis (<3 weeks). 100 uL was taken

from the remaining cell suspension and placed in a cuvette to measure optical density and cell count. pH was recorded in the remaining cell suspension prior to flash-freezing in liquid nitrogen.

Cell fixation and growth measurements

Optical density was measured in a 1:10 dilution on a NanoDrop Spectrophotometer as absorbance at 730 nm in a 1mm path length cuvette. Optical density of cyanobacterial cultures is measures at absorbance at 730 nm with the intent to measure total biomass accumulation with the exception of photosynthetic pigments and common molecular entities (*e.g.* protein at 280 nm [223], nucleic acids at 260 nm [224], chlorophyll *a* at 465 nm and 666 nm [186]). However, the summed contributions of cellular components to the absorbance measurement at 730 nm are to our knowledge unknown and therefor subject to misrepresentation of balanced accumulation of biomass. In this study, we assume that optical density provides a measure of total biomass accumulation and therefore an approximate measure for cell volume. After measurement, this cell suspension was fixed in glutaraldehyde in PBS at a final concentration of 0.25% and incubated at room temperature for ~15 minutes prior to flash-freezing in liquid nitrogen and storage at -50°C for <3 weeks. Fixed cells were thawed and 200 uL was transferred to a 5 mL flow cytometry tube. 20 uL of CountBright™ beads was added to the cell sample. Cell counts were performed on a CyAn1 flow cytometer. Cells per uL were calculated and multiplied by a dilution factor of 10. Pellets were lyophilized for 48 hours and the pellet dry cell weight (DCW) was measured.

Biphasic extraction of metabolites

Metabolites were extracted from dried pellets via a modified MTBE:MeOH:H_2O biphasic extraction, based on the protocol developed by Salem *et al*. [138]. Briefly, 6 mL of 75% methanol (MeOH) was added to pellets, vortexed, and transferred to glass vials. 9 mL of 100% methyl *tert*-butyl ether (MTBE) was added, vortexed for 30 seconds, placed on automatic shaker for 1.5 hours at 4 ºC, and sonicated for 15 minutes. 3.75 mL of water was added, each extraction was vortexed by hand for 1 minute, and centrifuged for 10 minutes at 3,270*g* at 4ºC. A biphasic solution with a pellet formed: the top, green MTBE layer and the bottom, clear MeOH:H_2O layer were separated into separate tubes and dried under $N_{2,gas}$ overnight. The pellet was stored at -80 ºC. After drying, the MTBE layer was resuspended in 100 uL 1:1 toluene:MeOH, transferred to a LC-MS vial insert, and stored at -80C for <1 month prior to MS analysis. The MeOH:H_2O layer was resuspended in 1 mL of 1:1 H_2O:MeOH, transferred to a 1.7 mL centrifuge tube and spun at 15,000*g* for 2 minutes at 4 ºC. The supernatant was split into two 465 µL aliquots—one for GCMS and one for LC(HILIC)MS—in glass vials and dried under $N_{2,gas.}$ The protocol outlined above is suitable for filter-quenched cyanobacteria samples and centrifuged cell pellets.

The polar methanol/water fraction resulting from the biphasic extraction was processed for analysis by hydrophilic interaction liquid chromatography (HILIC) LC-MS. Dried samples were resuspended in 100 µL 1:1 H_2O:MeOH and 10 µL were aliquoted into a pooled QC sample. Samples were stored at -80 ºC until analysis. The pooled QC sample was mixed and aliquoted into twelve vials. A QC injection was run every tenth injection. The dried polar fraction for analysis by GC-MS was stored at -80 ºC until derivatization, immediately prior to MS analysis. Samples were derivatized in 30 uL methoxyamine HCl and 30 uL MSTFA, as specified in the following section. Ten microliters were removed from each sample to create a pooled QC sample, mixed, and aliquoted into thirteen vials. A QC sample was run after every sixth

injection. The non-polar MTBE phase was processed for non-targeted LC-MS analysis. Twenty microliters from each sample were pooled, mixed, and aliquoted into thirteen pooled QC samples. QC injections were placed after every sixth injection.

Derivatization of metabolites for GC-MS detection

All samples analyzed by GC-MS were derivatized using the following protocol. Briefly, dried-down samples were resuspended in 30 µL or 50 µL (indicated in sample preparation section) of warm methoxyamine HCl, incubated at 60 ºC for 45 minutes, vortexed briefly, and sonicated for 10 minutes at maximum intensity. Samples were again incubated at 60 ºC for 45 minutes, 30 µL or 50 µL of derivatization reagent was added (indicated in sample preparation section), vortexed briefly, and incubated at 60 ºC for 40 minutes. Samples were centrifuged briefly to collect precipitates and the supernatant was transferred to a GC-MS vial insert for analysis.

Hydrolysis of pellet-bound content to amino acids, nucleic acids, and polysaccharides

An acid hydrolysis protocol was developed for identification of amino acids, nucleosides, and carbohydrates bound in insoluble pellet of protein, DNA/RNA, and polysaccharides, respectively. Pellets remaining from the biphasic extraction were removed from storage at -80 degrees C and residual solvent was evaporated under nitrogen gas. Pellets were re-suspended in 3 mL of 6 M hydrochloric acid (HCl) using vigorous vortexing and pipette re-suspension. The resulting suspension was a bright teal. The suspension was transferred equally two three separate glass vials for hydrolysis of the separate polymer constituents.

Hydrolysis of proteins to amino acids was completed with a hydrochloric acid (HCl) hydrolysis, based on previously published protocols [refs]. Briefly, vials were incubated at 110

degrees C for with a loose cap seal. After 4 hours, the acid in each vial was entirely evaporated; 1 mL of 6 M HCl was added to each vial, vortexed, sealed tightly, and returned to 110 degrees C. After a total of 24 hours, vials were removed, and remaining acid was evaporated under nitrogen gas. Samples were resuspended in 150 µL of 1:1 MeOH:H2O, 20 uL was removed to create a pooled QC sample, the pooled QC was aliquoted into fourteen vials, and the solvent was evaporated under nitrogen gas. Amino acid samples were derivatized in 30 uL of methoxyamine HCl in pyridine and 30 uL of MTBSTFSA as described in *Section 3.6.* The peak integration of each amino acid in each sample was manually checked and curated in the software Chromeleon™. Spectral abundance for each amino acid was exported and normalized to the cell number in each sample and multiplied by 1E8. Amino acids that were not detected include the biochemically fragile asparagine and tryptophan. Aspartic acid with 2- and 3-derivitization agent modifications were detected and summed for the total spectral abundance. Threonine with 2- and 3-derivitization agent modifications were detected and summed for the total spectral abundance.

Polysaccharides and nucleic acid polymers were hydrolyzed to nucleosides using a modified protocol from Huang *et al.* (2012) [144]. Briefly, vials were incubated at 130 degrees C for 10 minutes, removed and allowed to cool, and 100 uL was transferred to a glass teardrop vial. The remaining pellet was incubated at 160 degrees C for 40 minutes, removed and allowed to cool, re-suspended in 100 uL LC-MS grade water, vortexed for 15 seconds, centrifuged at 1,500*g* for 2 minutes, and the supernatant was transferred to the glass teardrop vial which contained purines from the 130 degree C incubation. (Note: There were significant insoluble chunks in the resulting suspension from the 160 degree C incubation that were not transferred on to the next processing step). The acid was evaporated under nitrogen gas. Samples were removed after the 10 minute 130ºC incubation to preserve purines (guanine and adenine) from degradation during

the 40 minute 160ºC incubation. Huang *et al.* (2012) demonstrated that less than ~5-10% of the purines remained after a 40 minute incubation at 160ºC.[12] Therefore, calculations of purine content were assumed to be representative of hydrolysis from half of the sample. However, the presence of pyrimidines in the first hydrolysis product (130ºC for 10 minutes) remains a concern. Huang *et al.* (2012) demonstrated that after 5 minutes at 130ºC less than 5% of the pyrimidines were hydrolyzed, but after 10 minutes at 130ºC nearly 100% of the pyrimidines were hydrolyzed. Therefore, it can be assumed that the pyrimidines (cytosine, thymine, and uracil) are over-represented in the data set. To account for these over-representations, corrective actions were taken (see *Section 3.11.2.*). Huang *et al.* (2012) found that utilizing this protocol resulted in minor deamination of pyrimidines and purines (11% and 16%, respectively). Removal of an amino group from cysteine via deamination (addition of water and release of ammonia) would result in an artificially high abundance of uracil at the expense of cysteine. Deamination of guanine results in the formation of xanthine and deamination of adenine results in the formation of hypoxanthine. Huang *et al.* (2012) describe observing deamination into xanthine, hypoxanthine, and other compounds as a result of the 160ºC hydrolysis step.

Mass spectral detection of metabolites

For non-targeted GC-MS experiments, metabolites were detected using a Trace 1310 GC coupled to a Thermo ISQ mass spectrometer. Samples (1 μL) were injected at a 10:1 split ratio to a 30 m TG-5MS column (Thermo Scientific, 0.25 mm i.d., 0.25 μm film thickness) with a 1.2 mL/min helium gas flow rate. GC inlet was held at 285°C. The oven program started at 140°C for 1 min, followed by a ramp of 15°C/min to 330°C, and 5 min hold. Masses between 50-650

[12] Calculated as [(ending concentration)/(starting concentration)]*100% from Huang *et al.* (2012) data presented in Figure 2.

m/z were scanned at 5 scans/sec under electron impact ionization. Transfer line and ion source were held at 300 and 260°C, respectively. Pooled QC samples were injected after every 6 actual samples.

For non-targeted LC-MS experiments, two microliters of extract were injected onto a Waters Acquity UPLC system in discrete, randomized blocks with a pooled QC injection after every 6 sample injections and separated using a Waters Acquity UPLC CSH Phenyl Hexyl column (1.7 µM, 1.0 x 100 mm), using a gradient from solvent A (2mM ammonium hydroxide, 0.1% formic acid) to solvent B (Acetonitrile, 0.1% formic acid). Injections were made in 100% A, held at 100% A for 1 min, ramped to 98% B over 12 minutes, held at 98% B for 3 minutes, and then returned to starting conditions over 0.05 minutes and allowed to re-equilibrate for 3.95 minutes, with a 200 µL/min constant flow rate. The column and samples were held at 65 °C and 6 °C, respectively. The column eluent was infused into a Waters Xevo G2 Q-TOF-MS with an electrospray source in positive mode, scanning 50-2000 m/z at 0.2 seconds per scan, alternating between MS (6 V collision energy) and MSE mode (15-30 V ramp). Calibration was performed using sodium iodide with 1 ppm mass accuracy. The capillary voltage was held at 2200 V, source temp at 150 °C, and nitrogen desolvation temp at 350 °C with a flow rate of 800 L/hr.

Pre-processing of MS data

Raw data was converted to *.CSV with Waters® Databridge. For idMS/MS (RP-LC-MS runs), a file was converted for low-collision, high-collision, and LockSpray for each sample. Peaks were detected within the XCMS workflow using the Centwave algorithm [145]. Centwave detects a feature as a bounded, two-dimension signal (*m/z* and retention time) composed of one-dimensional *m/z* peaks (centroids) [146]. In this way, regions of interest (ROIs) are regions where at least p_{min} centroids with a deviation of less than μ_{min} (ppm) occur. This avoids the

'binning' approach which is subject to skewing and disruption of peak detection. The result is a list of mass traces with mass traces of differing lengths, where each ROI may have 0, 1, or >1 distinct chromatographic peaks. Peaks are then detected using a matched filter approach with user-defined peak width ranges (*minpw, maxpw*), mass deviation (μ in ppm), and signal-to-noise ratio (*snthresh*) with continuous wavelet transform (CWT) method, which accommodates peaks of differing widths. Features were detected via XCMS as follows: group peaks with wide bandwidth (3), correct retention time, re-group with narrow bandwidth (1.5), and fill peaks. The output is an XCMS object (*xset*) for each feature (*m/z* and RT) for each sample for both low- and high-collision energy. Class information (biological replicates) were added to the phenotype slot and XCMS was run again using *minfrac=1,* which specifies that a called peak must be present in all three biological replicates of at least one time-point.[13] This is a different approach than the standard *minfrac=0.25,* or a called peak must be present in 25% of all samples. I adopted the class-based approach because we are specifically interested in peaks which are only present at select time-points, perhaps less than 25% of the samples. The requirement than the peak must be present in all three biological replicates should remove noise.

Each feature's spectral abundance was normalized to the cell number in the sample at that time point. The next step, feature clustering, did not perform properly with very small spectral abundances, such as those which have been divided by ~10^9 cell numbers. Therefore, after biomass normalization each feature was multiplied by $1x10^{10}$ to achieve numbers compatible with RAMClustR.

[13] The *pmfxcms* function was run with *regroup=TRUE* after the addition of class information; the resulting *xset* was used for all downstream processing.

Compounds were created by clustering features using RAMClustR [147]. RAMClustR uses a similarity matric which calculates feature correlation across samples and retention time correlation between features. Hierarchical clustering of the similarity matrix was computed via the fastcluter package [148]. The resulting clustered dendrogram is cut using DynamicTreeCut and spectra are created with clusters and features abundances from input data [149]. The abundance for each mass in spectra is a weighted mean of feature intensity. The RAMClustR outputs are compounds (clusters of correlated features) and intensities for each sample; spectral abundance intensities reflect weighted mean of all features within the compound.

For targeted HILIC TQS analysis of metabolites, the integration for spectral abundance of each compound was assessed within the Skyline software package (https://skyline.ms/project/home/begin.view?). Compounds which were not detected and excluded from further analysis, and absent ions were removed from spectral integration. Integration windows for each detected compound and each sample were manually assessed for accuracy within Skyline.

Annotations of compounds for non-targeted MS analyses

For non-targeted GC-MS analyses, putative metabolites identified via RamClustR were annotated using the Golm database via the RamSearch GUI (https://vpr.colostate.edu/pmf/metabolomics-resources/). Head-to-tail spectral matches were manually assessed, and annotations were accepted which provided adequate spectral matching and retention time index matches. The Golm Database removes fragments at masses 73 and 147 due to widespread prevalence. However, actual data does not have these ions removed and therefore a good spectral match can be achieved with the presence of ions at mass 73 and 147 but absence in the reference spectra. Annotations were iteratively assessed using the retention time

index for reference spectra/metabolites in the Golm Database. This was achieved by manually annotating metabolites in the RamSearch software, importing annotations into R (into the RC object), and plotting the retention time index from the Golm database against the retention time measured in my experiment. For amino acid analysis via GC-MS of the pellet fraction, the peak integration was manually assessed for each detected compound in each sample using the Chromeleon software package (https://www.thermofisher.com/order/catalog/product/CHROMELEON7).

APPENDIX C – DATA PROCESSING AND ANALYSIS APPROACHES FOR TIME-COURSE MULTI-PLATFORM MASS SPECTROMETRY METABOLOMICS DATASETS

SUMMARY

The time-course nature of the mass spectrometry metabolomics experiments conducted in this thesis required a slightly different approach to data pre-processing, processing, and normalization than in standard in the field. The goal of this Appendix is to address the data processing and statistical analyses used for the metabolomics study, and provide supplemental analyses to those presented in Chapter 4 and Appendix B. First, data pre-processing steps are presented with corresponding justifications. Next, data analysis strategies are defined. An unsupervised overview of the variation in datasets is presented for each MS platform. Finally, the assessment of measures of significance and justification of cut-off values are discussed for statistical assessment of linear and quadratic regression analysis.

DATA PRE-PROCESSING

Overview

Cultures were subjected to varying light intensities over the course of a day/night cycle, which may cause variation: this is the *induced variation* we seek to understand. All other sources of variation were accounted for using pre-processing steps in order to appropriately assess the inducted variation. The data processing steps are summarized in Table C-8.

Table C-8. Summary of variation and pre-processing step in metabolomics data.

	Source	Type of variation	Processing step
1	Biomass	Collection variation	Per-cell normalization
2	Extraction efficiency	Work-up variation	Not required
3	Instrument sensitivity	Analytical variation	Not required
4	MS platform	Platform variation	Autoscaling / Pareto scaling

5	Day/night time-course	Induced variation	Not applicable

A constant volume of cell culture was sampled across a growth curve such that metabolites were extracted from a different number of cells resulting in variation in the collected biomass; to account for this variation, the spectral abundance of each compound was normalized on a per-cell basis. Samples were injected at different times on a mass spectrometer which is subject to sensitivity drift, resulting in variation in the analytical methods; we find no evidence of significant drift in instrument sensitivity over the course of the short analytical runs and therefore take no action. This assertion was exhaustively explored, as total ion current (TIC) and quantile normalization are routinely applied in the mass spectrometry community to account for variation in instrument sensitivity across batches.

Due to the time-series nature of this study, we observe small variability within groups (biological replicates, time-points) and large variability across groups. This is clearly visible by considering the distribution of select metabolites across samples: in some cases, the distribution is bimodal, in some cases shifted right. For datasets such as this, both TIC and quantile normalization approaches are inappropriate [225]. To illustrate this point, we applied quantile normalization to the data-sets; post-normalization, bi-modal metabolite distributions have been artificially converged to normal distributions. In fact, we are very interested in metabolites which exhibit bimodal distributions. Thus, no normalization approach was applied to correct for instrument variability as all tested methods construed the data.

Different mass spectrometry platforms provide different levels of sensitivity and detections, resulting in variation in sensitivity based on the platform used; we apply pareto scaling. Pareto-scaled data enables assessment of variation within a single data set and includes variation with large abundances. All data used in the following analyses were per-cell normalized and pareto-scaled, unless otherwise noted.

Biomass normalization

A constant volume of cells was sampled at each time point. Over the course of the experiment, the cell density was increasing and thus each sample will have a slightly different biomass. Biomass was measured as cell density (absorbance at 730nm), cell number (cell counts via flow cytometry), and pellet dry cell weight (mg biomass post-lyophilization).

Dry cell weight (DCW) was measured as the weight of the flash-frozen and lyophilized pellet prior to extraction. This biomass measure is the only one which does not rely on the assumption that the 10 mL sample volume was consistent across every sample; if the sampling volume was different for one sample was different than 10 mL, it will be reflected in the pellet DCW. There are two critical limitations with using DCW to biomass-normalize. First, the pellets were separated from media post-centrifugation by a simple decanting step; no aspiration equipment was used. This subjects the pellet to residual media and therefore salts. White crystals were observed above the pellet post-lyophilization; this was assumed to be residual salt and the presence/absence was not consistent across samples. Therefore, using DCW could be an inaccurate measure of biomass due to the weight of residual media which was weighed. Second, the inherent technical and user error associated with the DCW measurements is far greater than the other two methods. DCW was recorded by labeling fresh 50 mL tubes, pre-weighing them (~400 mg), collecting and processing cells in the tube, and post-weighting the tube for cell biomass content. All pellets were between 0.08-1.80 mg, ~0.02% of the total weight of the sample (Figure C-5-19). Also worth noting is the lyophilization batches were not randomized, and six samples received an extra 4 hours of lyophilization as compared with the others (time points 1 hr. and 2 hr.). It is unclear if that had any effect on the DCW measures.

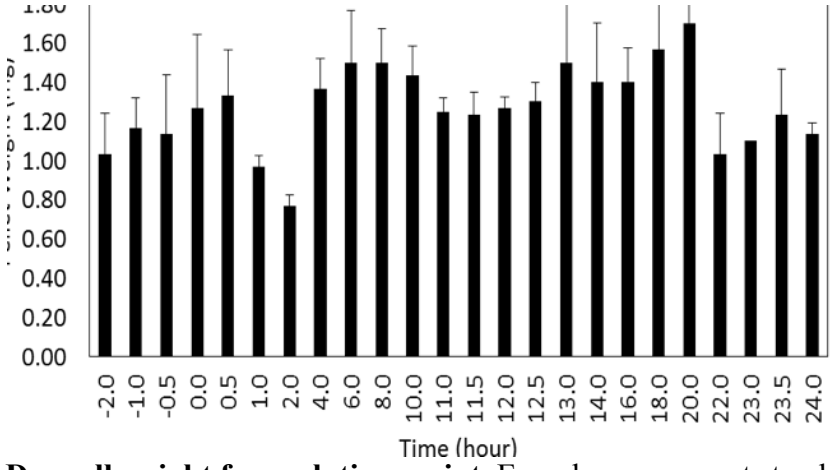

Figure C-5-19. Dry cell weight for each time-point. Error bars represent standard deviation across three biological replicates.

The difference between OD and cell number is biologically interesting (discussed in previous section) and the choice between OD and cell number for biomass normalization will be consequential. The choice is dependent on the biological questions this study was designed to assess:

(1) How does the biomass composition of a cyanobacterium change over the course of a day/night cycle?

(2) How does metabolism of a cyanobacterium change over the course of a day/night cycle?

To answer these questions, per-cell normalization was chosen. The feature spectral abundance was normalized to the cell number prior to clustering into putative metabolites with the RamClustR package. RamClustR was not able to accurately cluster features with such a low abundance post-cell normalization; thus, per-cell spectral abundance was multiplied by 1E10 to obtain higher numbers to enable accurate clustering. All reported metabolite abundances are per-cell unless noted otherwise.

The total signal per cell across time behaves similarly for each platform, with small differences and biological interpretations (Figure B-5-17). Total signal per cell as a function of

time follows the same trend observed in the other platforms, but the drop in signal upon cell division/transition to darkness is not as severe. Furthermore, the noise in the data is more pronounced, suggesting either more analytic error with this platform or more variability across cells in the population.

For non-polar metabolites in the organic phase, total signal per cell as a function of time displays a very sharp decrease in signal at 9-12 hours, corresponding both with a period of rapid cell division and a transition to a dark phase. The biological meaning behind this could simply be dilution of non-polar metabolites into multiple cells post-division. However, there could also be an aspect of pigmentation degradation associated with the drop. Further analysis will be employed to attempt to differentiate these trends.

For amino acids, signal per cell versus time shows a clear trend mimicking the biomass accumulation/cell division trends (Figure C-5-20). Total amino acid content of the pellet increases during the day and sharply decreases at 6-12 hours when the cells are actively dividing. During the night, the amino acid content of the cells remains relatively stable. Total amino acid content per OD displays a more consistent level throughout the day/night cycle (Figure C-5-20).

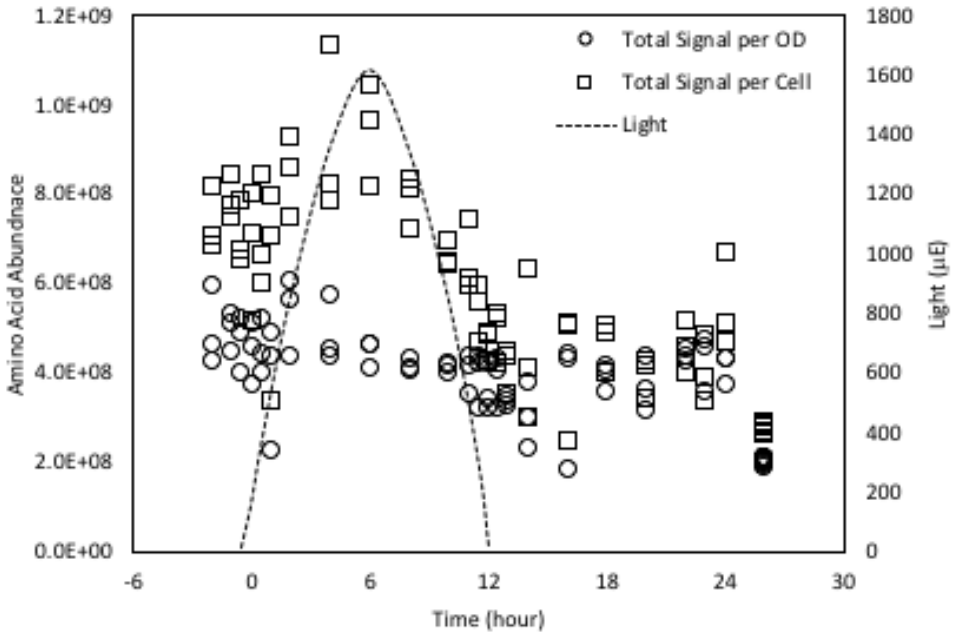

Figure C-5-20. Total amino acid content per OD (circle) and per cell (square). Each biological replicate is presented as a unique data point. The light profile is indicated as a dotted line. All 13 of the QC injections are assigned an arbitrary time value of 26 hours to demonstrate the small effect of run order on total detected amino acid abundances.

Extraction efficiency

Seventy-two samples were randomly assigned one of six extraction batches. To assess whether all batches exhibited the same extraction efficiency total non-normalized signal from each sample in the (RP) LC-TOF-MS experiment was compared by batch (Figure C-5-21). There is not a significant trend in the total signal across extraction batches which indicate batch number did not influence overall extraction efficiency.

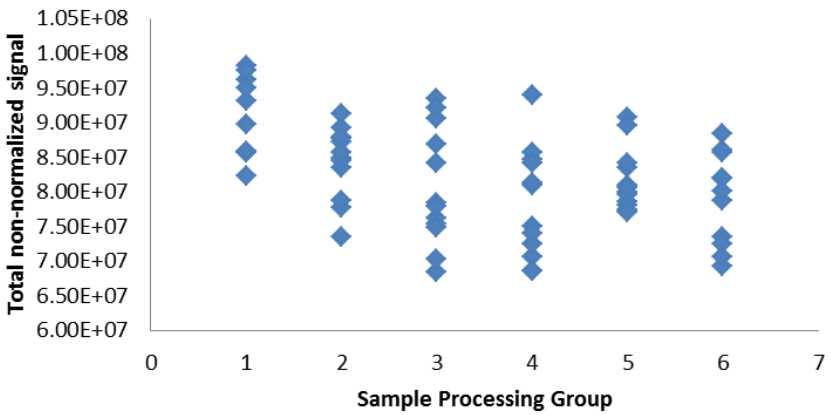

Figure C-5-21. Extraction efficiency assessment by batch number for non-targeted LC-TOF-MS. Total non-normalized signal for each sample is shown. Samples are grouped with other samples processed at the same time.

Labeled glucose (^{13}C-glucose) was added to the lyophilized biomass prior to extraction and analysis. Neither glucose nor ^{13}C-glucose were detected via targeted HILIC-MS despite including the appropriate SRM parameters, and thus no calculation of recovery efficiency could be conducted.

Instrument sensitivity

Mass spectrometry instrument sensitivity is subject to drift over time. Drift is affected by many factors including saturation by sample detection, especially across large batches which require many days of experimental runs. To assess instrument drift, pooled quality control (QC) samples were run between every six injections. The pooled QC sample is composed of a fraction of each individual sample and therefore should be representative of the metabolite composition of all samples.[14]

[14]It should be noted that particularly low-abundance metabolites that are only present in a small fraction of samples will likely be too dilute for detection in QC samples.

In all the MS platforms used in this study, including both GC and LC, the total QC signal did not significantly drift across the batch (Figure B-5-17). The observation was surprising at first, as instrument drift is well documented in the metabolomics literature and field. However, two considerations could explain the consistent signal observed. First, the low sample number in this study only necessitated one batch so the total run time was relatively short (~48-72 hours total). Often mass spectrometry experiments can last weeks, which requires several batches with intermittent instrument conditioning. Here, all samples were run consecutively without any conditioning between injections. Second, the series of fractionation steps (organic phase, aqueous phase, insoluble pellet, pellet hydrolysis and washes) reduced the complexity of the injection to any one instrument. This reduction of complexity reduced the "dirtiness" of the samples, and therefore likely reduced build-up on the instrument detection system which would ultimately improve the robustness of detection across multiple injections.

Still, QC drift was subjected to a QC-based normalization approach to assess the value of such a normalization technique to the data. MetaboQC is an LCMS normalization method developed to correct for instrument sensitivity drift using the behavior of QC samples [226]. Specifically, the MetaboQC method uses a mathematical fit to the QC sample drift across run order and batches to normalize all other metabolite abundances in all samples. MetaboQC was run on the LC-TOF-MS data and assessed pre- and post-normalization (Figure C-5-22). Total signal from QC injections were relatively stable pre-normalization. However, after MetaboQC LOESS normalization, new variability has been induced by the normalization technique. This imposed variability pattern observed for Compound 3 was consistent across all metabolites in the data set, often more pronounced. Indeed, it appears that over-fitting the data resulted in a LOESS-like pattern in QC total signal. Taken together, these assessments indicate that a QC-

based correction method is not suitable for this data set because the QC signal is stable across the relatively-short experiments.

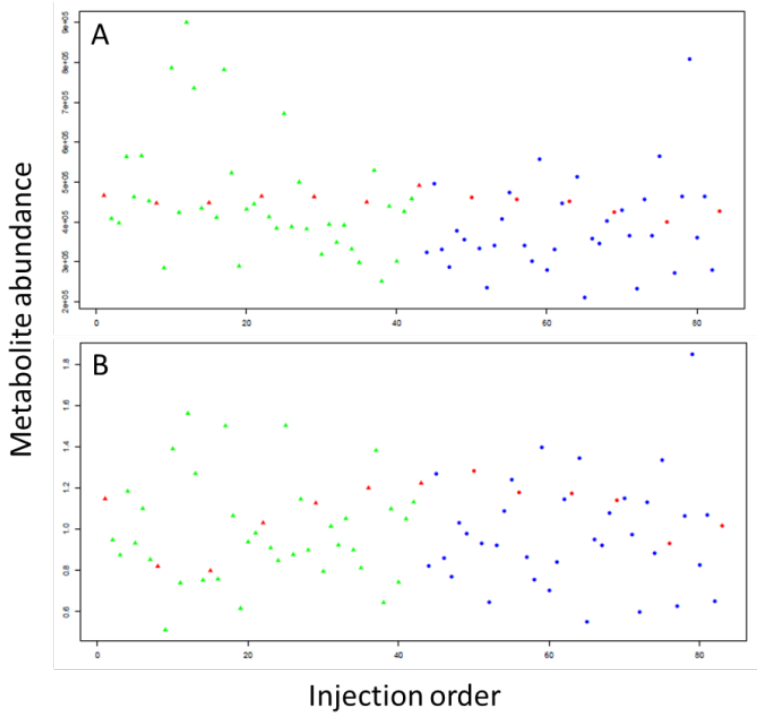

Figure C-5-22. Metabolite 3 abundance vs. injection order (A) without normalization and (B) with MetaboQC normalization. QC injections are shown as red triangles. Sample injections are shown as green (Day 1) and blue (Day 2) circles.

TIC is a widely applied normalization technique in metabolomics experiments and is used to correct drift in total signal uniformly on the sample set. In the case of our experimental design, this approach is not suitable due to the potentially differing metabolite *quality* (*i.e.,* what type of metabolites are present) and *quantity* (*i.e.,* how many metabolites per cell are present) across the many time-points. Indeed, we hypothesize that, for example, wholly different metabolites and metabolite abundances will be observed at 6:00 AM as compared with 6:00 PM. Diverse metabolites are detected at different levels due to difference in chemistries; some are ionized more readily or generate larger detection signal at the same unit concentration. Due to this difference, normalization by TIC would bias data toward time points were the quality of

metabolites are more readily detected, as well as time points were the quantity of metabolites is lower. Therefore, this method was deemed unsuitable for application to this data set.

Quantile normalization is a common normalization approach in microarray transcriptomics experiments and normalizes signals based on signal distribution [227]. Quantile normalization was also applied and analyzed in this study. However, this method relies on the key assumption of normality. In the case of this experimental design, the distribution of metabolites across all samples (corresponding to time points) can display non-normal trends (Figure C-5-23). Instead, the distributions may contain biological meaning. Distribution of metabolite abundance across time (samples) exhibits different behavior: glyeric acid 3-phosphate and butanoic acid display a bi-modal distribution while adenosine 5-phosphate and ribose 5-phosphate display mostly low abundance but very high levels at select times. Application of quantile normalization to this data set disrupts bimodal distributions and imposes a normal distribution. Therefore, the application of this method is inappropriate and would construe the true nature of metabolite abundance over time. This assertion is supported with ample evidence in recent work by Stephanie Hicks and Rafael Irizarry [225]: when global changes are driven by *biological variability across group (time),* as opposed to global technical variability or batch effects, quantile normalization is not appropriate. Simply stated, when there are global changes in distributions across groups—small variability within groups, large variability across groups—quantile normalization is inappropriate.

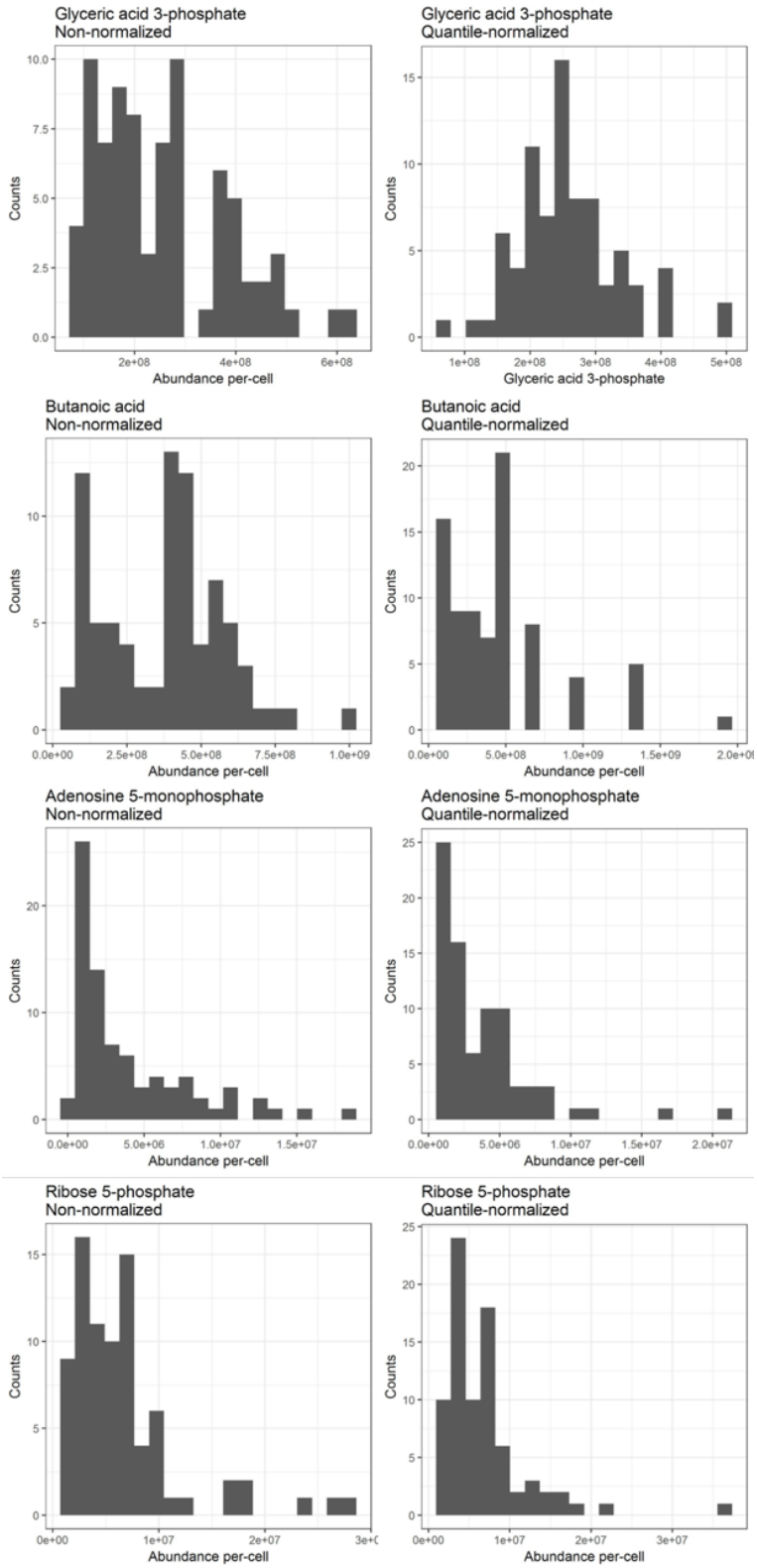

Figure C-5-23. Histograms of select metabolite distributions across all samples from analysis of the aqueous fraction via non-targeted GC-MS with no- and quantile-normalization.

MS platform (data scaling)

We seek to compare the metabolite abundance *change* over time for metabolites detected in five different mass spectrometry experiments, including targeted and non-targeted campaigns on high- and low-resolution instruments equipped with gas and liquid chromatography separation systems. Such an analysis requires substantial pretreatment of the data.

Centering and scaling pretreatment addresses several aspects of metabolomics data, including differences in the magnitude of abundance and metabolite heteroscedasticity. Pretreatment is required to meet the assumptions of downstream analysis (*i.e.* the assumption of heteroscedasticity for regression analysis) and enable metabolite change comparison across platforms. However, the choice of centering and scaling technique depends highly on the questions you seek to ask and the type of data you have and will dramatically affect the results from downstream statistical inquiries. Thus, this section has been dedicated to the task of data centering and scaling and overviews of metabolite data pre- and post-treatment are provided for each platform.

Centering converts concentrations to fluctuations around zero. This step enables comparison of changes from both high- and low-abundance metabolites. Mean-centering is most common. Scaling converts fluctuations in data to differences relative to a scaling factor. This step enables comparison of metabolites which had large differences in fold-change. There are many approaches to scaling, each appropriate for a different question and data set. These approaches can be broadly grouped into classes based on the type of scaling factor: a measure of data dispersion (*e.g.* standard deviation) or a measure of data size (*e.g.* mean).

Autoscaling is applied such that metabolite abundances are centered by the sample mean and scaled by the sample standard deviation [228]. Pareto scaling is another scaling method commonly applied in metabolomics work [229]. Pareto scaling is similar to autoscaling except

the square root of the standard deviation is used as the scaling factor. This enables retention of information on the relative abundance of metabolites which was lost with autoscaling. Both methods retain the original unit (as opposed to dimensionless data).

Autoscaling was performed using the *scale()* function in the base R package such that values were centered by the metabolite mean value and scaled by the metabolite standard deviation. Pareto scaling was performed using the *parteoscale()* function in the RFmarkerDetector package (https://cran.r-project.org/web/packages/RFmarkerDetector/index.html). Pareto-scaled data is used for all analyses in this work, unless stated otherwise.

MS DATA ANALYSIS STRATEGIES

Extracting information from large datasets generated with high-throughput methodologies—such as MS—is a major challenge facing postgenomic era science. Among the many fields of 'omics, metabolomics lags behind in terms of databases availability and curation, data processing algorithms, and statistical approaches to assess the information. Still, the past two decades have developed a common data analysis workflow for metabolomics data, with many methods taken from transcriptomics (mostly microarray) methodologies and not all appropriate for the analysis of time-series data. The goal of this section is to provide background on the approaches utilized in the task of extracting information in Chapter 4.

The statistical approaches employed fall broadly into two categories: unbiased and biased (also referred to as unsupervised and supervised). The identity of the sample is provided to supervised analyses, while it is not provided to unsupervised analyses. For this study, unsupervised methods do not use the quantitative variate—time in this case—as an input to the

analysis, which supervised methods do use time as an input. Utilizing supervised analyses alone can be misleading, and often a mix of both supervised and unsupervised analyses is appropriate to extract information from a large dataset. Here, principal component analysis (PCA) and hierarchical clustering analysis (HCA) are applied as unsupervised methods. Partial least squares regression analysis (PLS), analysis of variance (ANOVA), artificial neural networks (ANN), and support vector machines (SVM) are supervised methods applied here to gain understanding of metabolic interaction with time. Methods used to interpret the results from both unsupervised and supervised methods—referred to as interpretation methods here—include metabolite set enrichment analysis (MSEA) and pathway analysis. These final methods require tight integration with metabolite databases. Each approach mentioned above is introduced and a brief background is provided.

Principal component analysis (PCA)

PCA is an unbiased multivariate statistical method used to provide an overview of large trends in variation. PCA is singular-value decomposition or eigenvalue decomposition method which develops a set of linearly uncorrelated variables termed 'principal components' which maximize variation in the data set. The outputs of PCA are component scores associated with each sample and loadings scores associated with each metabolite within a sample. Optimal interpretation of PCA utilizes both the scores and loadings to determine whether the variation in the data is drive by intended biological variation or analytical variation, processing variation, or unintended biological variation.

PCA is commonly employed to detect outliers in metabolomics experiments. D.M. Hawkins defines an outlier as "an observation which deviates so much from the other

observations as to arouse suspicions that it was generated by a different mechanism" [231]. Outlier detection and treatment is an active area of research, with many different methods available for different applications. In this study, the goal of outlier assessment is to determine if any sample deviates from its biological replicates in a way so great it may disrupt, construe, or mislead the interpretation of downstream analyses. The ability to detect outliers is reduced in this study because the induced variation falls along a continuum due to the continuous time-course experimental design. This will likely make outlier detection more difficult as opposed to a classification study, where two groups differ in a binary way and therefore the data should separate more clearly, thereby aiding the identification of outlier samples which do not separate in the commonly observed way. Furthermore, the use of only three biological replicates in this study limits our ability to remove outliers. Therefore, the approach to outlier assessment used here is very conservative. PCA scores plots will be assessed for the presence of blatant outliers, and further probing with tests on the raw data will be applied for cases where an outlier may be present.

Hierarchical cluster analysis (HCA)

HCA and heat mapping can be used to visualize the overall structure of the data as well as probe relationships and/or correlations between samples or variables. HCA is an unsupervised cluster analysis method which seeks to cluster similar objects into groups, often represented as a dendrogram [232]. Within HCA, there are many choices which affect the cluster analysis, including metric for dissimilarity and linkage criteria. Metric affects the shape of the clusters and linkage criteria determines the criteria which are to be considered for feature similarity. Commonly, Euclidean distance is used as the metric for HCA on metabolomics data [233].

Several linkage criteria are popular, including centrodial, average, single (nearest neighbor) and complete (farthest neighbor) [234].

Other cluster analysis methods exist for the interpretation of metabolomics data. *K*-means is a variant where the number of clusters is user-defined, and Fuzzy *k*-means is a variant of *k*-means where a sample may belong to more than one cluster [234]. Because these methods require user-specification of the cluster number, they require validation and are used in combination with other methods. Commonly, data input to HCA are restricted to metabolites deemed to be significant by other methods (*e.g.* ANOVA), or dimensionally reduced data sets (*e.g.* PCA loadings). In this way, insight regarding un-annotated metabolites can be gained by assessing the annotated metabolites in the co-clusters.

Heat maps often accompany clustering dendrograms to visualize the patterns of metabolic abundance ordered by the clusters. Heat maps are visual representations of the data as colorized cells in a matrix, with samples and metabolites ordered according to clustering output. Because the coloration of heat maps greatly influences the interpretation, color schemes associated with heat maps is an area of active discussion. The convention in transcriptomics work is green coloration for down-regulation and red coloration for up-regulation.

Regression analysis

A primary goal of this study was to identify *which metabolites significantly change* over the course of a day/night cycle, and then assess *how they change.* Paramount among the analyses useful toward this goal is regression analysis. Regression analysis is commonly employed in a broad range of fields—not just time-course studies—as it is useful toward understanding the relationship between a dependent variable and independent variables. Models derived from

regression analysis are commonly employed toward the prediction and forecasting of data, as seen in the fields of financial forecasting. Time-series analysis comprises a field of statistics which seeks to meaning from the response variables (*e.g.* measured metabolite abundances) to characterize time. The goals of such analyses overlap significantly with machine learning techniques such as artificial neural networks (ANN) or support vector machine (SVM). While useful, ANN and SVM techniques are not discussed within this section.

The regression problem can be stated as "how to model one or several dependent variables, responses, **Y**, by means of a set of a predictor variables, **X**" [235]. Said another way, regression analysis seeks to mathematically describe the relationship between a dependent variable (*e.g.* the observed abundance of a metabolite) and an independent variable (*e.g.* time). Linear and quadratic regression analyses were run on all identified metabolites—both annotated and non-annotated—from all platforms.

MSEA & Pathway Analysis

Metabolite set enrichment analysis (MSEA) "directly investigates if a group of functionally related metabolites are significantly enriched" [236]. This approach, contrasted with statistical methods described above, has the ability to detect subtle coordinated changes in metabolites within a given biological pathway. MSEA is the metabolomics analog of the popular transcriptomics Gene Set Enrichment Analysis technique [237]. MSEA can be run with continuous phenotypes (*e.g.* time). Quantitative enrichment analysis (QEA) is a mode of running MSEA which tests against a null hypotheses described in *globaltest():* "none of the covariates in the tested group is associated with the response" [238]. This method is based on a regression model of the response variables as a function of the covariates, and thus is less prone to reject the null based on a few small changes. MSEA QEA computes a *Q*-stat, Bonferroni-corrected p-

values, and FDR estimates (discussed in next section) for interpretation of the significance of the results. The *Q*-stat is a measure of the correlation between a compound concentration profile and the phenotype [238].

An obvious requirement for meaningful MSEA results are appropriate metabolite sets upon which to search. MSEA contains five libraries of metabolite sets (~1000 groups) and the ability to specify use of non-mammalian metabolite sets. Metabolite sets in MSEA are defined to be: "involved in the same biological process, changed significantly under the same pathological conditions, and present in the same locations" [236]. This limitation—and potential source of bias—should be kept in mind during MSEA interpretation. Furthermore, annotation is a requirement for input and thus insight into the identity or function of un-annotated compounds is not gained through MSEA.

UNSUPERVISED OVERVIEW OF VARIATION IN THE METABOLOMICS DATA

Pellet content: amino acid, nucleic acid, polysaccharide and other metabolites

PCA of amino acid content of the pellet

PCA was run on pareto-scaled per-cell amino acid abundances and the first four components are summarized in Table C-9. The first four components capture 97% of the variation in the data, with PC1 capturing 85% of the variation alone.

Table C-9. Summary of the first four principal components for GC-MS analysis of the amino acid content of the insoluble pellet.

Factor	PC1	PC2	PC3	PC4
Proportion of Variance	0.85	0.07	0.03	0.02
Cumulative Proportion	0.85	0.92	0.95	0.97
Standard deviation	61248	17873	11220	8988

PC1 vs. PC2 is explored further to gain an understanding of the underlying variation in the data. Along PC1, separation into day- and night-time samples is achieved (Figure C-5-24A). Interpretation of the loadings plot indicates that this separation is driven by many factors, with day-time samples richer in amino acids such as alanine and serine, while night-time samples are richer in amino acids such as threonine, aspartic acid, methionine, and glutamine (Figure C-5-24B). Separation along PC2 is not easy to interpret, as it does not clearly separate based on time. The QC injections cluster very tightly in the upper right-hand corner of the plot, indicating that there is little variation among the injections but the content is not a perfect hybrid of all the injections because it is not centered on zero. The total signal of QC injections, while steady across injection order, was substantially lower than the rest of the samples (see previous section). The lower signal intensity could be a factor in the deviation of QC samples away from [0,0] in the PC1 vs. PC2 scores plot.

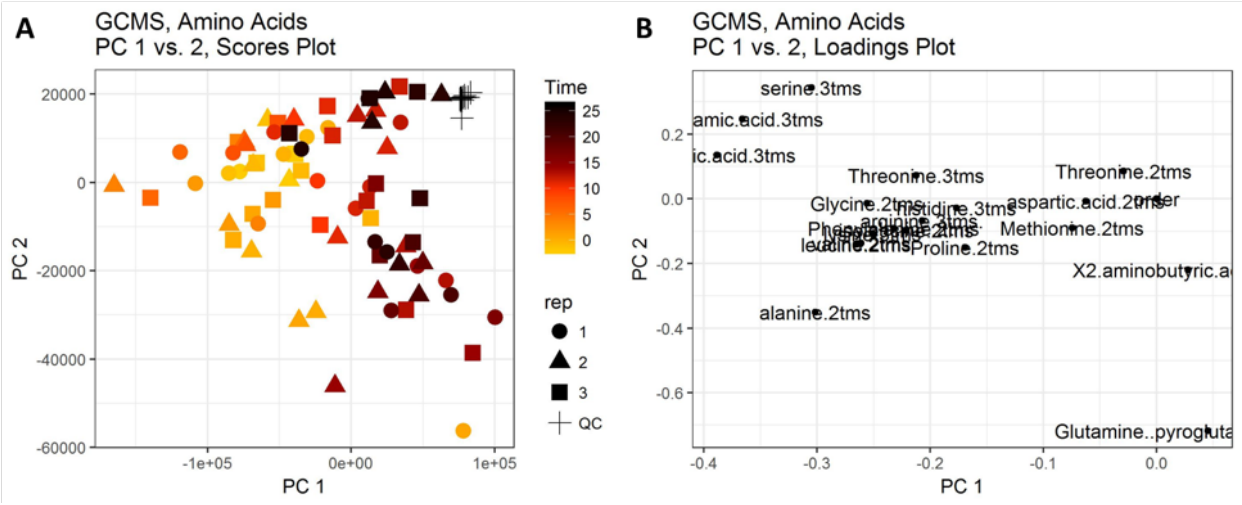

Figure C-5-24. PC1 vs. PC2 (A) scores and (B) loadings plot for GC-MS analysis of the amino acid content of the insoluble pellet.

Sample 1 (lower right-hand corner to scores plot) stands out as a potential outlier, with the source of variation likely attributable to glutamine, which was detected as pyroglutamic acid (+2 MTBSTFA). Sample 1 was collected at the very first time-point, *t*=-2, and is a biological

replicate of samples 2 and 3. The amount of glutamine detected in sample 1 does not stand out as disproportionately high, nor does the distribution of amino acids detected in sample 1 as compared with biological replicates (Figure C-5-25A). Instead, the histograms of amino acids in samples 1-3 suggests that the absence of a key amino acid may be the underlying cause of variation in sample 1 as compared with samples 2 and 3 (Figure C-5-25B). The underlying cause for variation between samples 1-3 is not so clear as to warrant removal of sample 1 as an outlier.

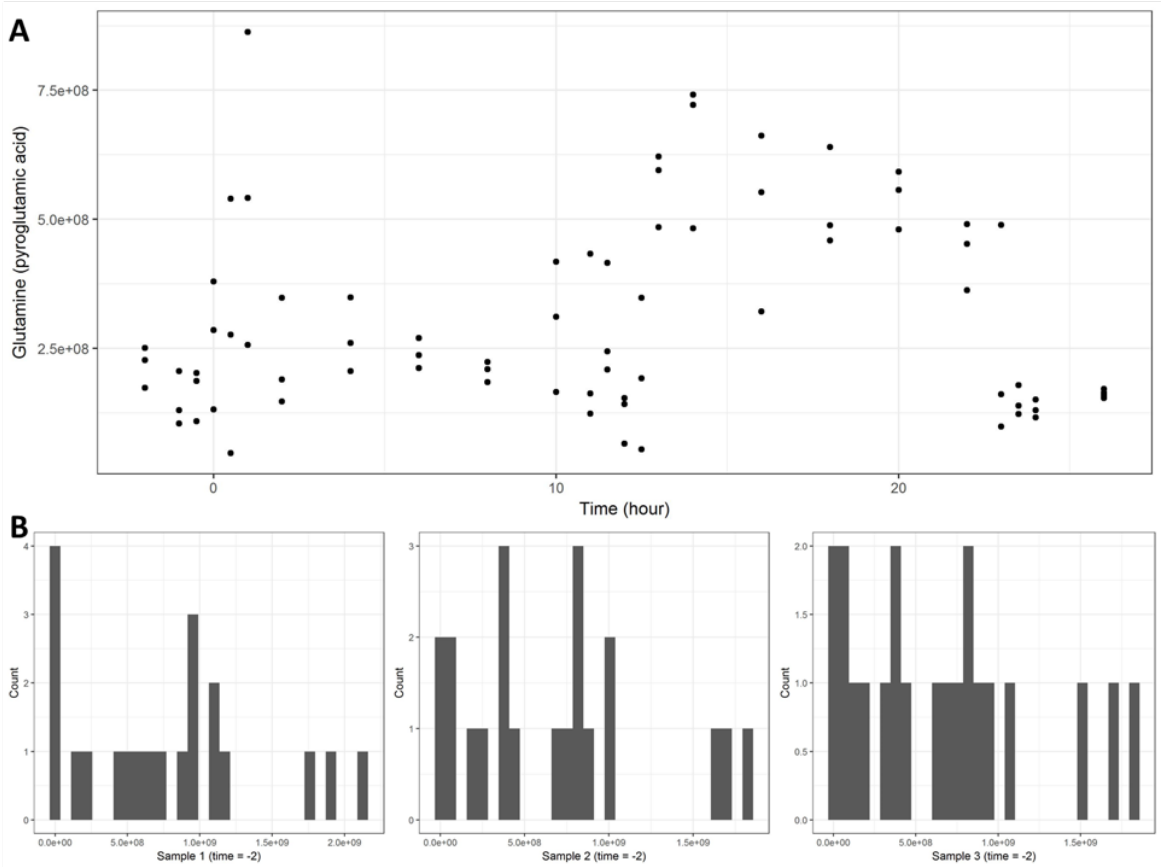

Figure C-5-25. Glutamic acid in the pellet. (A) Glutamic acid over the course of a day/night cycle. (B) Distribution of glutamic acid across samples 1, 2, and 3.

PCA of nucleic acids, polysaccharides, and other components of the pellet

PCA was run on per-cell normalized and pareto-scaled metabolite abundances for non-targeted GC-MS analysis of the pellet and the first four principal components are summarized in Table

C-10. The first four components capture 87% of the variation, with PC1 capturing 59% of the variation alone.

Table C-10. Summary of the first four principal components for non-targeted GC-MS analysis of the insoluble pellet.

Factor	PC1	PC2	PC3	PC4
Standard deviation	146511	77055	53777	36959
Proportion of Variance	0.59	0.16	0.08	0.04
Cumulative Proportion	0.59	0.75	0.83	0.87

Early-morning time-points separate from the later samples along PC 3 (Figure C-5-26A). Interpretation of the loadings plot indicates that C682 is largely responsible for this separation (Figure C-5-26B). The QC injections cluster quite closely in the center of the plot, as expected.

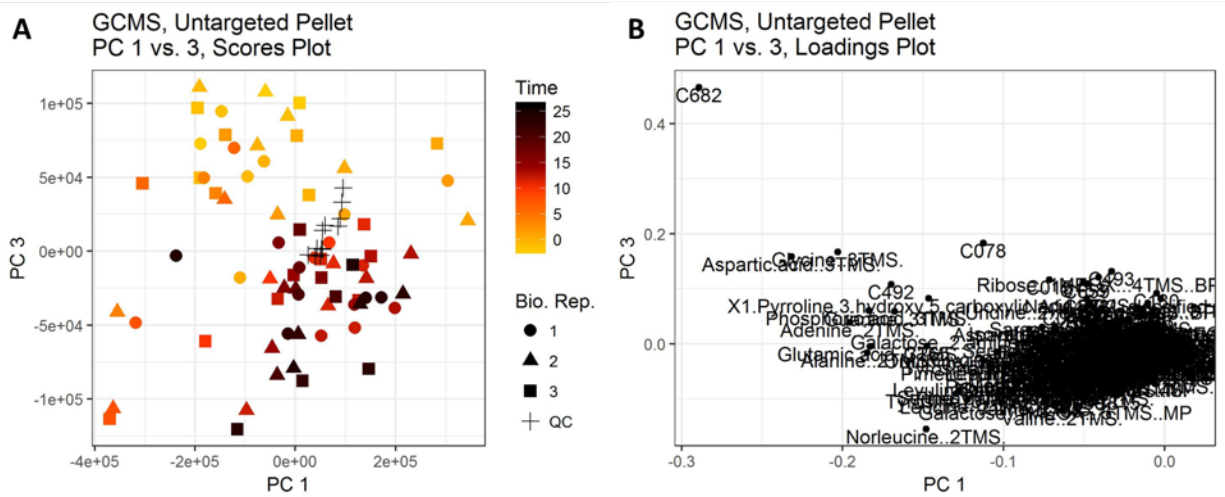

Figure C-5-26. PC1 vs. PC3 (A) scores plot and (B) loadings plot for non-targeted GC-MS analysis of the insoluble pellet.

Polar and semi-polar metabolites in the aqueous phase

PCA of small polar compounds detected by GC-MS

PCA was run on pareto-scaled per-cell metabolite abundances and the first four components are summarized in Table C-11. The first four components capture 81% of the variation in the data, with components PC1 and PC2 capturing 65% of the variation alone.

Table C-11. Summary of the first four principal components for non-targeted GC-MS analysis of the aqueous phase.

Factor	PC1	PC2	PC3	PC4
Proportion of Variance	0.39	0.26	0.11	0.05
Cumulative Proportion	0.39	0.65	0.76	0.81
Standard deviation	72265	58892	38377	26635

Separation along PC4 is visualized as mid-day samples clustered near the top, late-night samples clustered in the middle, and early-morning samples clustered near the bottom of the y-axis (Figure C-5-27A). Interpretation of the loadings plot indicates that glycolic acid, lactic acid, and glyceric acid 3-phosphate contribute to the mid-day sample separation, while butanoic acid and C0724 contribute to the early-morning sample separation (Figure C-5-27B). Again, phosphoric acid contributes to the separation of early- and late-time points along PC2. The QC injections cluster quite closely in the center of the plot, as expected.

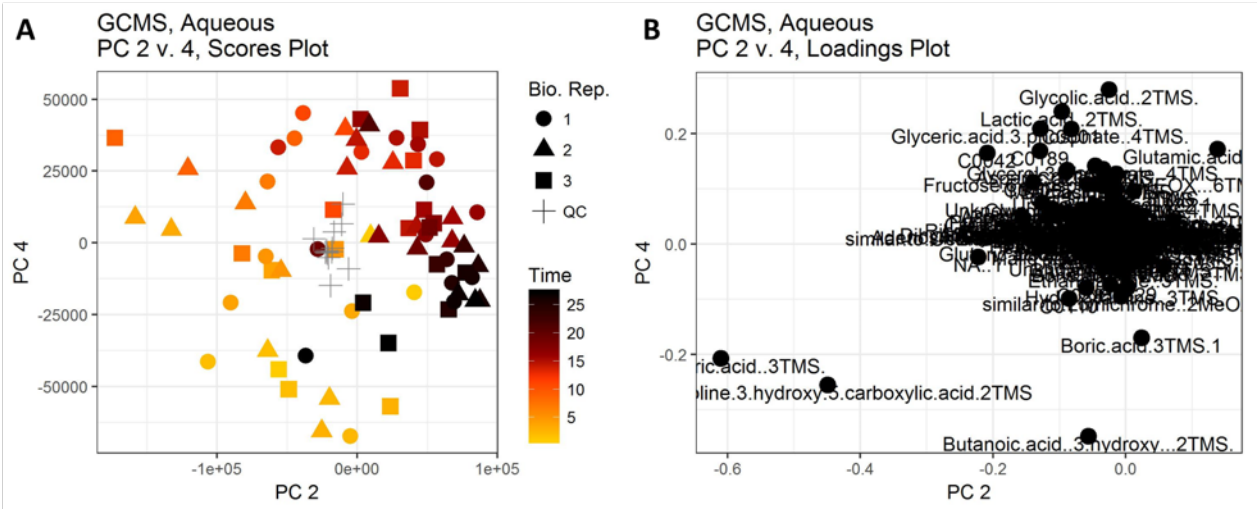

Figure C-5-27. PC2 vs. PC4 (A) scores plot and (B) loadings plot for non-targeted GC-MS analysis of the aqueous phase.

PCA of large polar and semi-polar compounds detected by targeted HILIC-TQS

PCA was run on pareto-scaled per-cell metabolite abundances and the first four components are summarized in Table C-12. The first four components capture 92% of the variation in the data, with the first component capturing 72% alone.

Table C-12. Summary of the first four principal components for targeted HILIC-MS analysis of the aqueous phase.

Factor	PC1	PC2	PC3	PC4
Proportion of Variance	0.72	0.10	0.07	0.03
Cumulative Proportion	0.72	0.82	0.89	0.92
Standard deviation	36005	13476	11269	7620

Both PC1 and PC2 provided separation by time (Figure C-5-28A). Separation of earlier time-points along PC1 is driven mainly by adenosine, adenine, and S-adenosylmethionine (Figure C-5-28B). Separation of mid-day time-points along PC2 is driven mainly by oxidized glutathione (GSSG) and arginine (Figure C-5-28B). Reduced glutathione (GSH) supports several essential cellular functions, including quenching reactive oxygen species and thereby relieving oxidative stress caused during high-light stress [239]. The ratio of GSH to oxidized glutathione (GSSG) can be used as a proxy for assessing oxidative stress [240]. Observing GSSH driving separation of mid-day samples is consistent with a hypothesis that the pool of GSH is depleted in cells experiencing high oxidative stress during the middle of the day under high light exposure.

The QC injections cluster quite closely in the center of the plot, as expected. The overlap of black-colored time points ~ 25 hours with yellow-colored time points ~ 0 hours is expected as all of these samples correspond to the onset of dawn. The variation in these samples may be attributed to glutamate and glutathione.

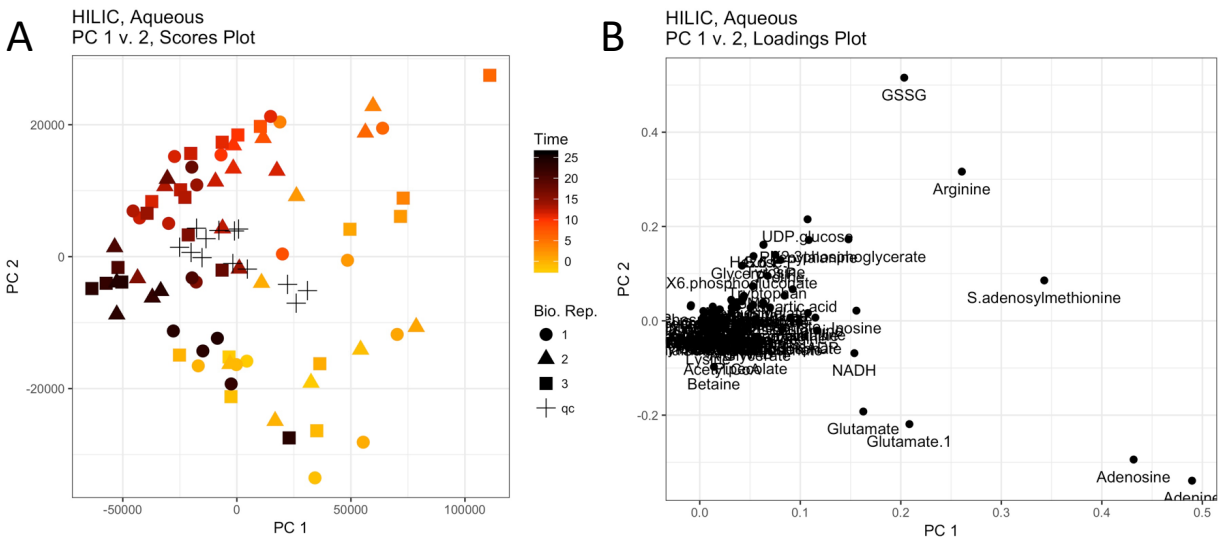

Figure C-5-28. PC1 vs. PC2 (A) scores and (B) loadings plots for targeted HILIC-MS analysis of the aqueous phase.

Non-polar metabolites in the organic phase

PCA and HCA of non-polar compounds detected by non-targeted LC-MS

PCA was run on pareto-scaled per-cell metabolite abundances and the first four components are summarized in Table C-13. The first four components capture 63% of the variation in the data.

Table C-13. Summary of the first four principal components for non-targeted LC-TOF-MS analysis of the organic phase.

Factor	PC1	PC2	PC3	PC4
Proportion of Variance	0.44	0.10	0.06	0.04
Cumulative Proportion	0.44	0.54	0.59	0.63
Standard deviation	8641	4016	3099	2578

PC1 provided excellent sample separation by time (Figure C-5-29A), which seems to be largely driven by C0165, with several other metabolites including C0017, C0487, C0016, and C0001 contributing to the separation of early and late time-points. Annotation of RP-MS data using InterpretMSSpectrum and MSFinder identified C0017 as associated with the molecular

formula $C_{44}H_{72}N_4O_6$, C0487 as Eschscholtzxanthin, C0016 as 2-[(9Z)-9-Hexadecenoyloxy]-3-(hexopyranosyloxy)propyl (9Z,12Z,15Z)-9,12,15-octadecatrienoate, and C0001 as associated with the molecular formula $C_{24}H_{51}N_7O_6$. Due to the increased uncertainty of RP-MS annotations, the biological meaning of RP-MS data will be discussed further in the following section. The QC injections cluster quite closely in the center of the plot, as expected.

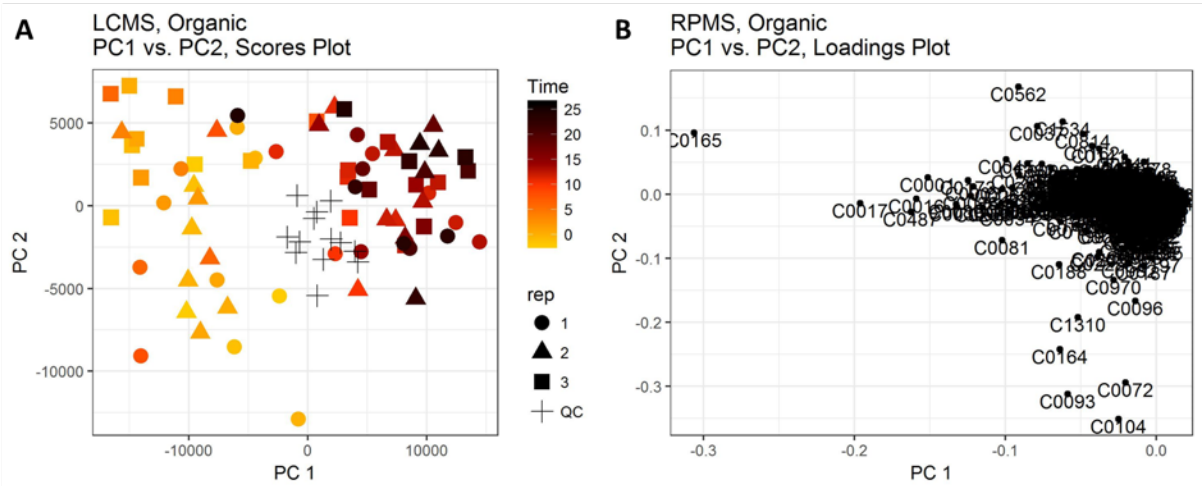

Figure C-5-29. PC1 vs. PC2 (A) scores and (B) loadings plots for non-targeted LC-TOF-MS analysis of the organic phase.

Due to the large number of compounds and the difficulty associated with determining suitable annotations in the non-targeted RP-MS dataset, HCA was also run to identify the key metabolites for further focus. HCA was run unbiased: both the samples (columns) and metabolites (rows) were clustered without restriction. In this way, two key interpretations are made possible. First, analysis of the sample clusters can be informative as to whether biological replicates and temporally similar time-points are truly similar at the metabolite level. Second, analysis of the metabolite clusters can be informative as to which metabolites contribute to variation over time, and—when a handful of the metabolites are annotated—can be informative as to the identity of un-annotated metabolites. The metabolites used for HCA were the top 50

most important metabolites as identified by PLS-DA. Without this filter to reduce the number of metabolites included, the heat-map was 2,000+ metabolites long and very difficult to interpret. HCA output accompanied with a heat map indicated that biological replicates and samples of similar circadian time cluster relatively tightly (Figure B-5-9). This observation serves as an additional validation that the global variation in the data structure is driven by time. Metabolites cluster into two main groups, with both groups displaying decreased abundance during the night period (classes O-U).

Analysis of the heat-map results is more intuitive when the samples (columns) are ordered by time (Figure C-5-30). Interestingly, the top 50 metabolites identified as significantly changing by PLS-DA and visualized by heat-map all display roughly the same behavior: up-regulation during the light period, down-regulation during the night period. Interpretation of the heat-map results indicate that up-regulation of metabolites begin *prior* to the onset of light, indicated by the red coloration of metabolites within sample classes A-C, the 10^{th}, 11^{th}, and $11½^{th}$ hour of the 12-hour night phase. Abundance levels remain relatively steady throughout the day, and a shift to decreased abundance begins roughly own-regulation begins just before transition to night, at ~the 10^{th}, 11^{th}, and $11½^{th}$ hour of the 12-hour day phase. A shift back toward up-regulation is observed in the Y-L24 class, corresponding to the moment just prior to a shift back into light, which supports the notion that these metabolite accumulation patterns follow a diurnal pattern. However, it is worth noting that the up-regulation shift is not observed in classes V and W as one would expect based on the behavior of classes A-C.

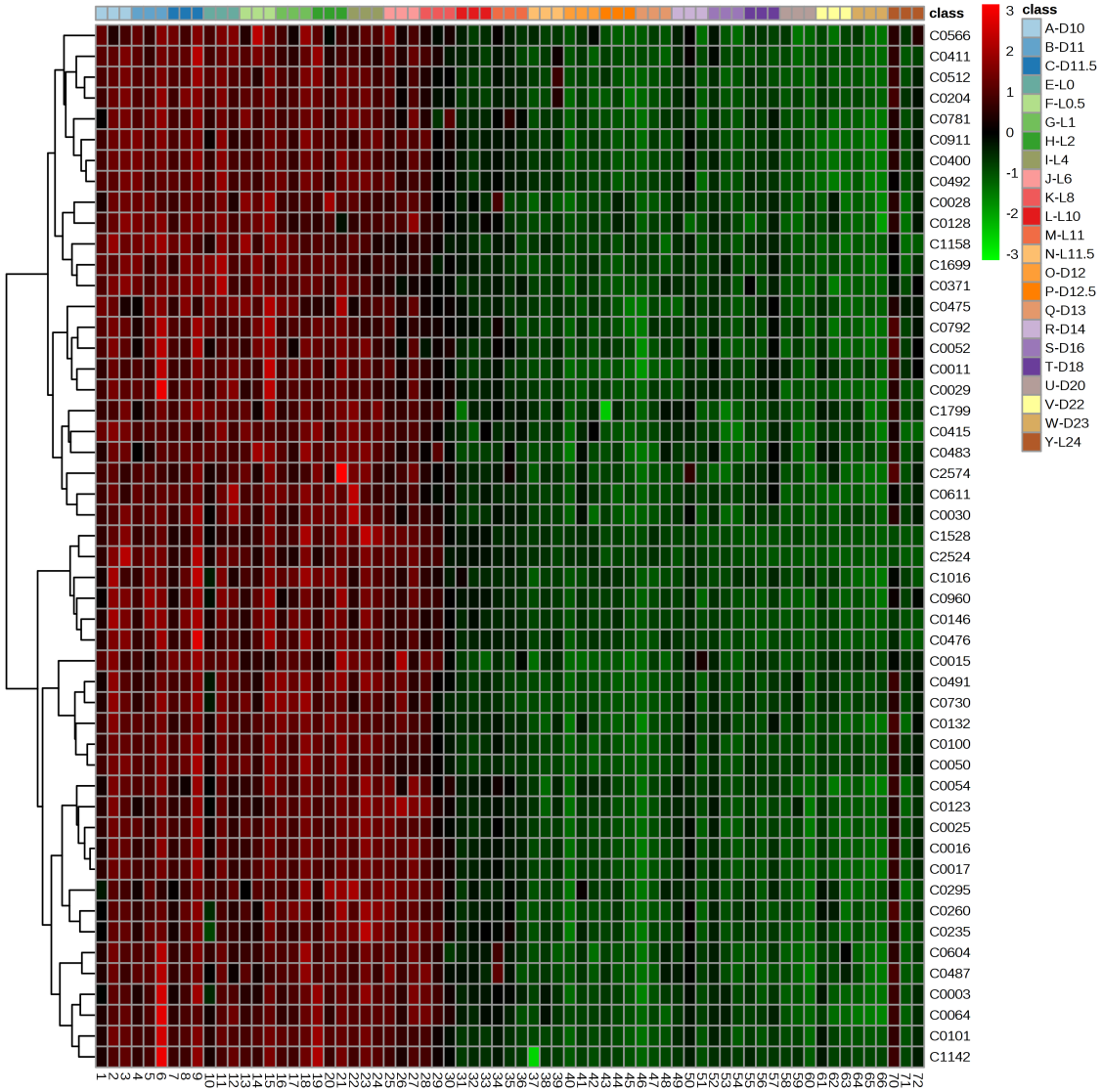

Figure C-5-30. HCA and heat-map of changing non-polar metabolites identified from the organic phase by non-targeted RP-MS. Samples (columns) were not clustered; only metabolites (rows) were clustered. Metabolite up-regulation is depicted in red, while down-regulation is depicted in green. Class A, B, and C correspond to last hours of darkness before transition to light. Classes E-N correspond to hours during the day. Classes O-Y corresponds to hours during the night.

METABOLOMICS REGRESSION ANALYSIS MEASURES OF SIGNIFICANCE CALCULATION AND CUT-OFF VALUE SELECTION

Non-targeted GC-MS analysis of the aqueous phase

For GC-MS analysis of the aqueous phase, 1,329 tests were completed simultaneously and therefore a multiple testing correction adjustment is necessary. For linear regression analysis, a *q*-value cut-off of 0.01 gives 1,163 significant tests with <5 of those tests expected to be false positives (Figure C-5-31A-B). For quadratic regression analysis, a *q*-value cut-off of 0.05 gives 285 significant tests with <15 of those tests expected to be false positives (Figure C-5-31C-D). *Q*-value distributions were roughly uniform(0,1) (Figure C-5-32). BH adjusted *p*-values were not selected because the distribution for quadratic regression was poor (Figure C-5-33).

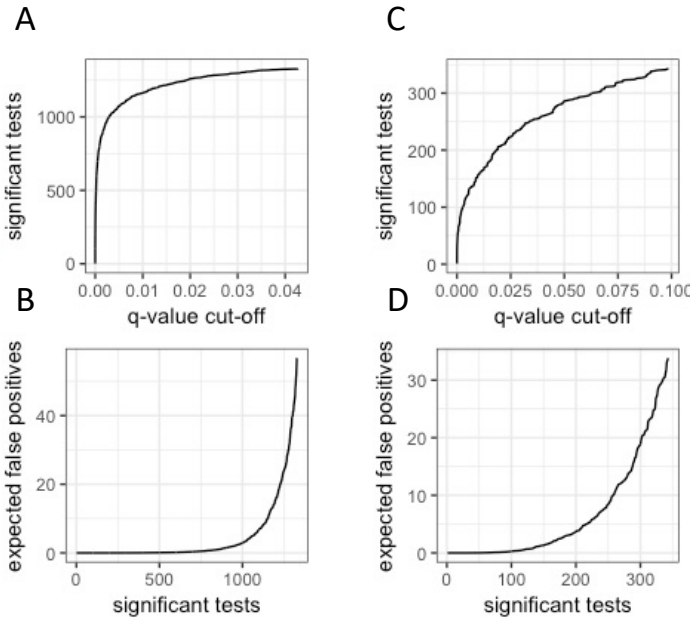

Figure C-5-31. q-value and FDR estimations for GC-MS analysis of the aqueous phase. (A-B) Linear regression analysis. (C-D) Quadratic regression analysis.

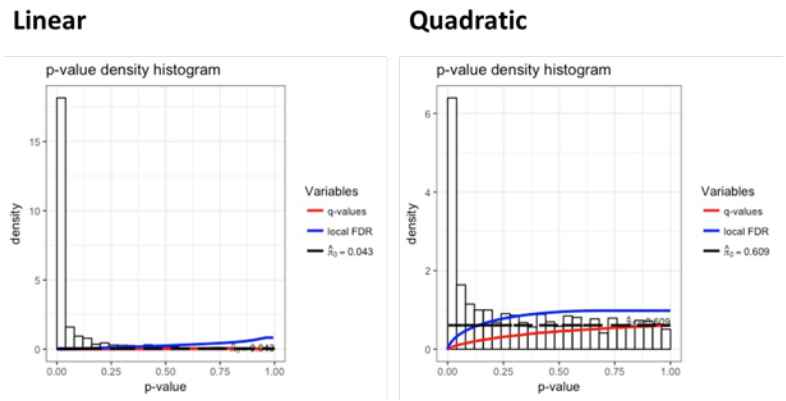

Figure C-5-32. Histograms of Q-values from an ANOVA of linear and quadratic regression analysis of metabolites in the aqueous phase detected by non-targeted GC-MS.

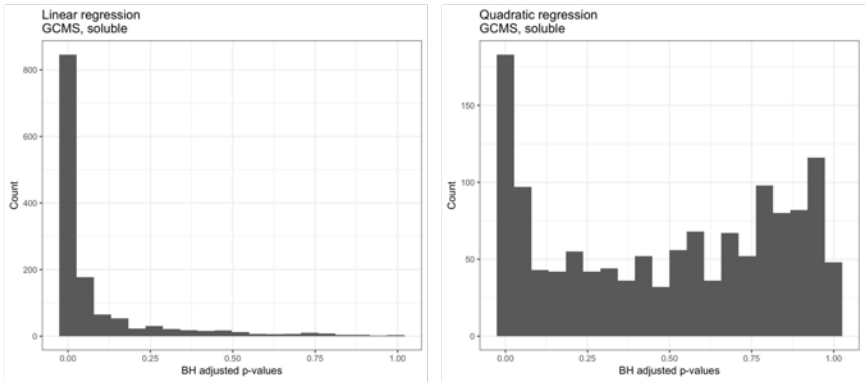

Figure C-5-33. Histograms of Benjamini-Hochberg corrected p-values from an ANOVA of linear and quadratic regression analysis of metabolites in the aqueous phase detected by non-targeted GC-MS.

Targeted HILIC analysis of the aqueous phase

For targeted HILIC-MS analysis of the aqueous phase, 103 tests were completed simultaneously. Without any adjustment, a *p*-value cut-off of 0.05 would suggest <6 false positives and therefore no adjustment was deemed necessary. A *p*-value of 0.01 would capture the potentially alternative linear regression analysis tests (Figure C-5-34A). A *p*-value of 0.125 would capture the potentially alternative quadratic regression analysis tests (Figure C-5-34B). BH adjusted *p*-values were also calculated and display roughly the same distribution as raw *p*-values (Figure C-5-35).

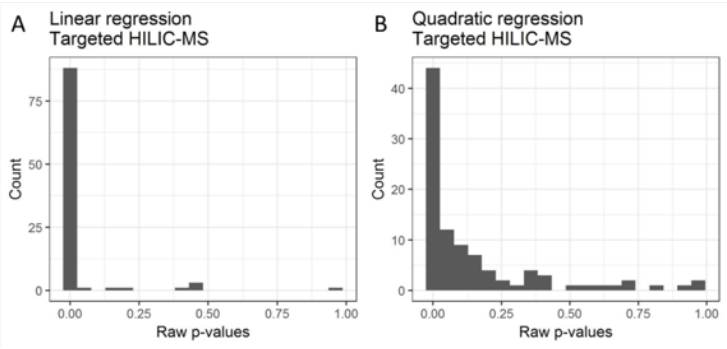

Figure C-5-34. Raw p-value histograms for (A) linear and (B) quadratic regression analysis of the aqueous phase by targeted HILIC-MS.

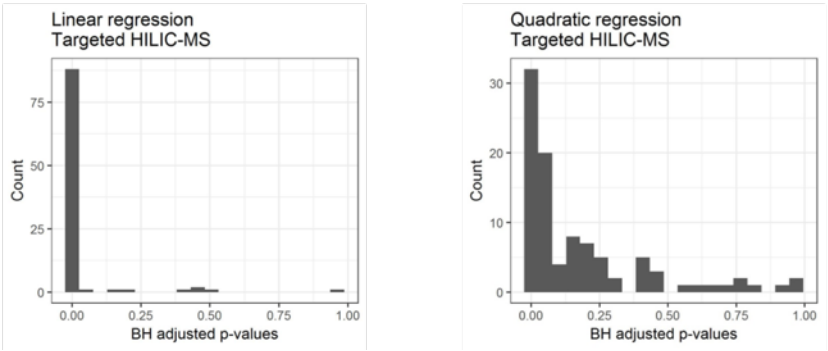

Figure C-5-35. Histograms of Benjamini-Hochberg corrected p-values from an ANOVA of linear and quadratic regression analysis of metabolites in the aqueous phase detected by targeted HILIC-MS.

The *q*-value calculation fit for both linear and quadratic regression of the HILIC data was very poor (Figure C-5-36). Observing the *q*-value histograms for both linear and quadratic analysis indicate a large proportion of tests are significant (Figure C-5-37). However, because the tuning parameter fit is not acceptable, *q*-values were not used. The reason for the poor fit is not well understood, but the low number of tests could be a contributing factor (103 tests); *q*-value application is more appropriate for situations with high numbers of test, such as those seen in high-throughput genomics, or non-targeted metabolomics, experiments.

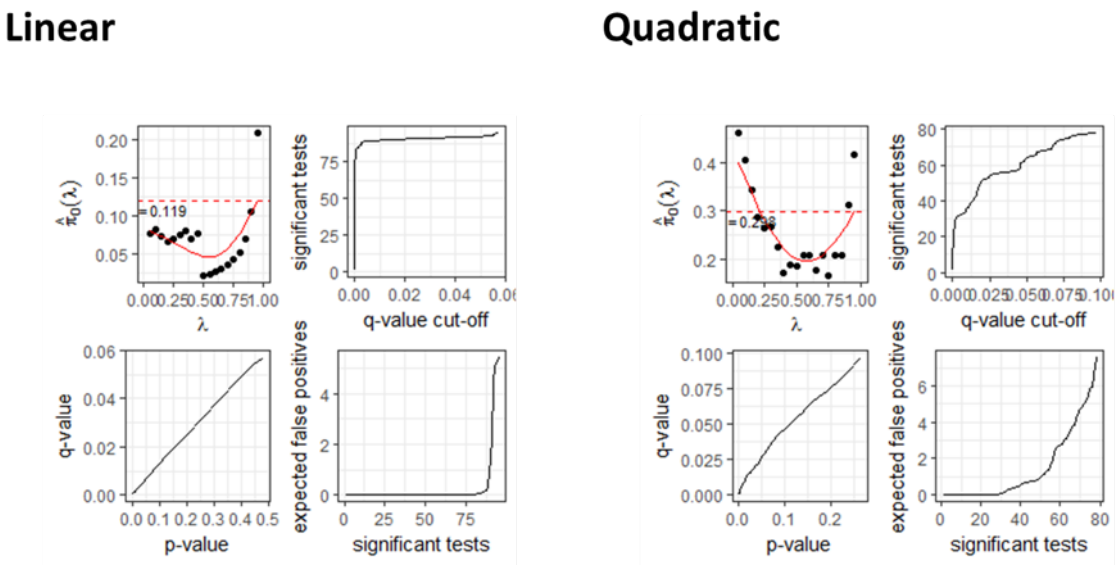

Figure C-5-36. Q-value and FDR estimation for targeted HILIC-MS analysis of the aqueous phase.

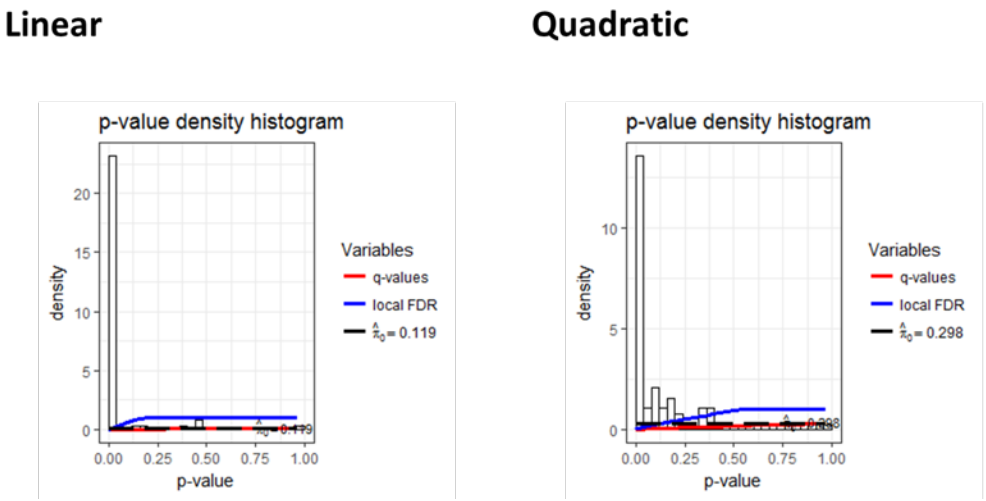

Figure C-5-37. Q-value histograms for targeted HILIC-MS analysis of the aqueous phase.

Non-targeted RP-MS analysis of the organic phase

For RP-MS analysis of the organic phase, 2,683 tests were completed simultaneously and therefore a multiple testing correction adjustment is necessary. For linear regression analysis, a *q*-value cut-off of 1E-4 gives 1,327 significant tests with <5 of those tests expected to be false positives (Figure C-5-38A-B). A BH adjusted *p*-value of 0.05 would capture the potentially alternative tests as depicted in the histogram of BH adjusted *p*-values (Figure C-5-39A). For

quadratic regression analysis, a *q*-value cut-off of 0.001 gives 492 significant tests with <5 of those tests expected to be false positives (Figure C-5-38C-D). A BH adjusted *p*-value of 0.05 would capture the potentially alternative tests as depicted in the histogram of BH adj. *p*-values (Figure C-5-39B).

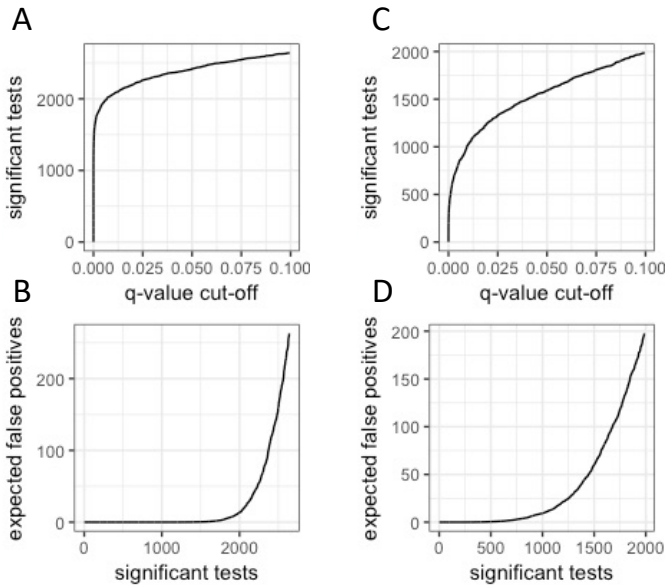

Figure C-5-38. q-value and FDR estimation for RP-MS analysis of the organic phase. (A-B) Linear regression analysis. (C-D) Quadratic regression analysis.

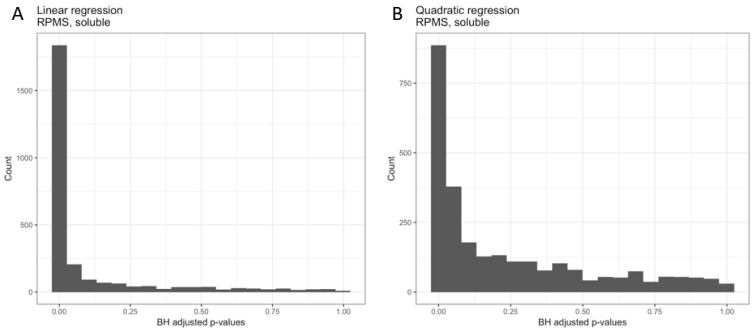

Figure C-5-39. BH adjusted p-value histograms for (A) linear and (B) quadratic regression analysis of the organic phase by RP-MS.

GC-MS analysis of the pellet

For GC-MS analysis of the pellet amino acid content, only 20 tests were completed simultaneously and therefore a multiple testing correction adjustment is not necessary. A *p*-value cut-off of 0.05 is selected by convention for both linear and quadratic regression analyses.

For GC-MS analysis of the pellet nucleic acid, polysaccharide, and other compound content, 862 tests were completed simultaneously and therefore a multiple testing correction adjustment is necessary. For linear regression analysis, a *q*-value cut-off of 0.01 gives 173 significant tests with <5 of those tests expected to be false positives (Figure C-5-40A-B). For quadratic regression analysis, a *q*-value cut-off of 0.01 gives 54 significant tests with <1 of those tests expected to be false positives (Figure C-5-40C-D). BH adjusted *p*-values were not selected because the distribution for quadratic regression was poor (Figure C-5-41).

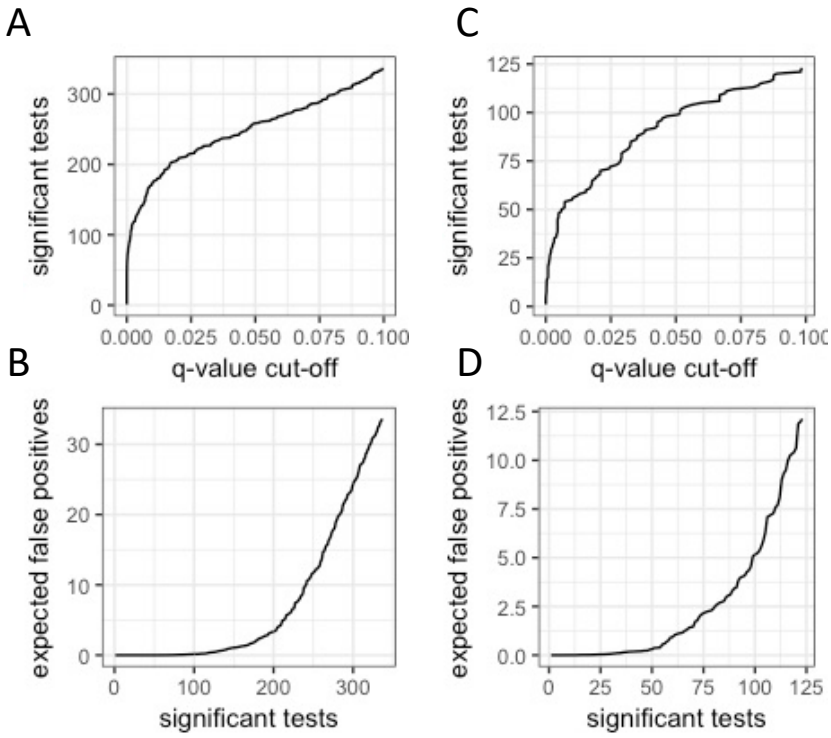

Figure C-5-40. q-value and FDR assessment for GC-MS analysis of the pellet nucleic acids, polysaccharides, and other components. (A-B) Linear regression analysis. (C-D) Quadratic regression analysis.

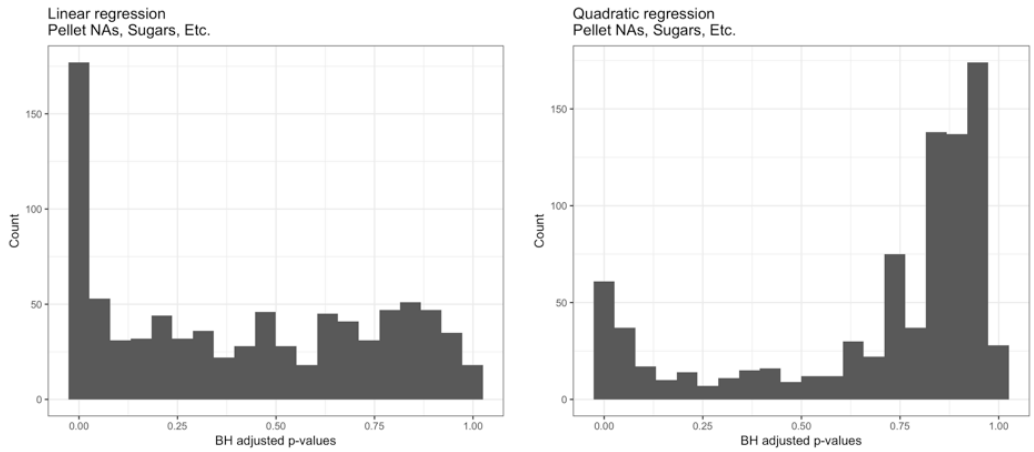

Figure C-5-41. Histograms of Benjamini-Hochberg corrected p-values from an ANOVA of linear and quadratic regression analysis of metabolites in the pellet detected by non-targeted GC-MS.

APPENDIX D –EXPERIMENTAL DESIGN FOR LIGHT-ENTRAINED HEPTADECANE PRODUCTION [15]

RATIONALE AND EXPERIMENTAL OVERVIEW

Well-characterized biological parts for controlling heterologous gene expression in cyanobacteria are needed to improve strain design. Regulatory events unique to cyanobacteria require genetic tool development separate from those developed for heterotrophic chasses, such as *E. coli* [126]. We hypothesize that light-entrained biochemical production will increase chemical yield in LD cycles by minimizing the metabolic burden. We propose to test this hypothesis by comparing constitutive (always "on") and light-entrained alkane production in the model cyanobacterium *S.* 6803 cultivated in diurnal LD cycles. Energy cofactors (ATP, NADPH) will also be analyzed to assess metabolic burden.

Task 2.1. Promoter development: Robust, loci-independent, light-synchronized promoters will be developed. Toward this goal, we aim to (i) identify circadian-entrained genes through the re-analysis of previously published transcriptomics data, (ii) construct promoter-probes with the bioluminescent reporter *luxAB*, and (iii) characterize expression in diurnal LD cycles. Toward task (i), we have re-analyzed diurnal LD cycle microarray data [85] and selected genes with robust circadian-entrained transcript accumulation patterns. Preference was given to genes with support for circadian entrainment in other studies, or similar circadian phase in *S.* 7942 [109]. Promoter DNA was selected 500 bp upstream of the translation start site or up-to the adjacent gene translation stop site; this strategy should capture activator and operator regions but avoid repressor regions [30]. Bacterial luciferase (*luxAB*) from pPAM1580 was used as the probe

[15] Excerpt from A. Werner's Preliminary Exam Proposal

[241]. Promoter-probes were inserted into the pIGA plasmid for homologous recombination into *slr0168* chromosomal neutral site as previously described [32]. Notably, P_{psbA2}, a popular "light-entrained" promoter, does not show light entrainment in *slr0168* neutral site (data not shown). Thus, chromosomal loci-dependence will be assessed by characterizing each promoter:probe from the RSF1010 replicative plasmid pPMQAK1 [242]. *S.* 6803 transformants were sequence verified and initial bioluminescent characterization in diurnal LD cycles was performed on all promoter:probes integrated into the chromosome. P_{hlic}, P_{rbp1}, $P_{slr0006}$, P_{ndhF3}, and P_{rbcL} are promising light-entrained promoters (see *Appendix A*). Future work will characterize these promoter:probes from *slr0168* chromosomal integration and pPMQAK1 over 5-day sinusoidal LD cycles. This will enable assessment of possible growth phase dependence, robustness, calculation of circadian parameters, determination of loci dependence on temporal expression, and comparison of replicative plasmid versus chromosomal *slr068* neutral site strength.

Task 2.2. Alkane production: Alkanes are attractive biofuel targets due to high energy density and, for alkanes such as heptadecane, octane ratings superior to many current fuels [206]. *S.* 6803 naturally synthesizes alkanes from acyl-ACP via fatty acyl-ACP reductase (*sll0209, aar,* Aar) and aldehyde-deformylating oxygenase (*sll0208*, *ado,* Ado) [207].[16] Two distinct promoters have been identified upstream of *ado* and one has been identified between *aar* and *ado* [243]. Our re-analysis of transcriptomics data [85] indicates *ado* is natively light-activated while *aar* levels are relatively constant. Natively, Aar and Ado form a 1:1 complex, which significantly improves Ado turnover *in vitro* likely by substrate channeling [244]. Debate surrounds Ado's mechanism of action [245], oxygen (in)dependence [246,247], and by-product generation [248].

[16] Standard notation of a transcript is lower-case italics and a protein is capitalized; this notation is used throughout the proposal to specify mRNA levels (*e.g. aar*) and protein levels (*e.g.* Aar).

Still, engineering efforts have achieved alkane yield increases in *E. coli* [249,250] and cyanobacteria [248,251]; the details of these efforts will be discussed in the following sections.

This task will apply promoters from Task 2.1. toward light-entrained alkane production to test the hypothesis that light-entrained production will increase alkane yield, titer, and/or productivity as compared to constitutive ("always on") production. Here, we propose an experimental approach to construct and validate strains, assess metabolic burden, and quantify chemical production in the light-entrained and constitutive alkane strains. Our engineering approach (Figure 1) is broken into four steps, detailed below.

(A) Construct and validate constitutive (S_C) and light-entrained (S_L) alkane production strains: Previous efforts suggest Ado is the rate-limiting step in alkane synthesis. Ado will be tagged with six histidines (His_6) to enable a direct assay of Ado protein abundance. Ado and Ado-His_6 activity will be compared with a previously developed aldehyde dehydrogenase assay to ensure the His_6 tag does not interfere with enzymatic activity [252]. A constitutive promoter will drive expression of Aar and Ado-His_6 in S_C. Light-entrained promoters from Task 2.1. will drive light-entrained expression of Aar and Ado-His_6 in S_L. Constructs will be inserted into the *slr0168* chromosomal neutral site via markerless homologous recombination [253]. RT-q-PCR and anti-His6 Western blotting will validate temporal expression of *aar/ado* transcript and Ado protein abundances, respectively, in *S.* 6803.

(B) Evaluate S_C and S_L: Wild-type *S.* 6803, S_C, and S_L will be cultivated in diurnal LD cycles. Metabolic burden in both strains will be evaluated by quantifying ATP, ATP/ADP, and NADPH/$NADP^+$ with commercially available kits. Alkane yield, titer, and productivity will be measured by GC-flame ionization detector (FID) throughout a diurnal LD cycle. Wild-type, S_C, and S_L growth rates will be compared as an additional measure of burden on cellular growth.

(C) Construct and validate a base strain (S_B) with improved alkane production capacity:

We propose to improve alkane production by increasing cofactor (NADPH) and precursor (acetyl-CoA) availability. Genetic modifications conferring these advantages will be constructed in a 'base strain' (S_B) and assessed separate from S_L to allow assessment of the underlying hypothesis without the potentially confounding effects of additional over-expression targets. The proposed S_B genetic modifications are supported by four key observations. First, fatty aldehyde accumulation inhibits cell growth [254]. Cao and coworkers demonstrated that fatty alcohol production is, counterintuitively, synergistic with alkane production, likely by balancing fatty aldehyde accumulation with the slow Ado turnover rate [255]. We therefore do *not* propose to knock-out fatty alcohol synthesis pathways. We will also perform an aldehyde toxicity assay in *S.* 6803. Second, increased expression of fatty acid synthesis (FAS) genes can improve alkane yield [248,256]. Yet, feedback inhibition of acyl-ACP on upstream FAS enzymes [257] supports *not* over-expressing FadD, which may increase acyl-ACP accumulation and down-regulate FAS genes essential for precursor generation. Thirdly, improved NADPH cofactor generation improves alkane production [248,255]. Lastly, Ado is inhibited by H_2O_2, a possible byproduct of alkane synthesis; however, inhibition can be relieved by expression of Ado-Catalase fusion protein [252]; therefore, we propose to over-express *S.* 6803 catalase peroxidase (KatG) to lessen H_2O_2 toxicity. Together, these observations support over-expression of RibF, PetH, Fd, FabD, FabH, FabG, FabI, FabF, and KatG (see Figure D-5-42 for full enzyme descriptions). All engineering targets will be expressed from the replicative plasmid pPMQAK1. RT-q-PCR validation and energy cofactor measurements will be performed as described for S_C and S_L. Alkane, fatty acids (including free fatty acids, FFAs), and fatty alcohols will be quantified at several time-points across the circadian day. We anticipate that alkane increases may be

accompanied by fatty alcohol and fatty acid/FFA increases because all three benefits from fatty aldehyde precursor availability.

(D) Combine S_B and S_L and assess alkane production: To further improve S_L alkane production levels, the base strain modifications (S_B) will be combined with light-entrained Aar and Ado (S_L). We anticipate this strain design to have further improved alkane yields due to large acetyl-CoA and NADPH availability and decreased hydrogen peroxide inhibition of Ado. Strain evaluation will be performed as described for S_L.

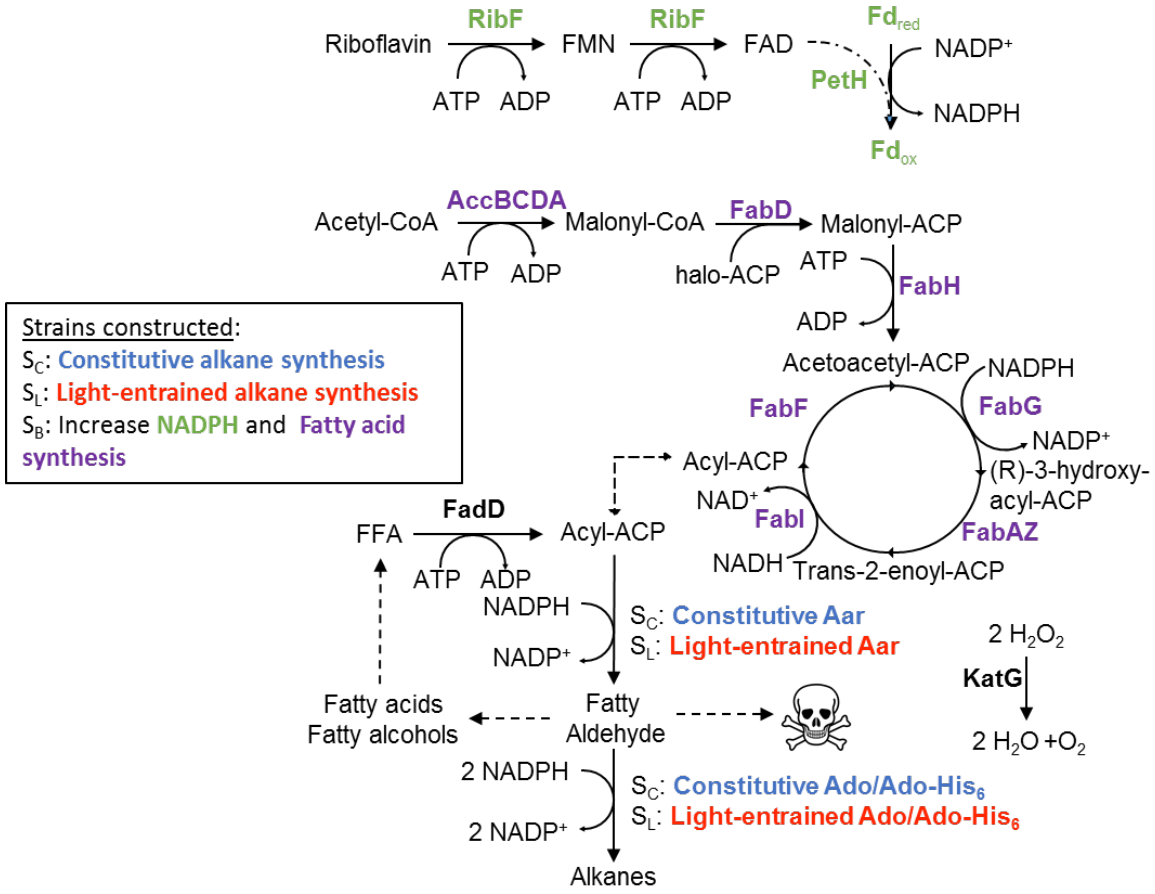

Figure D-5-42. Metabolic engineering strategies for alkane production in LD cycles. Alkane production will be driven via constitutive (SC, blue text) and light-entrained (SL, red text) expression of fatty acyl-ACP reductase (Aar) and aldehyde-deformylating oxygenase (Ado). Genetic targets for increased NADPH (green targets) and fatty acid synthesis (purple targets) will be incorporated into the base strain (SB). Abbreviations: RibF, riboflavin biosynthesis protein; PetH, ferredoxin-NADP oxidoreductase; Fd, ferredoxin I; FabD, malonyl-CoA-ACP transacetylase; FabH, 3-oxoacyl-ACP synthase III; FabG, 3-oxoacyl-ACP reductase; FabA/Z,

(3R)-hydroxymyristol-ACP dehydrase; FabI, enoyl-ACP reductase; FabF, β-ketoacyl-ACP synthase I; KatG, catalase peroxidase (KatG).

Expected outcomes and alternative approaches: Unexpectedly, none of the selected promoter candidates had dark-entrained bioluminescent profiles, including those with dark-entrained transcript accumulation in circadian microarray studies (*e.g.* P_{lrtA}) [85]. Although we have not yet measured transcript accumulation, previous work in our laboratory demonstrates that P_{lrtA}:*gfp* confers dark-entrained transcript accumulation but *not* fluorescence. One hypothesis to explain the transcript-to-protein disconnect is ncRNA interactions within the coding region which regulate translation but not transcription.

We expect to see increased alkane production in S_{BL} as compared to S_{BC} due to decreased metabolic burden during the energy-poor night-phase. If this is not observed, we will assess the alternative hypothesis (H_3) that night-entrained production is superior, which may be supported if ATP/NADPH regulation is relaxed at night. To assess H_3, we will explore a light-entrained siRNA, genetic inverters, or screen additional putative dark-entrained promoters to construct dark-entrained alkane production. Pathway fine-tuning may further reduce any observed aldehyde toxicity [255]. Antisense RNA gene silencing can be utilized to minimize "leaky" off-phase gene expression. Anti-6xHis Western blotting is an alternative approach to quantify Ado if levels are below the limit of SDS-PAGE detection. Previous efforts indicate *Nostoc punctiforme* natively produces higher amounts of alkanes [207]; therefore, *Nostoc* Ado will replace *S.* 6803 Ado if aldehyde dehydrogenase activity appears to be the bottleneck toward alkane production.

APPENDIX E – DATA TO SUPPORT A DYNAMIC BIOMASS EQUATION

SUMMARY

Identified metabolites which are assigned as biomass equation constituents as described by Nogales *et al.* (2014) [76] are presented in this chapter. This model was updated and expanded recently by work from the lab of Dr. Ashok Prasad [77]. Future work in the Prasad group plans to incorporate the data presented in this Appendix into the genome scale model enabled dynamic flux balance analysis (DFBA). DFBA promises to improve the method by which engineering targets are identified for strain design which will be successful in diurnal light:dark (LD) cycles. Towards this goal, metabolites which are referenced in the biomass equation have been identified over the course of a 24-hour diurnal sinusoidal light:dark (sinLD) cycle with a multi-platform mass spectrometry metabolomics workflow. A summary of metabolite coverage by macromolecule is presented in Table E-14. A list of biomass equation components not detected in the present study is presented in Table E-15. Subsequent tables present the metabolites for each macromolecule, including references for mass spectrometry platform and compound number. This Appendix should serve as a short-cut guide for locating relevant biomass equation components in future modelling efforts.

SUPPLEMENTAL TABLES

Table E-14. Summary of biomass equation coverage. In cases where a compound was detected from multiple phases (*e.g.* adenine from the pellet and soluble phase) or in multiple forms (*e.g.* adenine, adenosine, ATP) the number of compounds detected for a macromolecule class may exceed the number of molecules listed for that class in the Nogales *et al.* (2012) biomass equation.

Macromolecule class	# of constituents in		Reference table for identified metabolites	Comments
	Nogales *et al* (2012) Equation	% coverage in this study (# compounds, redundant)		
DNA/RNA	8	100% (25)	Table E-16	Detected all bases (A, T, U, G, C) from insoluble pellet, likely attributable to chromosomal DNA. Detected all nucleotides, most NMPs, most NTPs, and some dNTPs from soluble phase.
Protein	20	95% (50)	Table E-17	Detected all amino acids except asparagine, cysteine, glutamine, methionine, tryptophan from the insoluble pellet. Insoluble amino acids are likely attributable to protein. Detected all except cysteine from the soluble aqueous phase.
Lipids	36	100% (35)	Table E-18	Detected MGDG, DGDG, SQDG, and PG lipid species from the soluble organic phase. These are the species listed in the biomass equation with varying alkyl chain lengths; MS is ill-equipped to determine the distribution of double bonds on each chain (*e.g.* MGDG 32:2) but can only assign total carbon in the alkyl chains (*e.g.* MGDG 32).
Photosynthetic pigments	9	Not determined	Table E-19	The biomass equation lists 4 tocopherols, 3 xanthophylls, and 2 carotenoids. We detected metabolites from each of these classes; confident structural annotations are difficult, so the total coverage of the exact biomass equation constituents cannot be determined.
Soluble pool	25	28% (11)	Table E-20	Spermidine was only detected from the insoluble pellet; unclear is this the appropriate phase for accurate representation of the biomass equation.
Inorganic ions	14	0	N/D	Did not use technique(s) for detection of inorganic ions
Glycogen	1	100% (6)	Table E-21	Detected glucose/galactose mono- and di-saccharides from the insoluble pellet fraction; based on previous work, it is very likely attributable to glycogen. However, we did not use MS technique which can determine linkages. Further work needed to determine entities to which insoluble carbohydrates are associated. Biomass equation provides molecular formula of C6H10O5 for glycogen.

Peptidoglycan	1	Not determined	N/D	Sugar and amino acid content of GC-MS analysis of the insoluble pellet likely captured this content; further protocol development required to determine the exact content.
H_2O	1	0	N/D	Did not us techniques(s) for detection of intracellular water content.

Table E-15. Molecules in biomass equation not detected in present study.

Macro-molecule name from Nogales *et al.* (2012)	Compound name	Molecular formula provided in Nogales *et al.* (2012)
Soluble pool	10-Formyltetrahydrofolate	C20H21N7O7
	5,10-Methylenetetrahydrofolate	C20H21N7O6
	5,6,7,8-Tetrahydrofolate	C19H21N7O6
	5-Methyltetrahydrofolate	C20H24N7O6
	Adenosylcobalamin	C72H100CoN18O17P
	chorismate	C10H8O6
	Coenzyme A	C21H32N7O16P3S
	Heme O	C49H56FeN4O5
	Malonyl-CoA	C24H33N7O19P3S
	Protoheme	C39H30FeN4O4
	Putrescine	C4H14N2
	Pyridoxal 5'-phosphate	C8H8NO6P
	Riboflavin	C17H20N4O6
	Succinyl-CoA	C25H35N7O19P3S
	Undecaprenyl diphosphate	C55H89O7P2
Protein	Cysteine	C3H7NO2S
Peptidoglycan	Peptidoglycan	C40H62N8O21
Inorganic ions	k	K
	nh4	H4N
	mg2	Mg
	ca2	Ca
	fe2	Fe
	fe3	Fe
	cu2	Cu
	mn2	Mn
	mobd	MoO4
	cobalt2	Co
	zn2	Zn
	so4	O4S
	pi**	HO4P
	na2	Na

Table E-16. Metabolites associated with DNA and RNA macromolecules.

Phase	Class	Source*	Cmpd. ID	Annotation	Metabolite
Pellet	Base	GC-MS (NA)	C006	Adenine 2TMS	Adenine, pellet
	Base	GC-MS (NA)	C160	Cytosine 2TMS	Cytosine, pellet
	Base	GC-MS (NA)	C005	Guanine 3TMS	Guanine, pellet
	Base	GC-MS (NA)	C299	Uracil 2TMS	Uracil, pellet
	Ribonucleoside	GC-MS (NA)	C022	Uridine 4TMS	Uridine, pellet
	Deoxyribonucleoside	GC-MS (NA)	C029	Uridine 2 deoxy 3TMS	Deoxyuridine, pellet
Soluble	Base	HILIC-MS	N/A	Adenine	Adenine
	Ribonucleoside	HILIC-MS	N/A	Adenosine	Adenosine
	Ribonucleoside	GC-MS (EI)	C0025	Adenosine 3TMS	Adenosine
	Nucleoside diphosphate	HILIC-MS	N/A	ADP	ADP
	Nucleoside monophosphate	HILIC-MS	N/A	AMP	AMP
	Nucleoside triphosphate	HILIC-MS	N/A	ATP	ATP
	Ribonucleoside	HILIC-MS	N/A	Cytodine	Cytodine
	Deoxyribonucleoside	HILIC-MS	N/A	Deoxycytidine	Deoxycytidine
	Nucleoside monophosphate	HILIC-MS	N/A	CMP	CMP
	Nucleoside triphosphate	HILIC-MS	N/A	CTP	CTP
	Base	GC-MS (EI)	C0361	Guanine 3TMS.	Guanine
	Ribonucleoside	GC-MS (EI)	C0017	Guanosine 5TMS.	Guanosine
	Deoxyribonucleoside	HILIC-MS	N/A	Deoxyguanosine	Deoxyguanosine
	Nucleoside monophosphate	HILIC-MS	N/A	GMP	GMP
	Nucleoside diphosphate	HILIC-MS	N/A	GDP	GDP
	Nucleoside triphosphate	HILIC-MS	N/A	GTP	GTP
	Ribonucleoside	HILIC-MS	N/A	Uridine	Uridine
	Deoxyribonucleoside	HILIC-MS	N/A	Deoxyuridine	Deoxyuridine
	Nucleoside monophosphate	HILIC-MS	N/A	UMP	UMP

*GC-MS (NA) refers to the non-targeted GC-MS run on the nucleic acid content of the insoluble pellet. GC-MS (EI) refers to the non-targeted GC-MS run on the soluble aqueous phase of the biphasic extraction. HILIC-MS refers to the targeted HILIC LC-MS run on the soluble aqueous phase of the biphasic extraction.

Table E-17. Metabolites associated with protein or other amino acid-containing macromolecules.

Phase	Source*	Compound ID	Annotation
Pellet	GC-MS (AA)	N/A	Alanine
	GC-MS (AA)	N/A	Arginine
	GC-MS (AA)	N/A	Aspartic acid
	GC-MS (AA)	N/A	Glutamic acid
	GC-MS (AA)	N/A	Glycine
	GC-MS (AA)	N/A	Histidine
	GC-MS (AA)	N/A	Isoleucine
	GC-MS (AA)	N/A	Leucine
	GC-MS (AA)	N/A	Lysine
	GC-MS (AA)	N/A	Phenylalanine
	GC-MS (AA)	N/A	Proline
	GC-MS (AA)	N/A	Serine
	GC-MS (AA)	N/A	Threonine
	GC-MS (AA)	N/A	Tyrosine
	GC-MS (AA)	N/A	Valine
Soluble	HILIC-MS	N/A	Alanine
	GC-MS (EI)	C0039	Alanine..3TMS.
	GC-MS (EI)	C0536	Alanine..3TMS..1
	HILIC-MS	N/A	Arginine
	GC-MS (EI)	C0154	Arginine...NH3...3TMS.
	HILIC-MS	N/A	Asparagine
	HILIC-MS	N/A	Aspartic.acid
	GC-MS (EI)	C0049	Aspartic.acid..2TMS.
	GC-MS (EI)	C0089	Aspartic.acid..3TMS.
	GC-MS (EI)	C0060	Aspartic.acid..4TMS.
	HILIC-MS	N/A	Glutamine
	HILIC-MS	N/A	Glutamate
	HILIC-MS	N/A	Glutamate.1
	GC-MS (EI)	C0084	Glutamic.acid..2TMS.
	GC-MS (EI)	C0166	Glutamic.acid..2TMS..1
	GC-MS (EI)	C0009	Glutamic.acid..3TMS.
	GC-MS (EI)	C0033	Glutamic.acid..3TMS..1
	GC-MS (EI)	C0275	Glycine..3TMS.
	HILIC-MS	N/A	Isoleucine
	HILIC-MS	N/A	Leucine
	HILIC-MS	N/A	Lysine
	HILIC-MS	N/A	Methionine.SO
	HILIC-MS	N/A	Methionine

	HILIC-MS	N/A	Phenylalanine
	HILIC-MS	N/A	Proline
	HILIC-MS	N/A	Serine
	GC-MS (EI)	C0195	Serine..2TMS.
	GC-MS (EI)	C0231	Serine..4TMS.
	HILIC-MS	N/A	Threonine
	GC-MS (EI)	C0518	Threonine..2TMS.
	GC-MS (EI)	C0254	Threonine..3TMS.
	HILIC-MS	N/A	Tryptophan
	HILIC-MS	N/A	Tyrosine
	GC-MS (EI)	C0190	Tyrosine..3TMS.
	HILIC-MS	N/A	Valine

Table E-18. Metabolites associated with lipid membranes or other glycerolipid containing macromolecules.

Class*	Source**	MS-Finder structural annotation	Annotation based on MW	Molecular Formula
MGDG	MGDG, No acyl chains			C11H16O10R2
	C0902	(2S)-3-(β-D-Galactopyranosyloxy)-1,2-propanediyl (7Z,10Z,7'Z,10'Z)bis(-7,10-hexadecadienoate)	'MGDG 32:4'	C41H70O10
	C0133	2,3-dipalmitoyl-1-α-D-galactosyl-sn-glycerol	'MGDG 32:0'	C41H78O10
	C0611	(2S)-3-(β-D-Galactopyranosyloxy)-2-(palmitoyloxy)propyl (9Z,12Z)-9,12-octadecadienoate	MGDG 34:0'	C43H78O10
	C0026	(2S)-3-(β-D-Galactopyranosyloxy)-2-(palmitoyloxy)propyl (9Z,12Z)-9,12-octadecadienoate	'MGDG 34:2'	C43H78O10
	C0152	2-O-oleoyl-3-O-palmitoyl-1-O-α-D-galactosyl-sn-glycerol	'MGDG 34:1'	C43H80O10
	C0050	(2S)-3-(β-D-Galactopyranosyloxy)-2-(palmitoyloxy)propyl (5Z,8Z,11Z,14Z,17Z)-5,8,11,14,17-icosapentaenoate	'MGDG 36:5'	C45H76O10
	C0504	3-(β-D-glucopyranosyloxy)propane-1,2-diyl (9E,9'E)bis-octadec-9-enoate	'MGDG 36:2'	C45H82O10
	C0943	β-D-Galactopyranoside, 2,3-bis((1-oxooctadecyl)oxy)propyl	'MGDG 36:0'	C45H86O10
DGDG	DGDG, No acyl chains			C17H26O15R2
	C0099	(2S)-3-{[6-O-(α-D-Galactopyranosyl)-β-D-galactopyranosyl]oxy}-2-(tetradecanoyloxy)propyl (9Z,12Z,15Z)-9,12,15-octadecatrienoate	'DGDG 32:3'	C47H82O15
	C0292	(2S)-3-{[6-O-(α-D-Galactopyranosyl)-β-D-galactopyranosyl]oxy}-2-(tetradecanoyloxy)propyl stearate	'DGDG 32:0'	C47H88O15
	C0473	(2S)-3-{[6-O-(α-D-Galactopyranosyl)-β-D-galactopyranosyl]oxy}-2-(tetradecanoyloxy)propyl (5Z,8Z,11Z,14Z,17Z)-5,8,11,14,17-icosapentaenoate	'DGDG 34:5'	C49H82O15
	C0144	(2S)-3-{[6-O-(α-D-Galactopyranosyl)-β-D-galactopyranosyl]oxy}-2-(tetradecanoyloxy)propyl (5Z,8Z,11Z,14Z,17Z)-5,8,11,14,17-icosapentaenoate	'DGDG 34:5'	C49H82O15
	C0472	(2S)-3-{[6-O-(α-D-Galactopyranosyl)-β-D-galactopyranosyl]oxy}-2-[(7Z,10Z)-7,10-hexadecadienoyloxy]propyl (9Z,12Z)-9,12-octadecadienoate	'DGDG 34:4'	C49H84O15
	C0913	(2S)-3-{[6-O-(β-D-Galactopyranosyl)-β-D-galactopyranosyl]oxy}-2-(palmitoyloxy)propyl (9Z)-9-octadecenoate	'DGDG 34:1'	C49H90O15
	C0724	KDYAPQVYJXUQNY-JTADZIJASA-N	'DGDG 36:6'	C51H84O15

	C0131	(2S)-3-{[6-O-(α-D-Galactopyranosyl)-β-D-galactopyranosyl]oxy}-2-(palmitoyloxy)propyl (5Z,8Z,11Z,14Z,17Z)-5,8,11,14,17-icosapentaenoate	'DGDG 36:5'	C51H86O15
SQDG	SQDG, No acyl chains			C11H16O12SR2
	C1670	2-Hydroxy-N-{(2S,3R,4E)-3-hydroxy-1-[(3-O-sulfo-β-D-threo-hexopyranosyl)oxy]-4-octadecen-2-yl}octadecanamide	'SQDG 33'	C42H81NO12S
	C0509	(2R)-3-[(9Z,12Z,15Z)-9,12,15-Octadecatrienoyloxy]-2-(palmitoyloxy)propyl 6-deoxy-6-sulfo-α-D-glucopyranoside	'SQDG 34'	C43H76O12S
	C0828	2-[(9Z)-9-Octadecenoyloxy]-3-(palmitoyloxy)propyl 6-deoxy-6-sulfo-α-D-glucopyranoside	'SQDG 34'	C43H80O12S
PG	PG, No acyl chains			C8H13O10PR2
	C2086	phosphatidylglycerol 32:1	'PG 32:1'	C38H73O10P
	C2062	phosphatidylglycerol 32:1	'PG 32:1'	C38H73O10P
	C0158	phosphatidylglycerol 32:1	'PG 32:1'	C38H73O10P
	C2094	phosphatidylglycerol 34:3	'PG 34:3'	C40H73O10P
	C1742	phosphatidylglycerol 34:3	'PG 34:3'	C40H73O10P
	C1321	phosphatidylglycerol 34:3	'PG 34:3'	C40H73O10P
	C1020	phosphatidylglycerol 34:3	'PG 34:3'	C40H73O10P
	C0809	phosphatidylglycerol 34:3	'PG 34:3'	C40H73O10P
	C0319	phosphatidylglycerol 34:3	'PG 34:3'	C40H73O10P
	C0023	phosphatidylglycerol 34:3	'PG 34:3'	C40H73O10P
	C2080	phosphatidylglycerol 34:2	'PG 34:2'	C40H75O10P
	C0851	phosphatidylglycerol 34:2	'PG 34:2'	C40H75O10P
	C0666	phosphatidylglycerol 34:2	'PG 34:2'	C40H75O10P
	C0542	phosphatidylglycerol 34:2	'PG 34:2'	C40H75O10P
	C0084	phosphatidylglycerol 34:2	'PG 34:2'	C40H75O10P
	C1419	phosphatidylglycerol 36:2	'PG 36:2'	C42H79O10P

Table E-19.Metabolites identified which are similar to photosynthetic pigments contained in the biomass equation.*

Class	Subclass	Cmpd. ID	MS-Finder annotation	Molecular Formula
Carotenoids	Xanthophyll	C0037	Mutatochrome	C40H56O
Carotenoids	Xanthophyll	C0058	Anhydrolutein-II	C40H54O
Carotenoids	Xanthophyll	C0064	adonixanthin	C40H54O3
Carotenoids	Xanthophyll	C0089	(3'E)-1'-Hydroxy-3',4'-didehydro-1',2'-dihydro-β,ψ-caroten-2'-one	C40H54O2
Carotenoids	Xanthophyll	C0226	3',8',19-Trihydroxy-3-methoxy-7,8-didehydro-β,κ-caroten-6'-one	C41H56O5
Carotenoids	Xanthophyll	C0260	Alloxanthin	C40H52O2
Carotenoids	Xanthophyll	C0295	Alloxanthin	C40H52O2
Carotenoids	Xanthophyll	C0331	(9cis,9'cis,11cis,11'cis,13cis)-3,3',4'-Trihydroxy-β,β-caroten-4-one	C40H54O4
Tocopherols	Vitamin K	C0403	Vitamin K1 2,3-Oxide	C31H46O3
Carotenoids	Xanthophyll	C0487	Eschscholtzxanthin	C40H54O2
Carotenoids	Xanthophyll	C0508	Cyclodiadinoxanthin	C40H54O3
Carotenoids	Xanthophyll	C0742	Rhodoxanthin	C40H50O2
Carotenoids	Xanthophyll	C0746	(3R,3'S,6'S,9cis)-4',5'-Didehydro-5',6'-dihydro-β,β-carotene-3,3'-diol	C40H56O2
Carotenoids	Xanthophyll	C1085	adonixanthin	C40H54O3
Carotenoids	Xanthophyll	C1097	adonixanthin	C40H54O3
Carotenoids	Carotene	C1138	Apocarotenal	C30H40O
Carotenoids	Xanthophyll	C1292	(3R,3'S,6'S,9cis)-4',5'-Didehydro-5',6'-dihydro-β,β-carotene-3,3'-diol	C40H56O2
Carotenoids	Xanthophyll	C1382	adonixanthin	C40H54O3
Carotenoids	Xanthophyll	C1399	Cyclodiadinoxanthin	C40H54O3
Carotenoids	Xanthophyll	C1438	(3R)-3',4',7',8'-Tetradehydro-β,β-caroten-3-ol	C40H52O
Carotenoids	Xanthophyll	C1463	(3R)-3',4',7,7',8,8'-Hexadehydro-β,β-caroten-3-ol	C40H50O
Carotenoids	Carotene	C1491	Methyl (3S,4S,21R)-14-ethyl-13-formyl-4,8,18-trimethyl-20-oxo-3-(3-oxo-3-{[(2E,7R,11R)-3,7,11,15-tetramethyl-2-hexadecen-1-yl]oxy}propyl)-9-vinyl-21-phorbinecarboxylate	C55H72N4O6
Carotenoids	Xanthophyll	C1644	(3R)-3',4',7',8'-Tetradehydro-β,β-caroten-3-ol	C40H52O

Carotenoids	Xanthophyll	C1832	(3S)-3-Hydroxy-β,β-caroten-4-one	C40H54O2
Tocopherols	Phylloquinone	C1854	Phylloquinone	C31H46O2
Carotenoids	Carotene	C1974	7',8'-Didehydro-phi,phi-carotene	C40H46
Carotenoids	Xanthophyll	C2452	(3R,3'E,9'cis)-3-Hydroxy-1'-methoxy-3',4'-didehydro-1',2'-dihydro-β,ψ-caroten-19'-al	C41H56O3

Chlorophyll *a* may have been detected as [M+] (*m/z*=892.56) from C0345 in RP-MS dataset with main fragment ion (*m/z*=614) with the loss of a phytyl tail [258]. Further work with standards is necessary to confirm.

Table E-20.Metabolites identified which are similar to those assigned to the biomass equation 'soluble pool'.

Source	Cmpd. ID	Annotation
GC-MS (NA)	C072	Spermidine...CO2...5TMS.
HILIC	N/A	Acetyl CoA
HILIC	N/A	FAD
HILIC	N/A	NADH
HILIC	N/A	NAD
HILIC	N/A	NADPH
HILIC	N/A	NADP
HILIC	N/A	Glutathione*
HILIC	N/A	S.adenosylmethionine
HILIC	N/A	Thiamine pyrophosphate
HILIC	N/A	ADP
HILIC	N/A	ATP

*Biomass equation specifies reduced glutathione.

Table E-21. Metabolites identified which are similar to those assigned to biomass equation macromolecule glycogen. Insoluble C5 carbohydrates and carbohydrates from the soluble phase are not included in this table as they would not be attributable to glycogen.

Phase	Source	Cmpd. ID	# C6 sugars	Annotation
Insoluble	GC-MS (NA)	98	1	Glucose, 1,6-anhydro-, beta- (3TMS)
	GC-MS (NA)	20	1	Galactose (1MEOX) (5TMS) MP
	GC-MS (NA)	12	1	Galactose, 2-amino-2-deoxy-, D- (1MEOX) (5TMS)
	GC-MS (NA)	62	2	Isomaltose (1MEOX) (8TMS) MP
	GC-MS (NA)	2	2	Melibiose (1MEOX) (8TMS) MP
	GC-MS (NA)	39	2	Trehalose, alpha,alpha'-, D- (8TMS)